Imperial College London

IMPERIAL COLLEGE BUSINESS SCHOOL

Understanding Superstar firms, market dynamics and intangibles

Jenniffer Solorzano Mosquera

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Imperial College London, May 2023

Abstract

This thesis adds to the debate whether superstar firms, defined as top-income distribution firms (e.g., Amazon, Microsoft, Google, Facebook (Meta)) are good for economies. Rather than attributing market power to observed trends in reduced business dynamics and increased concentration, I propose a more optimistic explanation stemming from an economic model transformation due to a more intangible economy, technological advancements, and globalization.

The first chapter describes U.K. business dynamics (e.g., entry and exit of firms) and its connection with intangible investments (e.g., brands, management practices, corporate culture). It builds on an original longitudinal database using business register data and the application, for the first time in the U.K. in decades, of the OECD's DynEmp3 software routine to obtain comparable and matchable business dynamics indicators with intangibles' data. Findings expose a negative relationship between employment growth rates and intangible investments, likely arising from frictions generated by a more intangible economy; skilled over unskilled labor is preferred.

The second chapter addresses market power increasing concerns. Evidence of increasing measured markups, especially for Superstars, is debunked by the chapter, demonstrating it highly depends on the estimation approach used. After correcting some estimation inconsistencies, measured markups in the U.S. and Europe remained stable since the Great Recession. Findings support notions of rising industry concentration resulting from technological changes rather than lax regulation permitting market power.

In the third chapter, we introduce unexplored innovation and intangible asset measurements for empirical research connecting markups and business dynamics. We find that in the U.K. and some Latin American countries, social media data can serve as proxies for these concepts at the firm level. We create measurements *Twitnovation*, representing a firm's innovation drivers, and *Aboutnovation*, representing a firm's investment in customer engagement. They explain a substantial portion of labor productivity variation and correlate strongly with marketing expenses. Social media data can offer a more cost-effective and immediate alternative to traditional firm innovation surveys.

Contents

Declaration of Originality 9

Copyrights Declaration 10

Acknowledgements 12

Disclaimer 13

Introduction 14

1 Business Dynamics and Intangible Investments 19
- 1.1 Introduction . . . 19
- 1.2 Business Dynamics: Definitions . . . 22
  - 1.2.1 Business Dynamics Indicators . . . 22
  - 1.2.2 Unit of analysis of the firm . . . 22
  - 1.2.3 Level of aggregation for analysis . . . 22
  - 1.2.4 Job creation and destruction . . . 26
  - 1.2.5 Job Reallocation . . . 27
  - 1.2.6 Entry and exit rates . . . 28
  - 1.2.7 Dispersion and skewness . . . 29
  - 1.2.8 High-Growth Firms . . . 30
  - 1.2.9 Definitions of firms' demographic groups . . . 32
- 1.3 Intangible Investments: Definitions . . . 32
- 1.4 Conceptual Framework . . . 32
  - 1.4.1 Main trends of business dynamics indicators . . . 32
  - 1.4.2 Declining dynamics: Are intangibles the reason? . . . 35
- 1.5 Data . . . 40
  - 1.5.1 Business Dynamics Data . . . 41
  - 1.5.2 Intangibles Investments Data . . . 46
- 1.6 Empirical Analysis . . . 50
  - 1.6.1 Business Dynamics and Intangible Investments: Stylized facts . . . 50
  - 1.6.2 Dynamic Panel Model Estimations . . . 77
  - 1.6.3 General regression and estimation methods . . . 78
  - 1.6.4 Regression Results for Dispersion Differentials . . . 80
- 1.7 Motivating model: employment dynamism and intangibles . . . 86
  - 1.7.1 Discussion: Dispersion Differential 90th-10th . . . 89
- 1.8 Conclusions . . . 92

2 Market Power and Technology[1] 99
- 2.1 Introduction . . . 99
- 2.2 Conceptual Framework . . . 104
  - 2.2.1 Market Power and Superstar Firms . . . 105
  - 2.2.2 Micro-founded research on market power . . . 106
  - 2.2.3 Stylized facts in market power . . . 107
- 2.3 Methodology . . . 114
  - 2.3.1 Model 1: De Loecker et al. . . . 115
  - 2.3.2 Model 2: New . . . 122

2.3.3 Model 3: New, more flexible model . . . 125
2.3.4 Implementation: Model 2 (New) . . . 126
2.3.5 Implementation: Model 1 (De Loecker et al.) . . . 129
2.4 Data . . . 129
2.5 Results . . . 133
2.5.1 Production function estimates . . . 134
2.5.2 Aggregate markups . . . 135
2.5.3 Sectors . . . 143
2.5.4 Percentiles by sector . . . 151
2.5.5 Global . . . 152
2.6 Conclusion . . . 153

3 Twitnovation and Aboutnovation[2] 159
3.1 Introduction . . . 159
3.2 Conceptual Framework . . . 164
3.3 Data . . . 166
3.3.1 Latin America . . . 166
3.3.2 United Kingdom . . . 168
3.4 Social Media based indicators . . . 170
3.4.1 Twitnovation . . . 170
3.4.2 N-Gram Based Measures of Technologies and Intangibles . . . 175
3.4.3 Measuring Customer Service . . . 177
3.4.4 The relation between social media indicators and firm performance 181
3.5 Aboutnovation . . . 186
3.5.1 Aboutnovation Language . . . 186
3.5.2 Aboutnovation Types . . . 187
3.5.3 Aboutnovation sentiment . . . 190
3.5.4 Aboutnovation Types and Labor Productivity . . . 194
3.5.5 Aboutnovation as a Proxy for Intangible Investment . . . 201
3.6 Conclusion . . . 203

4 Conclusions 206

References 211

Appendices 232

Appendix A UK data comparability 232
A.1 BSD and comparability with UK official demography reports . . . 232
A.2 Assumptions applied to BSD to run DynEmp3 . . . 238
A.3 Comparability between U.K. and U.S. data for business dynamics analysis 239
A.4 Definitions of Entry and Exit Rates for the U.S. . . . 241

Appendix B DynEmp3 242
B.1 Fact BD: Uni-personal firms gains terrain in size distribution . . . 242
B.2 Fact BD: Job Creation and Destruction stable except during the financial crisis and Brexit vote . . . 244
B.3 Fact BD: Births and Deaths of Firms Recover after financial crisis but Decline After Brexit vote . . . 245

B.4 Fact BD: The financial sector is more dynamic than the rest of the economy 246
B.5 Fact BD: Brexit's vote cuts up short business dynamics recovery 247
B.6 Job Creation and Destruction 249
B.7 Excluded sectors 250

Appendix C Regressions **253**
C.1 Arellano Bond 253
C.2 Blundell Bond 255
C.3 Long difference estimations 255
C.4 Dispersion differentials 257
C.5 Job Reallocation Rate 258
C.6 Entry and Exit Rates 259
C.7 Shares of employment 261

Appendix D Intangible Investments Data Matching with Business Indicators By Sector **263**

Appendix E Additional definitions for U.S. estimations of business dynamics indicators **264**

Appendix F Dictionary of keywords for tech and intangible variables **265**

Appendix G Twox Latin America **271**

Appendix H Twox UK **273**

Appendix I Labor Productivity, Twox and Patents **276**

Appendix J Aboutnovation **277**
J.1 Descriptives 277
J.2 Types of mentions - Examples 278
J.3 Language 282
J.4 Sentiment of Mentions 283
J.5 N-grams Aboutnovation 284
J.6 Labor Productivity and Mentions 285

Appendix K Summary statistics: Twitnovation and Aboutnovation **288**

Appendix L Assumptions of (De Loecker et al., 2020) for estimation of elasticities using COMPUSTAT dataset **289**

Appendix M Additional assumptions for estimation of elasticities for markups **291**

Appendix N Estimations of markups **291**

Appendix O Forlani Method: Cobb-Douglas **291**

Appendix P Forlani method: Translog **293**

Appendix Q Descriptives of markups **293**

Appendix R Version for US (Markups): All Rolling Vs. Non-Rolling Both Models as in Figure 30b **295**

Appendix S Weighted version for 131 countries (Markups) **296**

Appendix T Weighted version for 131 countries (Markups) - By Industry **297**

Appendix U Weighted version for 131 countries (Markups) - By Tech Sector **299**

List of Figures

1 Firm categories . . . 24
2 Dynamics . . . 26
3 Unit of Analysis. Source: Author based on Business Structure Database User Guide. . . . 44
5 Mean size growth by age class . . . 51
9 Entries and exit rates . . . 56
18 High-Growth Firm's participation over time . . . 67
21 Share of intangible investments and dispersion of employment growth . . . 70
22 Employment Growth Rate Rank . . . 70
23 Employment growth by intangible intensity level . . . 72
24 Average Intangible Investment Intensity by Industry Sectors (Level 1) . . . 73
25 Average Intangible Investment Intensity by Industry Sectors (Level 2) . . . 73
26 Sectors . . . 74
27 Share of Employment of High Growth Firms . . . 75
28 Employment Growth by Intangible Intensity and Firm Size . . . 77
29 Average U.S. Markup (weighted by sales). De Loecker et al. . . . 137
30 Average U.S. Markup (weighted by sales) - Adjusted Sample (1955 - 2016) 138
31 Average European Markup (weighted by sales) - (1998 - 2016) . . . 138
32 Average U.S. Markup (Weighted vs. Unweighted) - Adjusted Sample . . . 139
33 Average European Markup (Weighted vs. Unweighted) . . . 139
34 U.S. Weighted vs. Unweighted markups: Autor et al. 2020 . . . 140
35 Average markup percentiles (weighted by sales). COMPUSTAT (U.S. only) 141
36 Average markup percentiles (weighted by sales). ORBIS (Europe only) . . . 141
37 Elasticities and Variable Input Shares. (U.S.: 1955 - 2016, EU: 1999 - 2016) 143
38 U.S. Manufacturing . . . 144
39 U.S. Wholesale . . . 145
40 U.S. Retail . . . 146
41 U.S. Information . . . 147
42 Europe Manufacturing . . . 148
43 Europe Wholesale . . . 149
44 Europe Information . . . 150
45 U.S. Average Markups for Information by percentile . . . 151
46 U.S. Average Markups for Manufacturing by percentile . . . 151
47 European Average Markups for Information by percentile . . . 151
48 European Average Markups for Manufacturing by percentile . . . 152

49 All sectors. Rolling within the sector . . . 153
50 Trend Over Time for the Average Firm: Latin America . . . 172
51 Examples of Twitnovative Words: Latin America . . . 173
52 Variation at the Firm-Week Level: Latin America . . . 174
53 Firm-Level Response Rate: Latin America . . . 179
54 Regressions of Labor Productivity on Customer Engagement - Latin America - . . . 186
55 Examples of Aboutnovation Words: Positive and Negative Sentiment . . . 192
56 Regressions of Labor Productivity on Customer Engagement (Mentions) United Kingdom . . . 200
57 Regressions of Labor Productivity on Customer Engagement (Mentions) Latin America vs. United Kingdom . . . 200
58 Correlation between investment data and Aboutnovation - United Kingdom - . . . 201
B.8 Growth in the number of firms by size, 2000-2019 . . . 244
B.10 Entries and exits . . . 246
H.18 Trend Over Time for the Average Firm . . . 273
H.19 Twitnovative words according to Twinovation Success and correlation with Relative Productivity: United Kingdom . . . 274
J.20 Response Time and Number of Mentions: Latin America . . . 277
J.21 Examples of Aboutnovation Words . . . 284
J.22 Regressions of Labor Productivity Quantiles on Customer Engagement United Kingdom . . . 286
R.23 Average U.S. Markup (weighted by sales). De Loecker et al. vs. New - (1955 - 2016) . . . 295
R.24 Average European Markup (weighted by sales). De Loecker et al. vs. New (1999 - 2016) . . . 295
S.25 De Loecker Style vs. Model 2 . . . 296
S.26 De Loecker Style vs. Model 2, technologies change . . . 296
T.27 De Loecker Style (Model 1) - Manufacturing and Wholesale . . . 297
T.28 De Loecker Style (Model 1): Accommodation and Information . . . 297
T.29 Model 2: Manufacturing and Wholesale retail . . . 298
T.30 Model 2: Accommodation and Information . . . 298
U.31 Tech sector by sub-sector: Model (1) . . . 299
U.32 Tech sector by sub-sector: Model (2) . . . 300
U.33 Tech sector by sub-sector: Model (3) . . . 301

List of Tables

1 Definitions . . . 25
2 Industry aggregation levels . . . 42
3 Industry Sectors Correspondance . . . 50
4 Dispersion of employment growth rates . . . 81
5 Dispersion Differentials 90-10 . . . 82
6 Dispersion Differentials 90-10 and 90-50 . . . 83
7 Dispersion 90-10 Long Differences . . . 84
8 Share of firms by age*size class . . . 85

9 Summary Statistics U.S.: De Loecker et al. sample . . . 130
10 U.S. COGS share: Original De Loecker et al. . . . 131
11 U.S.: De Loecker et al. sample after additional assumptions . . . 131
12 Summary Statistics Europe: De Loecker et al. sample after add. assumptions . . . 133
13 U.S. Output Elasticities (Weighted by sales) . . . 135
14 U.S. Markups (Weighted by sales) . . . 136
15 EU Markups (Weighted by sales) . . . 136
16 Handles - Sample Composition . . . 167
17 Comparing Mean Financials of Handle Sample with All Orbis Firms . . . 168
18 Top 10 Sectors (2-digit NAICS) (in %) . . . 168
19 Summary Financials UK: All UK Orbis Firms with a Handle and Non-Missing Turnover or Employees . . . 168
20 Mean Financials of Handle Sample with All Orbis Firms: UK . . . 169
21 Correlations between Twox and Tech Bigrams - Latin America - . . . 176
22 Correlation between Twox Success, Customer Engagement and Bigrams - Latin America - . . . 181
23 Regressions of Labor Productivity - Latin America - . . . 183
24 Regressions of Labor Productivity on Customer Engagement - Latin America - . . . 185
25 Mentions by type - Latin America - . . . 188
26 Mentions by type - UK - . . . 189
27 Mentions by type and subtypes - Latin America - . . . 189
28 Mentions by type and subtypes - UK - . . . 190
29 Sentiment of mentions in Spanish - Latin America (Bello (2021) methodology) . . . 191
30 Sentiment of mentions in English - UK (Go et al. (2009) - GBH methodology) 191
31 Average sentiment score of mentions by reference type and language: Latin America - Vader (Hutto, 2020; Hutto and Gilbert, 2014) and Sentiment-Spanish (Bello, 2021) - . . . 194
32 Average sentiment score of mentions by reference type and language: United Kingdom - Go et al. (2009) - GBH methodology - . . . 194
33 Regressions of Labor Productivity on Customer Engagement by type - Latin America - . . . 195
34 Regressions of Labor Productivity on Customer Engagement - United Kingdom - . . . 197
35 Regressions of Labor Productivity on Customer Engagement - United Kingdom - - Sentiment (GBH Methodology) - . . . 199
A.1 Differences in definitions . . . 232
C.2 Dispersion Differences: Long Differences Regression 90th-50th . . . 257
C.3 Dispersion Differences: Long Differences Regression 50th-10th . . . 258
C.4 Job Reallocation Rates . . . 258
C.5 Entry rates (Number of companies) . . . 259
C.6 Entry rates: Number of employees . . . 259
C.7 Exit rates: Number of companies . . . 260
C.8 Exit rates: Number of employees . . . 260
C.9 Young and small firms . . . 261
C.10 Mature and large . . . 262

D.11 Sector Classification . . . 263
D.12 Intangibles Dataset Sector Classification . . . 264
F.13 Technology bigrams . . . 265
F.14 Intangible Bigrams . . . 270
G.15 Twitnovative Tweets in 2022: Latin America . . . 271
G.16 Correlation between Twox Share, Customer Engagement and Bigrams - Latin America - . . . 272
I.17 Regressions of Patent Counts - Latin America - . . . 276
J.18 User Mentions - Latin America - . . . 277
J.19 Mentions by type and subtypes - Latin America - English Examples . . . 278
J.20 Mentions by type and subtypes - Latin America - Spanish Examples . . . 279
J.21 Mentions by type and subtypes - United Kingdom - English Examples . . . 280
J.22 Sample mentions - Latin America . . . 281
J.23 Sample mentions - UK . . . 282
J.24 Mentions by language - Latin America - . . . 282
J.25 Mentions by language - UK - . . . 283
J.26 Sentiment of mentions in English - Latin America (Vader) . . . 283
J.27 Sentiment of mentions in English - UK (Vader) . . . 283
J.28 Regressions of Labor Productivity on Customer Engagement - United Kingdom - - Likes and Retweets - . . . 285
J.29 Regressions of Labor Productivity on Customer Engagement II - United Kingdom - . . . 287
K.30 Summary Statistics - Latin America - . . . 288
K.31 Summary Statistics - United Kingdom - . . . 289
N.32 Comparing Variables for Prodest Estimation: De Loecker et al. . . . 291
Q.33 Comparing Raw Variables: De Loecker et al. . . . 293
Q.34 Comparing Raw Variables: De Loecker et al. . . . 294
Q.35 Cogshares percentiles . . . 294
Q.36 Descriptives: Global sample . . . 294

Declaration of Originality

I hereby certify that this thesis constitutes my own and my coauthors' work and that all other material has been properly acknowledged and referenced. Jenniffer Solorzano Mosquera

Copyrights Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC). Under this license, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the license terms clear to others by naming the license and linking to the license text. Where a job has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this license or permitted under UK Copyright Law.

Acknowledgements

Words cannot express my gratitude to my fantastic supervisors, Jonathan Haskel and Ralf Martin, for their invaluable kindness, patience, and feedback. I could not have undertaken this journey successfully without Jonathan inspiring me to pursue intriguing and impactful research subjects such as the highly transformative power of intangible assets and their interaction with technological change and globalization. Without him sharing his extraordinary insight and expertise with me, plus his incredible wisdom and practicality acquired from dealing daily with the most challenging realities of one of the top economies in the world, I would not have been able to complete a program at such a leading institution. Thanks, Jonathan, for all your time, dedication, inspiration, and guidance throughout this journey and for facilitating all the opportunities you have opened and continue to open for me to advance in my research and career.

With his staggering intelligence, creativity, and wit-genius level, Ralf inspired me to push harder, think faster, and, most importantly, ask, why not? My supervisor's excitement in generating and developing new ideas is inspirational and helped me persevere and address research questions with innovative perspectives, proposing novel methodologies and using unconventional data. Thank you, Ralf, for inspiring me to be creative and make the work of economists so much fun and exciting.

This endeavor would not have been possible without the generous support from the OECD, including the Productivity, Innovation and Entrepreneurship team led by Chiara Criscuolo, who supported me with the software routines for the DynEmp and direct access to their economists and experts that helped me to develop my research. I am also grateful to the Inter-American Development Bank, especially Matteo Grazzi, and the Competition, Technology, and Innovation team led by Gonzalo Rivas, who supported my research. Thanks to their willingness to engage in new and exciting ideas, we could explore alternative data sources for traditional economic questions. Thanks, Matteo and Gonzalo, for supporting me throughout this process and trusting my abilities to deliver research significantly and urgently needed for Latin America. I also have to thank Eugenie Dugoua for her incredible teachings on discipline and processing techniques to analyze big data. Thanks, Eugenie; without your insights, it would have been like navigating murky waters when exploring such controversial data as social media.

I am also grateful to my friends and PhD mates, especially Moritz, Federica, Massimo, Petra, Mobeen, Dheeya, Robert, Antoine, Prashant, and Adelina, for their support in so many ways throughout these years. From moral backing to coding and editing help, I was having fun and chatting about life. All of you helped me enormously to keep going. Dheeya, thank you for reminding me about the importance of patience and resilience during the hardships of completing such an academic program. Prashant, thank you for sharing your enthusiasm and curiosity

for research and trying new approaches.

Special thanks to Charlotte and Dennis for being an inspiration, a voice of reason, and valuable insights for completing shared research projects. But importantly, for being an example of how to complete a Ph.D. with enthusiasm and joy despite all not-so-joyful moments.

Thanks should also be given to Carol Propper and Renata Kosova for their enormously insightful comments on my thesis throughout the different stages until the final examination. Your views helped me to tailor my approach for more valuable results. I also thank the administrative team of the Doctoral Program at the Imperial College Business School for their incredible support throughout these years in all aspects of the program and my life in London—mainly thanks to Katherine Campbell, Nikki Hann, Catherine Lester, and Pamela Monteith.

Lastly, but more importantly, to my family. First of all, I need to thank my husband, Robert. Embarking into a PhD was a tough, life-altering decision. But thanks to his incredible support and sacrifices to keep us strong as a couple and the decisions he had to shape to make my dream come true, not only for my professional growth but also for the contribution it would make to us as a family, I was able to go through this journey. This thesis coming to fruition is a testament to the power of love and partnership. It is a testament to our hardships, which, although challenging, made us stronger and paved an even better future for our family. Thanks, my love, for inspiring me with your drive, intelligence, resilience, sense of responsibility, and dedication to complete this Ph.D.

I also dedicate this thesis to my savvy mother, Elvira. I would not have been able to finish this work without her multidimensional and unconditional support during the Ph.D. and all the steps that led me to start my journey at Imperial. Her teachings on hard work, faith, and consistency were critical. Thanks to her, I learned that there are no limits except the ones we impose on ourselves. Mom, this work is all yours as it is mine. Thank you for teaching me how to get things done no matter what. I dedicate this thesis to my brave sister, Erika. Thanks to her, I opened my eyes to a world outside Colombia and made me want to explore it. Thanks, sister, for teaching me to always be hungry for more and not conform. I dedicate this thesis to my brilliant father, Efrain. His profound curiosity and passion for literature, philosophy, and science inspired me to pursue knowledge continuously. Thanks, Dad, for encouraging me constantly to ask why. I dedicate this thesis to my lovely nieces, Paula and Isabella, who reminded me to laugh and remember that life is a unique, quick ride we must take advantage of while alive. Thank you, beautiful little ladies; you relieved me of many stressful days during this PhD journey.

Disclaimer

The views expressed in this thesis are my own and do not necessarily represent the views of the OECD, the Inter-American Development Bank, or the Bank of England staff.

Introduction

This thesis aims to contribute to a better understanding of the intertwining between observed trends in economies worldwide over the past few decades. In particular, the slowdown in business dynamics, the rise of intangible capital, the rising market power, and the emergence of the "superstar firms" - firms that have a substantially higher income than peers, pulling away from them further over time. Each of these topics has developed its unique stream of literature. More recently, explanations behind them imply that those trends are deeply interconnected. Nonetheless, a lack of data and appropriate conceptual and methodological approaches have paved the way to blame superstars as the drivers of low dynamism and rising market power. This thesis partially questions this narrative by proposing original empirical and theoretical methods.

A heated global debate in policy and academia clashes about a relentlessly increasing market power (MkP), especially in industrialized economies. Several authors indicate that what seemed to be a purely microeconomic phenomenon of specific industries and markets has now extended to the overall economy (Syverson, 2019). The implications of increased average levels of market power across the board are significant since they can degrade key metrics in economy-wide well-being, including investment, innovation, total output, aggregate productivity, and income distribution.

This discussion has led to a laser focus on superstar firms. The most commonly known superstars are those of the type of the FAANGs (Facebook, Amazon, Apple, Netflix, and Google) or, as more recently called, the MATANAs (Microsoft, Apple, Tesla, Amazon, and Nvidia), depending on who is at the top of the distribution. Although superstar firms can be more productive than other firms (Autor et al., 2017; Bessen, 2017; Autor et al., 2020) -which is something positive-, the standard argument is that over the last two decades, superstars have led to increased concentration and profits in most industries especially in the U.S., while all the typically negative consequences of market power have been on the rise (Furman, 2015; Grullon et al., 2019; Barkai, 2016; Philippon and Gutierrez, 2017). Thus, superstars are, in essence, bad for economies. Why is this argued in concrete?

First, according to various studies on industrialized economies, as superstars have emerged, firm-level markup measurements (price over marginal cost) have consistently increased—especially for larger firms. Higher markups are considered unrefutable evidence of market power (De Loecker and Warzynski, 2012; De Loecker et al., 2018, 2020; Loecker and Eeckhout, 2021). The most influential studies date from 2018, when economists De Loecker, Eeckhout, and Unger found astonishing increases in average markups in the U.S. from 18% in 1980 to over 67% over costs in 2016 (De Loecker et al.,

2018). They saw similar rises in European countries but less prevalence in Latin America and Asia (Loecker and Eeckhout, 2021). Fundamentally, the story is that superstars may be exercising market power by setting higher prices, disincentivizing consumer demand, and leading firms to demand fewer factors of production, such as labor. This triggers workers' income share in revenues to decline. Subsequently, production decreases as it does the incentives to make further investments to innovate, eventually ushering to a reduced percentage of capital in revenues, leaving owners taking a more significant chunk of the revenue pie.

Second, over a little more than the last two decades, business dynamics experienced a secular decline, especially after the dot com bubble in the 2000s (Decker et al., 2016c, 2017, 2019; Bijnens and Konings, 2017, 2018a; Calvino et al., 2020). Business dynamics (B.D.) refers to the entry and exit of firms in and out of the market and the reallocation of inputs, for example, jobs, from less to more productive companies/establishments. In contrast to market power, B.D. seems to have been declining in industrialized economies regardless of size. Lower dynamics are concerning because there is less "creative destruction" (Schumpeter, 1942). On average, entry and job reallocation rates declined by about three and five percent in OECD countries in the last two decades Calvino et al. (2020). In the U.S. only, business entry rates fell from 10.9% in 1980 to 7.3% in 2018, equivalent to a 36% drop (Bessen, 2022). Declining business dynamics are also often linked to stagnated Total Factor Productivity (TFP)[3] growth and the widening of the within-industry TFP gap between frontier and laggard firms over time (Haskel and Martin, 2002; Andrews et al., 2015, 2016a,b; Akcigit and Ates, 2018; Van Reenen, 2018; Calvino and Criscuolo, 2019).

The predominant argument that superstars trigger these trends is based on the analysis of authors like Grullon et al. (2019); Philippon and Gutierrez (2017); Gutiérrez and Philippon (2017a, 2018); Cabral (2018). Generally, their premise is that rising market power belongs to a selected group of dominant firms with price-setting power that has been granted to them due to weak competition regulation, including less antitrust enforcement and exacerbated permissiveness to acquisition deals, particularly by tech giants. Subsequently, they have fueled higher concentration and prevented fluent competition and reallocation of factors. For instance, Philippon and Gutierrez (2017) links the increase in concentration with the decreasing antitrust enforcement in the United States. The authors compare it with more stable concentration levels in Europe during periods of vigorous enforcement and decreasing entry barriers.

However, are the "immoral" superstars the reason behind rising market power and lower business dynamics? More importantly, is market power indeed increasing? Or is it that bad that business dynamics have slowed down? Maybe it is essential to remember that arguments such as Philippon and Gutierrez (2017)'s are suspicious since they assume that

more concentration means less competition, which is not always true [4]. Yet, the discourse is so powerful that audiences overlook critical problems with the methodologies backing up such conclusions. Furthermore, more efficient firms may not need to reallocate resources, so naturally, dynamics are low. In contrast to connotating negatively the superstars, like intrinsically anti-competitive and less pro-employment, this thesis aims to look more even-handedly at these questions. I build on recent work to propose new theoretical and empirical strategies that support the idea that behind these observed trends, the primary explanation can be a more laudable one.

The premise of this thesis is that the superstars have emerged more likely due to the rise of Intangible Capital (Corrado et al., 2005, 2009; Haskel and Westlake, 2018) in combination with two transcendental events, the ICT revolution -and its complementarity with intangibles- and globalization -which propelled the incumbency advantage of larger firms. Intangible capital refers to assets that one cannot easily touch or measure (e.g., innovative-intellectual property, software, data, research, design, or organizational processes, among others)(Corrado et al., 2009, 2017; Haskel and Westlake, 2018). A notable surge in the proportion of intangible investments worldwide has been evident in the last decades. In contrast, tangible capital (e.g., machines or equipment) has declined. It is estimated that intangible assets represented 35% of the total investments made across the U.S. and 10 prominent European economies in 2019. This figure is a substantial increase from the 26% recorded in 1995. Projections indicate a sustained trajectory, with intangible assets anticipated to reach 45% of the aggregate investment portfolio by 2030 (Hazan et al., 2021).

In this thesis, we argue that as the intensity of intangible assets increases, this deepens the implications of intangibles' economic properties, following the 4 S's approach of scalability, sunkness, synergies, and spillovers described in Haskel and Westlake (2018). The link is that high intangibility increases firms' likelihood to experience increasing returns of scale while simultaneously reducing their elasticity of employment demand when there are productivity shocks -in alignment with Decker et al. (2016c)'s business dynamics work.

Let's show an example. Imagine a manufacturing firm that develops a new way of organizing its production line, which can be considered as becoming more intensive in an intangible asset: organizational processes. The process can be efficiently replicated or expanded as the firm grows without significant cost increases (high scalability). It can also be firm-specific, implying potentially more challenges for competitors to reproduce it (high sunkness). It can also lead to increased efficiency in other areas of the firm learning from it, such as logistics (high spillovers). Finally, the process may synergize well with other intangibles or resources within the firm, enhancing decision-making and improving efficiency.

Thus, a more intangible manufacturing firm can experience less downtime, improved quality control, and reduced costs, boosting its sales. Simultaneously, more intangibles might not require hiring more workers, triggering lower business dynamics behavior. However, the same properties may alienate other firms from these benefits. Either because they cannot invest in such novel processes or cannot create them. Thus, as superstars can become more intangible, they may exhibit higher markups without being anti-competitive. Meanwhile, competitors stay behind, and since they are more, in the aggregate, overall dynamics decline.

Although it is hard to test all the parts of this storyline, with each chapter of this thesis, I aim to contribute to addressing some critical aspects of each.

In the first chapter, ***Business Dynamics and Intangible Investments***, I create an original dataset of businesses registered in the U.K. to show the negative and significant correlation between intangible investment intensity and business dynamics at the industry level. The chapter starts by characterizing the firms (e.g., by age-size-industry cell) contributing the most to the business dynamics decline. For this, I apply the well-known OECD-DynEmp data harmonization software routine that permits comparing countries' statistics on business dynamics, an exercise not done for the U.K. until now[5].

Later, the chapter introduces dynamic panel regressions, demonstrating that such groups of firms (e.g., young[6], small[7], high-growth[8], high-tech[9] firms versus mature, large firms) are influenced differently by intangibles and that given their relative importance, the aggregate effect can be damaging on dynamics. The chapter concludes with a model demonstrating that the more intangible a firm is, and depending on the demographics to which the firm belongs, the less responsive it becomes to hire or fire workers when an unexpected shock to its productivity occurs (i.e., a new organizational process), impacting further business dynamics.

The second chapter, ***Market Power and Technology***, questions the consensus methodology proposed by De Loecker et al. (2020) and shows how aggregate markups are not increasing as much as the authors claim. In this co-authored piece, we contrast De Loecker's estimates for the U.S. and Europe (including the U.K.) with an original application of Forlani et al. (2023)'s method, which, in contrast to De Loecker's, accounts for extra omitted firm heterogeneities (e.g., product demand). Furthermore, we contribute to introducing the feature of time-changing returns to scale.

The findings support biases in estimated markups with the De Loecker and co-authors' estimation method. More interestingly, the results support that the industry sectors that exhibit a rise in markups, like the high-tech sector, are more likely due to them experiencing higher returns to scale over time.

Without proper measures of intangibles, it is harder to argue that these investments are directly triggering low business dynamics or high markups in specific sectors. Thus, the third and final chapter, ***Twitnovation and Aboutnovation: Can social media explain firm performance?*** Aims to contribute directly to this empirical problem of capturing the elusive-to-measure but key firms' performance drivers: innovation and intangible investments. Although substantial macro literature estimates intangible assets by capitalizing expenses from national accounts (Corrado et al., 2005), measures at the firm level remain a challenge.

In this co-authored piece, using an original collection of Twitter posts from and about U.K. firms and other countries, we formulate original text-based, firm-level indicators that constitute proxies of these concepts. We found a positive and significant correlation, especially between the indicators using "about" tweets and labor productivity [10] and investments in marketing and advertising[11].

The findings suggest that we can capture these crucial drivers of firms' success and that non-conventional data sources may offer cheaper and more immediate alternatives than innovation and management surveys. This method could enable connecting firm-level business dynamics and markup data with intangibles' data and, therefore, more directly exploring their links in future analyses.

1 Business Dynamics and Intangible Investments

1.1 Introduction

Schumpeterian endogenous growth models suggest that business dynamics, or more explicitly, reallocating inputs and resources from less to more productive businesses, facilitates "creative destruction" and leads to economic growth and innovation. Over the past few decades, however, the recent abundance of firm-level panel data has allowed researchers all across the globe to make puzzling discoveries. Findings suggest that the vital link between business dynamics and productivity has weakened.

First, business dynamism has been slowing down in similar circumstances, regardless of economy size or level of self-reliance and regulations burdens. During the last 20 to 40 years, from Belgium to the U.S., or OECD countries more generally, have been experiencing sluggish decline, significantly steeper after the dot com bubble burst in the 2000s. See Decker et al. (2016c); Bijnens and Konings (2018b, 2017, 2018a); Calvino and Criscuolo (2019). Interestingly, and despite all the technological progress of the last few decades, the firms most affected tend to be the same across studies: the number of small, young, high-growth firms, especially from the high-tech sector, has declined. High-growth firms are more extensive and older on average now.

Second, despite declining dynamism, productivity disparities across producers of the same good or within narrowly defined industries are persistent or even increasing over time, as revealed by increasing within-industry productivity dispersion measures[12] (Andrews et al., 2016a; Akcigit and Ates, 2019; Decker et al., 2020). However, under the creative destruction premise, one would expect that less dynamism would mean less productivity dispersion; after all, with less dynamism, there is less reallocation and less transfer toward more productive business (See Andrews et al. (2016a); Akcigit and Ates (2019); Decker et al. (2020)).

The explanations for this phenomenon are multiple[13], but many of them circle the same premise. That "something" triggered the "misallocation" of productivity-enhancing production inputs like labor, so firms are not motivated or face obstacles to expand or contract. In this chapter, I dig into one of the explanations associated with input reallocation, which most likely can be linked with this thesis's central premise: that rising intangible investments are one of the primary triggers of slowing business dynamics.

Decker et al. (2019) suggests that what is happening is an emerging weakness in the covariance between business dynamics and productivity that arises from a lack of responsiveness of firm size to productivity shocks as it used to. Building on this, the work of Hopenhayn (1992) on firm dynamics and of Bartelsman et al. (2013) on distortions in

the allocation of resources across firms, I suggest that weakness in covariance and the lack of responsiveness emerge from intangibles because these assets may trigger frictions in labor reallocation that stall or prevent companies from adjusting via employment when a productivity shock occurs.

More concretely, I propose a model that illustrates how, in equilibrium, intangibles facilitate maintaining simultaneously high within-industry productivity dispersion and low employment growth. The model also permits analytical and empirical (using the U.K. case) examination of how intangibles may affect firms depending on their demographics, for instance, how it affects the high-growth firms[14] different from those at the bottom of the size employment growth distribution. It also allows an understanding of the allocation of resources across existing firms, the selection of firms staying in the market, and the degree of churning. Ultimately, the model illustrates how the affectation of specific groups of firms and their relative importance in the economy explains aggregated trends in business dynamics.

This chapter contributes to the literature in various ways. First, it addresses some empirical limitations that have constrained researchers' ability to test intangibles as the reason for the weakened covariance. Measuring intangibles is complex, and matching it to business dynamics data to explore their correlation is even more challenging. In this chapter, I match intangible investment data with business dynamics at the industry level and pave the way for firm-level matching given the structure of the original dataset created, leaving the ground ready for future research. The dataset leverages the progress made in intangible measurements at the industry sector level from the INTAN project[15]. The dataset built on business dynamics indicators can be connected via the firms' unique identifier with innovation and intangible firm-level surveys in the U.K.

Second, this study contributes to the empirical efforts to understand business dynamics globally and in key developed economies still lacking evidence. Previous U.K. studies rely on cross-sectional analysis, describe dynamics or are not updated(Anyadike-Danes et al., 2009; Mason et al., 2016). Yet, for the first time in over a decade for the U.K., this study applies a legitimated statistical routine imperative for high-quality cross-country analyses, such as the OECD's DynEmp3 to compute business dynamics indicators. DynEmp is a software routine designed by the OECD that produces a consistent framework and standardized analyses of business dynamics across countries[16] (Calvino et al., 2018). The software routines require a longitudinal database of business registry data.

The benefits of this application are twofold. I created a longitudinal database from scratch for the U.K. using the cross-sectional versions of the yearly renditions of the U.K. Business Structure Database (BSD)[17]. The effort resulted in a checklist of assumptions that could be applied for future reconstructions of the longitudinal version and a ready-to-

use panel dataset of firms. Subsequently, the study promotes high-quality cross-country comparisons by adding a prominent economy such as the U.K. Although the ideal would be that national statistical offices take advantage quickly and officially of the increasing amount of available micro-data, limited resources, and scope often preclude valuable data from being delivered on time, hampering researchers' capability of uncovering urgent issues.

Finally, this study contributes to still limited within-cell analysis, another issue that adds to the lack of recent evidence for the U.K. Recent findings showed that it is essential to make within-industry, within-firm size, or firm-age class analysis to understand wide-economy trends. Studies abound for the U.S., but for the U.K., are limited. Independent or government official reports are merely descriptive, and often trends lack within-class analysis depth (See, for example, Prashar and Hart (2019), from the U.K. Enterprise Research Center (ERC) or the Office of National Statistics official annual reports on business dynamics such as Office for National Statistics (2018)). Furthermore, this study makes it possible to connect intra-group analysis more directly with intangible investments, adding to the business dynamics-intangibles link research strand, still in its infancy for the U.K. (See UK Statistics Authority (2020)).

The chapter faces some caveats, however. Still, although the built business dynamics database is ready to connect with firm-level data on intangibles in future studies, the present one is limited to the industry level. That is why at least exploiting the intra-cell analysis in the chapter is imperative. Furthermore, the assumptions to build the database imply careful interpretations of indicators of business dynamics towards the end of the sample period, primarily related to firms' exit. Although the estimated trends generally match the U.K.'s statistical offices' reporting, it is helpful to be aware of the implications of potential differences.

The chapter is organized as follows. After presenting essential definitions to understand what business dynamics indicators are used throughout the chapter (sections 1.2 and 1.3) and a recap of the available literature (section 1.4), I characterize the firms (e.g., by age-size-industry cell) contributing the most to the business dynamics decline. For this, I apply the OECD-DynEmp data harmonization software routine and match the resulting indicators with the INTAN dataset of intangibles at the industry level (section 1.5). Later, the chapter introduces dynamic panel regressions, demonstrating that such groups of firms (e.g., young[18], small[19], high-growth[20], high-tech[21] firms versus mature, large firms) are influenced differently by intangibles and that given their relative importance, the aggregate effect can be damaging on dynamics (section 1.6). Finally, the chapter introduces the model demonstrating that the more intangible a firm is, the less responsive it turns to hire or fire workers when an unexpected shock to its productivity occurs (i.e., a new organizational process), impacting further business dynamics (section 1.7). The

final section concludes (section 1.8).

The main findings reveal that in the U.K., Business Dynamics also declined as in other developed economies. Still, in contrast, there was a recovery after the financial crisis and a new slowdown after the Brexit vote. Yet, more importantly, I find that rising intangible intensity may be associated with a closing gap in the employment growth rate between the high-growth and bottom firms[22]. Also, fewer small and young firms are growing at high growth rates, declining labor reallocation, fewer startups, and more companies are reducing their size and getting older.

1.2 Business Dynamics: Definitions

1.2.1 Business Dynamics Indicators

The literature on business dynamics commonly focuses on observing three fundamental indicators: firm entry, exit, and job reallocation, i.e., the simultaneous job creation and job destruction of firms. I focus on both the absolute values of births and deaths of firms and the number of jobs (employees) created (hired) or destroyed (fired) in each period and on their rates. The latter informs on the relative behavior of current stocks.

1.2.2 Unit of analysis of the firm

For this study, I only focus on one unit of analysis: the firm. I use this term interchangeably with enterprise, company, and business. Following the definition of a firm as considered in the primary data source of this study, a firm is a unit that carries out one or more business activities (the production of goods or services) through one or more local units or "plants"—for example, a retail outlet, an accounting office, or a factory that produces finished goods. The plant or local unit is the source of the business activity. Suppose the enterprise is only located in one site. In that case, the enterprise is the local unit Office for National Statistics (2017)[23]. Later in section 4, I describe the different units of analysis available in the primary dataset used in this study and how the unit of the firm is defined and fits in the overall structure.

1.2.3 Level of aggregation for analysis

The DynEmp3 program obtains different levels of analysis by class. Let's call c a specific class or group of aggregation at which any business dynamics indicators can be estimated. Subscript c can be a single class or a combination of classes. In this study, we focus on the following classes: firms' demographic status group g (entering, incumbent, or exiting[24]), firms' age class ag, firms' size class sz, firms' percentile class p, firms' industry s and for all UK firms in the whole economy UK. The equation below sets out the distinct levels of disaggregation used in this study.

$$c = \begin{cases} \text{whole economy } (UK) = \text{ all firms in the economy} \\ \text{industry sector } (s) \\ \text{firm's industry sector } (s) * \text{ firm size class } (sz) \\ \text{firm's industry sector } (s) * \text{ firm age class } (ag) \\ \text{firm's industry sector } (s) * \text{ firm size class (sz)} * \text{ firm age class } (ag) \\ \text{firm's percentile class } (p) \\ \text{firm's industry sector } (s) * \text{ firm's percentile class } (p) \\ \text{firm's demographic status } (g) \end{cases} \tag{1}$$

For example, the following sums all employment at year t of any firm i belonging to industry sector c to estimate the total employment level at year t for industry sector c in the U.K. as in 1.2.3. This includes all employment added to that year:

$$E_{c,t} = \sum_{i \in c} \Delta^{+} E_{i,t} \tag{2}$$

Where $E_{c,t}$ is the number of employees in cell c in year t. With $\Delta^{+} E_{i,t} = 1_{(E_{i,t} - E_{i,t-1}) > 0} . (E_{i,t} - E_{i,t-1})$, where $1_{(E_{i,t} - E_{i,t-1}) > 0}$ is a dummy variable equal to one when the firm level change in employment is positive and 0 otherwise.

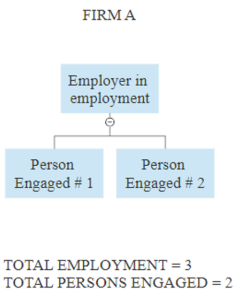
1.2.3.1 Self-employment and persons engaged A firm can be composed of one or more people. I follow the EU KLEMS[25] jargon to distinguish between these two types of firms. As depicted in Figure 1, firm B is defined as a one-person or self-employment firm whose employer (proprietor) is in employment. No persons are engaged, so the total number of people in employment for this firm is 1. Firm A is a firm whose employer is in employment along with another two persons; these latter two persons are referred to as persons engaged. Thus, the total employment for firm A is 3, corresponding to three people in employment. Most of the figures presented in this study are estimated, including all firms, meaning that firms of type A and B are included. When all types of firms are considered simultaneously, meaning types A and B together, the group is tagged as "all employees in employment". In cases where only type A firms are considered, the tag is "excluding employers in employment". It refers to firms with one or more persons engaged.

Available datasets and official UK employment publications consider different definitions of self-employment. See Table 1 for a brief comparison between this present study's definitions and the ones used in those other publications. Notice that the title of "employer in employment", although not an official name to call these types of situations,

is aligned with terminology commonly used in employment in employment law and tribunals where the liability of one employer over another employer is addressed, depending on how much influence has one over the other. [26]

FIGURE 1
Firm categories

(a) Case A



(b) Case B

See the explanatory section in Table 1 for observing the different definitions of employment used in the BSD, the IDBR, and the ONS in business demographics publications.

TABLE 1
Definitions

Source / Definitions	Employment	Employees
EU KLEMS	Total employment = Employers in employment (or self-employed) + persons engaged	Total number of employees = Employers in employment (or self-employed) + persons engaged
Business Structure Database	Total employment = Employees + working proprietors (as defined for the IDBR).	Total number of employees = Employees It counts each employee as one unit irrespective of the hours worked within the business (ILO definition)
UK Business Demography (ONS Publication)	Total employment = Employees + the number of working owners (also called, owner-directors) who receive drawings or a share of the profits.	Total number of employees = Employees An employee is defined as anyone (different to the owner-director) aged 16 years or over who is paid directly from the payroll, in return for carrying out a full-time or part-time job or being on a training scheme. Full-time is defined as working more than 30 hours per week and part-time is defined as working 30 hours or less per week.
DynEmp3 (OECD's business demography program)	Total employment = Total number of persons engaged = owner + employees (PERSENG option).	Total number of employees = Persons engaged One employee means one unit irrespective of the hours worked within the business. Also, called headcount.
This study	**All employees in employment = Employer in employment (or self-employed) + persons engaged**	**Excluding employers in employment = Employees**

Note:
Sources: Office for National Statistics 2006
Office of National Statistics 2019; Calvino, Criscuolo, and Menon 2018

1.2.3.2 Entering, exiting and incumbent firms In this paper, I base the analysis on three main groups of firms in a given period t, entering firms (I_t^{ent}), exiting firms (I_t^{ext}) and incumbent firms (I_t^{incb}). A firm is an entrant in t if it is not present in the microdata in year $t-1$ but is present in t. Each entrant is a birth. An exiting firm is a unit not present in t but in $t-1$. Each exiting firm is a death, and the death year is t. An incumbent firm is a unit present in $t-1$ and t. A firm's demographic status or group refers to the demographic event experienced at a particular time or reference period. For instance, the same firm can be an entering firm in the year 2012, can be an incumbent firm in the year 2014, and an exiting firm in the year 2016. A firm cannot belong to more than one group in the same reference period. Positive employment means that at least one person, employer or other people engaged are in employment in the firm i at time t or the reference period of interest. I^{ent}, I^{ext}, and I^{incb} can be firms with or without persons engaged at a point in time t; thereby, firms of type A or the type B as explained in the preceding subsection can belong to any of these three groups. I define

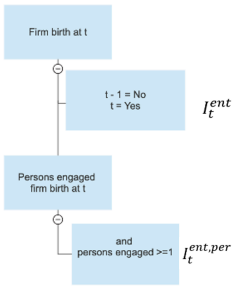
the following subgroups to identify the subset of firms of type A in each of these groups, as illustrated in Figure 2.

$I_t^{ent,per}$ as the number of entering firms excluding those with only employers in employment. To be part of this subset of entering firms, the firm must comply with the entry condition as presented in the above paragraph and enter year t with at least one person engaged. This definition is compatible with the "employer birth" definition used by the ONS Business Demographics publication. An employer birth is considered when the enterprise is not present in $t-1$ or $t-2$, but once it enters the register at year t, it does so with at least one person engaged. $I_t^{\text{ext, per}}$ as the number of exiting firms excluding those with employers in employment. To be part of this subset of exiting firms, the firm must comply with the condition of exit presented earlier and exit year t with at least one person engaged; the death year has to have at least one person engaged besides the employer.

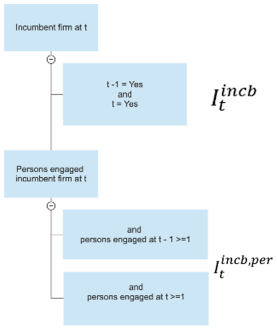
$I_t^{\text{incb, per}}$ as the number of incumbent firms excluding those with employers in employment. To be part of this subset of incumbent firms, the firm has to comply with the conditions of incumbents as presented earlier and has to have had at least one person engaged in years $t-1$ and t.

FIGURE 2
Dynamics

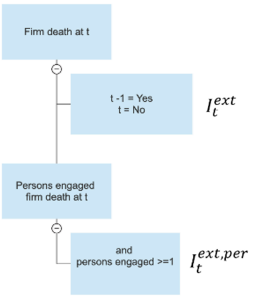
(a) Birth or entering of a firm



(b) Incumbent firm



(c) Death or exiting firm



1.2.4 Job creation and destruction

Gross job creation (JC) is defined as the sum of all positive firm-level job variations between $t-1$ and t in cell c:

$$JC_{c,t} = \sum_{i \in c} \Delta^{+} E_{i,t} \tag{3}$$

Where $E_{i,t}$ is the number of employees of firm i in year t. With $\Delta^{+} E_{i,t} = 1_{(E_{i,t}-E_{i,t-1})>0}.\,(E_{i,t} - E_{i,t-1})$, where, $1_{(E_{i,t}-E_{i,t-1})>0}$ is a dummy variable equal to one when

the firm level change in employment is positive, and 0 otherwise.

Note that this definition is the sum of firms in a particular cell, with the cell to be defined later (e.g. it might be a country-year-industry cell).

Gross job destruction (JD) is defined as the absolute value of the sum of all negative firm-level job variations between $t-1$ and t in cell c :

$$JD_{c,t} = \left| \sum_{i \in c} \Delta^{-} E_{i,t} \right| \tag{4}$$

With $\Delta^{-} E_{i,t} = 1_{(E_{i,t} - E_{i,t-1}) < 0} . (E_{i,t} - E_{i,t-1})$, where, $1_{(E_{i,t} - E_{i,t-1}) < 0}$ is a dummy variable equal to one when the firm level change in employment is negative, and 0 otherwise. Net job creation (NJC) is defined as:

$$NJC_{c,t} = JC_{c,t} - JD_{ct} \tag{5}$$

I can also express these figures regarding rates to understand their relative performance. The Job Creation Rate (JCR) is the ratio of gross job creation over average employment in that cell in the two years, times 100:

$$JCR_{c,t} = \frac{JC_{c,t}}{\frac{1}{2}(E_{c,t} + E_{c,t-1})} * 100 \tag{6}$$

Where, $E_{c,t} = \sum_{i \in c} E_{i,t}$, the total number of employees of all companies i that belong to the level of aggregation (cell) c at year t. Similarly, the Job Destruction Rate (JDR) is the ratio of gross job destruction over average employment in the two years, times 100:

$$JDR_{c,t} = \frac{JD_{c,t}}{\frac{1}{2}(E_{c,t} + E_{c,t-1})} * 100 \tag{7}$$

Note again that these rate definitions refer to a particular cell.

1.2.5 Job Reallocation

Job Reallocation Rate (JRR) summary states JC and JD rates. Hence, JRR is defined as the sum of gross job creation and gross job destruction rates over the average of total employment in cell c in period t and total employment in period $t-1$:

$$JRR_{c,t} = \frac{JC_{c,t} + JD_{c,t}}{\frac{1}{2}(E_{c,t} + E_{c,t-1})} * 100 \tag{8}$$

The DynEmp3 calculates this rate separately for all firms and only incumbent firms. These distinct levels of partition permit inferring organic reallocation of employment (produced exclusively by incumbent firms) since it can be isolated from the effect produced by the entry and exit of firms, which can create and destroy employment as well. Accordingly, the Job Reallocation Rate of Incumbents $\left(JRR^{incb}\right)$ is defined as the sum of job creation and job destruction rates of incumbents. Notice that this rate does include only the job creation and destruction or changes in employment of incumbents, meaning companies who are not appearing in the sample for the first time. The JRR from equation 6 considers the employment created and destroyed by entrant and exiting firms. Thus, the JRR for incumbents is defined as the job creation and job destruction by only stayers over an average of total employment in cell c between period t and $t-1$:

$$JRR^{incb}{}_{c,t} = \frac{JC^{incb}{}_{c,t} + JD^{incbit}{}_{c,t}}{\frac{1}{2}\left(R^{incb}c_{c,t} + E^{incb}b_{c,t-1}\right)} * 100 \tag{9}$$

1.2.6 Entry and exit rates

The entry rate (ENR) is defined as the proportion of entering firms over the active stock of firms at period t; in other words, the number of entering firms I^{ent} over the total number of incumbent firms I^{incb} and entering firms I^{ent} in a given cell c at year t, times 100 :

$$ENR_{c,t} = \frac{I^{\text{ent}}{}_{c,t}}{I^{\text{incb}}{}_{c,t} + I^{\text{ent}}{}_{c,t}} * 100 \tag{10}$$

Where I is the total number of firms i with none or more people engaged in the firm in cell c at time t, the denominator is the number of active firms in t. Conversely, exit rate (EXR) is defined as the proportion of exiting firms over the active stock of firms until t-1 or the number of exiting firms I^{ext} over the total number of incumbent firms I^{incb} and exiting firms I^{ext} in a given cell c at year t, times 100. The denominator is the number of active firms in $t-1$:

$$EXR_{c,t} = \frac{I^{ext}{}_{c,t}}{I^{\text{incb}}{}_{c,t-1} + I^{\text{ext}}{}_{c,t-1}} * 100 \tag{11}$$

The definition of growth is based on Bartelsman et al. (2005). The entry rate in terms of employment ($ENRE$) is defined as the proportion of employment created by entering firms over the active stock of employment at period t; in other words, total employment of entering firms E^{ent} over the total stock of employment of incumbent firms E^{incb} and entering firms E^{ent} in a given cell c at year t, times 100 :

$$ENRE_{c,t} = \frac{E^{ent}{}_{c,t}}{E^{\text{incb}}{}_{c,t} + E^{ent}{}_{c,t}} * 100 \tag{12}$$

Where E is the total employment for a particular group of firms i (ent, ext, or incb) with none or more people engaged in the firm in cell c at time t. Conversely, the exit rate based on employment ($EXRE$) is defined as the proportion of total employment destroyed by exiting firms over the active stock of employment until t-1 or which is the same, the total employment destroyed by exiting firms E^{ext} over the total employment of incumbent firms E^{incb} and total employment of exiting firms E^{ext} in a given cell c at year t, times 100 :

$$EXRE_{c,t} = \frac{E^{ext}{}_{c,t}}{E^{\text{incb}}{}_{c,t} + E^{ext}{}_{c,t}} * 100 \tag{13}$$

Exiting firms at time t should have 0 employment at year t. However, to compute this rate DynEmp3 assumes the value of E for exiting firms for the year t as equal to the value of gross destruction of employment of firms between t $-$ 1 and t. In other words, it accounts for the employment that existed in t-1 but that was destroyed in t. For this reason, EXRE can be larger than ENRE.

1.2.7 Dispersion and skewness

To explore growth rates over time, the literature on business dynamics typically analyses the dispersion and skewness of the distribution of growth rates in terms of employment or turnover. Common practice is to study the dispersion differentials between the 2 tails of the distribution; the 90th and the 10th percentiles of the distribution of growth rates. To explain the underlying reasons for changes in overall dispersion and its 90-10 differential, researchers study the skewness of the distribution by estimating the differentials of, say employment or turnover growth rates, between the top percentile (90th) and the median (50th) and, this against the top bottom percentile (10th). Job reallocation indicators, although more about the churn rate of employment over time, is an alternative measure of business volatility but are highly correlated with the 90th-10th dispersion measure, so both indicators are frequently presented simultaneously.

To estimate this differential, the DynEmp3 program recreates the following steps. First, it estimates the employment growth rate between t and t-1 for each firm i in the sample, $e_{i,t}$. Second, it ranks each employment growth rate $e_{i,t}$ from the lowest to the highest and classifies each firm's employment growth rate in five intervals of growth distribution. Firms in the bottom 10% of the distribution of employment growth rate (p=1), firms between the 11th and 25th percentile (p=2), firms between the 26th and 50th percentile (p=3), firms between the 51st and 75th percentile (p=2), firms between the 76th and

90th percentile (p=2), and firms in the top 10% of the distribution (p=6). Third, it estimates the weighted-average employment growth rate $\overline{e}_{c,t}$ by cell class c at time t, where $c =$ (firm's industry sector $s *$ firm's percentile class p). Thus, for each aggregation level c, we have the weighted average employment growth rate in the form:

$$c = \{ \text{ firm's industry sector } (s) * \text{ firm's percentile class } (p)$$

Thus, for each aggregation level c, we have the weighted average employment growth rate in the form:

$$\overline{e}_{c,t} = \sum_{i \in c} e_{i,t} w_{i,t} \times 100 \tag{14}$$

Where $e_{i,t}$ is the employment growth rate of firm i defined as $e_{i,t} = \frac{E_{i,t} - E_{i,t-1}}{\frac{1}{2}(E_{i,t} + E_{i,t-1})}$) and $w_{i,t}$ is the weight defined as $w_{i,t} = \frac{E_{i,t}}{\sum_{i \in c} E_{i,t}}$. Fourth, the program estimates the **dispersion differentials** in the form:

$$D_{a,b} = \left(\min_{x \in S} \overline{e}_{a,t} - \min_{x \in S} \overline{e}_{b,t} \right) \times W \tag{15}$$

Where $D_{a,b}$ is the employment-weighted dispersion differential for sector s (or for the whole economy), the difference between the employment growth rate of percentile group $p = a$ and the employment growth rate of percentile group $p = b$ in the sector, with the condition that $a > b$. W is a weight used to adjust aggregate employment differentials between the original database sector classification (UK SIC 1992/2003/2007) and the OECD's DynEmp3 A38 industry classification (See Table D.11 in Appendix D. Later in the document, we focus on weighted dispersion differentials.

1.2.8 High-Growth Firms

A dynamic economy depends not only on a more significant number of new entrants but also on the chances of growth and expansion that such new entrants initially have and develop over time. High-growth firms' research focuses on describing the level of prevalence of these firms in the economy over time (Bijnens and Konings, 2018b; Hart, 2008; Buss, 2002; Decker et al., 2016c, 2017; Aghion and Akcigit, 2015; Acemoglu et al., 2018; Akcigit and Ates, 2018, 2019). High-growth firms are important for various reasons. One of them, prevalent in the literature, is their role in job creation. Only in the U.S., do they account for up to 50 percent of gross job creation. Most entering start-ups exit within the first five and ten years and a substantial fraction of those who survive remain small. However, the few young firms that survive reach very high growth levels and even

make up for all the shrinking and job losses that the losers from their cohort created (Decker et al., 2016c; Bijnens and Konings, 2017; Haltiwanger et al., 2013). High-growth firms are generally disproportionally young; this is the case for countries such as the U.S. and Belgium, despite the difference in economies' size.

Another important subject in the literature on High-Growth Firms in business dynamics is related to the heterogeneity of entrepreneurship and its contribution to aggregated productivity growth. Based on the distinction between "transformational" versus "subsistence" or "lifestyle" entrepreneurship made by studies such as (Schoar, 2010), high-growth firms are argued to emerge from the former group. These firms reach these levels of growth because they are created with the intent for innovation and growth, to make a major product or process innovation (see Acemoglu et al., 2013 and Hurst and Pugsley, 2012, 2014 cited in Decker et al. (2016c)). Furthermore, although the contribution of HGFs in the short term to overall employment is small, these firms attract much policy interest because they tend to produce spillover effects, as in the case of fast-growing technology-based small businesses. Organisation for Economics Co-operation and Development (2020). In contrast, subsistence entrepreneurs create their businesses out of necessity, contributing less to major outcomes in innovation and productivity growth (Shane, 2010). Hence, the evolution of HGFs in countries is an important indicator of healthy business dynamics.

To study this phenomenon, the OECD defines a high-growth firm as an incumbent or exiting firm with over ten employees and with an annual employment growth rate above 20 percent over three consecutive years:

$$HGF_{i,t} = \begin{cases} 1, & \left(\sqrt[3]{\frac{E_{i,t+3}}{E_{i,t}}}\right) - 1 > 0.20; E_{i,t} \geq 10 \\ 0, & \text{otherwise} \end{cases}$$

Where $E_{i,t+3}$ is the number of employees of firm $i = \{$ ext, incb $\}, 3$ years after current year t. The literature has multiple definitions for what a High-Growth Firm (HGF) is. The measure is considering only firms with ten or more employees because it is susceptible to the initial size of the firm, particularly when measuring high growth by employment Bravo-Biosca (2016a). Just doubling a unipersonal firm will constitute a growth rate of 100 percent, while if a 100-employee firm increases another employee, the rate is much lower. However, what is relatively common amongst them is that there is a property for these firms to grow faster relative to peers and in shorter periods in relevant performance variables. The most common are employment or turnover. It is also common to refer to these firms as the fastest growers in the distribution of the chosen performance variable. For instance, firms in the 90^{th} percentile of the employment growth distribution.

1.2.9 Definitions of firms' demographic groups

Throughout the chapter, I refer to different types of firms, meaning they belong to particular demographics of the firms' population.

A young firm is considered less than or equal to 6 years old, following OECD definitions in DynEmp. A mature firm is older than 6 years. A small firm has 249 employees or less, following OECD definitions in DynEmp. A large firm refers to a firm with 250 employees or more.

1.3 Intangible Investments: Definitions

Intangible capital or intangible investments refer to non-physical assets such as software, databases, research, marketing and advertising, intellectual property, and organizational processes (Haskel, 2020), with distinct economic properties to tangible capital, like machines or equipment (Haskel and Westlake, 2018). We define investment intensity as any of the four specifications below:

$$\text{Intangible intensity} = \begin{cases} ShIntan^{x}_{i,t} & = \frac{P^{I}Iintan_{i,t}}{P^{v}V_{i,t}} \\ ShTan^{x}_{i,t} & = \frac{P^{I}Itan_{i,t}}{P^{v}V_{i,t}} \\ ShIntan^{y}_{i,t} & = \frac{P^{I}Iintan_{i,t}}{P^{I}Iintan_{i,t}+P^{I}Itan_{i,t}} \\ ShTan^{y}_{i,t} & = \frac{P^{I}Itan_{i,t}}{P^{I}Iintan_{i,t}+P^{I}Itan_{i,t}} \end{cases} \tag{16}$$

Where $ShIntan^{x}_{i,t}$ is the share of nominal intangible investments over the gross nominal value added of the economy for the industry sector i and year t, and $ShIntan^{y}_{i,t}$ is the share of nominal intangible investments over the total nominal value of the investment, namely nominal investments in intangible plus the value of nominal tangible investments, similarly, for tangible investments with $ShTan^{x}_{i,t}$, and $ShTan^{y}_{i,t}$.

1.4 Conceptual Framework

1.4.1 Main trends of business dynamics indicators

Before the technology revolution of the 2000s, the relationship expected between Business Dynamics and productivity was positive. Under the creative destruction logic, the greater (lower) the dynamics, the greater (lower) the productivity. However, as Decker et al. (2016b) shows for the U.S., lower dynamism did not necessarily indicate lower productivity at some point in history. On the contrary, lower dynamism signaled greater efficiencies that made inessential the entrance of start-ups or the exit of incumbent companies to improve aggregate productivity. This was evident between the years 1980 and 2000 when the retail trade and services sectors led the decline in indicators of business

dynamism in the U.S. These sectors shifted from single-unit establishment firms ("Mom and Pop" businesses) to large, national chains, thereby with more efficient networks and processes which led to improved industry-level productivity. This transformation was relatively benign, and lower dynamics just reflected the shift in the sector's business model Decker et al. (2016a).

However, the U.S. High-Tech sector (ICT-producing and intensive ICT-using sectors) did not experience the same patterns. Despite exhibiting high dynamism between 1990 and 2000 and contributing the most to the U.S. improvements in aggregate productivity (see Fernald, 2014 cited in Decker et al. (2016b)), the sector displayed the most significant declines in dynamics after the year 2000, dragging industry-level productivity.

This "productivity puzzle" stresses that despite significant technological advances, the high-tech sector has slowed down, and productivity has stagnated. The evidence above suggests that the economy-wide trends masked variations across sectors. Consequently, researchers have turned to conducting within-cell analyses to detangle the underlying factors of the decline and shed light on changes to what was supposed to be a strong business dynamics-productivity link.

Decker et al. (2016c) shows that when various compositional effects are taken into account simultaneously (e.g., firm size class, firm age class, industry sector, multi-unit status, the transformation of an economy from manufacturing to a services-based), they can only explain about 15% of the reduction in job reallocation in the U.S. The contribution is low because the effects offset each other. For instance, the shift towards more mature firms accounts for 26% of the decrease in job flows[27], whereas there is an increase of 13% due to the shift towards more volatile industry sectors[28].

Evidence from the U.S., Belgium, and other studies suggests that three particular cell effects drive the whole economy's trend in B.D.: age, industry sector, and percentile class. In particular, the declines in dynamics are more significant for young, High-Tech, and High-Growth firms. Historically, young firms have had a significant share of activity in any given year (Decker et al., 2016c), and make a long-lasting contribution to aggregate employment despite high failure rates (Haltiwanger et al., 2013; Bravo-Biosca et al., 2016). However, the share of employment in young firms has declined Decker et al. (2016c); Bijnens and Konings (2018b). Furthermore, the type of young firms declining the most are those with high-growth potential (Decker et al., 2016c). Firms are likelier to be "high-type" or have higher innovative capacity when young rather than later in their life cycle Acemoglu et al. (2018). These firms are also referred to as "transformational" entrepreneurs, in contrast to "subsistence" or "lifestyle" entrepreneurs (Schoar, 2010). The former is conceived as young firms to make significant innovations that will yield over time to a more prominent firm that generates income and substantial employment

fewer (Decker et al., 2016c). This is evident with the decline in the 90th percentile of the young firm-growth rate distribution in the U.S. (Decker et al., 2016a).

The prevalence of High-growth young firms is vital because they have disproportionately accounted for job creation and productivity growth in the past Bravo-Biosca et al. (2016); Bravo-Biosca (2016b). Furthermore, the sector that exhibits greater innovation capacity and subsequent significant impacts on aggregate productivity is the high-tech sector; at least, it has been fundamental for developed countries such as the U.S. The issue is high-tech has experienced the most significant decline in overall dynamics and high-growth firms' prevalence Calvino and Criscuolo (2019).

Transformational entrepreneurs with higher innovative capacity are likelier to integrate the most innovative and dynamic sectors. Thus, they exhibit a greater prevalence of high-growth firms. The high-tech sector has been shown, in particular, to be likelier to be composed of transformational entrepreneurs and, subsequently, to have a more considerable prevalence of young and high-growth firms (Acemoglu et al. (2013) cited in Decker et al. (2016c); Acemoglu et al. (2018)).

Because of the above, the young firms' group and the high-tech sector tend to exhibit more significant skewness in their growth rate distributions than other age and sector classes that are more mature and have less innovative or transformational power. Skewness, as described in the definitions section, refers to the difference between the employment growth rates of the firms in the top 10 percentile of the distribution and those from the firms in the median, called the 90-10 differential.

Researchers assessing the case of the U.S., Belgium, and OECD countries have found that the retail and trade sectors seem to have led to a decline in the whole's economy business dynamics before the 2000s (Bijnens and Konings, 2017; Decker et al., 2016b). They were the most significant contributors to the decline in the dispersion of growth rates. Since they were naturally not skewed due to their low transformational capacity, their decline impacted the whole's economy skewness very little. Thus, the economies' decrease in dispersion was not accompanied by a significant decrease in overall skewness. However, after the 2000s, a naturally highly-skewed sector, such as the high-tech sector, experienced decreased dispersion, which resulted in an overall decline in the economy's growth rate distribution skewness (Goldschlag and Miranda, 2020; Decker et al., 2016c). The 90-50 differential decreased more abruptly after the 2000s for young and high-tech firms, which resulted in a decline in the average growth rates of firms over time and within the overall prevalence of high-growth firms (Goldschlag and Miranda, 2020; Decker et al., 2016c).

1.4.2 Declining dynamics: Are intangibles the reason?

Although developed countries have found common ground, finding similar business dynamics trends (as described in section 1.4.1), the plethora of potential explanations and the lack of evidence have prevented experts from reaching a consensus on a reason. Examining in detail or testing all those potential alternatives is beyond the scope of this study. However, in this section, I review the evidence relevant to justifying the business dynamic trends with rising intangible investment.

I argue that the recent increase in intangible investment intensity can be one of the reasons behind the weakening of the relationship between business dynamics and productivity. More specifically, following Decker et al. (2019) and prior work, I argue that such a link has become weaker due to the declining responsiveness of firms via employment when a productivity shock occurs. More importantly, my angle is that such declined responsiveness is due to frictions to employment adjustments, potentially triggered by the increasing intangible intensity of firms. Likewise, I suggest that lower dynamics can signal alternative, non-employment expansion or contraction means—for example, more acquisition or production of intangible capital instead of hiring more workers.

Decker et al. (2019) argues that a potential explanation for the U.S. case is a decreased responsiveness to productivity shocks after 2000. In alignment with Hopenhayn (1992) and Hopenhayn and Rogerson (1993), the authors suggest that firms hire (fire) employees in response to higher (lower) productivity, and in the absence of adjustment costs, productivity shocks immediately alter the firm size. The shock changes the potential relative payoff between profits, entry, and exit costs, giving solid incentives to grow or contract. Consequently, the firm's growth depends on the dispersion and persistence of productivity shocks.

However, the authors show that despite increasing productivity volatility and dispersion after 2000 in the whole U.S. economy, especially in its high-tech sector[29], start-up rates and job reallocation declined. Thus, shocks kept occurring but without evident impact on hiring or firing. Yet, their Hopenhayn-style model suggests that firms can become unresponsive when introducing adjustment costs. Firms may decide not to destroy jobs in the face of lower expected productivity or create jobs in the face of higher expected productivity. Thus, low-productivity firms are less likely to exit the market, and high-productivity firms are less likely to expand.

The reasons for the reduced responsiveness in itself are also multiple. Some authors say businesses might expand or contract by transitioning from national to overseas operations or via outsourcing and globalization, which has naturally transformed how productive firms expand (Decker, Haltiwanger, Jarmin, and Miranda, 2016a). Also, policies that reduce incentives for job readjustments, such as unlawful discharge regulation (Davis and

Haltiwanger, 2014); financial frictions such as financial risk (Caballero and Hammour, 2005); or transition from general to specific purpose technologies that reduced demand for companies products (Calvino and Criscuolo, 2019; Aghion and Akcigit, 2015; Byrne, Byrne)[30].

Nonetheless, this section aims to highlight some studies supporting the idea that increased intangible intensity is one credible factor responsible for the decline in responsiveness of firms in terms of employment, mainly because of intangibles' economic properties. I classify how intangibles may alter responsiveness and dynamics in four channels. Although I do not pretend to provide evidence for all of the factors associated with them, they supply a framework analysis for discussion.

First, firms kept being responsive to productivity shocks but via other things except employment, like acquiring more intangible assets. Or indeed responding via employment, but in the opposite direction, substituting workers for intangibles. Second, intangible assets may disincentive market entry or encourage market exit, further altering net job creation. Third, if firms reacted via employment (job reallocation), their responsiveness may have declined because their intangibleness' triggered frictions to contraction or expansion via employment. In particular, such frictions can disincentive hiring certain workers but incentivize them to hire others. Overall, the net impact on employment growth can be harmful depending on what channel is dominant and what type of workers are more prevalent in the economy. Fourth, intangible intensity economies imply slower response in employment, if there is any, due to the "sunkness" of intangible investments.

1.4.2.1 Intangibles replace labor as a means of response. First, intangible-intensive firms may respond flexibly to labor rigidities and respond to shocks via means different from employment. Bartelsman et al. (2016) findings suggest that higher exit costs stifle firms' willingness to grow via employment. For example, firing costs make firms less willing to fire unproductive workers, reducing their capacity to invest in and take advantage of productivity-enhancing assets (e.g., algorithms or data) instead. Intangible-intensive companies who foresee this would avoid enduring labor market restrictions by bypassing growing in size.

The latter related to intangibles' advantages over tangibles for expansion (scaling up) and enhanced productivity without hiring more workers (Decker et al., 2016a). Intangibles exhibit greater *Scalability* since they can be used simultaneously in multiple places. For example, licensed software, a non-rival good, whose value increases with the use of others (network effects). Thus, a firm that faces a positive shock does not need to hire more workers but can decide to buy or build new software in-house (Haskel and Westlake, 2018). This may increase within-plant productivity growth without hiring more workers.

Yet, that would also imply that adjustments via employment may still occur for firms that cannot respond by acquiring more intangibles. For example, firms without the capacity to build in-house or purchase software. Precisely, scalability fosters "winner-takes-all" scenarios, where a few firms can gain substantial market shares due to intangibles' critical differentiators. Companies capable of building, managing, or acquiring intangibles possess higher scalability, enabling them to capture market share without a proportional increase in employment. This phenomenon aligns with the "superstar firms" story exemplified by IT platforms like Amazon towering the online retail market (Akcigit and Ates, 2019; Calvino and Criscuolo, 2019; Bessen, 2017), or Alphabet (Google), whose software investments contribute significantly to market capitalization (USD 354 billion according to Bloomberg in 2020). Phenomena like this enable higher concentration in IT sectors, lower labor share (Autor et al., 2017), and rising mark-ups in digital-intensive sectors (De Loecker et al., 2020, 2018; Calligaris et al., 2018). Nonetheless, only a few of these ideas have been empirically tested (Decker et al., 2017).

1.4.2.2 Intangibles halt market entry. The above also hints at intangibles' property of *synergies*, which amplifies their impact on business dynamics. Synergies arise from the effective combination of ideas. For instance, Walmart's success in the '90s resulted from combining tangible investments in computers with intangible investments in processes and supply chain development (Decker et al., 2016c; Haskel and Westlake, 2018). Those types of models contributed largely to U.S. productivity in the '90s while business dynamics declined. The synergistic capacity of intangibles also enhances intangible firms' value while encouraging market exit and associated employment destruction. For example, Microsoft's ownership of PowerPoint made it more valuable than when held by its original start-up developer, Forethought. Microsoft had more complementary, critical assets, such as more robust marketing or infrastructure; thus, selling it to Microsoft could enhance its value (Cabral, 2018).

Here, the second channel is precisely how intangibles may trigger fewer dynamics. However, Digital technologies and other intangible examples have a double sword since their prevalence in an economy may shift market entry incentives in opposite directions. On the one hand, they may encourage entry, as a single investment can be reused without additional costs; furthermore, it dwindles search costs while increasing quality search, lowering replicability, transportation, and tracking costs, resulting in greater experimentation, upscaling, and productivity-enhancing creative destruction.

On the other hand, increased competition may reduce entry incentives after all. According to the "shadow of Google" concept, young firms may perceive industry giants as substantially powerful competitors, discouraging them from contending and looking to grow just enough to be acquired as early as possible (Decker et al., 2016b; Azoulay

et al., 2020). In turn, gaps between the frontier and laggard firms can exacerbate, increasing market concentration and reducing more innovative startups (Corrado et al., 2019; Andrews et al., 2015). Furthermore, the property of intangibles causing *spillovers* decreases business dynamics or employment growth. This is because spillovers allow firms to benefit from the innovations of others without having to invest in them themselves. This can lead to decreased investment in innovation and slow economic growth (Haskel and Westlake, 2018; Corrado et al., 2017)[31]. Some studies suggest that the net effect of these two opposite outcomes tends to be negative on market entry (Haskel and Westlake, 2018).

Furthermore, in highly intangible industries, startup costs are elevated. For example, biotechnology needs extensive R&D for novel drugs and therapies. Their investments, such as intellectual property and regulatory approvals costs, are elevated. This pushes up the productivity threshold for production decisions and the average productivity of incumbent firms, fostering larger sizes and economic concentration (Eisfeldt and Papanikolaou, 2014; Chiavari and Goraya, 2021). Smaller firms, unable to reach this threshold, may still adapt through employment changes unless significant labor restrictions impact their firm-size category.

Finally, Haskel and Westlake (2018)'s concept of *spillovers* of intangible investments refers to the idea that the benefits of investing in intangible assets can extend beyond the firm investing. Although intangibles' *spillovers* can bring about positive externalities and contribute to innovation and productivity gains, there are scenarios in which these spillovers might have nuanced or even adverse effects on employment growth rates in the aggregate economy. For example, spillovers related to technological advancements, particularly in automation and artificial intelligence, could lead to job displacement for routine or low-skilled tasks. While the highly intangible-intensive firms benefit from increased efficiency, the broader economy might experience reduced demand for certain types of jobs, leading to negative employment growth rates in specific sectors.

Furthermore, if the spillovers are concentrated in specific industries that are highly automated or have low labor intensity, the positive impact on employment growth rates may be limited. The rest of the economy, particularly sectors with a higher reliance on labor, might not experience the same positive spillover effects, potentially leading to an overall negative impact on employment growth rates (Acemoglu and Autor, 2011).

1.4.2.3 Intangible-intensiveness demands more skilled than unskilled labor: complementary and substitution effects The third channel through which firms may have become less responsive via employment to shocks relates to the changing role of workers in a more intangible economy. It transitioned to workers being required to be more skilled and with multi-tasking abilities (Calvino and Criscuolo, 2019). OMahony

(2020) distinguishes intangibles between innovative property (e.g., R&D) and economic competencies (e.g., brands, firm-provided training) and shows that intangibles interact differently with high-skill[32] and low-skill workers. R&D complements labor, whereas brands, firm-provided training, and organizational capital investments substitute labor. This pattern is similar to what other studies have described when the ICT revolution triggered inequality among workers; only high-skilled workers attuned to new tech were more in demand. However, the pattern is less clear for intangibles[33].

Suleyman and Ozkara (2021); Chiavari and Goraya (2021) for instance, divide labor into skilled and unskilled workers to investigate intangible capital's role in labor reallocation. They find that the more intangible firms are (or, the more knowledge-intensive a sector is), they require more skilled labor to complement such capital. In contrast, there is an almost perfect substitution effect of low-skilled workers with more intangible capital in less intangible industries. This implies that intangibles' intensity creates friction to adjust to certain types of labor, affecting overall employment growth. If the prevalence of skilled workers in an economy is not substantial, aggregate employment demand is reduced.

The implications for aggregate employment growth are amplified when high-intangibles-intensive industries demand more skilled labor and attract it easily. In contrast to small firms, larger firms (which also tend to be the most intangible) have an advantage in attracting more skilled workers Suleyman and Ozkara (2021). For example, a large pharmaceutical not only demands workers with doctorates because of its complex laboratories but also beats smaller pharmaceuticals in attracting skilled workers because it can provide exceptional employee training. Again, this, combined with evidence showing that larger firms tend to expand and contract less in employment on average, can result in vapid or null employment growth.

Corrado, Criscuolo, Haskel, Himbert, and Jona-Lasinio (2021) study on OECD countries supports how workers' demand has changed in a more intangible economy. They find a significant and positive correlation between industry-productivity dispersion and intangible intensity. Size alters the relationship between intangible intensity and within-industry productivity dispersion only in low-skill-intensive sectors. The one-for-one substitution of intangibles per labor is more prevalent in less intangible and knowledge-based sectors. In contrast, for high-skilled sectors to scale up and keep up with the demands of their economic activity and intangibleness, they need more workers. Still, only of the skilled type [34]. A similar story occurs with increased IT capital intensity (Faggio et al., 2007).

Even recent studies on *Work From Home* (WFH) policies due to the COVID-19 pandemic hint at differences in workers' demand due to intangibles. Eberly, Haskel, and Mizen

(2021) finds a positive and significant correlation between changes to WFH policies and intensity in intangible capital across industries. They argue that years before the pandemic, homes were getting equipped with more and faster computers, so once the pandemic hit, workers at home just untapped the potential of unused capital[35]. Firms' capacity to exploit this increased labor productivity (e.g., improved communications and teamwork collaboration). It enabled their production network more generally overseas or cushioned their activities when they had to fire workers.

1.4.2.4 Intangibles imply slower responses to shocks. Finally, another characteristic of intangibles, *sunkness*, alters overall employment growth. Chiavari and Goraya (2021) demonstrates there have been increases in input share of intangible capital at the expense of labor in the U.S. due to an intangible capital-biased technological change (IBTC)[36]. They show that intangible capital is much more misallocated than tangible capital because intangibles require hard-to-recover investment costs (e.g., developing software implies sunk costs like research, and if the software underperforms due to market competition, sunk costs cannot be recovered). Higher sunkness implies that an intangible-intensive economy is slower to respond to productivity shocks[37].

Given the evidence above, this chapter contributes to digging deeper into the implications of intangibles for employment growth and business dynamics by building on the work on frictions to job reallocation by Bartelsman, Haltiwanger, and Scarpetta (2013). In a general equilibrium setting, Bartelsman et al. (2013) explores the impact of distortions, such as taxes to job destruction or labor overhead costs, on firms' employment responses to productivity shocks. Their model shows homogeneous employment change across firms without distortions, whereas, with distortions, the change depends on the firm's and friction's sizes. Using this context throughout the chapter, intangible intensity is considered a source of friction for job reallocation.

Consequently, the stylized facts, regression analysis, and formulated model presented later in this chapter (sections 1.6,1.7) are tailored to support evidence of intangible intensity associated with a decline in business dynamics indicators, including entry of firms, the share of young high-growth firms, and employment growth dispersion.

1.5 Data

This chapter uses two primary data sources—first, the product of the DynEmp3 business dynamics variables computed for the U.K. The definitions of such variables are in section 1.2. These include variables on entry and exit rates, dispersion differential in employment growth rates, and share of employment by age*size class combinations. Second, the intangibles investments and capital dataset, INTAN dataset. I match these two sources of data, the business dynamics indicators and the intangible investment variables at

the year and the 2-digit industry level. This is the lowest level of disaggregation for the business dynamics indicator we count on since it is impossible to match firm by firm directly. Hence, the primary dataset comprises the match between the business dynamics result and its match by sector and year on intangible investments information.

1.5.1 Business Dynamics Data

1.5.1.1 The DynEmp3 Code The estimations presented in this paper are the product of the intermediate outcome of version 3.0 of the DynEmp STATA program (referred to as DynEmp3 from now on) designed by the OECD Directorate for Science, Technology, and Innovation Calvino et al. (2015). Thanks to Chiara Criscuolo, Flavio Calvino, Sara Calligaris and Rudy Verhac, and other members of the Directorate and the OECD's DynEmp team, we had access to the codes, explanatory documents, and incredible guidance on the program's details, and data preparations.

The DynEmp project is based on a distributed data collection exercise to create a harmonized cross-country micro-aggregate database on employment dynamics from confidential micro-level data, whose primary sources are firms and establishment data from national business registers.

The DynEmp3 codes are explicitly developed by the OECD DynEmp team. They are codes applied decentralized by national experts from statistical agencies, academia, ministries, or other public institutions who have access to the national micro-level data.

DynEmp3 requires as input a longitudinal micro-level dataset at the enterprise or establishment level from each country. The dataset needs to be in the shape of a panel, firm $\times$ year[38]. Individual units must be identified by a unique longitudinal identifier (e.g., the business reference number for a specific company) that has to be constant over time. If the unit exits the business register, its identifier must not reappear again since the job dynamics statistics for a given year are also based on information on contiguous years. The DynEmp3 code also allows users to specify a variable with the firm's birth year. For units whose birth year is missing, it assumes the birth of the firms the first time the team appears with non-zero employment in the register unless this year coincides with the first year of the business register (in which case the birth year is left missing) (Calvino et al., 2018). This paper constructs a longitudinal micro-dataset at the firm level for the U.K. covering 1998-2018.

The industry aggregation used in the codes allows the OECD's STAN - STructural ANalysis Database[39] industry classification at the macro sector level (A7), with seven major macro sectors, and, at a finer industry level (A38), with 38 two-digit sectors. The present study has three types of industry aggregations (see Table 2).

First, we use the OECD's A7 macro sectors aggregation, but we only focus on manufacturing and non-financial market services (level 1 in Table 2). We also refer to the latter macro sector as just Services. Second, we use a sector aggregation using the A38 two-digit level to aggregate into six macro sectors following the classification used in Decker et al. (2016a) for the U.S. to facilitate cross-sector comparisons with that country. The latter aggregate industries into Manufacturing, Retail, Services, Information, FIRE (finance, insurance, and real estate), and others (level 2 in Table 2). Finally, we use the OECD classification for ICT vs. Non-ICT intensive sectors, also derived from a more granular industry-level category, the ISIC rev.4, equivalent to the UK SIC 2007 industry classification at the 3-digit level. These sectors are divided into ICT manufacturing, ICT services, content and media, other manufacturing, non-financial market services, and other sectors (Level 3 in Table 2). See Table D.12 Appendix for a detailed description. All levels of aggregation except for level 2 are estimated by default by DynEmp3.

TABLE 2
Industry aggregation levels

Level 1	
Sector	**Ind A38 (2-digit codes)**
Manufacturing	10-33
Non-Financial Market Services a.k.a. Services	45-63, 68-82

Level 2	
Sector	**Ind A38 (2-digit codes)**
Manufacturing	10-31
Retail	45
Services	49, 55, 69-77
Information	58,61
FIRE (Finance, Insurance and Real Estate)	64, 68
Others	01, 05

Level 3	
Sector	**ISIC Rev. 4 (3-digit codes)***
ICT manufacturing	261-64; 268
ICT services	582; 611-19;620;631;951
Content and media	581;591;592;601-602;639
Other manufacturing	10-33 excluding ICT manufacturing
Other non-financial market services	45-62, 68-82 excluding ICT services
Other sectors	Other sectors in none of the groups above (except for STAN aggregation excluded sectors: finance and insurance (64-66) and, public sector (84)

(a) *UK SIC 2007 industry classification is directly comparable with ISIC Rev. 4 classification at the 3rd digit level. Hence, 2-digit codes refer to the SIC 2007 (2-digit codes) classification.

(b) STAN aggregation based on ISIC rev. 4 classification (compatible with SIC 2003 and SIC 2007 at the 3-digit level) is available in Appendix.

DynEmp3 produces a series of intermediate output databases according to these different levels of industry classifications, among other c classes. For most figures in this study, we use DynEmp3 output level 2, which corresponds to level 1 in Table 2, and DynEmp3 output level 3, which corresponds to level 2 in Table 2. For figures and analysis on Information, Communication, and Technologies (ICT), we use the DynEmp3 output level 8, and for the dispersion analysis, we use a customized combination of DynEmp3 levels 3, 7, and 72.

1.5.1.2 UK Business Demographics Data This chapter focuses on the use of the Business Structure Database (BSD) distributed by the UK Data Service and maintained by the UK statistics authority, Office for National Statistics (ONS) (See Office for National Statistics (2018)). The BSD consists of a micro-level series of annual snapshots (taken every March) of the UK's Inter-Departmental Business Register (IDBR). The IDBR relies heavily on data collected by Her Majesty's Revenue and Customs (HRMC) on firms and plants that register as VAT (Value Added Tax) traders or PAYE (Pay As You Earn) employers [40]. It also collects information on ownership and control and is considered a "live" register. The latter is the sense that it is updated more or less continuously, and the data is then created by a "snapshot" of every March. The IDBR is the key sample frame for UK business statistics.

The ONS says, "The consistency of IDBR's reference numbers for its various micro-level units of analysis throughout time enables these BSD files to form a panel data set" (for National Statistics, 2006). Thus, the BSD in the shape of a panel has the potential to reflect various business demographics over time, although with specific problems, as explained later. For this paper, I have linked the BSD's yearly snapshots for the 1998-2018 period to form a longitudinal dataset.

The IDBR reports three different types of units of analysis 3. Administrative, statistical, and reporting units (also called observation units). The administrative units are those reported as VAT traders or PAYE employers. The reporting units hold the mailing address for the business and are the units for which registered businesses report their survey data to ONS. However, the BSD's primary focus and, thereby, the primary structure are the statistical units.

Three categories of statistical units aggregate different levels of information for the businesses: the enterprise, the enterprise group, and the local unit.

FIGURE 3
Unit of Analysis. Source: Author based on Business Structure Database User Guide.

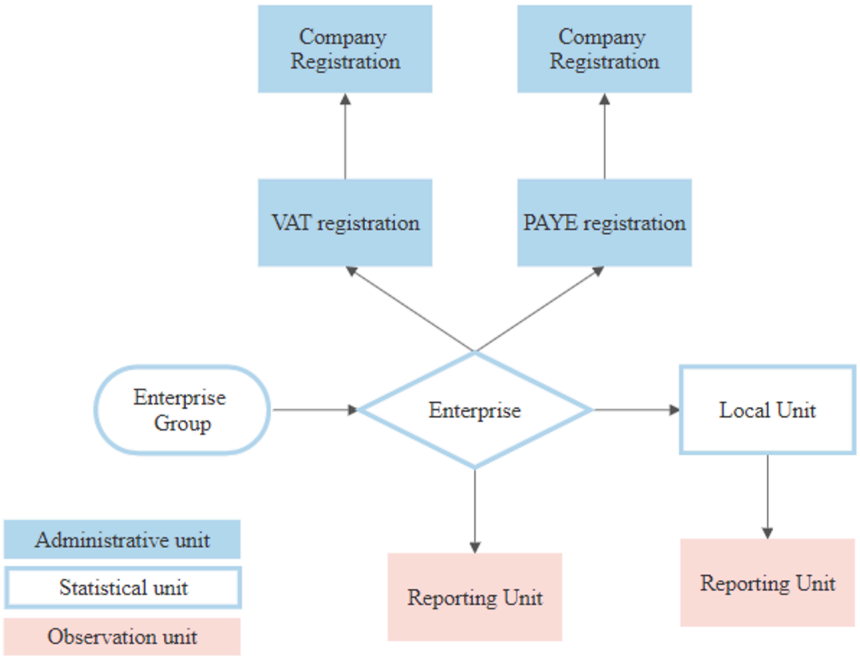
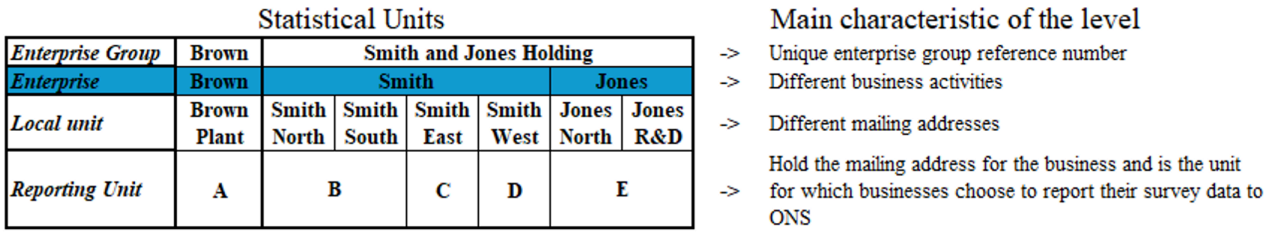


Figure 3 illustrates these different units of analysis for the businesses. "An enterprise holding group is an association of enterprises bound together by legal and/or financial links; it does not have to have a unique decision-making center, and it constitutes an economic entity that is empowered to make choices, particularly concerning the units it comprises" (Office for National Statistics 2006). Each of the enterprises it controls might develop different business activities. An enterprise "is the smallest combination of legal units and an organizational unit producing goods or services. It benefits from a certain degree of autonomy in decision-making, especially in allocating its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit." "The local unit is an enterprise or part thereof (e.g., a workshop, factory, warehouse, depot, etc.) situated in a geographically identified place. Economic activity is carried out at or from this place – save for certain exceptions - for which one or more persons work for the same enterprise" (for National Statistics, 2006).

FIGURE 4
Example of units of analysis



Note: (1) Smith and Jones Holding is a holding company registered in London that owns two enterprises, Smith and Jones, which develop different business activities. Both firms produce different plants located at different mailing addresses. Enterprises Smith and Jones each are subdivided into different reporting units, which in turn are composed of the appropriate local units within each enterprise. The enterprise decides at what level it aggregates its local units into reporting units. (2) Examples of how to sum employment for each unit of analysis: Employment of reporting unit B = Employment of local unit Smith North + Employment of local unit Smith South. Employment of reporting unit A = Employment of local unit Brown = Employment of enterprise Brown.
Sources: Author illustration based on Criscuolo et al. (2003) cited in Business Structure Database User Guide (2006).

The reporting unit is the same as the enterprise in general, but in some cases, the structure is more complex when various local units are under the same enterprise. For instance, the separate classification of local units is located within Great Britain and Northern Ireland. We only focus on the enterprise level—the last report aggregates information about its corresponding local units. The final raw panel data built for this study resulting from linking the BSD annual snapshots has 8.733.568 reporting units at the enterprise level for the whole analysis period.

The resulting panel dataset also includes a unique IDBR identifier per firm. Such identifier allows the dataset to be connected with other firm-level data sources available within the U.K. Data Service and the Office of National Statistics. The identifier can be connected to the annual versions of the U.K. Innovation Survey (Office for National Statistics, 2023b)[41] and the 2009-2010 issue of the U.K. Intangibles Survey (Haskel et al., 2010)[42]. This is possible because the BSD is sampled from the population of firms in the IDBR, and both surveys were also sampled out of that registry. However, creating the dynamics indicators took most of the time, which prevented reaching the point of matching the datasets for the present analysis. The construction of the business dynamics indicators required a particular preparation to apply the DynEmp3 routine and ensure that the resulting indicators could be consistent with those released by official ONS releases.

Access to the BSD data from the IDBR is through a secure lab system since the information in the registry may lead to the identifiability of firms, which can be dangerous for businesses. Thus, matching this data requires specific steps to ensure the process protects confidentiality, which also takes considerable time. However, as the matching is possible, it would undoubtedly allow researchers to relate levels of intangible assets

intensity or propensity to innovate or types of innovation to the likelihood of entry and exit or decisions on employment in future analysis.

In the appendix A, I include a section explaining the comparability between the results obtained using the original database created for this study and the dataset used and the official business dynamics indicators reported by the U.K.'s Office for National Statistics (ONS) (Section A.1). In general, I show that by making certain assumptions (Section A.2) and being aware of certain caveats by following ONS' general statistical guidelines on business registry data management, I obtain results with the database I build that correspond to ONS' official releases. The suggested sections above provide more detail about this subject and the assumptions used to build the business dynamics dataset for the present chapter.

1.5.2 Intangibles Investments Data

Despite the rising importance of intangibles, a comprehensive measurement of the magnitude of such investments and their relationship with other micro and macro variables in innovation and productivity remains elusive. This is particularly true at the firm level and in developing countries.

The reasons for such difficulties vary. For example, the rapid pace of newly emerging tech and its augmented complexity makes it hard to keep up with new forms of intangibles associated with such new developments. Also, intangibles are not always seen as an asset because of the difficulties of valuing them. It is easy to assign a value to intangibles based on the price at which they can be purchased on the market (e.g. a Microsoft software program). Still, intangibles generated internally (e.g. an employee writing software) are usually not included in firms' balance sheets. Instead, the resources used to make that internal intangible possible are considered costs with amortization or as part of other inputs or expenses. Subsequently, liquidating these intangibles is cumbersome as generally done with a physical asset.

Moreover, there is no available public market data on the previous valuations or intangibles purchases, making it challenging to find comparable values and, therefore, make valuations by proxy. Finally, there is a lack of standardization and uniformity in intangible accounting. The standards for accounting intangibles differ by industry; they also can be old and do not reflect intangibles as income generator assets, which, besides proper measurement, compromises regulation and comparative analyses.

Although less progress has been made to have intangible investment measurements at the firm level, improvement at the macro level is much more noticeable. Corrado, Hulten, and Sichel (2009)(CHS) implemented a methodology in which they capitalized costs that otherwise would have been considered merely amortizable and considered

investments. These authors addressed the conceptual problem of defining intangible assets using the idea of symmetrical treatment of investment. Any expense that reduces current consumption to increase it in the future qualifies as an investment. Therefore, an investment in knowledge capital is as much as an investment made in plant and equipment. With this approach, and focusing on the U.S., they created an empirical strategy to identify expenses that could enhance producers' future capacity. They classified them into investment categories: computerized information, innovative property, and economic competencies. In this way, these intangible investments could be added to the other investments in countries' national accounts. At the time, only software and R&D were recognized as assets and measured in the national accounts (Corrado et al., 2019).

However, incorporating these methods into national accounts has moved slowly, depending on the country. For instance, obtaining comprehensive estimates for developing countries remains challenging, where national accounting systems are outdated, poorly standardized, and lack uniformity within and across industries, preventing the creation of fully comparable data between countries.

Fortunately, the UK has data on intangible investments applying the Corrado et al. (2009)(CHS) methodology, with which we can build an empirical strategy to test some correlations with business dynamics. This is possible thanks to these and other authors' cumulative efforts in measuring intangible investments at the macro level. For instance, the INTAN-Invest project[43] created these measures for business and industry sectors for twenty-eight European member states plus the U.S. Corrado et al. 2017, under the SPINTAN project[44] later included a modified version of their CHS framework to estimate investments in intangibles from the public sector. Also, Fukao et al. (2009) created intangible measurements for other countries such as Japan.

Thanks to a merger between the INTAN-Invest and the SPINTAN projects, there is full coverage of intangible investment by industry sectors for about 21 countries, including the U.K. This merger allows the generation of total economic growth accounts with intangibles as productive assets.

In the present study, we use this database to find evidence about the relationship between intangible investments and business dynamics indicators. To the best of our knowledge, no current work is connecting the U.K.'s business dynamic indicators produced by the DynEmp3 codes and this dataset. Thanks to Professor Jonathan Haskel, a member of the Monetary Policy Committee at the Bank of England, for facilitating access to this dataset.

This database (INTAN-SPINTAN, from here onwards) contains annual information from 1995 to 2017 about the "productive" expenses converted into different intangibles

investment categories. The authors take the CHS expenditure-investment approach in which they transform the expenses into measurements of the value of an investment in intangible assets. Then, they produce estimated values comparable across countries and industry sectors (corresponding to the sections of the NACE rev. 2 classification). INTAN-invest provides all the information for the private and business sectors. This can be mapped to comparable investments from the non-market or public sectors coming from the SPINTAN source. See the mapping in Table D.11, and for a complete description of the sectors in the original SPINTAN source, see Table D.12 in the Appendix.

The primary variables of interest for our present analysis are their measurements of the nominal value of intangible investments and tangible capital for each sector. Their measurements of intangible investment correspond to the estimation of the following equation:

$$\begin{aligned} P^N N_t = \sum_{i=1}^{N} \sum_{s=1,t=1}^{s} & \gamma_{i,s,t}^{\text{own-account}} \lambda_{i,s,t}^{\text{own-account}} \text{ OwnCost}_{i,s,t}^{\text{Indicator}} \\ & + \gamma_{i,s,t}^{\text{Purchased}} \lambda_{i,s,t}^{\text{Purchased}} \text{ Purchased}_{i,s,t}^{\text{Indicator}} \end{aligned} \tag{17}$$

Where P^N N is the nominal expenditure/investment in the intangible N at the moment in time t, i is the industry sector, and s is the sector to which each sector can be mapped (i.e., market vs. non-market sector). As you will see in Table 5, there are different investment categories for tangible and intangible capital. Each intangible N can be already accounted for in National Accounts (i.e., Software, see Table in Appendix) or not (i.e., Design). Simultaneously, each intangible is assumed to be composed of two parts: Own-account intangibles and Purchased intangibles. In the equation, they are time-series indicators of the own account and purchased components of intangible investment, respectively. Parameter γ is a time series indicator to transform intermediate expenditure on intangibles into a sector-industry gross output (own account) or gross spending measure varying over time. Parameter λ represents the capitalization factor; it adjusts a spending measure to a measure of investment (Corrado et al., 2019).

Own-account intangibles are measured via the standard approach of gross-fixed capital formation for own final use based on production costs (the sum of compensation to employees, intermediate consumption, and the cost of capital). For example, the cost of a group of employees developing software within the firm for the use of the firm. Also, the cost of training employees within the firm. Or the costs of branding, market research, and design produced within the firm. The availability of these exact costs is limited. Therefore, the authors utilize proxies. For instance, they use the opportunity costs of

employees doing firm-specific training for training. For in-house software development, they use the average wage of software developers outside the software-producing industry (Corrado et al., 2016).

Purchased intangibles are expenses on externally generated intangibles. Intangibles, such as design, training, market research, or branding, are not accounted for in national accounts, so the authors mostly use current price lists per industry. These are usually available from most national statistics offices in the countries of study. In some cases, purchased intangibles are the only information available regarding an intangible category when no viable alternative proxies are found for the own-account intangibles.

The present study uses the NACE Rev. 2 industry classification (through its correspondence with the ISIC Rev 4 classification used by DynEmp3). For this reason, at the moment, we assume we do not need to consider the distinctions between market and non-market or public sector components. We assume that since the investment measurements reflect the investments in the overall economy, as well as that our business dynamics measures reflect patterns in the overall economy. Thus, we can focus on the correspondence between the two via the industry sector.

In Table 3, we present the mapping we built between the INTAN-SPINTAN dataset and the levels of industry sector disaggregation that we manage throughout this study. As seen in D.11 from the Appendix, each core section from the NACE Rev. 2 industry classification system can be matched with the corresponding core section belonging to the industry classification used to estimate business dynamics indicators. We only have specific business dynamics indicators available at the industry sector level, so not all the estimated indicators can be matched with the intangibles dataset.

TABLE 3
Industry Sectors Correspondance

Level 1		Industry sectors correspondence the INTAN-SPINTAN dataset	
Sector	Ind A38 (2-digit codes)	ISIC codes Rev 4 NACE Rev.2	ISIC Core Sector NACE Core Sector
Manufacturing	10-33	10-33	C
Non-Financial Market Services a.k.a. Services	45-63, 68-82	45-47, 49-53,55-56,58-60,61, 62-63, 68,69-71, 72,73-75,77-82	G,H,I,J,L,M,N
Level 2		Industry sectors correspondence the INTAN-SPINTAN dataset	
Sector	Ind A38 (2-digit codes)	ISIC codes Rev 4 NACE Rev. 2	ISIC Core Sector NACE Core Sector
Manufacturing	10-31	10-33	C
Retail	45	45-47	G
Services	49, 55, 69-77	49-53, 55-56, 69-71,72,73-75,77-82	H, l, M, N
Information	58,61	58-60, 61-63	J
FIRE (Finance, Insurance and Real Estate)	64, 68	64-66, 68	KL
Others	01, 05	01-07	A,B

Notes: Levels 1 and 2 correspond to Table 2, industry classifications for DynEmp3 analysis
L is not included in the INTAN-dataset
A is included in SPINTAN but no not in INTAN
After the matching between Dynemp variables, we obtain a sample of 18 years for 16 industry sectors. Add table.

1.6 Empirical Analysis

1.6.1 Business Dynamics and Intangible Investments: Stylized facts

This section presents some crucial stylized facts on business dynamics and intangible investments in the U.K. at different disaggregation levels: national, industry sector, age, and firm size class. Appendix B includes additional facts and business indicators.

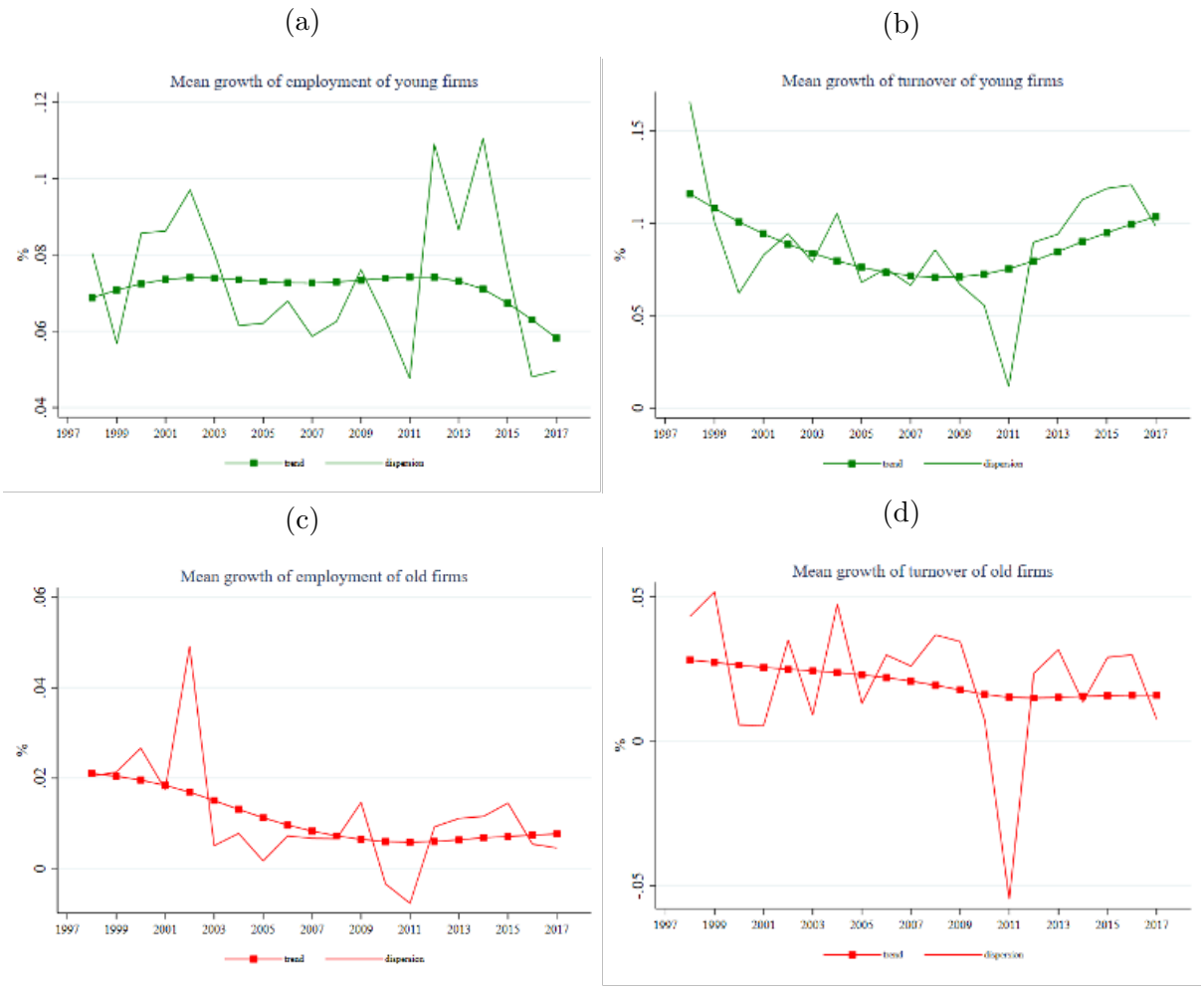
We combine the aforementioned primary data sources, 1) business dynamics indicators and 2) investment in intangible assets, to have a panel dataset (year × industry sector) composed of 16 industries and 21 years from 1998 to 2018, totaling 336 observations [45].

1.6.1.1 Fact BD1. Firms are scaling up less in terms of employment It might be argued that a more dynamic economy has many start-ups (young firms) entering the market and (or) many small firms being able to upscale to larger sizes over time. To illustrate whether this is the case in the U.K., Figure 5 presents the evolution of the average employment growth of firms by age class. The figure demonstrates that young firms have not scaled up in employment. Instead, their average employment growth rate declined some years after the financial crisis.

However, around the same time, they scaled up in turnover. In contrast, old firms have continued to decline in their capacity to scale up in both metrics, more noticeably in

terms of employment after the 2000s boom and more stable towards the end of the sample period. However, the number of self-employment firms in the U.K. has increased over time (See appendix B.1) despite larger firms employing most people (almost 50%). In that case, if unipersonal, self-employment firms are the types of young firms that are entering the economy, such firms are the most likely to contribute the least to employment growth.

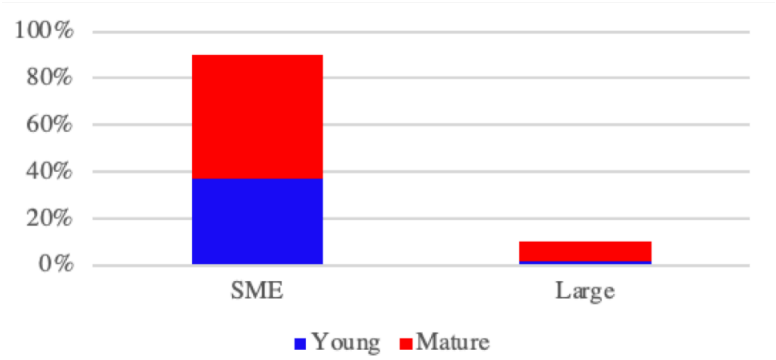
FIGURE 5
Mean size growth by age class



Note: Notes: Figures 5a and b refer to the mean growth of employment or turnover of all young firms. Figures c and d correspond to the same variables but for mature firms.

1.6.1.2 Fact BD2. Young SMEs contribute more to net job creation One of the positive signs of dynamism is that the small firm's pool mainly comprises young firms that can provide positive net job creation. However, in the U.K., for the sample period and grouping self-employment and SMEs, Figure 6 shows that the young firm's pool is the minority. 90% of the U.K. pool of firms are small or self-employed firms. Of this 90%, only 37% are young, whereas 53% are mature.

FIGURE 6
Share of Employment by firms' age×size class



Notes: This figure reports the average for the 1998-2018 period of the yearly share of employment in total employment by size class (*sz*: with a cut-off at >=250 employees) or age class (*ag*: young means below six years old) in a given year. Total of companies: 2,765,150, including self-employment. This updated aggregate is based on Office for National Statistics (2021), which differs from other bulletins [46]. The figure includes the count of VAT or PAYE-based firms in the United Kingdom.

When considering self-employment firms only, they are prevalently young, with 60% out of the total. However, only 41% of SMEs are young, and 59% are mature. Thus, since SMEs are more prevalent than self-employment (around 3.3 million firms in total (60%) against 2.2 million (40%), respectively), the U.K. exhibits a less dynamic environment where young firms do not constitute almost half the pool of the most prevalent firm-size group.

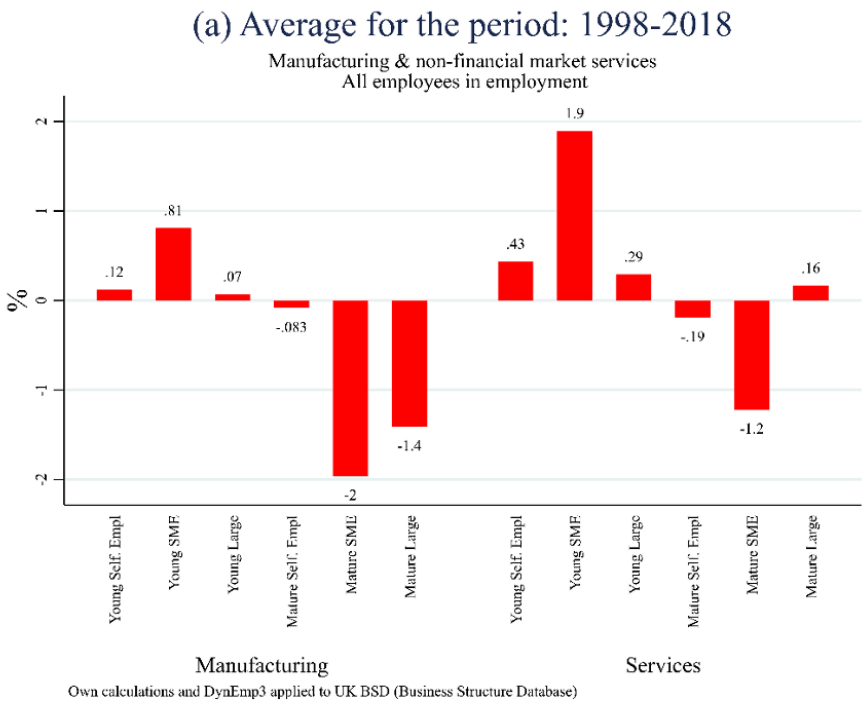
The proportion of young firms in each size group also differs by industry sector. While in manufacturing, only 28% of SMEs are young, in the services sector, these constitute 42%. Each size-age group differs in employment contribution as well. Figure 7(a) displays the net job creation share in total employment by size-age class. The figure reveals that in both sectors, Young SMEs are the size-age group contributing the most to net job creation during the period. While Young SMEs represent two percent of the share of the positive net job creation in total employment in the services sector, they contribute just 0.02 percent in the manufacturing sector. Mature SMEs contribute around 2 percent to net job creation; but by destroying jobs; the sign is negative.

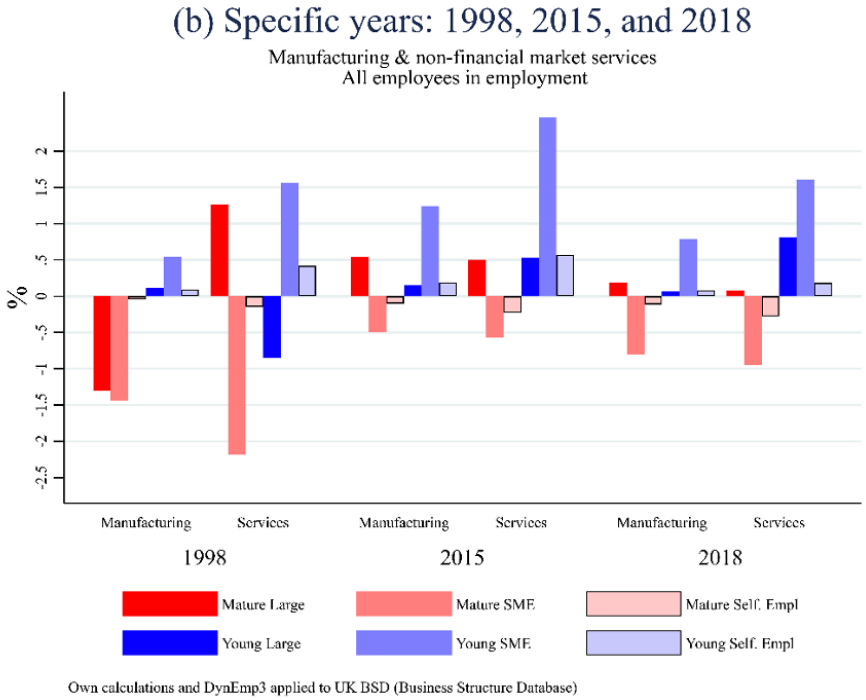
The findings align with previous discoveries where mature firms destroy jobs in the manufacturing sector (Criscuolo et al., 2014, cited in Calvino and Criscuolo (2019)). The arrival of new technologies and the increasing competitiveness, particularly in this sector, makes it easier to take out of business those firms with a lower capacity to keep up with such advancements, so when they leave, they destroy jobs.

Figure 7(b) shows the net job creation of these age×size groups for specific years. Although the previous figure showed a negative picture on average, over time, this figure reveals that the contribution to negative net job creation of mature SMEs decreased for both sectors from 1998 to 2015. Simultaneously, the positive contribution of young

SMEs to net job creation increased, going from 1.5 percent in 1998 to 2.5 percent of total employment in 2015. Notice, however, that the figures for 2018, two years after the Brexit vote 2016, present a reversion to this trend. Mature SMEs returned to destroy jobs in both sectors. While in 2015, job destruction of mature SMEs had decreased its share of total employment by around 0.5 percent, in 2018, this size-age class's negative net job creation impact increased again to around minus one percent.

FIGURE 7
Share of Net Job Creation by Size and Age





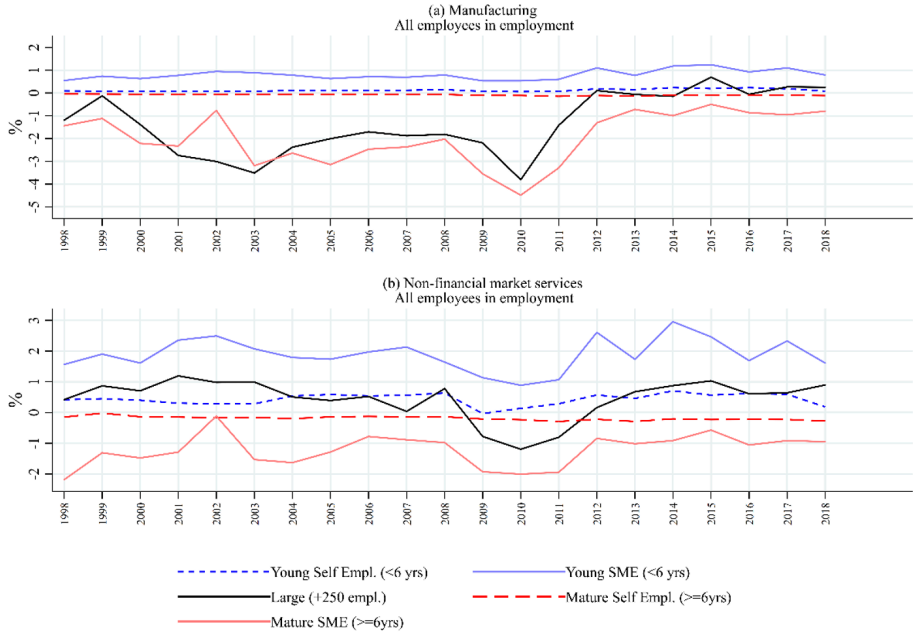
Notes: 1) This figure reports the contribution of net job creation to aggregate employment change by size and age class. Size: 1 employee in employment (Self Empl.); between 2 and 249 employees in employment(SMEs) and large (>=250 employees in employment)). Age: Young means below six years old. It is defined as net job creation (i.e., The difference of total employment at time t and t-1) of the particular group over total employment in the macro-sector (a) on average over the period 1998-2018 or (b) for specific years. 2) (a) and (b) exclude the public sector (ind_{a38} code = 84) and households and extraterritorial activities (ind_{a38} code =97). 3) Manufacturing includes (ind_{a38} codes 10-33) and services include (ind_{a38} code = 45-63, 68-82). The OECD DynEmp team recommends dropping from manufacturing, coke, refined petroleum products (19), and services real estate activities (68). Here, they were not omitted. 4) Includes only firms with non-missing information on employment and age.

Figure 8 shows more changes explicitly year by year in net job creation (NJC). In the manufacturing sector, Mature SMEs have turned to destroy fewer jobs after the financial

crisis, with a NJC contribution going from negative to close to zero in recent years. In contrast, in the services sector, although NJC noticeably declined during the crisis, mature SMEs' contribution seems relatively stable throughout the analysis.

The opposite is true when assessing the young SMEs' NJC contribution. After a period of stability, Young SMEs from manufacturing slightly increased their contribution towards the end of the sample, whereas the contribution of young SMEs in services noticeably recovered after plumming during the financial crisis. However, such recovery seemed to have stopped in years close to the Brexit vote in 2016.

FIGURE 8
Share of Net Job Creation by age×size class in manufacturing and services over time



Own calculations and DynEmp3 applied to UK BSD (Business Structure Database)

Notes: • This Figure reports net employment creation or destruction in each group (size-age class) relative to aggregate employment (average over year's t and $t-1$ in non-financial market services. It is defined as the ratio of net job creation of the particular group over total employment between t and $t-1$. Source: DynEmp3 level 3 output. • Exclude public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code =97). • Services include (ind_{a38} codes 45-63, 68-82). The OECD DynEmp team recommends dropping services and real estate activities (68). This sector is not omitted here. • Includes only firms with non-missing information on employment and age.

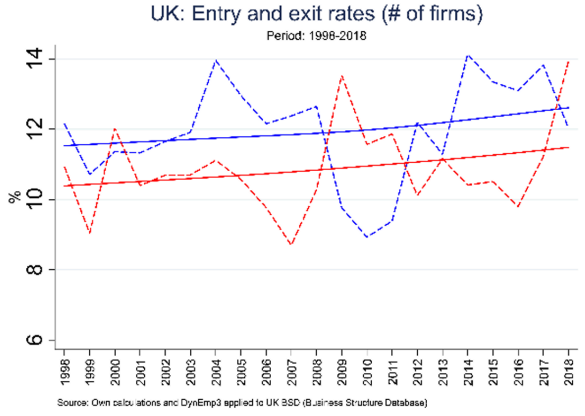
1.6.1.3 Fact BD3. Fluctuations in Entry, Exit, and Job Reallocation Rates: A Post-Financial Crisis Recovery Leading Up to the Brexit Year To understand the U.K.'s business dynamics relative to other countries, this section compares entry, exit and job reallocation trends with the U.S.

The U.S. Business Dynamics Statistics (USBDS, called USBDS from now on to avoid confusion with the U.K.'s BSD dataset) is a research dataset administrated by the U.S. Census Bureau. It is based on the Longitudinal Business Database (LBD) that describes

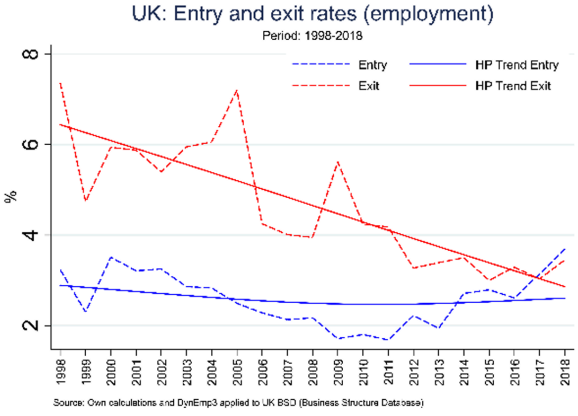
the United States' business dynamics across various measures. Disclosure analysis is performed before releasing to the public to protect the confidentiality of LBD data. Figure 9 presents entry and exit rates in terms of firms (called ENR and EXR, respectively) and employment (ENRE and EXRE) for both countries. Figures 9a and 9b illustrate the case for the U.K. and figures 9c and 9d for the U.S. These plots show that for the comparable analysis period (1998-2016), the U.K. experienced a similar secular decline to the U.S. in both types of entry rates after the 2000s. Still, the U.K.'s entry rates recovered after the 2009 financial crisis. However, such recovery slowed down right before the Brexit vote in 2016 when entry rates started to decline again. See details on the computations of entry and exit rates and comparability with U.K. data in Appendices A.3 and A.4.

FIGURE 9
Entries and exit rates

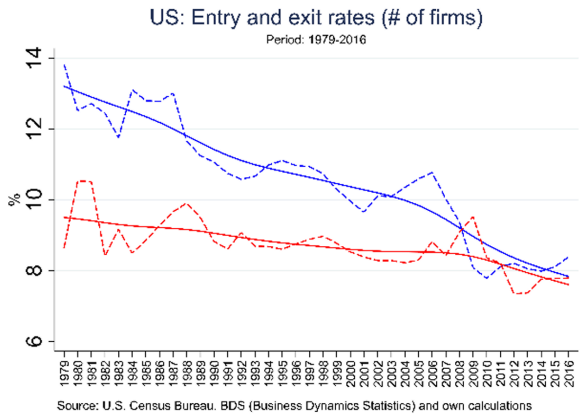
(a) U.K.: In terms of firms



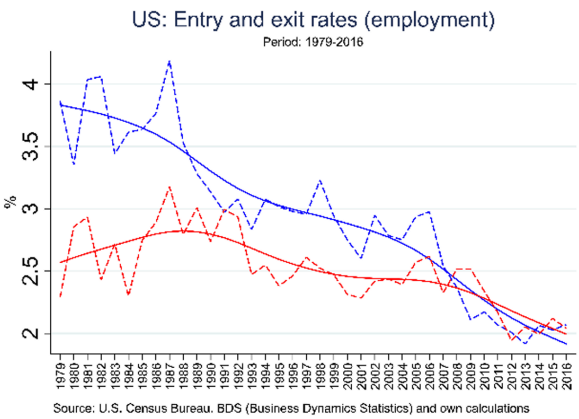
(b) U.K.: In terms of employment



(c) U.S.: In terms of firms



(d) U.S.: In terms of employment



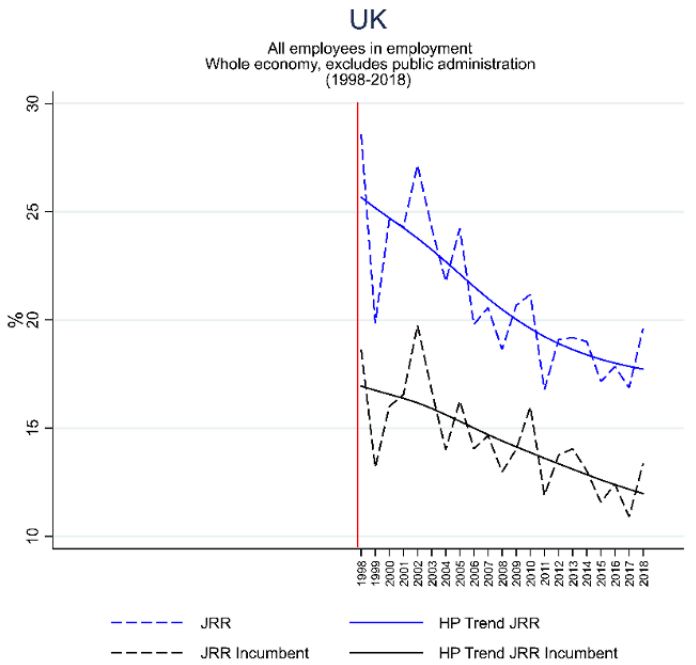
Note: Figures a and b refer to U.K. data, which includes the whole economy (excludes public administration) and considers all employees in employment. Figures c and d refer to U.S. data which covers the whole economy and excludes employers in employment.
The red lines report exit rates and the blue lines report entry rates. Owing to methodological differences, figures may deviate from officially published national statistics. Exclude public sector (ind_{a38} code = 84), households, and extraterritorial activities (ind_{a38} code =97).

Job reallocation is a comprehensive measure of how jobs are created and destroyed in a given year relative to the average employment stock available. For the analyses, consider the equations of job reallocation rates for the U.S. as follows:

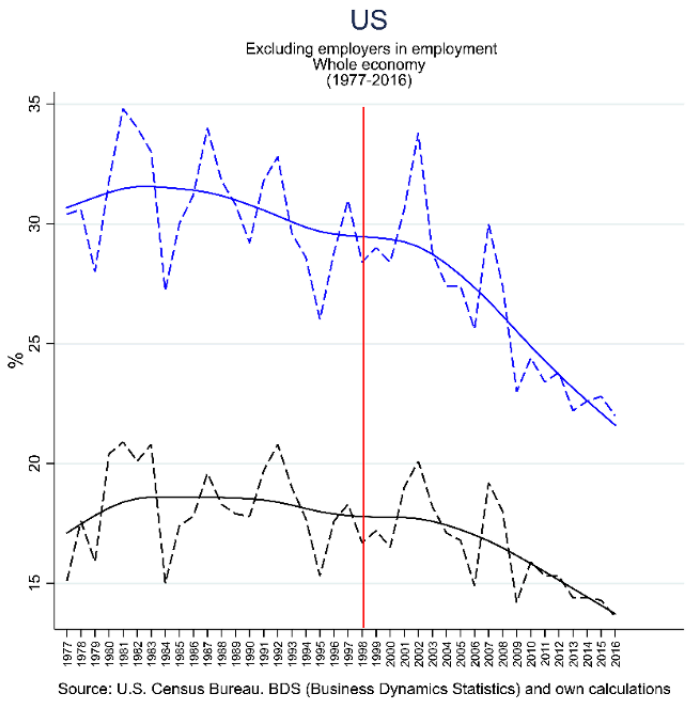
$$\begin{aligned} USJCR_{c,t} &= \frac{USJC_{c,t}}{\frac{1}{2}\left(E_{c,t} + E_{c,t-1}\right)} * 100 \\ USJDR_{c,t} &= \frac{USJD_{c,t}}{\frac{1}{2}\left(E_{c,t} + E_{c,t-1}\right)} * 100 \end{aligned} \tag{18}$$

US $JCR_{c,t}$ indicates the Job Creation Rate for the U.S. and $USJDR_{c,t}$ displays the job destruction rate for the U.S. Figure 10 shows job reallocation rates for the U.K. (1998-2018) and the U.S. (1979-2016). As shown in Decker et al. (2016c), this figure reveals that U.S. job reallocation declined until 2000, when the downward trend accelerated; a pattern also exhibited for the U.K. Figure 10's upper panel presents the U.K. job reallocation for only firms and incumbents, and both display the same downward trend. The pattern shared by incumbents and the general pool of firms indicates that the slowdown was not exclusively associated with fewer entries and exits. It was due to declines in organic growth and the growth pace of incumbent firms.

FIGURE 10
Job Reallocation Rate Declining



Source: Own calculations and DynEmp3 applied to UK BSD (Business Structure Database)



Source: U.S. Census Bureau. BDS (Business Dynamics Statistics) and own calculations

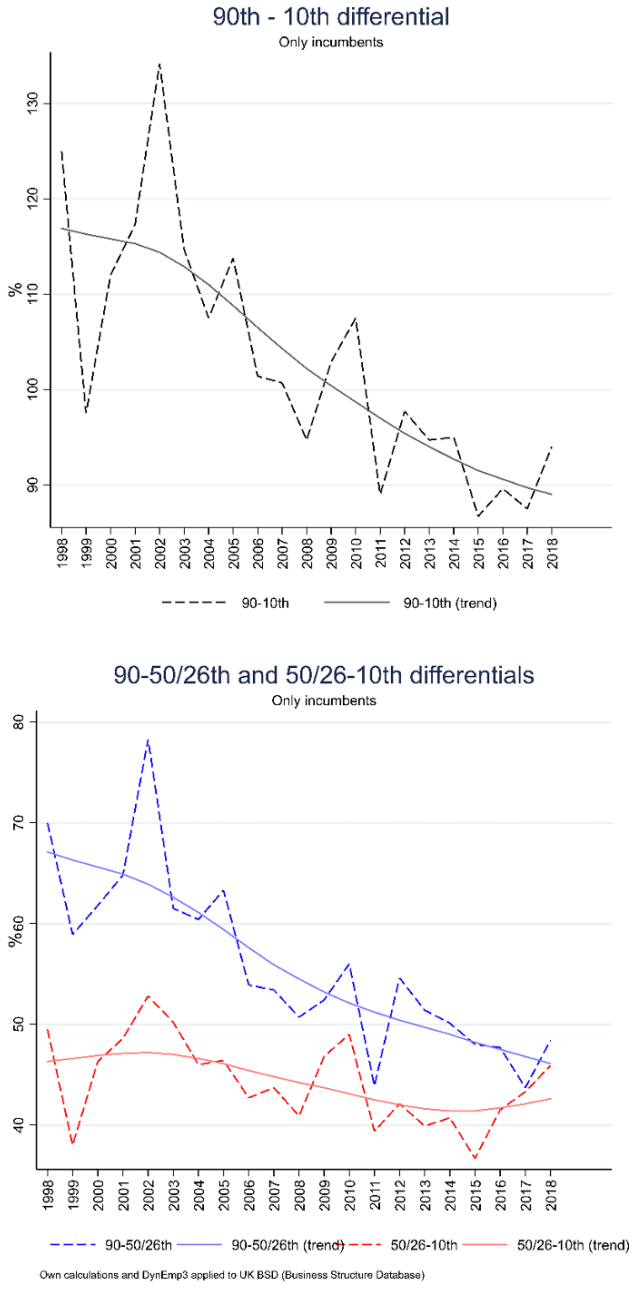
Note: UK: • Trends are HP with a smoothing factor of 100. • Employment: Employment figures are an average of any firm's last four quarters of the financial year. If the average cannot be taken, numbers for the previous quarter are used. Employment is the sum of persons engaged and employers in employment. • Persons engaged: It counts each person hired as one unit, irrespective of the hours worked within the business. • Job reallocation rate is the sum of all positive and negative changes in employment by entrants, exitors, and incumbent firms; employer firms only. Job reallocation rates for incumbents only account for this firm group's positive and negative changes. • Exclude public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code = 97), coke and refined petroleum products (19), real estate activities (68), and construction (41).
US: • Trends are HP with a smoothing factor of 100. • Employment – Paid employment consists of full and part-time employees, including salaried officers and executives of corporations, who were on the payroll in the pay period, including March 12. Included are employees on sick leave, holidays, and vacations; proprietors and partners of unincorporated businesses are not included. • Payroll – Total annual payroll includes all forms of compensation, such as salaries, wages, commissions, bonuses, vacation allowances, sick leave pay, and the value of payments in kind (e.g., free meals and lodgings) paid during the year to all employees • Job Creation (JC) – Job creation is the sum of all employment gains from expanding establishments from year t–1 to year t, including establishment startups. Note that the contribution of firm births can be measured by using the job creation from establishments with healthy ages equal to zero. • Job Destruction (JD) – Job destruction is the sum of all employment losses from contracting establishments from year t–1 to year t, including establishments shutting down. • Source: U.S. Census Bureau. BDS Glossary. https://www.census.gov/programs-surveys/bds/documentation/glossary.html

1.6.1.4 Fact BD4. Asymmetric Patterns in Employment Growth Rates

1.6.1.4.1 Fact BD4a. Dispersion differentials shrunk Figure 11 illustrates three employment growth rate dispersion differentials for the U.K. using the DynEmp3 protocol. Unlike previous studies, DynEmp3 analyzes the skewness of the distribution of employment growth rates by comparing it against the distribution's median and adjacent firms. The presented dispersion differentials, such as 90th-26th/50th and 50/26th-10th, deviate from the conventional 90-50 and 50-10 used in prior literature such as Decker et al. (2016c). Moreover, DynEmp3 focuses solely on incumbent firms, offering insights into organic growth, and ignores dispersion analysis for the entire firm universe in the relevant group (e.g., entrants and exiting firms). However, as presented earlier, there is evidence of a secular decline in organic growth in the U.K. beyond fewer startups or exiting firms.

In Figure 11's top panel, the 90th-10th dispersion differential for incumbent firms consistently decreases, indicating a convergence in growth rates between top-tier and bottom firms. The bottom panel highlights that the decline in overall dispersion is mainly driven by high-growth firms (90th-50th/26th percentiles), which are slowing down their annual growth rate, aligning more closely with the median firm.

FIGURE 11
Dispersion differentials of employment growth rates
Whole economy, excludes public administration
All employees in employment

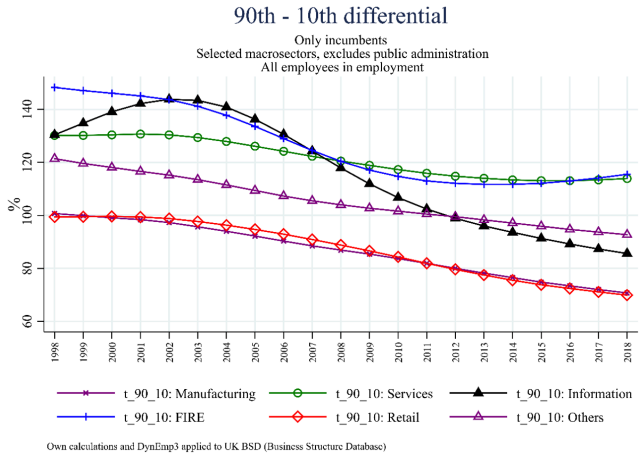


Notes: • The figures show the trend of three employment-weighted dispersion differentials: 90-10, 90-26/50, and 26/50-10th. • Trends are Hodrick–Prescott with a smoothing factor of 50.

1.6.1.4.2 Fact BD4b. Puzzling Dispersion Differentials: A Sectoral Analysis with a Focus on the Information Sector Decline Figure 12 depicts selected sectors' 90th-10th dispersion differential trends. Retail Trade and Manufacturing show consistent declines, while the sectors of Information, Services, Finance, Insurance, and Real Estate (FIRE) sharply declined after 2000. Despite industry-specific variations, the

U.K. experienced a generalized decline in dispersion post-2000, converging to lower levels over time, akin to trends observed in the U.S. by Decker et al. (2016a).

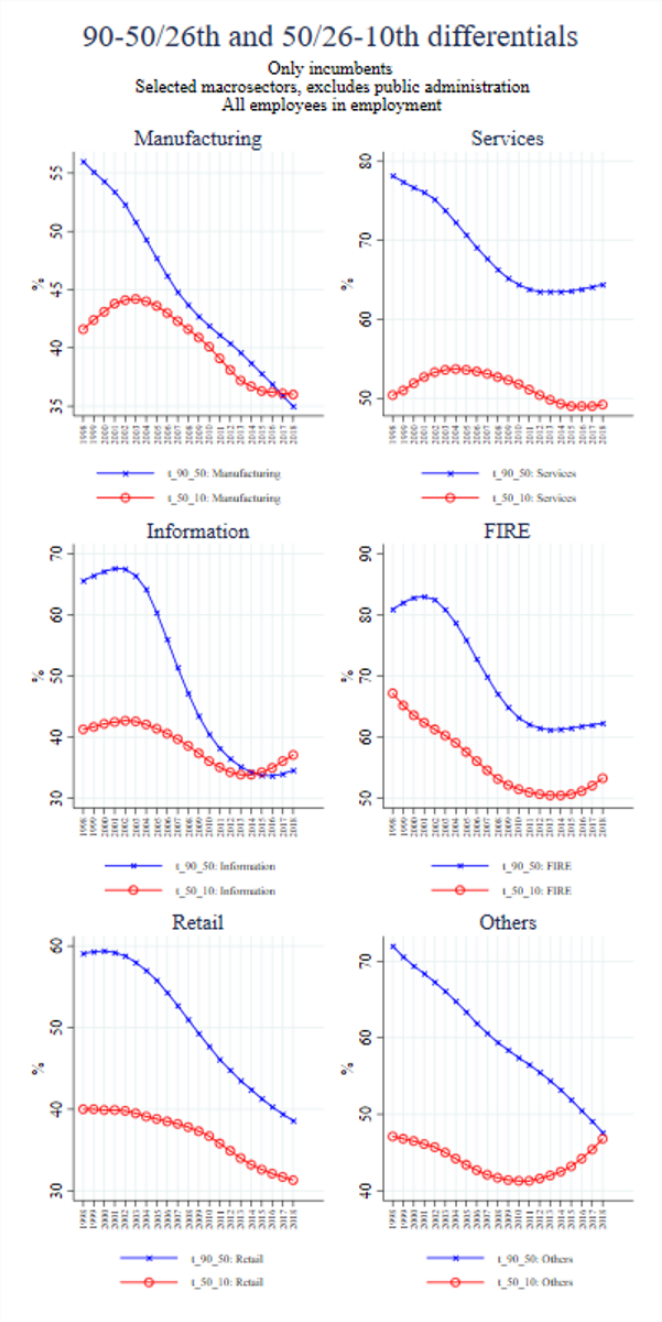
FIGURE 12
Employment growth rates dispersion differentials by sector



Notes: Trends are Hodrick–Prescott with a smoothing factor of 50.

1.6.1.4.3 Fact BD4c. Divergent Sectoral Skewness: Information Sector Top Growers Shrinking Faster Figure 13 illustrates skewness in employment growth rate distributions for various sectors. The graph reveals that decreased dispersion primarily results from reduced growth rates in the top percentile firms. Over time, top-grower firms resemble median firms more closely, while low-growth firms maintain slow growth. Notably, sectors differ in the magnitude of the decline in the top tier. Information and FIRE sectors saw the sharpest declines in the 90-50/26th differential after 2000, while manufacturing, retail, and services experienced a smaller decline in the same differential, with manufacturing undergoing a smaller decline before 2000.

FIGURE 13
Employment Growth Rates Skewness

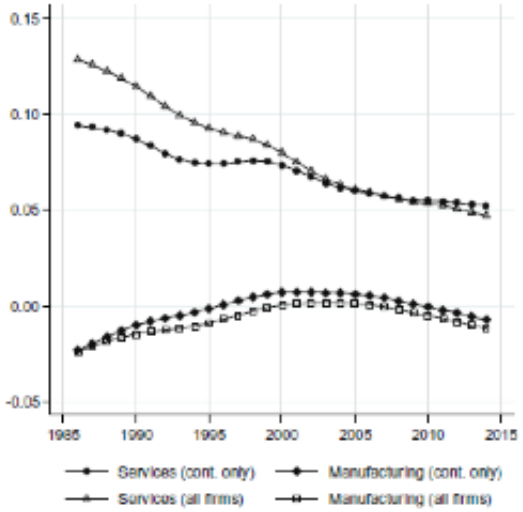


Notes: Trends are Hodrick–Prescott with a smoothing factor of 50.

Previous research on the U.S. and Belgium informs the understanding of analogous patterns in the U.K. The authors propose that post-2000 tech revolution, a firm's survival probability was influenced by its ICT intensity. For instance, Bijnens and Konings (2018b) contends that smaller ICT-intensive firms in Belgium's smallest companies were more prone to low growth or market exit due to challenges in covering fixed ICT costs such as acquiring ICT knowledge and infrastructure. This aligns with Decker et al. (2016c, 2017)'s argument for the U.S., highlighting high-skewness sectors experiencing the most impact from the decline. Similar trends may be inferred in the U.K. In Belgium, Bijnens

and Konings (2017, 2018b)'s analysis (see Figure 14) shows the manufacturing sector maintaining low skewness from 1985-2014, while the services sector, initially more skewed, steadily declined, nearly converging with manufacturing sector skewness levels.

FIGURE 14
Belgium's skewness of the employment growth rate distribution by sector

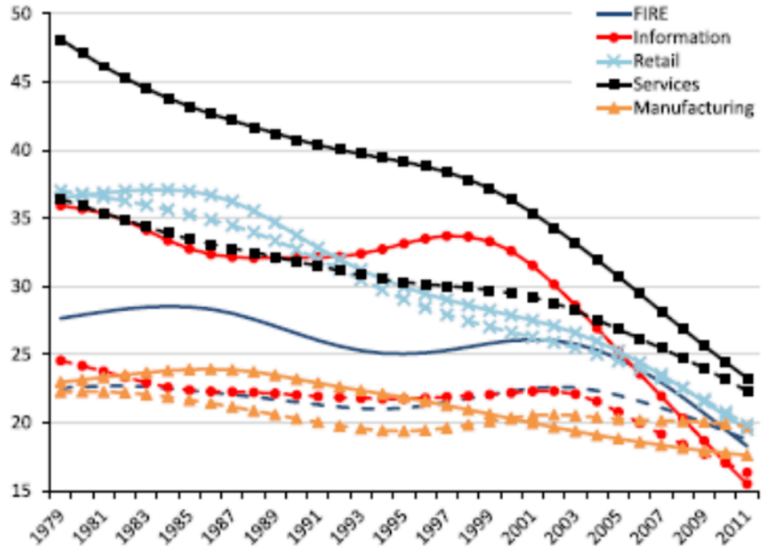


Note: Only incumbents and all firms Manufacturing vs. Services.
Source: Bijnens and Konings (2018b). "Skewness" or difference between 90–50 and 50–10 differentials for the services and manufacturing sectors. The skewness of the services sector significantly decreases. For manufacturing, it initially increases and comes down post-2000." Bijnens and Konings (2018b)

Figure 15 reveals a parallel trend in the U.S., where the Information and Services sectors, initially high in dispersion due to top-tier growth rates, sharply declined after 2000, converging with low-skewed sectors like Manufacturing and FIRE. Figure 13 illustrates that U.K. dispersion shifted towards the lower tail with specific differences from the other countries before 2000. Until the early 2000s, the U.K. manufacturing sector exhibited a highly skewed employment distribution, surpassing the U.S. and Belgium. Also, the U.K. services and FIRE sectors did not converge as much to the lower end over time, suggesting these sectors were notably resilient amidst an overall decline.

The above reveals that countries exhibit some differences in highly skewed sectors. However, some patterns are similar. The U.K.'s retail sector's 90-50 dispersion differential converged towards its 50-10, suggesting the same story that big box stores are pushing out smaller firms. Nonetheless, the retail sector seemed much more highly skewed in the U.K. than in other countries during the late 90s and early 2000s. Likewise, the collapsing pace in the dispersion differential of retail has been slower than in the information sector, indicating slower displacement out of the market of smaller firms.

FIGURE 15
Skewness of the employment growth rate distribution by sector in the U.S.

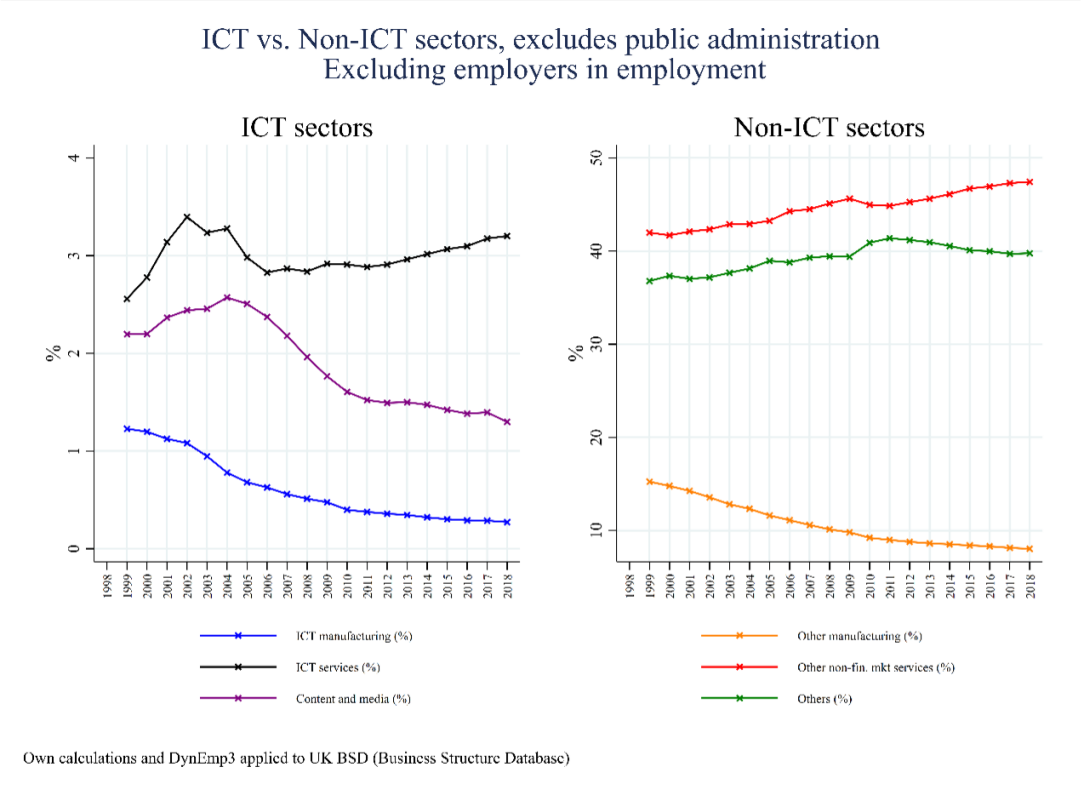


Note: All firms. Selected macrosectors. Figure 11 "90-50 and 50-10 differentials, selected industries. The y-axis does not start at zero. Solid lines indicate a 90-50 differential; dashed lines indicate a 50-10 differential. The 90-50 differential and the 50-10 differential are the difference between the 90th and the 50th percentile and the 50th and 10th percentile, respectively, of the employment-weighted distribution of firm employment growth rates. Data are H.P. trends using a parameter set to 100. Industries are defined on a consistent NAICS basis. Data include all firms (new entrants, continuing, and exiters)." (Decker et al., 2016a).
Source: Decker et al. (2016a) and author calculations from the U.S. Longitudinal Business Database.

1.6.1.5 Fact BD5. Information, Communications, and Technology (ICT)-Intensive Sectors: An Uphill Battle with Decline The level of ICT intensity in each industry sector may have a role in how different sectors behave. DynEmp3 can compute the share of employment in ICT and non-ICT-intensive industries as described in Table 2, and results are depicted in Figure 16.

The figure shows that, on average, the share of employment in ICT-intensive sectors declined right after the 2000s ICT revolution, but the decline slowed after the 2009 financial crisis. Content and media and ICT manufacturing's share of employment declined after 2000, with the former experiencing an even sharper decline. In contrast, non-ICT sectors recovered employment share over time, except for non-ICT intensive manufacturing, which also slowly declined, similarly to what is depicted on the left panel of the figure.

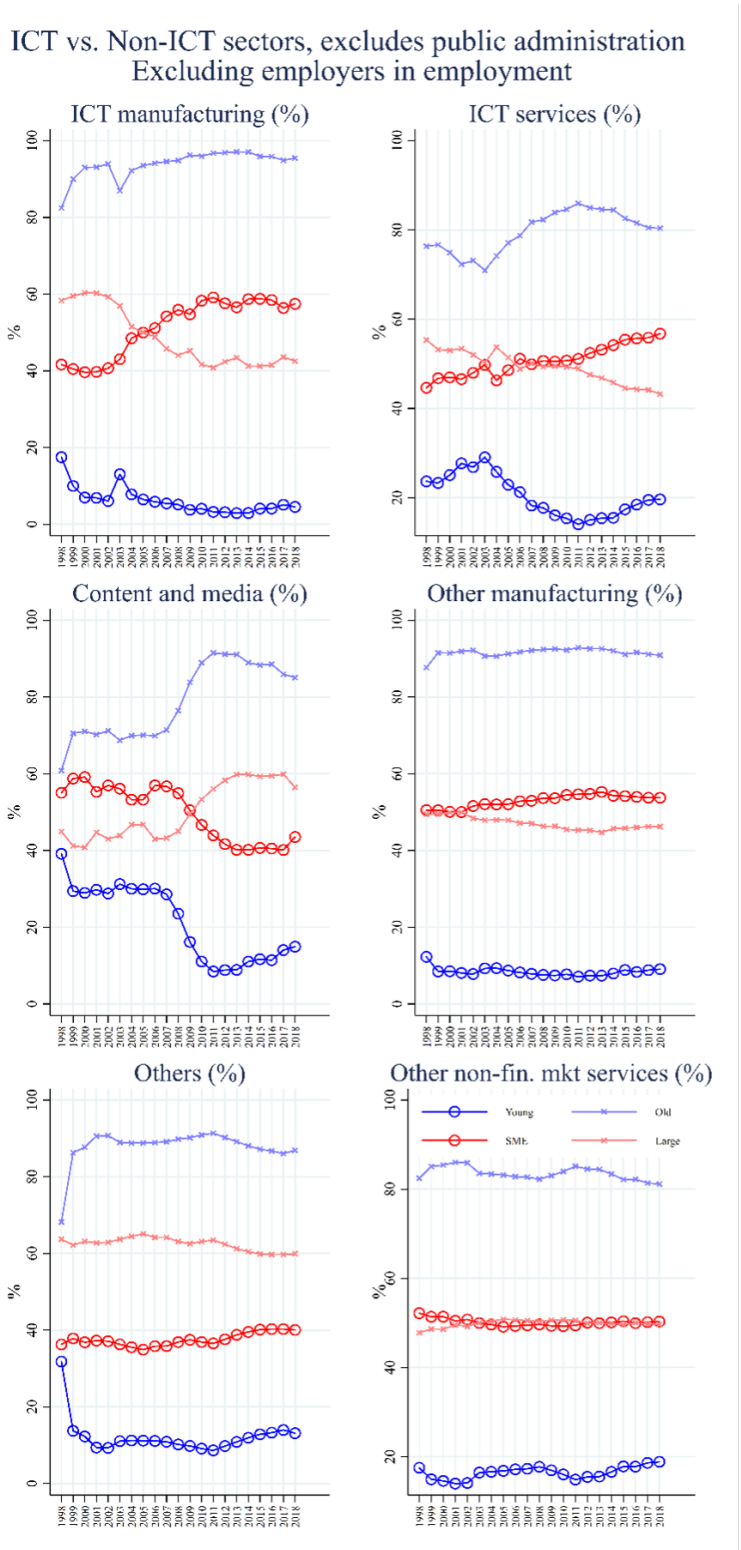
FIGURE 16
Share of Employment



Source: DynEmp3 output level 8

Figure 17 presents the share of employment by ICT intensity by age and size class. The most interesting pattern of the plot is that content and media, the most intangible industry sector of the group, experienced a noticeable decline in the share of employment of young and small firms. Interestingly, the pattern is slightly different with the ICT services sector; despite exhibiting similar patterns of age (young firms declined participation as older firms reclaimed participation), it shows opposite trends considering size. In other words, firms in ICT services would be more prevalently composed of smaller and older firms, indicating a lack of dynamics.

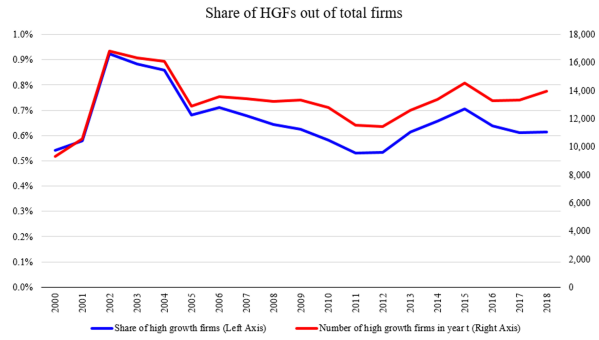
FIGURE 17
Share of employment by age class and size class



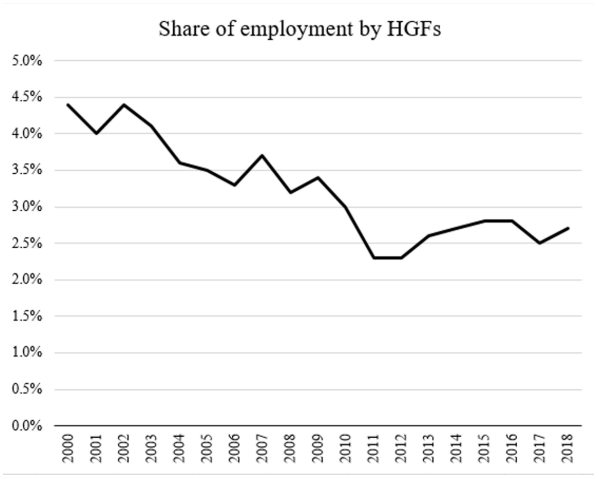
Note: This figure reports the share of employment in total employment by size class (*sz*: with a cut-off at >=250 employees) or age class (*ag*: young means below six years old) in a given year for each industry group under the ICT vs. Non-ICT classification.

1.6.1.6 Fact BD6. The Diminished Momentum of High-Growth Firms: A Shift Towards Smaller and Older Entities High-growth firms (HGFs) are firms whose growth pace is higher than peers over a short time frame[47]. On average, for the 2000-2018 period, HGFs were approximately below five percent of the total number of firms in the economy. Still, they contributed to around half of the employment growth of the U.K. economy. After the 2000s boom, the number of HGFs declined substantially until the floor of the financial crisis, but they picked up again, a recovery halted by the Brexit vote in 2016 (see Figure 18). However, HGF's employment contribution has declined over time, with a slight recovery after the Brexit vote year (see figure 18a). This means that the fastest-growing firms in the U.K. decreased their contribution to employment and have not yet recovered the peaks in numbers they reached in the early 2000s.

FIGURE 18
High-Growth Firm's participation over time



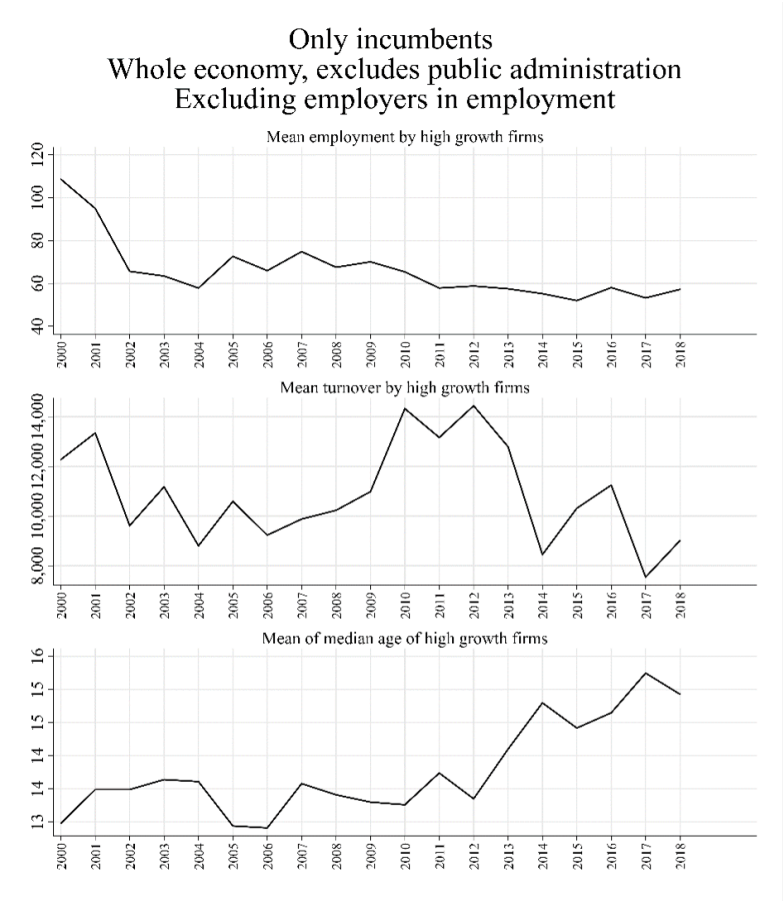
(a)



(b)

HGFs' lackluster performance is illustrated again in Figure 19, where HGFs' average size fell and remained low after 2000 despite a slight recovery towards the end of the sample period. Interestingly, they also have become older. The average of their median age has declined. Therefore, HGFs in recent years for the U.K. tend to be smaller and older.

FIGURE 19
High-Growth Firms mean employment over time

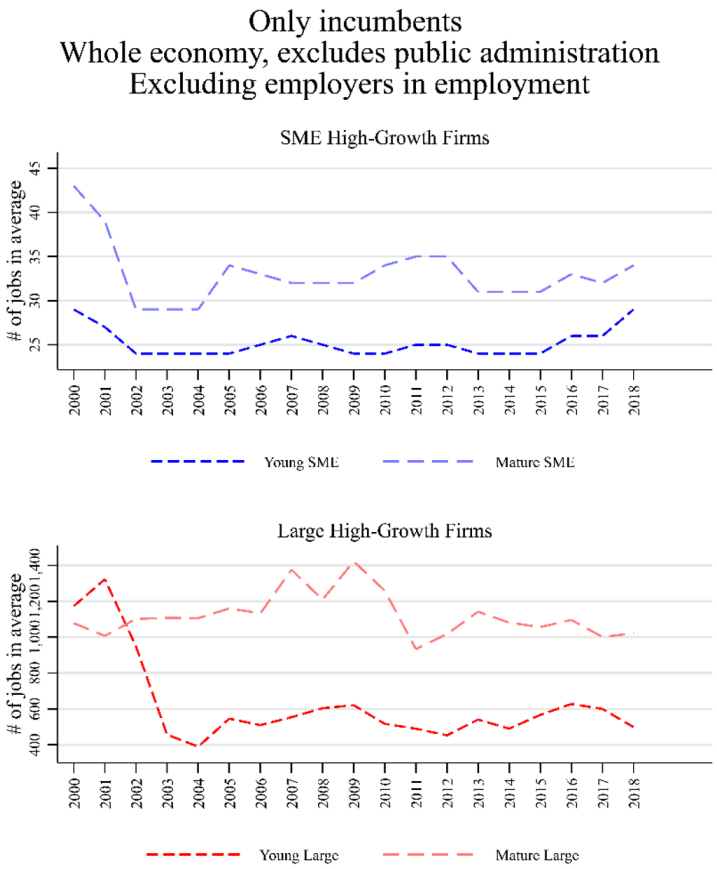


Notes: The mean of the median age is calculated across the A7 macrosector*year class median age of high-growth firms in the size group.

These patterns correspond to previous research where the skewness of the employment and turnover growth rates distribution has significantly declined. For instance, earlier studies for European countries and the U.K. show that the average, the median, and the variance of the employment growth rates fall with size and age (See Bravo-Biosca (2016a)), in contradiction with Gibrat's law where distribution and size are not correlated. Thus, the larger and more mature firms get, the more likely their average growth rate in employment tends to decline.

Figure 20 aligns with these patterns when observing the mean employment in the U.K. firms by age × size group. As illustrated in the top panel, Small HGFs, young and mature, declined in size, and although, towards the of the sample period, they experienced a slight recovery, their contribution to employment remains low. The bottom panel suggests that young and large high-growth firms are not creating as many jobs as they used to before the 2000s. Mature and large firms did not exhibit an abrupt decline after the 2000s; their average size has remained relatively stable.

FIGURE 20
High-Growth Firms mean employment by age × size class



Notes: This figure reports the share of employment contributed by high-growth firms in total employment by size*age class, with a cut-off at >=250 employees for size large and age above five years old for mature) in a given year.

Exploring patterns of HGFs is comparable to exploring the firms in the 90th extreme of the 90-10 dispersion differential. Hence, from the previous analyses, it is possible to infer that young firms are no longer in the top 90th of the employment growth distribution since HGFs are less likely to be composed of young firms.

Regarding firms growing at the slowest pace of the employment growth distribution, previous analyses on dispersion suggest bottom firms keep growing at the same pace. We can infer this since the 50-10 dispersion differential has remained relatively constant. In other words, HGF firms are becoming more like the median firms regarding employment growth.

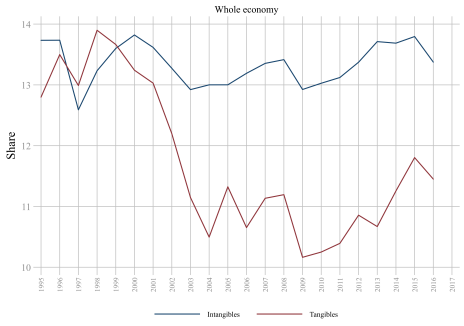
1.6.1.7 Fact INT1. As Intangible Intensity Rises in the UK, Dispersion Differential of Employment Growth Declines Figure 21a, presents the yearly shares between 1995 and 2016 of intangible intensity in the U.K., as defined in section 1.3, $ShIntan^{x}_{i,t}$ and $Shtan^{x}_{i,t}$. Except for the setback in trends towards the end of the sample

period, the share of tangible investment had slowed down as intangible investment was slightly increasing.

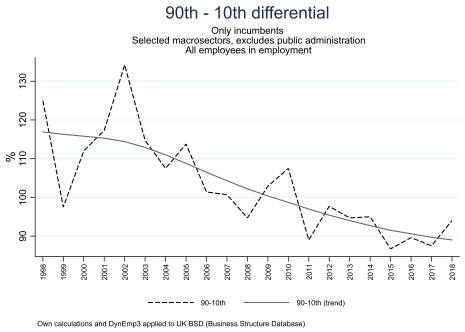
For the same period, except the most recent years available in the sample, the dispersion differentials of employment growth rates exhibited a downward trend (see Figure 21b). Notice that by the end of the period, as the intensity of intangibles starts declining, the inverse occurs with the dispersion differential.

FIGURE 21
Share of intangible investments and dispersion of employment growth

(a) Intangibles intensity



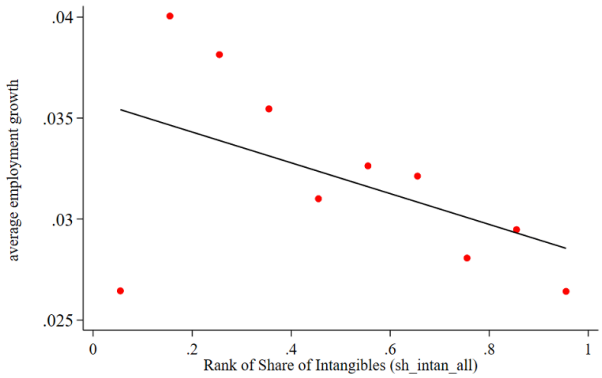
(b) Dispersion



Note: Author's calculations

A simple correlation associates these two variables negatively. Figure 22 shows a rank regression for levels of Intangibles intensity against average employment growth. The effects of intangible intensity are plotted in ascending rank order. As expected, there is a negative correlation between the dispersion differential and the intangibles' intensity.

FIGURE 22
Employment Growth Rate Rank



Notes: Binned scatterplots provide a non-parametric way of visualizing the relationship between two variables. It groups the x-axis variable into equal-sized bins, computes the mean of the x-axis and y-axis variables within each bin, and then creates a scatterplot of these data points.

1.6.1.8 Fact INT2. Intangible industries exhibit unaltered employment growth dispersion differentials As discussed in section 1.4.2 of the conceptual framework, it could be argued that substitution and complementary effects may exist between intangibles and employment, varying across sectors. For instance, highly intangible industries could have needed to hire workers to implement their intangible investments. For instance, assets in Artificial Intelligence, big data, and customer relationship management, among others, can require specialized skills for their operation. Similar to firms that would have increased labor demand in the face of a positive productivity shock, they could have started to demand more labor, but only of the skilled type. Furthermore, specific industries that are highly intangible could have started to absorb positive shocks to productivity through other mechanisms entirely, except with labor.

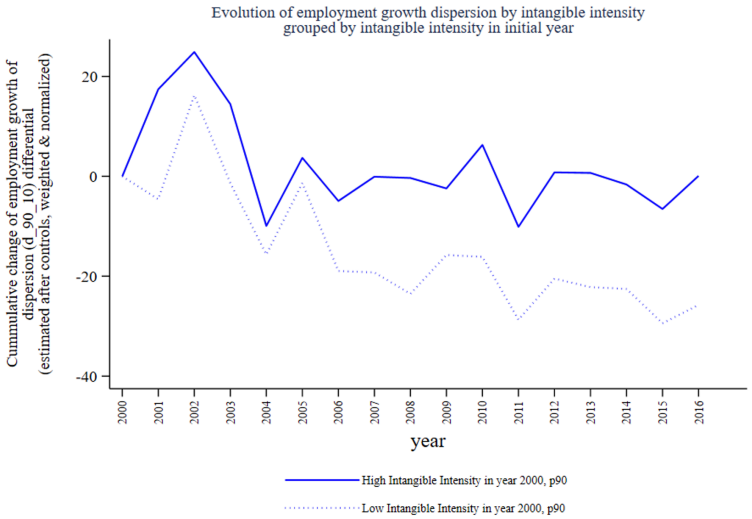
To explore this intuition, apart from what has been already hinted by Figure 12 for the Information Sector, Figure 23 plots the evolution of the dispersion differential over time by the sectors' level of intangible intensity. Panel a considers industries as highly intensive in intangibles, as those with a share of intangibles investment in the 90th percentile of the distribution in 2000. The threshold in panel b is just above the median. The figure shows that the dispersion differential has remained very similar for highly intangible firms. In contrast, firms in the lowest intensity percentiles have exhibited a decreasing dispersion differential over time.

When considering the 90th percentile as the threshold to classify an industry as highly intangible intensive, the only sector that qualifies is the professional services sector. When the 50th is the threshold, besides professional services, the group is also integrated by industries such as administration and support, education, finance and insurance, information, manufacturing, other services, and wholesale and retail.

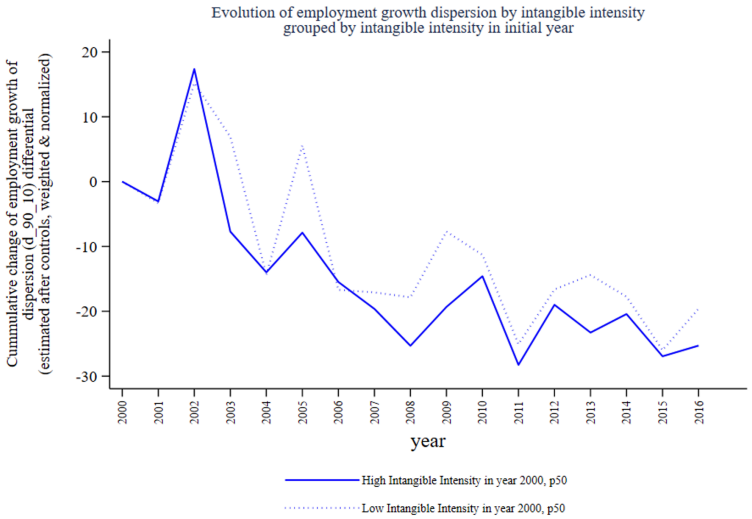
This suggests that a knowledge-intensive, services-oriented industry, such as professional services, must keep hiring labor since this sector is, in essence, worker-intensive. The rest of the economy, including the information sector, does not have to keep increasing its workforce since it may find intangibles to substitute workers (i.e., software) or respond via investment in complementary tangible assets such as computers and other hardware. Furthermore, if they respond via more workers, they tend to hire more of the skilled kind since those possess the needed skills to operate such knowledge. The effect of hiring more workers in these sectors can be, however, small. Most of the workforce comprises unskilled workers, so the aggregate impact on employment growth is not visible. Indeed, evidence in the U.S. shows that the share of skilled labor income shares out of the total revenues pie of companies is increasing relative to unskilled labor. Thus, since unskilled labor is affected and is the majority of the workforce pie, aggregate employment growth rates decline.

FIGURE 23
Employment growth by intangible intensity level

(a) Percentile 90



(b) Percentile 50



Notes: The graph plots the evolution of employment growth dispersion differentials between the 90-10th percentiles for high and low intangible-intensive industries after controlling for other factors driving employment growth dispersion, including wages, industry sector, and capital-labor ratios. Industries are ranked by their intensity of intangible investment in the year 2000. Industries above the median or the 90th percentile are classified as *High intangibles intensity*, and industries below that median or 90th percentile as *Low intangibles intensity*. Averages weighted by gross output across two-digit industries are shown for both groups, normalized to 0 in the starting year. The period is 2000-16. Employment growth dispersion differentials are measured as the 90-10 difference in employment growth rates, the difference in employment growth between firms at the 90th percentile of the employment growth distribution in the industry, and firms at the 10th percentile. The vertical axes represent log-point differences from the starting year: for instance, in figure a, employment growth dispersion differentials in the high intangible intensity group have just around 0 in the final year, which corresponds to approximately 0% higher employment growth dispersion differential in 2016 compared to 2000. In contrast, the low intangible intensity percentage has increased by 18%. Source: Authors' calculations based on DynEmp3 created indicators for the UK (October 2020) and INTAN-Invest database.

1.6.1.9 Fact INT3. High-growth firms shrink in sectors where intangible intensity rises What about the relationship between intangibles intensity and HGFs?

Figures 24a and 24b present the average intangibles intensity by Level 1 (Table 2) of macro sector disaggregation from 1997 to 2016. Panel a shows that the share of investment in intangibles over value-added is greater for manufacturing (blue boxes) than for services excluding FIRE and information. Panel b exhibits similar patterns for the share of investment in intangibles over total investment. Figures 25a and 25b show similar patterns at Level 2 of industry aggregation, where services include FIRE and information.

FIGURE 24
Average Intangible Investment Intensity by Industry Sectors (Level 1)

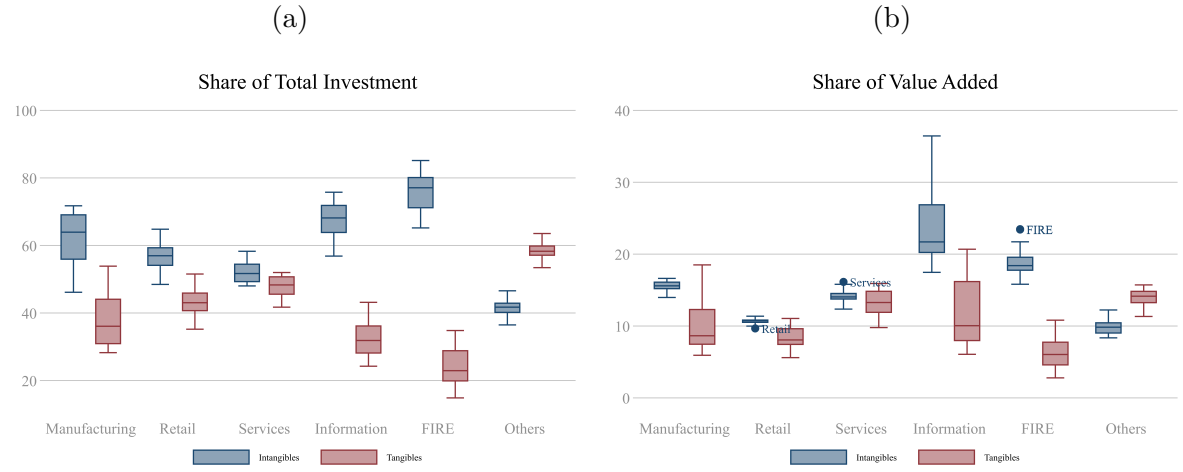
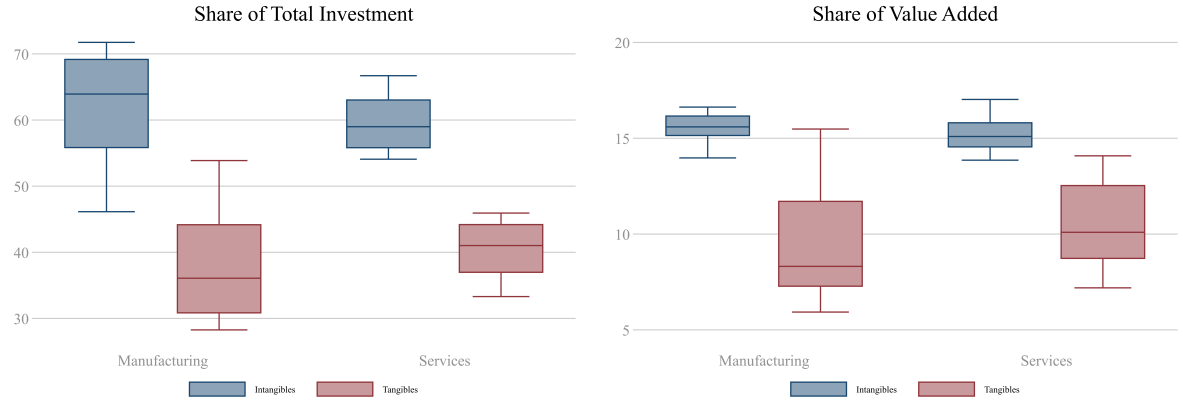


FIGURE 25
Average Intangible Investment Intensity by Industry Sectors (Level 2)

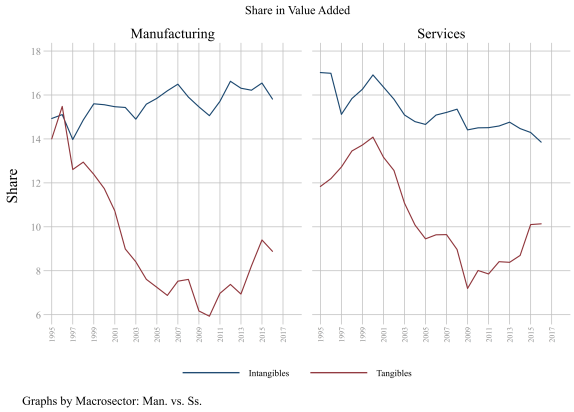


Note: The figures present box plots representing the average intangible intensity, following the definitions of equation 16 for the 1998-2016 period and by levels 1 and 2 of industry sector disaggregation as presented in Table 2. The mean and variation are estimated across years for each aggregated sector. Lines represent confidence intervals at 97.5%.

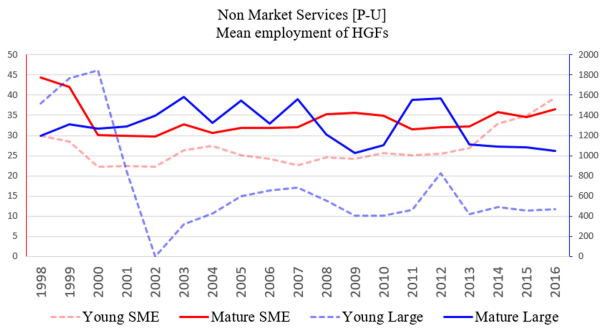
Likewise, Figure 26a shows that manufacturing's intangible intensity has been rising, whereas, in services, it has declined.

FIGURE 26
Sectors

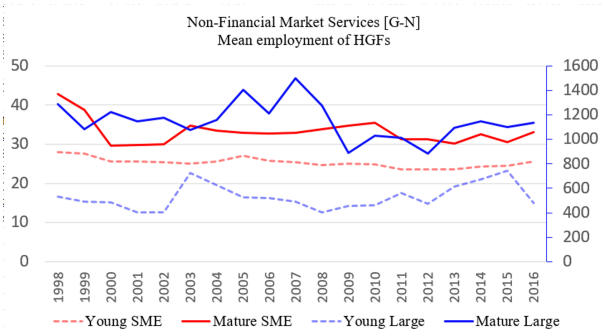
(a) Intangible intensity by sector



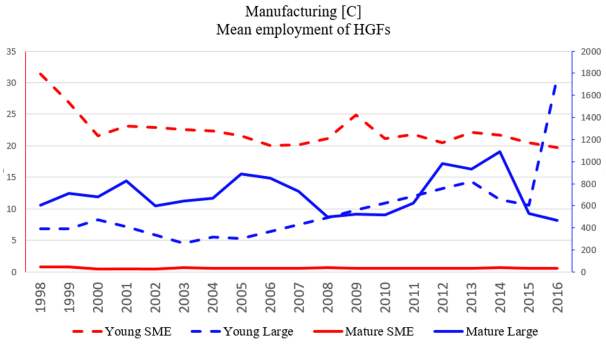
(b)



(c)



(d)

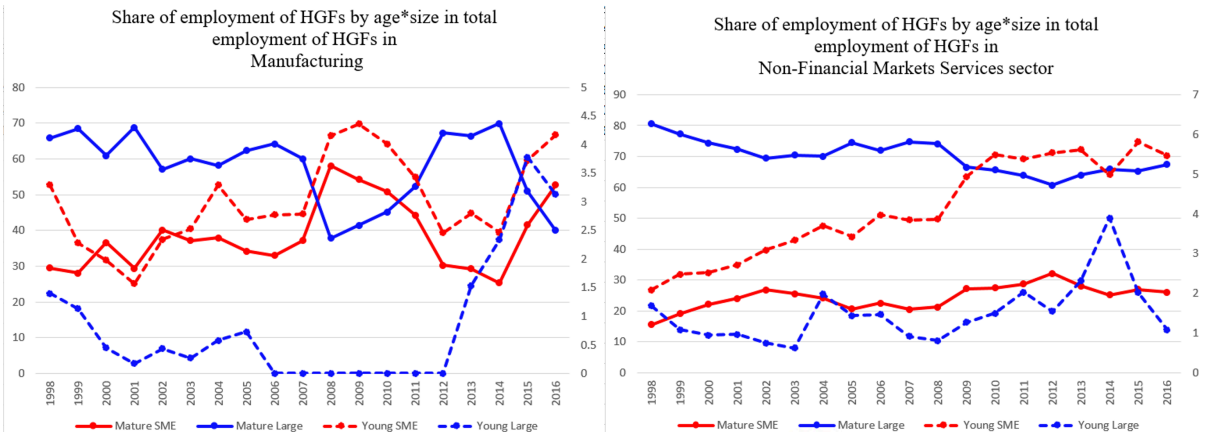


Note: Panel (a) the shares of intangible intensity over time for manufacturing and services. Panels (b), (c), and (d) show levels of mean employment of High Growth Firms for manufacturing and services. Points of Panel d from 2006 to 2012 contain linear interpolations of actual data.

Furthermore, the mean employment size of high-growth young SMEs in manufacturing (Figure 26d) seems to be shrinking more than in the services sector (Figures 26b and 26c). Thus, in industries where intangibles are rising (like in manufacturing as shown in Figure 26a), their high-growth young SMEs are becoming smaller and losing participation in the total employment contribution of HGFs.

FIGURE 27
Share of Employment of High Growth Firms

(a) (b)



Note: The Y-axis is expressed in %; the primary axis on the left represents percentages for young firms. The secondary Y-axis on the right represents values for large firms. Values equal to zero indicate that the sector did not have HGFs in that age×size group for that particular year, so it didn't contribute to any employment.

The same is observed when checking the evolution of the share of employment by age×size group of the high-growth firms. Figure 27 shows that in manufacturing, high-growth young SMEs' share has fluctuated around the same levels (see panel a), whereas, in the services sector, where the intangibles intensity has slightly declined, the share has increased (see panel b). This aligns with evidence from the Office of National Statistics that reports manufacturing, along with information and financial services, as the sectors consolidating as the largest investors in intangibles by 2020, especially in R&D, own-account software, and own-account organizational capital (Office for National Statistics, 2022).

Such findings may indicate a shift from labor to intangible investments, but the impact on aggregate employment growth is still weak if there is some complementarity between the two. Manufacturing might need more skilled workers as it becomes more intangible. However, given the low proportion of skilled workers relative to unskilled workers in the overall economy and that manufacturing represents just about 20% of the U.K. economy, the substitution effect kills the complementarity. Furthermore, it does not matter much that the services sector exhibits increases in the employment share or the

average employment size of this sector's high-growth young SMEs. Despite this sector constituting most of the U.K. economy and employing most people, this is not a positive sign of substantial and positive dynamics in employment.

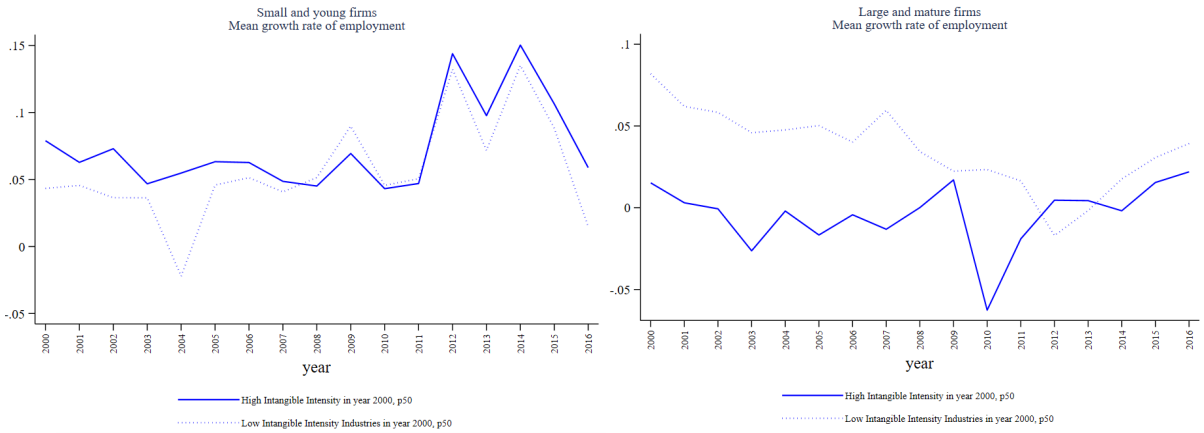
As shown in section 1.6.1.5, the increasing trends in employment share are only observed in the ICT-intensive or knowledge-intensive portion of the services sector, but this still constitutes the minority of the sector. Most of the sector provides essential services; despite having turned to more Knowledge-Intensive Services (KIS) in recent years, as the Office on National Statistics reported, essential services still represent 61% percent of the services sector[48]. KIS refers to "services which require a high proportion of high-level skills and inputs of research and development (R&D), such financial services, business services, and professional, scientific and technical services (Office for National Statistics, 2023a). Hence, the complementarity and, therefore, the positive effect that the ICT services sector might trigger is insufficient to push aggregate figures.

Figure 28 aligns with this story. The average employment growth rate of larger and mature firms has declined for large and mature firms with low intangible intensity. In contrast, it has remained stable for high-intangible intensity over recent years. Combining the fact that highly intangible firms are required to complement their assets with skilled labor and that larger and mature firms tend to grow less in employment, their employment growth has been more prone to respond to shocks to productivity by churning jobs (e.g. replacing unskilled jobs with more intangibles, because they can keep investing in them).

FIGURE 28
Employment Growth by
Intangible Intensity and Firm Size

(a) Mean growth of employment by sector's intangibles intensity

(b) Mean growth of employment by sector's intangibles intensity



Note: The graph plots the evolution of the mean employment growth rates for high and low intangible-intensive industries by company size after controlling for other factors driving employment growth dispersion, including wages, industry sector, and capital-labor ratios. Industries are ranked by their intensity of intangible investment in the year 2000. Industries above the median or the 90th percentile are classified as *High intangibles intensity* and industries below that median or 90th percentile as *Low intangibles intensity*. Averages weighted by gross output across two-digit industries are shown for both groups, normalized to 0 in the starting year. The period is 2000-16.
Employment growth rates are presented on axis Y. The vertical axes represent log-point differences from the starting year: for instance, panel a shows that employment growth rates for small and young firms in the high intangible intensity group are around 0 in the final year, which corresponds to approximately 0% higher employment growth rates in 2016 compared to 2000; similar to the low intangible intensity group. In contrast, panel b shows that the employment growth rates for the low intangible intensity percentage had declined by 10%. Source: Authors' calculations based on DynEmp3 created indicators for the UK (October 2020) and INTAN-Invest database.

1.6.2 Dynamic Panel Model Estimations

This section explores the impact of intangible intensity on business dynamism indicators in a more formal setting by applying econometric techniques. The regression analysis consists of putting DynEmp3's business dynamics measures as dependent variables against shares of intangible investments and control variables.

The expectation is that employment-based business dynamics variables (e.g., dispersion of employment growth rate) adjust with delay to changes in intangible investment because hiring or firing workers can be costly. Likewise, intangibles represent a certain degree of sunkness themselves. Changes to employment may arise from changes to firms' intangible assets, capital stock, or total investment ratios to labor and(or) wages.

The employment adjustment may depend on the passage of time. For instance, a company might take time to adapt to new investments in artificial intelligence (AI) since it needs

to train the existing workers or hire more workers with relevant skills to operate such assets. Similarly, the firm may take some time to decide to fire workers since it first requires implementing the new investment in AI and evaluating whether the new systems can replace manned labor.

To capture these types of delays or frictions, the specification of the empirical model includes lagged versions of intangibles and other factors of production as regressors. The first difference of the employment-based business dynamics variable is also included. The latter argues for a dynamic model in which the dependent variable's lags are also incorporated as regressors.

1.6.3 General regression and estimation methods

The following is a general expression to introduce the concept above of delayed adjustments in business dynamic variables based on a general dynamic panel model:

$$Y_{c,t}^{s} = \sum_{j=1}^{p_y} \alpha_j y_{c,t-j}^{s} + \sum_{j=0}^{p_x} \overrightarrow{\mathrm{x}}_{c,t-j}^{s} \overrightarrow{\beta}_j + \epsilon_{c,t}^{s} \tag{19}$$

For $s = \{1, ..., N\}$ and $t = \{1, ..., T\}$, $\epsilon_{c,t}^{s} = \mu_s + \vartheta_{c,t}^{c}$, and we assume that $E(\mu_s) = E(\vartheta_{st}) = E(\mu_s \vartheta_{c,t}^{c}) = 0$. The estimated parameters $\alpha_j, ..., p_y$ and p_x correspond to the lagged dependent (or independent) variable; β_j is a vector of parameters to be estimated of dimensions, $k_1 \times 1$; and, $x_{c,t}^{c}$ is a vector of $1 \times k_1$ of covariates.

$y_{c,t}^{s}$ takes the value of a series of **business dynamics variables** by industry sector s and year t:

$$y_{c,t}^{s} \begin{cases} D_{c,t}^{s} : \text{Dispersion differential from equation 15 section 1.2.7} \\ ShEmpl_{c,t}^{s} : \text{Share of employment from equation 21 below} \\ JRR_{c,t}^{s} : \text{Job Reallocation Rate from equation 8} \\ ENR_{c,t}^{s} : \text{Entry Rate in terms of number of firms from equation 10} \\ EXR_{c,t}^{s} : \text{Exit Rate in terms of number of firms from equation 11} \\ ENRE_{c,t}^{s} : \text{Entry Rate in terms of employment from equation 12} \\ EXRE_{c,t}^{s} : \text{Exit Rate in terms of employment from equation 13} \end{cases} \tag{20}$$

For example, let's consider as the value of $Y_{c,t}^{s} = D_{c,t}^{s}$, so the dependent variable for this dynamic regression will be the employment growth dispersion differential for the industry sector s and year t, where c represents one of the types in the set $c = \{90-10, 50-10, 90-50\}$.

Another example is the share of employment by size and age at a given year t, following the definition for section 1.6.1.1:

$$ShEmpl_{c,t} = \frac{E_{c,t}}{E_t} * 100 \tag{21}$$

Where E_t is total employment in year t, $E_{c,t} = \sum_{i \in c} E_{i,t}$ or total employment correspondent to firms belonging to a combined category c of age and size at year t (e.g., $c = \{sz, ag, sz * ag\}$ where $ag = \{\text{age} < 6, \text{age} >= 6\}$ and $sz = \{\text{employees} < 250, \text{employees} >= 250\}$).

For variables $JRR^s_{c,t}$, $ENR^s_{c,t}$, $EXR^s_{c,t}$, $ENRE^s_{c,t}$ and $EXRE^s_{c,t}$, c indicates the whole set of companies for the sector. Some expressions omit the subscript c and add the s to simplify the notation to identify the sector.

The covariates in the regression expression of equation 19 can be of three types. First, **strictly exogenous**[49] where $E\left[\vartheta^s_{c,t} \mid \vec{x}^s_{c0}, \vec{x}^s_{c1,\cdots,}\vec{x}^s_{c,T}\right] = 0$, **predetermined or weakly exogenous** where $E\left[\vartheta^s_{c,t} \mid x^s_{c0}, x_{c1,\cdots,}x^s_{c,t}\right] = 0$ or, **endogenous** where $E\left[\vartheta^s_{c,t} \mid \vec{x}^s_{c0}, \vec{x}_{c1}, \ldots, \vec{x}^s_{c,t-1}\right] = 0$. Here the idiosyncratic error $\vartheta^s_{c,t}$ should be serially uncorrelated. The fixed effects μ_s can be correlated with the regressors $\vec{x}^s_{c,t-j}$.

In the present setting, the covariates in the vector of regressors $\overrightarrow{x}^s_{c,t-j}$, refer to the following: **intangible investment intensity** of sector s at year $t-j$, as defined in equation 16 where $ShIntan^{y,s}_{c,t} = \frac{PIIintan_{s,t}}{PIIintan_{s,t}+PIItan_{s,t}}$ as the intangible intensity or share of nominal intangible investments over the total nominal value of the investment, in percentage terms. **The investment-to-labor ratio** represents the sector's ratio of total investment to several employees as a proxy for substitution between capital and labor. **Wages** represent the sector's labor compensation costs (expressed in natural logarithm)[50].

Notice that the fixed effect μ_s in equation 19 is correlated by construction with the lagged dependent variables $y^s_{c,t-j}$. Consider the case in which there was only one lagged dependent variable to be included as a regressor, then by construction, $y^s_{c,t-1}$ is correlated with fixed effects in the error term, so there is *dynamic panel bias* (Nickell, 1981)—estimating Equation 19 via OLS yields biased and inconsistent results because the method omits the fixed effect μ_s. Introducing a dummy variable (fixed effect, FE) to control for unobserved (time-invariant) heterogeneity yields a biased coefficient. The standard fixed effects model ignores the short panel bias. $y^s_{c,t}$ is a function of the fixed effect so that the lagged dependent variable is correlated with the error term (Baltagi, 2008) and produces inconsistent estimators with T fixed.

There are two complementary ways to work around this endogeneity. First, transform

the data to remove the fixed effects. Second, instrument $y^s_{c,t-j}$ and any other similarly endogenous variables within the model with variables thought uncorrelated with the fixed effects. Difference and System generalized method-of-moments (GMM) models incorporate those strategies. Those are used here to correct the short panel bias and address fixed effects by the industry sector.

The empirical analysis includes the Arellano-Bond (AB), the Blundell Bond (BB) estimators and Long Differences (LD) Techniques plus other specifications as robustness checks. Sections C.1, C.2, and C.3 in the appendix report in detail the characteristics and differences between methodologies.

1.6.4 Regression Results for Dispersion Differentials

Table 4 describes regressions where the dependent variable is the $D^s_{c,t}$, the differential of employment growth rates between different percentile groups in the growth distribution. To find the appropriate number of lags for the dependent and endogenous variables as described for the general model of equation 19, after testing model specifications, the selected equation for this variable is:

$$\begin{aligned} D^s_{c,t} =& \beta_0 D^s_{c,t-1} + \beta_1 ShIntan^{y,s}_{c,t-2} + \\ & \beta_1 InvLabRatio^s_{c,t-1} + \beta_2 invLabRatio^s_{c,t-2} + \\ & \beta_3 Wages^s_{c,t} + \delta_c + \epsilon^s_{c,t} \end{aligned} \tag{22}$$

Where c represents the cell, in this case, is the whole set of firms for the industry s for a given year t. $ShIntan^{y,s}_{c,t-2}$ is the share of the intangible investment expressed as the ratio of intangibles to total investment (variable type y) for sector s at year t, $InvLabRatio^s_{c,t-1}$, and $InvLabRatio^s_{c,t-2}$ are the ratio of total investment to labor in $t-1$ and $t-2$, respectively, $Wages^s_{c,t}$ is the natural logarithm of labor compensation, δ^s_c are fixed effects of the industry and $\epsilon^s_{c,t}$ is an error term.

All regressors are assumed to be predetermined ($X^s_{c,t-1}$, $X^s_{c,t-2}$,... are allowed as instruments for $X^s_{c,t}$) so that the share of intangibles, the total investment to labor ratio and labor costs may not be completely exogenous. Predetermined implies that past disturbances can impact subsequent values of the regressors. For all the following AB estimations, the models are curtailed to avoid over-identification problems in such a way that only $y^s_{c,t-3}$ is used as an instrument for the first difference equation 76 and the lags $x^s_{c,t-1}$,$x^s_{c,t-2}$, $x^s_{c,t-3}$ are used as instruments for all predetermined regressors. For example, in the case of the AB model for the dispersion differential, $\Delta\text{ShIntan}^{y,s}_{c,t-2}$ is instrumented in the first-difference of equation 19, only with $\Delta\text{ShIntan}^{y,s}_{c,t-m}$ for $m = 3, 4, 5$ and not all other further available lags.

TABLE 4
Dispersion of employment growth rates

	(1) D 90th-10th	(2) D 90th-50th	(3) D 50th-10th
L.D 90th-10th	0.158* (0.0946)		
L.D 90th-50th		0.0318 (0.136)	
L.D 50th-10th			-0.0689 (0.0427)
L2.Share of intangibles	-0.415* (0.218)	-0.357** (0.175)	-0.156 (0.141)
Ratio total investment to labor	-0.0597 (0.178)	-0.0526 (0.0951)	-0.120*** (0.0253)
L.Ratio total investment to labor	-0.205** (0.0880)	0.0831 (0.0641)	-0.221*** (0.0549)
L2.Ratio total investment to labor	0.174** (0.0729)	-0.0851 (0.0882)	0.229*** (0.0368)
ln(Labor compensation)	-0.00697 (0.0752)	-0.129** (0.0613)	0.0183 (0.0383)
Constant	1.059 (0.921)	2.063*** (0.761)	0.352 (0.439)
Year FEs	X	X	X
Observations	304	304	304
Degrees of freedom	15	15	15
AR1 (p-value)	-2.525	-2.475	-2.390
AR2 (p-value)	-0.717	-0.177	-0.512
Sargan (p-value)	.	.	.
Chi-2 statistic	659.6	1752.4	311.5

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
* p<0.10, ** p<0.05, *** p<0.010

Note: - The table presents AB estimations for dispersion differentials for predetermined specifications with curtailed instruments.
- The dependent variable corresponds to the respective dispersion differential that appears as the column's title. The *xtdpd* command in STATA is used for the estimations.
- Sargan test appears null when assuming robust errors for estimations. Arellano and Bond (1991) show that the one-step Sargan test over-rejects in the presence of heteroskedasticity. Because its asymptotic distribution is not known under the assumptions of the vce(robust) model, the *xtdpd* command in STATA does not compute it when vce(robust) is specified.
- The error between predetermined and endogenous specifications does not differ much. We compare a full incorporation of GMM-type instruments into the difference equations (all lags available for each variable) versus a curtailed option limiting the lags used as an instrument per variable. The ar(2) tests for serial correlation amongst errors of order two and above, and the Sargan test for specification remain non-significant after this curtailing. The only one-step estimators with robust errors to control for heteroskedasticity are also reported. We employ xtabond in STATA for a second set of estimations to compare more easily before AB and BB estimators for all business dynamics variables. We also consider cases in which partial or entire endogeneity is allowed, and we use xtabond and xtdpds to allow for these differences.

Table 4 shows that the coefficient for $\text{ShIntan}_{c,t-2}^{y,s}$ is negative and significant for the 90-10 and the 90-50 differentials (columns 1 and 2) but not for the 50-10 differential (column 3). This result indicates a significant negative lagged effect of the investment in intangibles

on dispersion; thus, a higher level of intangibles-intensity is associated with a decline in dispersion differential. The 90-10 decline is attributed to a decline in the pace of growth of the fastest growth firms, which are growing more like the median firm because the coefficient 90-50 differential is significant and negative while the coefficient for the 50-10 differential is insignificant.

TABLE 5
Dispersion Differentials 90-10

	(1) Arellano-Bond (exog.) int.	(2) Arellano-Bond (exog.)	(3) Arellano-Bond (pred.)	(4) Arellano-Bond (endo.)	(5) ABv/BB (exog.)	(6) ABv/BB (pred.)	(7) ABv/BB (exog.)
L.D 90th-10th	0.214***	0.197***	0.207***	0.189***	0.290***	0.388***	0.389***
	(0.0586)	(0.0631)	(0.0544)	(0.0565)	(0.0570)	(0.0710)	(0.0670)
Share of intangibles	0.0445	-0.199	-0.282	-0.312	0.0513	-0.0104	0.0404
	(0.393)	(0.217)	(0.279)	(0.263)	(0.184)	(0.290)	(0.231)
L.Share of intangibles	0.211	0.0373	0.0434	0.0106	0.222	0.455	0.360
	(0.255)	(0.219)	(0.218)	(0.227)	(0.288)	(0.331)	(0.279)
L2.Share of intangibles	-0.622***	-0.442**	-0.543**	-0.606***	-0.351**	-0.188	-0.157
	(0.240)	(0.197)	(0.216)	(0.224)	(0.163)	(0.209)	(0.169)
Ratio total investment to labor		-0.0172	-0.0627	-0.0543	-0.0527	-0.190	-0.102
		(0.217)	(0.181)	(0.176)	(0.225)	(0.121)	(0.137)
L.Ratio total investment to labor		-0.214**	-0.228**	-0.240***	-0.229	-0.101	-0.140
		(0.106)	(0.0938)	(0.0911)	(0.155)	(0.105)	(0.0977)
L2.Ratio total investment to labor		0.217	0.234**	0.226***	0.233	0.270***	0.287***
		(0.138)	(0.0970)	(0.0869)	(0.142)	(0.0782)	(0.0924)
ln(Labor compensation)		-0.448***	-0.273*	-0.239*	-0.248	-0.220	-0.389**
		(0.151)	(0.152)	(0.145)	(0.201)	(0.209)	(0.191)
L.ln(Labor compensation)		0.136	0.144	0.0934	0.279	0.479	0.231
		(0.192)	(0.292)	(0.302)	(0.208)	(0.376)	(0.384)
L2.ln(Labor compensation)			-0.0109	0.0449		-0.312	0.129
			(0.172)	(0.177)		(0.238)	(0.256)
Constant	37.86***	17.50*	29.60***	32.66***	45.98***	35.32***	40.86***
	(6.281)	(9.754)	(6.778)	(7.916)	(6.668)	(8.696)	(7.667)
Year FEs	X	X	X	X	X	X	X
Observations	272	272	272	272	288	288	288
Degrees of freedom	15	15	16	15	15	15	15
AR1 (p-value)	-2.782	-2.747	-2.944	-2.826	-2.765	-3.066	-2.970
AR2 (p-value)	-0.877	-0.995	-0.948	-1.055	-0.626	-0.425	-0.243
Chi-2 statistic	494.6	1630.8	543.8	217.5	64.90	2190.6	1054.7

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods. Dependent variable is: D 90th-10th
ABv/BB: Arellano-Bover/Blundell-Bond
* p<0.10, ** p<0.05, *** p<0.010

Note: Robust standard errors are in parentheses.

Table 5 presents variations of the AB and BB models in a general setting (not excluding lags per the curtailed, primary model for dispersion as explained above). The table shows that the 90-10 differential declines as the share of intangibles increases in most models, except for models (5) and (6) for the predetermined and exogenous cases with the BB specification. Table 6 presents results contrasting dynamic panel estimations with basic methods such as OLS and Fixed Effects. The negative effect of the lagged share of intangibles on the 90-10 differential remains significant throughout the models, except for the BB with the endogenous assumption. The case is less evident for the 90-50 differential.

TABLE 6
Dispersion Differentials 90-10 and 90-50

(a) 90th-10th

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.D 90th-10th			0.550***	0.201***	0.179***	0.733***
			(0.0588)	(0.0611)	(0.0478)	(0.0918)
L2.D 90th-10th			0.270***	-0.0519		
			(0.0594)	(0.0610)		
Share of intangibles	0.205**	-0.357	0.128	-0.256	-0.160	0.514*
	(0.101)	(0.254)	(0.273)	(0.282)	(0.265)	(0.250)
L.Share of intangibles			0.374	0.253		
			(0.346)	(0.300)		
L2.Share of intangibles			-0.478*	-0.762***	-0.763**	-0.400
			(0.270)	(0.256)	(0.286)	(0.245)
Ratio total investment to labor	-0.265***	-0.00386	-0.0443	-0.0132	-0.0309	-0.102
	(0.0687)	(0.0895)	(0.0434)	(0.0864)	(0.229)	(0.282)
L.Ratio total investment to labor					-0.0451	0.0432
					(0.0775)	(0.290)
ln(Labor compensation)	-0.165***	-0.170	-0.0296	-0.187	-0.180	-0.0433
	(0.0272)	(0.116)	(0.0181)	(0.113)	(0.128)	(0.0336)
Constant	2.811***	2.881**	0.523**	3.111**		0.766*
	(0.302)	(1.277)	(0.221)	(1.265)		(0.366)
Year FEs		X	X	X	X	X
Observations	272	272	272	272	272	272
R-squared (adj.)	0.111	0.372	0.674	0.416		
Degrees of freedom	3	34	23	38	0	22
AR1 (p-value)					0.00432	0.00202
AR2 (p-value)					0.317	0.624
Sargan (p-value)					0.0896	0.710
Hansen (p-value)					1	1
F statistic	12.28	10.22	25.37	10.04	.	288.7

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. Indicates lagged variable 2 periods. Dependent variable is D 90th-10th
* p<0.10, ** p<0.05, *** p<0.010

(b) 90th-50th

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.D 90th-50th			0.326***	0.00851	0.0260	0.488***
			(0.0600)	(0.0627)	(0.0967)	(0.133)
L2.D 90th-50th			0.309***	0.00896		
			(0.0599)	(0.0626)		
Share of intangibles	0.0727	-0.467**	-0.265	-0.513**	-0.417	-0.0473
	(0.0668)	(0.217)	(0.236)	(0.250)	(0.251)	(0.135)
L.Share of intangibles			0.461	0.276		
			(0.299)	(0.266)		
L2.Share of intangibles			-0.180	-0.410*	-0.344	0.0851
			(0.234)	(0.226)	(0.200)	(0.151)
Ratio total investment to labor	-0.220***	-0.00286	-0.0711*	-0.00872	-0.0556	-0.184
	(0.0455)	(0.0764)	(0.0383)	(0.0768)	(0.152)	(0.180)
L.Ratio total investment to labor					0.0182	0.0584
					(0.0753)	(0.180)
ln(Labor compensation)	-0.110***	-0.141	-0.0341**	-0.153	-0.127	-0.0478
	(0.0180)	(0.0990)	(0.0158)	(0.100)	(0.114)	(0.0303)
Constant	1.817***	2.205**	0.614***	2.418**		0.827**
	(0.200)	(1.090)	(0.190)	(1.123)		(0.342)
Year FEs		X	X	X	X	X
Observations	272	272	272	272	272	272
R-squared (adj.)	0.117	0.256	0.449	0.254		
Degrees of freedom	3	34	23	38	0	22
AR1 (p-value)					0.00448	0.00159
AR2 (p-value)					0.757	0.299
Sargan (p-value)					0.151	0.696
Hansen (p-value)					1	1
F statistic	12.97	6.695	10.60	5.662	.	12.28

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods. Dependent variable is D 90th-50th.
* p<0.10, ** p<0.05, *** p<0.010

Note: The sample is the same for all models depending on model (6).

Finally, table 7 presents results for the long differences estimators for the 90th-10th dispersion differential. The LD4 Four Period Long Differences approach replicates the Huang and Ritter (2009) implementation of Hahn et al. (2007)'s estimator. These authors consider several differencing windows, but in the present study, only the 4-period variation is included in the simulations because it excludes the fewest sectors.

TABLE 7
Dispersion 90-10 Long Differences

	(1) D 90th-10th([t] - [t-17])	(2) D 90th-10th([t] - [t-16])	(3) D 90th-10th([t] - [t-15])	(4) LD4 = D 90th-10th([t] - [t-5])
L.D 90th-10th([t] - [t-17])	-0.210			
	(0.726)			
L.D 90th-10th([t] - [t-16])		0.621*		
		(0.322)		
L.D 90th-10th([t] - [t-15])			0.227	
			(0.159)	
L.D 90th-10th([t] - [t-4])				0.0574
				(0.134)
Share of intangibles ([t] - [t-17])	-1.568			
	(1.188)			
Share of intangibles ([t] - [t-16])		-2.891***		
		(0.912)		
Share of intangibles ([t] - [t-15])			-1.705**	
			(0.841)	
Share of intangibles ([t] - [t-4])				-0.207
				(0.262)
ln(labor compensation) ([t] - [t-17])	-0.117			
	(0.323)			
ln(labor compensation) ([t] - [t-16])		-0.296		
		(0.224)		
ln(labor compensation) ([t] - [t-15])			0.210	
			(0.178)	
ln(labor compensation) ([t] - [t-4])				-0.487***
				(0.144)
Ratio total investment to labor ([t] - [t-17])	-0.582			
	(0.364)			
Ratio total investment to labor ([t] - [t-16])		-0.584**		
		(0.229)		
Ratio total investment to labor ([t] - [t-15])			-0.177	
			(0.170)	
Ratio total investment to labor ([t] - [t-4])				-0.328***
				(0.111)
Constant	-0.0680	-0.0344	-0.259**	0.0221
	(0.328)	(0.164)	(0.131)	(0.0304)
Observations	16	16	16	224
R-squared (adj.)	0.0495	0.552	0.297	0.0677
Degrees of freedom	4	4	4	4
Chi-2 statistic	8.411	29.03	14.90	18.16

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods. Dependent variable is: d_90_10
* p<0.10, ** p<0.05, *** p<0.010

Note:
$D_{90th-10th}[t] - [t-17]$: Long Difference First Iteration. Difference between the dispersion differential in year $t = 2016$, which is the last year for which we have intangible investments information, and the dispersion differential in year $t = 2016 - 17 - 1 = 1998$.
$D_{90th-10th}[t] - [t-16]$: Long Difference Second Iteration.
$D_{90th-10th}[t] - [t-15]$: Long Difference Third Iteration.
$D_{90th-10th}[t] - [t-5]$: All years between t and $t-5$.

The long difference estimations display a negative relationship between intangible intensity and dispersion differentials in models described in columns (2) and (3), which contain the farthest iterations allowed by the sample size (only 18 years of lags). Appendix C.4 shows this negative relationship is more significant for the 90-50 differential. The findings also suggest the existence of a delay in the response of the dependent variable.

To explore whether this decline is associated with a particular group of firms that have become more prevalent than others, table 8 presents regression results for each age*size class of firms. Each column represents a model in which the dependent variable is the share that a specific age*class group of firms has of the total number of firms or from the total employment for the year. Panel 8a presents results using the AB method. It indicates only a slightly significant impact of intangibles intensity in column (3), a model corresponding to the share of firms in the mature*large group out of total firms for the

year.

TABLE 8
Share of firms by age*size class

(a) AB

	(1) Share of young*self-empl. firms	(2) Share of young*SME firms	(3) Share of mature*large firms	(4) Share of empl. by young*SME	(5) Share of empl. by mature*large
L2.Share of intangibles	-0.592	-1.447	-0.258*	-0.490	-5.612
	(4.761)	(5.677)	(0.149)	(1.802)	(5.459)
Ratio total investment to labor	2.021	1.848*	-0.0631	0.674	-0.341
	(2.016)	(0.991)	(0.107)	(0.507)	(1.162)
L.Ratio total investment to labor	-0.356	0.105	0.0452	-0.774*	4.598***
	(0.708)	(0.751)	(0.128)	(0.426)	(1.178)
L2.Ratio total investment to labor	0.000553	1.020	0.0125	1.698**	-3.216
	(1.291)	(1.400)	(0.0630)	(0.724)	(1.968)
ln(Labor compensation)	3.567***	3.331**	-0.0254	3.815***	-3.617*
	(1.012)	(1.678)	(0.0920)	(0.642)	(2.087)
L.Share of young*self-empl. firms	0.894***				
	(0.0755)				
L.Share of young*SME firms		0.425***			
		(0.146)			
L.Share of mature*large firms			0.956***		
			(0.0180)		
L.Share of empl. by young*SME				0.759***	
				(0.0779)	
L.Share of empl. by mature*large					0.423***
					(0.114)
Constant	-34.44***	-22.92	0.449	-38.40***	70.00***
	(8.557)	(17.34)	(1.010)	(7.016)	(21.35)
Year FEs	X	X	X	X	X
Observations	304	304	304	304	304
Degrees of freedom	15	15	15	15	15
AR1 (p-value)	-2.375	0.325	-1.479	-1.662	-1.753
AR2 (p-value)	-1.058	0.130	1.357	0.348	1.529
Sargan (p-value)	.	.	.	.	.
Chi-2 statistic	447.4	433.4	795.9	3155.2	75.15

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Method: Arellano-Bond, predetermined and curtailed

(b) BB

	(1) Share of young*self-empl. firms	(2) Share of young*SME firms	(3) Share of mature*large firms	(4) Share of empl. by young*SME	(5) Share of empl. by mature*large
L2.Share of intangibles	-5.017**	3.890	0.241	-0.338	-3.525
	(2.122)	(2.647)	(0.273)	(1.056)	(6.264)
Ratio total investment to labor	1.996	0.295	-0.0439	0.126	0.0206
	(1.420)	(1.160)	(0.135)	(0.360)	(1.304)
L.Ratio total investment to labor	-0.151	-0.767	0.0212	-1.274***	4.029**
	(0.624)	(0.789)	(0.156)	(0.381)	(2.018)
L2.Ratio total investment to labor	-0.293	2.085	0.132	1.401***	-5.626
	(1.008)	(1.547)	(0.0972)	(0.438)	(3.927)
ln(Labor compensation)	1.509***	0.0780	0.0444	0.175	-0.434
	(0.469)	(0.799)	(0.0412)	(0.243)	(0.880)
L.Share of young*self-empl. firms	0.988***				
	(0.0326)				
L.Share of young*SME firms		0.671***			
		(0.173)			
L.Share of mature*large firms			0.939***		
			(0.00959)		
L.Share of empl. by young*SME				0.990***	
				(0.0128)	
L.Share of empl. by mature*large					0.960***
					(0.0279)
Constant	-282.3***	-525.1***	14.80*	-362.2***	747.2**
	(88.62)	(152.2)	(8.910)	(99.47)	(319.0)
Year FEs	X	X	X	X	X
Observations	304	304	304	304	304
Degrees of freedom	15	15	15	16	15
AR1 (p-value)	-1.932	-0.226	-1.572	-1.747	-1.555
AR2 (p-value)	-1.435	-0.152	1.555	0.521	1.380
Sargan (p-value)	.	.	.	.	.
Chi-2 statistic	637.3	929.8	407.3	1.24332e+09	122.1

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Method: Blundell-Bond, predetermined and curtailed
* p<0.10, ** p<0.05, *** p<0.010

Note: The dependent variable for each model is as titled in each column, corresponding to the share of firms or employment of firms belonging to each age*size class in the total for the year.
Self-employment: firms (1 person)
Young: <6 years old
Mature: >= 6 years old
SMEs: 2 to 249 employees
Large: 250+ employees

This negative impact indicates that the more intangible the intensity in a sector, the lower the share of mature and large firms in the economy. This goes by the decline in the importance of big and mature firms in the total number of firms, as presented in the stylized facts. Panel 8b shows BB estimates for the same shares. It reveals that the only model with a negative effect is the model in column (2), which corresponds to the share of young (<6 years old) and self-employment firms (1 employer in employment = only the owner and zero persons engaged). Note that both of these effects are in terms of units of firms. It is not possible to observe a significant impact in terms of employment.

Appendix C.7 presents additional different estimations using more lags and generic methods.[51] In summary, the results for dispersion differentials tend to be more robust than for the rest of the business dynamics variables, as the results seem consistent across various specifications of the dynamic equations. The results for other business dynamics variables vary depending on the models' specifications. See appendices C.6 and C.7.

1.7 Motivating model: employment dynamism and intangibles

To discipline the empirical evidence in this study, this section details a proposed model that fits within the canonical models of firm dynamics with adjustment frictions. The model argues that when there are productivity shocks (TFP), there are *frictions* associated with the intensity of intangible investments of a firm that either delay or prevent adjustments of the company's size in terms of employment.

These models yield decision rules for company size growth (employment, capital, or investment) that are a function of the two key state variables in each period: the realization of productivity and the initial size of the firm. In a reduced form representation of the decision rule, we have size growth$_{i,t} = f_t\left(A_{i,t}, n_{i,t-1}\right)$, where size growth is any size growth, like the number of employees growth between t and t-1. $n_{i,t-1}$ is the initial size and $A_{i,t}$ is the realization of productivity or idiosyncratic productivity of firm i.

As in Decker et al. (2016c), the proposed model assumes that changes in the distribution of firm growth rates stem from either change in the distribution of $A_{i,t}$ (for example, measured as productivity or TFP dispersion) or in the responsiveness of the firm to a given realization of $A_{i,t}$. Suppose within-industry productivity dispersion has remained stable or increased, as in the case of the U.S., and employment growth variables have been declining (particularly in the high-tech sector and young firms). In that case, the size-productivity channel has also been weakened. Thus, I argue that the friction-producing weakening is intangibles-associated. Following the production function used by Bartelsman et al. (2013), the technology production in the proposed model is:

$$Y_{it} = A_{it}\varepsilon_{it}\left(n_{it} - f\right)^{\gamma-\alpha} k_{it}^{\alpha}, \gamma < 1 \tag{23}$$

Where Y_{it} is the output of firm i in period t, A_i is the time-invariant productivity component of the company i, k_{it} is the capital input of firm i at time t, n_{it} is employment at time t, f is intangible frictions on labor expressed in terms of overhead costs, namely the number of employees fixed for a company regardless of company size that does not directly contribute to production (i.e., number of employees in the marketing department). Let's assume for simplicity that f is the same across firms.

Firms are assumed to face a downward-sloping demand arising from a differentiated products environment, and the final good is supposed to be a CES aggregator of intermediate goods produced by individual firms:

$$Y_f = N_i^{(\rho-1)/p} \left(\sum_i Y_i \right)^{1/p} \tag{24}$$

The final goods sector is assumed to be perfectly competitive. This implies an inverse demand for good i was given by

$$P_{it} = P_t \left(\bar{Y_{it}} / Y_{it} \right)^{1-\rho} \tag{25}$$

The profits are to be maximized by each company as follows by combining equations on the final production function in 23 and demand curve from 25:

$$\pi_{it} = (1 - \tau_i - \theta_{it}) \, P_t \overline{Y}_t^{1-\rho} \, [Y_{it}]^{\rho} - w_{it} n_{it} - R_t k_{it} \tag{26}$$

Here, some of the additional distortions included by the original authors are kept; τ is a firm-specific but time-invariant distortion, and θ is a firm and time-specific distortion; both are idiosyncratic distortions that can impact the scale of a business; for instance, laws, regulations or taxes that apply only to specific company sizes, startup costs, etc.

Output Y is replaced using equation 23 in the profits equation 26 for maximization purposes. To simplify the analysis, I assume that:

- Equation 23 does not have a transitory productivity term, ϵ_{it}
- Therefore, the profits equation is specified without the transitory shock, and we can isolate the profit shock $S_i = (1 - \tau_i) \, A_i^{\rho}$ product of the idiosyncratic tax and the productivity shock A_i
- We suppress time subscripts for simplicity, where $K_x \, (w, Y)$ is a function of the equilibrium wage rate,
- Average output and model parameters do not vary across firms.

Then by re-expressing the first-order conditions of the maximization problem, the following is true:

$$(n_i^* - f) = S_i^{\frac{1}{1-\gamma\rho}} \frac{R}{W} \frac{\gamma - \alpha}{\alpha} K_x(w, Y) \tag{27}$$

$$k_i^* = S_i^{\frac{1}{1-\gamma\rho}} K_x(w, Y) \tag{28}$$

Then, the labor demand conditions are transformed in elasticities as in equation 27. Assuming that the parameters at the power of the productivity shock S are equal to a constant β and that the terms accompanying the right-hand side expression are equal also to a constant J, I rewrite the equation in this way:

$$\begin{aligned}
n_i \left(1 - \frac{f}{n_i}\right) &= S_i^{\beta} J \\
\ln n_i + \ln \left(1 - \frac{f}{n_i}\right) &= \beta \ln S_i + \ln J \\
\frac{\partial \ln n_i}{\partial \ln n_i} + \frac{\partial \ln (1 - f/n_i)}{\partial \ln n_i} &= \beta \frac{\partial \ln S_i}{\partial \ln n_i}
\end{aligned}$$

and by definition $\partial \ln n_i = \frac{1}{n_i} \partial n_i$ so,

so I can replace in the equality above and obtain:

$$\begin{aligned}
1 + \frac{\partial \ln (1 - f/n_i)}{\frac{\partial n_i}{n_i}} &= \beta \frac{\partial \ln S_i}{\partial \ln n_i} \\
1 + n_i \frac{\partial \ln (1 - f/n_i)}{\partial n_i} &= \\
1 + n_i \left(-\frac{f}{f n_i - n_i^2}\right) &= \\
1 + n_i \left(-\frac{f}{n_i (f - n_i)}\right) &= \\
1 - \frac{f}{f - n_i} &=
\end{aligned} \tag{29}$$

By reorganizing the last step in the process described in 29, the following expression applies at optimum equilibria:

$$El_{S_i}(n) = \frac{\partial \ln n_i}{\partial \ln S_i} = \beta \left(1 - \frac{f}{n_i}\right) \tag{30}$$

El_S is the elasticity or the relative (percentage) change in the employment level of firm i, n_i, to the relative shift in productivity shock s. $\beta = \frac{1}{1-\rho\gamma}$, is a constant equal to an

expression of all time and firm constant parameters in the model, where $\rho < 1$ and refers to the parameter in the CES aggregator equation of intermediate goods produced by individual firms, and $\gamma < 1$ allowing decreasing returns to scale that the most productive firm (with the best managerial ability for example) takes over the market.

Given equation 30 and according to assumptions, then $\beta > 0$, $\partial ln\left(n_i\right)/\partial ln\left(S_i\right) > 0$ if and only if (iff) $\left(1 - \frac{f}{n_i}\right) > 0$. Since the firm is not entirely intangible (e.g. an entirely artificially intelligent-operated company), overhead labor f cannot be greater than the total number of employees n. Thus, elasticity will never be zero, and any positive productivity shock will boost employment. Yet, the magnitude of the impact will depend on the prevalence of overhead over total labor. The more intangible a firm is, the response in the level of employment is smaller.

This model applies to employment levels in equilibrium, but what is the impact on employment changes over time? Does the distribution of elasticities of labor to productivity shock change over time? To explore these questions, I focus on one business dynamic variable (e.g., dispersion differentials) to conduct a comparative statics analysis at equilibrium and different scenarios.

1.7.1 Discussion: Dispersion Differential 90th-10th

This section presents some potential implications of the proposed model's predicted equilibrium (equation 30) on the shrinking 90th-10th dispersion differential in employment growth rate (equation 15)[52]. The goal of this section is not to provide confident inferences about the exact mechanisms which led to this closing gap. This is out of the scope of this chapter. However, the section aims to inspire future strands of empirical research that take advantage of the implications of a model that in the big picture, seems to fit many of the stylized facts found in business dynamics.

Let's assume that the β is the same for both groups of firms, but overhead costs f are firm-specific. Equation 30 suggests that firms that grow more in employment have a higher elasticity to productivity shocks (or are more responsive via employment) because they face fewer frictions to employment growth, meaning they are less intangible-intensive. Conversely, a firm facing more friction refers to one having higher levels of intangible intensity and it is less responsive to shocks via employment:

Firm at the 90th percentile $\equiv$ High labor elasticity $\equiv$ Low intangible intensity

Firm at the 10th percentile $\equiv$ Low labor elasticity $\equiv$ High intangible intensity

Therefore, it can be established that:

$$\beta\left(1-\left(\frac{f}{n}\right)_{90\text{th}}\right) > \beta\left(1-\left(\frac{f}{n}\right)_{10\text{th}}\right) \tag{31}$$

where $\left(\frac{f}{n}\right)_{90\text{th}}$ denotes the intangible intensity of a firm belonging to the 90th group, and $\left(\frac{f}{n}\right)_{10\text{th}}$, the intensity of a firm belonging to the 10th group. The inequality in this equation 31 holds if and only if:

$$\left(\frac{f}{n}\right)_{90\text{th}} < \left(\frac{f}{n}\right)_{10\text{th}} \tag{32}$$

From inequality 32, one can see, first, that there is an inverse relationship between employment growth and intangibles' intensity, as anticipated in the empirical section. Second, given the empirical findings on declining labor dispersion differential, the gap between $\left(\frac{f}{n}\right)_{90\text{th}}$ and $\left(\frac{f}{n}\right)_{10\text{th}}$ should be also shrinking. To put it differently, for inequality 32 to have turned into more like equality, the levels of intangible intensiveness of the two groups should have become more similar. One of the following three scenarios should have needed to occur for that to happen:

1.7.1.1 Scenario 1 Holding $\left(\frac{f}{n}\right)_{10\text{th}}$ constant, the intangibles intensity of firms growing at the top in employment ($\left(\frac{f}{n}\right)_{90\text{th}}$) increased over time. Why firms growing at the top in employment may experience an increase in intangible intensity?

From the empirical findings, firms at the top percentile of labor growth exhibit behaviours compatible with increasing intangibles in the aggregate economy. The average employment of high-growth firms (HGFs) has declined in recent decades. Furthermore, HGFs are growing employment at a lower rate and have become more mature. To put it differently, the companies that grew the most before, the young and small firms, are no longer as predominant in the group of HGFs. In addition, as aggregate intangible investments have increased, small and mature firms have become the age×size group contributing the most to employment in the U.K. economy.

If more mature, smaller firms tend to be now a more prominent age×size class composing the group of HGFs, growing slower, while intangibles are increasing in a small, mature-dominated economy, under what hypothetical circumstances could these HGFs become more intangible? Perhaps in a world where smaller and older HGFs face fewer frictions for increasing investment in intangibles. Evidence has shown that older firms may have higher credit acceptance rates than younger firms (Araujo and Hambur, 2018; OECD, 2021). Also, older firms can address other market failures that facilitate their access to credit. For instance, mature firms can reduce asymmetric information issues between them and investors because they can demonstrate a more extended history of credit

behaviour.

However, evidence shows that intangible investments are still a low pledgeability asset (except for patents in specific geographies) in the U.K. (Bank of England, 2022). Intangibles are challenging to use as collateral, and overcoming this obstacle has not happened. Furthermore, it was shown that more intangibility was negatively correlated with the prevalence of HGFs.

Moreover, HGFs are linked to higher labor-elasticity to productivity shocks. Thus, for such a firm to co-exist with being highly intangible, the complementarity effect of hiring more skilled workers due to greater intangible intensity would need to dominate the substitution effect of unskilled. However, as the evidence showed in the previous sections, that seems not to be the case.

1.7.1.2 Scenario 2 Holding $\left(\frac{f}{n}\right)_{90\text{th}}$ constant, the intangibles intensity of firms growing at the bottom decreased over time; $\left(\frac{f}{n}\right)_{10\text{th}}$ declines. Why firms growing at the bottom in employment may experience a decrease in intangible intensity?

As shown in the previous empirical findings, in contrast to highly intangible firms, which in aggregate kept churning jobs as in the old days when facing a shock, low-intensity firms showed reduced average employment growth throughout the years.

This may reflect that bottom-growth firms, or firms that tend to reduce their pace of employment growth, may struggle to keep up with the demands of a more intangible economy. In contrast to increasing their investments, evidence shows that bottom growth firms precisely do not count or cannot hire skilled workers to handle intangibles (Intellectual Property Office, 2022) potentially, so they abstain from investing in them or even reducing their current stock.[53].

This could be why the 50-10 dispersion differential remained the same throughout the sample period. This may also imply that bottom growers often comprise the most intangible, larger types of firms.

1.7.1.3 Scenario 3 $\left(\frac{f}{n}\right)_{90\text{th}}$ increases, and $\left(\frac{f}{n}\right)_{10\text{th}}$ decreases until the gap closes.

This scenario is also a possibility. As HGFs converted into a more mature segment of small firms, they may face fewer restrictions to access intangible investments, given the reduced information asymmetries they offer to potential lenders. Thus, the intangible intensity of top growth firms, $\left(\frac{f}{n}\right)_{90\text{th}}$, increases. Simultaneously, it could be the case that bottom firms start facing problems in keeping up with the labor demand they need with their high-intangible intensity (e.g., shortages of skilled labor, regulatory restrictions impeding qualified migration, among multiple others), so intangible intensity $\left(\frac{f}{n}\right)_{10\text{th}}$, decreases.[54].

The previous analysis considered a case when β is the same for both groups of firms, but overhead costs f are specific. What happens when overhead is the same for all firms? The inequality in 32 would turn out to be dependent only on the labor size for each group of firms:

$$\begin{aligned} \frac{f}{n_{90\text{th}}} &< \frac{f}{n_{10\text{th}}} \\ \text{so that,} & \\ n_{10\text{th}} &< n_{90\text{th}} \end{aligned} \tag{33}$$

If overheads are equal regardless of how much a company grows in employment, the elasticity would depend only on the company's labor size. According to equation 33, the assumption of highly labor-elastic companies with lower intangibles intensity holds if and only if the average labor size of firms that grow in employment the least is smaller than the average labor size of firms that grow the most. As mentioned earlier, we know that HGFs are becoming smaller and that firms at the bottom of the employment growth distribution are not changing their patterns much. In this case, it can be argued that in a context of equal fs, the gap between top and bottom firms is closing primarily because of the declines in employment growth of smaller firm sizes, which typically grew the most in employment.

1.8 Conclusions

This chapter provided insights into the recent and puzzling business dynamics decline in developed economies. Under the Shumpeterian "creative destruction" approach, a constant entry, exit, and resource reallocation between less to more productive firms (business dynamics) is needed to foster economic growth, innovation, and productivity. However, evidence in multiple countries has shown that this has not been the case, especially after the dot-com bubble of the 2000s, when business dynamics declined and within-industry productivity dispersion arose.

Among the multiple explanations that have emerged in the literature, Decker et al. (2016c) enlightened by the work of the canonical models of Hopenhayn (1992); Hopenhayn and Rogerson (1993) demonstrated that in the U.S., this could have occurred due to a decline in the responsiveness of firms via employment to productivity shocks. Furthermore, Bartelsman et al. (2013) demonstrated that by introducing frictions, such as firing costs or labor protection policies, adjustments in size and entry and exit decisions could be distorted.

Inspired by these findings and the lack of consensus and empirical evidence on the puzzle, this chapter explored how high-intensity intangible investments such as software,

organizational capital, marketing and advertising, research and development (R&D), and patents, among others, can be a source of lack of responsiveness due to frictions on labor, ultimately leading to low employment growth in an economy regardless of rising productivity dispersion.

The main argument was that intangible assets' economic properties activate diverse incentives to the economies' players. Based on prior literature, I discussed how intangibles' properties, such as *scalability*, *synergies*, *spillovers*, and *sunkness*, may lead to four types of trends that ultimately reduce firms' capability to respond via employment: Intangibles may substitute labor as a means of response to productivity shocks, impact firms' entry negatively, demand more skilled workers, so overall employment decreases due to their lower relative importance in the economy, and entail delays in responses to shocks if at all there is a response in employment.

The key to this argument is the complementary and substitution effects between intangible assets and labor. More intangible industries or firms demand higher-skilled workers since often these assets are knowledge-centric, while tend to demand fewer unskilled workers; therefore, such firms can be more willing to respond to a positive productivity shock by substituting unskilled workers or just getting more intangibles, not responding via employment at all. In contrast, less intangible ones could be more willing to keep reacting via employment changes, so they can decide to hire more workers in the presence of a positive productivity shock or, conversely, destroy jobs in the presence of a negative one. That may be why low-intangible-intensity firms can be less equipped when advanced technologies or market demand changes. Given their low productivity, they may take longer to learn how to integrate them into existing operations or lack the financial capacity to do so.

Since providing evidence for all these premises is challenging and out of the scope of this thesis, this chapter presented an empirical strategy using U.K. data for the 1998-2018 period, disciplined by an analytical model, to demonstrate at least a negative and significant relationship between an intangible economy and various definitions of dynamism, which could support some of these ideas.

To do so, the chapter started with aggregate and intra-group characterization and analysis of firms (e.g., by age, size, and industry cells or combinations). One of the most prominent results is that some elements set the U.K. apart from other countries while others did not. For instance, the U.K. recovered dynamism after the 2008 financial crisis compared to countries like the U.S. or Belgium. Also, the U.K.'s manufacturing firms decreased their traditional "job destruction" power (larger firms a bit more than SMEs) after the financial crisis; their net job creation switched from negative to close to zero. Moreover, despite sharing a declining trend since the early 2000s in market entries and organic job

reallocation while increasing exit rates, the U.K. economy recovered in these indicators after the financial crisis -this positive trend seemed to have stopped around the Brexit vote year, nonetheless-(section 1.6.1.3).

However, it shared many of the most concerning trends for policymakers in dynamics. The U.K. has in common with other countries that young SMEs, on average, regardless of sector, have contributed the most with positive net job creation throughout the period. Still, young SMEs, the most dynamic group in terms of employment, seemed to have lost vibrancy; young SMEs are the minority relative to the overall economy (most of the U.K. firms are SMEs, of which only 41% are young), and their rates of employment growth have declined in recent years (sections 1.6.1.1 and 1.6.1.2)[55].

Moreover, a critical pattern similar to other countries is the loss of "skewness" in the employment growth distribution. The gap in the pace of growth (or the employment growth dispersion differential) between the top and bottom growers in employment has narrowed over the years, mainly because the top growers, or High-Growth Firms (HGFs), are losing steam as they have become older and smaller on average (sections 1.6.1.4, 1.6.1.4.1 and 1.6.1.6). Similar intra-group facts explain the pattern. The information sector (a highly intangible intensive sector) experienced the sharpest decline of such a gap (section 1.6.1.4.3). High-skewed sectors, characterized by the high intensity of knowledge-related assets, generally experienced the sharpest declines. Some authors have attributed the phenomenon to the lack of ability of the most dynamic firms of the economy, the young SMEs, to catch up with the increased fixed costs of the ICT revolution of the early 2000s. Given their heightened synergies with technologies, this fact appeared more impactful for highly-intangible sectors (section 1.6.1.4.2).

These trends were confirmed when assessing the ICT-intensive sectors. During the analysis period, the share of employment in ICT manufacturing and the content and media sectors decreased, and the share of non-ICT manufacturing and ICT services increased. Simultaneously, the findings also showed that the manufacturing and content & media sectors have become more intangible-intensive (Section 1.6.1.9). Hence, manufacturing might have needed more skilled workers as it became more intangible. However, given the low proportion of skilled workers relative to unskilled workers in the overall economy and that manufacturing represents just about 20% of the U.K. economy, the substitution could have slayed the complementarity effect (Office for National Statistics, 2023a).

Additionally, it does not matter much that the services sector increased its employment share and its high-growth young SMEs' average employment size. Although services constitute most of the U.K. economy and employ most people, reclaiming employment contribution is not a substantial and positive sign regarding the HGF population in the

U.K. As shown in section 1.6.1.5, the increasing trends in employment share are only observed in the ICT-intensive or knowledge-intensive portion of the services sector, but this still constitutes the minority of the sector.

The Office of National Statistics reported that despite having shifted to more Knowledge-Intensive Services (KIS) in recent years, most of the services sector provides essential services; they still represent 61% percent of the services sector[56]. KIS refers to "services which require a high proportion of high-level skills and inputs of research and development (R&D), such financial services, business services, and professional, scientific and technical services" (Office for National Statistics, 2023a). Hence, the complementarity and, therefore, the positive effect that the ICT services sector might trigger is insufficient to push aggregate employment contribution figures.

These facts suggest that the more intangibles, the less aggregate employment growth. I confirmed this intuition once a negative relationship between intangible intensity and employment growth rate gaps between top and bottom firms was also found in the characterization of firms section. High intangible-intensive industries exhibit less change in employment growth (the 90-10 dispersion differential kept more stable throughout time) than low-intangible industries. The latter industries exhibited sharper declines in employment growth, more evidently—suggesting that less intangible industries kept churning employment at similar rates as always (see Section 1.6.1.8). In contrast, more intangible sectors potentially had to adapt by substituting unskilled workers, acquiring intangibles, and hiring skilled workers. This explains the overall effect where intangible sectors' employment growth is more similar to that of the median firm in recent years.

The exploratory analysis characterizing U.K. firms' trends over the period paved the way for the econometric exercise. Various regression specifications of dynamic panel models at the industry×year level were used to infer the statistical significance of the relationship between intangibles and dynamics. The regressions demonstrated that, as the intangible intensity of a sector increases, the 90th-10th dispersion differential of employment growth rates in that sector declines. The negative impact is more significant for the 90th-50th tail of the employment growth distribution dispersion differential than for the 50th-10th, confirming that intangible intensity has impacted HGFs the most (section 1.6.4).

According to the regressions, the age×size group of firms most affected by declining employment growth by increases of intangibles shares are those in the mature × large group. Thus, the participation of these firms in total employment decreases as industries invest more in intangible assets. Furthermore, I found that as intangibles share increases, the job reallocation rate decreases, and entry rates in terms of employment decrease. These regressions' results also confirmed that the response in dynamism comes with a delay, implying that intangible investments do not immediately alter companies'

employment or entry and exit decisions. The impact may hit companies' jobs until after two years of a change in intangible intensity.

Finally, to discipline these empirical findings, the chapter introduced a novel model that illustrated how, in equilibrium, intangibles facilitate maintaining simultaneously high within-industry productivity dispersion and low employment growth. The model permitted examining how intangibles may affect firms depending on their demographics, for instance, how it affects the high-growth firms, different from those at the bottom of the size employment growth distribution.

Ultimately, the model permits illustrating scenarios in which the degree of intangible intensity, the labor demand elasticity to productivity shocks, and the employment growth rate interact to explain observed trends in the data. The main conclusion of the model is that the labor demand elasticity to productivity shocks is higher, and the lower the firm's intensity of intangible investments. To put it differently, the more likely a firm is to respond in terms of employment, the less likely they are to be intensive in intangible assets.

The chapter makes different contributions to the literature. First, it addresses some empirical limitations that have constrained researchers' ability to test intangibles as the reason for the weakened relationship between business dynamics and productivity. Measuring intangibles is already complex, and much more linking the best estimates available at the firm level to business dynamics. To the best of my knowledge, the present study is the first independent and successful effort to build a panel dataset of real businesses in the U.K. ready to be connected with other official U.K. firm-level data (e.g., U.K. innovation and intangible investments surveys), that could eventually facilitate more robust inference of the intangibles-dynamics relation. Just recently, the U.K.'s Office of National Statistics launched efforts to create and make available for research longitudinal data to explore business dynamics evolution. However, it remains experimental. Furthermore, although there are official experimental estimations of intangible investments (see Office for National Statistics (2022)) by the ONS, they are not yet connected to their business dynamics experimental panel data. As shown in the data section, the primary input of this chapter is ready to be connected with relevant and official panel datasets via the firms' unique identifiers. Indeed, this will save time and pave the way for further studies.

Second, this chapter contributes to empirical efforts to understand business dynamics globally and in key developed economies still lacking evidence. The benefits of this chapter's empirical exercises are twofold. The panel database created required establishing a list of assumptions about business activity throughout the period, which can be reused. In addition, it applied standards recognized by the research and

policymakers community, like the OECD's DynEmp3 software routines, to compute business dynamics. Subsequently, the study promotes high-quality cross-country comparisons by adding a prominent economy, such as the U.K, while promoting reproduction and continuity of the analyses in further research efforts.

Third, this chapter contributes to still limited within-cell analysis, another issue that adds to the lack of recent evidence for the U.K. As seen throughout the analysis, the differences across industry sectors, firm size, and age groups or the speed of their employment growth were evident. This is imperative for effective and efficient policy making; identifying which groups of firms are the most affected by a rising intangible economy allows targeted action. For instance, this study proved that one of the U.K.'s weakest points in employment creation is the reduced pace of high-growth young SMEs in the most intangible sectors, despite an apparent recovery halted after the Brexit vote. These analyses could help policymakers design and implement programs that provide targeted and, therefore, practical support.

Lastly, although some recent research has procured full general equilibrium models explaining trends behind the intersection between businesses' labor dynamics and intangible capital, the model proposed here brings a practical interpretation of these interactions, directly connecting standardized Business Dynamism indicators with intangible capital trends. The canonical models on resource reallocation on which I based this study are not explicit about labor elasticity to shocks in the presence of frictions, so design policy intervention is made accordingly.

The chapter faces a few shortcomings, however. Although for this study, I built a panel dataset ready to connect with firm-level data on intangibles and innovation data, I only present whole economy or intra-group aggregate results, emerging from the matching between intangible data from the INTAN project and business dynamics at the industry level. The firm-level analysis could provide much robust insights. However, it was not possible then because of time constraints for conducting firm data confidentiality checks, and the pandemic hit, limiting access to secure lab data. Another issue can be that assumptions made to build business dynamic indicators differ slightly from official statistical office reports. They are not significantly different, but readers must be cautious when interpreting specific figures[57].

Ultimately, this study did not include much about business dynamics metrics apart from employment (or entry and exit). For instance, young and large decreased employment growth rates but increased revenue growth, which hints at differences in other dynamics metrics. Unfortunately, the DynEmp3 routine does not prioritize revenue metrics, given the often poor quality of revenue data in business registries. However, for future studies, the DynEmp routine could be fed with revenue data from other reliable sources, such as

the U.K. tax authority HRMC.[58]

2 Market Power and Technology[1]

2.1 Introduction

The recent debate over macroeconomic trends in investment, labor share, concentration, markups, and profits has prompted critical discussions due to the varied potential welfare and distribution implications (Akcigit and Ates, 2019)[59]. Market power, in particular, has emerged as the potential root of all evil, as multiple pieces of evidence indicate it has increased in the last couple of decades. The predominant argument is that globalization, lax regulation, and, most importantly, uncontrolled new technology have given some firms unprecedented leverage to control market entry and extract ever-larger rents from consumers and workers (Philippon and Gutierrez, 2017).

Precisely, the focus has been on whether Superstar firms, as they are called, which are growing too fast in profits and getting even farther away from peers over time, are triggering all of this. The most commonly known superstars are those of the type of the FAANGs (Facebook, Amazon, Apple, Netflix, and Google) or, as more recently called, the MATANAs (Microsoft, Apple, Tesla, Amazon, and Nvidia), depending on who is at the top of the distribution. It's a compelling story when faced with influential Superstars controlling key platform technologies in digital markets. There is little doubt that the last 20 years or so have seen the emergence of formidable companies on the back of new ICT technologies.

However, policy documents and academic research have overlooked the massive power of combining ICT technologies with novel business models, marketing strategies, organizational structures, research and development, and other intangible assets, which have reshaped the economic model entirely by reducing marginal costs and increasing returns to scale, which ultimately has facilitated the growth of superstar firms. Consequently, the exploration and provision of sound evidence on alternative explanations instead of purely market power has been limited.

Provided that more proper evidence could demonstrate that the macro trends that appear so concerning result from efficiency gains and technological progress rather than anti-competitive practices per se, policy actions would likely differ from current inclinations. Instead of harsh opposition to the existence of superstars, more nuanced policy approaches would be needed (Demsetz, 1973), like balancing anti-trust measures while leveraging the productivity benefits of intangibles.

Therefore, this chapter's primary goal is to deliver more evidence on the issues that prolong the ambiguities of the debate for policymakers. As of today, one of the biggest

[1]This chapter is co-authored with Ralf Martin, Imperial College London, and LSE.

obstacles is to confirm whether aggregate markups are rising or for which companies. Furthermore, if markups are rising, what are the reasons behind the increase? Is it technological change and globalization, market power or a combination? These are precisely some of the questions partially addressed here.

To assess to what extent Superstars are associated with market power, we must account for the interplay between technology and market forces. There has been much focus on markups as a measure of market power. The influential work of De Loecker and Warzynski (2012); De Loecker et al. (2018, 2020), and several studies that followed applying the same methodology (e.g., Syverson (2019); Manyika et al. (2018)) have suggested that we have seen a sharp increase in Lerner style markups in recent years. However, to measure markups, we have had to estimate marginal costs. We argue that current estimates of marginal costs suffer from two drawbacks:

Firstly, when estimating technology parameters, these are fixed over time. This, by construction, ignores critical technological changes that might have occurred in recent years and affected marginal costs. In particular, most studies impose that returns to scale and the composition of factor intensities remain fixed. An essential feature of new ICT technologies widely discussed is that they might offer returns to scale.

Secondly, there are inherent biases and inconsistencies in the estimation approaches used in most papers. The estimation of markups in most of the recent studies follows the De Loecker and Warzynski (2012); De Loecker et al. (2018, 2020)'s method. It relies on estimating a production function via the "control function" approach[60] and the cost-minimization process of firms to obtain the estimate of the elasticity output of the most flexible input (most susceptible to shocks, like "cost of materials"). The firm-level markup is obtained as the ratio of this elasticity over the share of the firm's revenues on that flexible input. The most prominent critiques emerge from their assumption that firms are price takers, which may lead to endogeneity problems combined with specific data constraints.

First, the lack of quantity data and the inability to control for additional firm heterogeneities bias estimates. Revenue data is often the only output data available. Since it conflates prices and outputs, it does not reveal exact information about output quantities, and prices go to the regression's residual. In the presence of market power and no price data to control for it, endogeneity problems emerge. Prices can be correlated with the variable input; for example, companies with price-setting power can reduce prices to sell additional output, biasing the elasticity estimate downwards. More generally, the method cannot control for other types of heterogeneity of firms beyond productivity (e.g., consumer preferences or, ironically, the firm's price-setting power), leading to further endogeneity problems.

Second, the "control function approach" in economic production function estimations involves introducing control variables to account for unobserved factors and improve estimates' accuracy. However, under this approach's maintained assumptions, prices are also a function of productivity, a demand shock and the quasi-fixed factor capital, as it is for the other flexible inputs of the production function, like materials, m, and labor, l. The flexible production factors are functions of the same variables, implying that we cannot find any instruments for this problem; there are no drivers of the variable input that are not correlated with the error term.

In this paper, we propose applying an alternative framework to address both of these issues, and we show that this matters for resulting estimated markups. Crucially, markups have not increased on average and are not as much at the top of the distribution (superstar firms) as predicted in recent years. We find that aggregate markups, once weighted by firms' sales in a given year, are lower than markups without weights (see Figure 32a). More importantly, we show that by applying the alternative methodology, the decline of the estimated elasticities is much more pronounced and seems similar to the decrease in variable input shares. Therefore, the markup increase is less noticeable as the numerator and denominator fall at similar rates. Therefore, we infer that the returns to scale or the technologies have changed over time rather than the difference between the revenues and costs (see Figure 37). Even more shockingly, we find that markups of the allegedly rent-seeking ICT sector are not increasing (see Figure 41 on the information sector) as neither those from the manufacturing sector in the U.S., where recent evidence shows (see chapter 1) its increased intangible intensity in recent years (see Figure 38).

To reach these conclusions, the chapter's methodology consists of computing firm-level markups replicating De Loecker's methodology on ORBIS[61] data from European countries and the same U.S. Compustat[62] data used in De Loecker et al. (2020). Then, we compare these estimates with those obtained by adapting a recently released alternative method, Forlani et al. (2023), which addresses some of De Loecker's main critiques. We call the De Loecker method applications model 1, which consists of a 2-stage GMM estimation a la Ackerberg. Then, we call model 2, or "new" the adaptation of the Forlani et al. (2023)'s method, which employs a much more straightforward econometric technique (a 2-stage IV regression).

The application of this method allows us to address two of the most controversial flaws of De Loecker's approach. Foremost, it addresses the evolution of production technologies over time by introducing time-changing output elasticities. Forlani et al. (2023) does not focus on this portion. Still, we built on the flexibility of this model to contribute to the literature by conducting markup estimations using 5-year rolling windows; we estimate output elasticities (the numerator of the markup formula) for each firm i at year t using all the panel data of five years before t, including t, and moving forward that window for the

$t+1$ estimates, and so on[63]. This differs from De Loecker's earliest applications, where the production function is estimated taking all years available, resulting in only one unique constant parameter for the entire analysis period. This is why we estimate model 1 and model 2 in two versions, the "no-rolling" (green, dashed lines on plots) and the "rolling" versions (red, continuous lines). The former refers to the markups computed using constant output elasticities over time, whereas the latter refers to markups computed with changing output elasticities.

It later addresses endogeneity and simultaneity problems by introducing an invertible demand function and an assumption that simplifies the regression. The advantage of this regression equation is that in addition to the primary regression of revenue used by De Loecker, it assumes a Dixit-Stiglitz demand system, so a linear approximation of demand can be combined with the revenue equation. Hence, it is possible to remove all unobservable variables from the econometrician—for example, markups. To put it differently, the alternative model's regression can put all information observable from the data on the left-hand side (LHS) and leave only on the right-hand side (RHS) the parameters related to the quasi-fixed production factors, such as capital and a decomposed shock that includes unknown firm's heterogeneities (e.g., demand or productivity shocks) and the random error term, for which is possible to assume an auto-regressive process. The procedure allows obtaining a "proxy" of elasticity output comparable to De Loecker's estimated elasticity.

In overview, the alternative model addresses three problems simultaneously with productivity estimation methods. Namely, demand shocks, markups, and Total Factor Productivity (TFP) heterogeneity (Forlani et al., 2016, 2023). This model improves on De Loecker and Warzynski (2012); De Loecker et al. (2018, 2020) measures because they deviate from the perfect competition usual assumption, which, as described, is problematic. In contrast, the alternative model makes explicit assumptions regarding market structure and uses a monopolistic competition approach; it uses the Dixit-Stiglitz demand system, a strategy that provides a tractable method of introducing imperfect competition and increasing returns to scale into the model.

Thus, this chapter contributes in various ways. First, it raises awareness about the flaws of a methodology that multiple policy and academic outlets accept as a fact. Besides highlighting methodological considerations, it makes evident inconsistencies in the authors' published work on De Loecker et al. (2020). Since we needed to replicate De Loecker's estimates to later compare them with those obtained with the alternative method, we noticed data management and econometric decisions that substantially could change results but were not mentioned in their article. For example, the authors use regular Olley and Pakes (OP) techniques to estimate the production for the first 20 years of their U.S. sample because their methodology would not converge, given the few data

points available.

Likewise, the authors' 2-stage GMM estimation software routine in STATA used in the paper was not published with their article, just their final estimates, making it harder for other researchers to confirm their alleged estimation steps and procedures. Similarly, the authors left substantial outliers in their final sample that, if removed, could considerably change the value of markups, like "impossible" values of shares of expenditures on the variable input, like 800% of the value of the revenue. The present work adds to the growing strand of research being more peaky and critical of the flaws in almost cult-like academic pieces of work. It hopes to prevent precarious policy recommendations.

Second, the chapter applies methodologies that respond more to the needs and realities of recent economic trends. Not only Forlani et al. (2016, 2023) inspired methods correspond more to current contexts because they introduce more realistic assumptions, but we also introduce estimation strategies like the rolling windows that are more aligned with a world where artificial intelligence, quantum computing, extended reality, blockchain, internet of things (IoT) have skyrocketed efficiency and reduced costs. By simply allowing changing output elasticities over time, the chapter exposes the transformative power that the combination of technological change and intangibles has had over returns to scale. We demonstrated that, as other recent evidence has shown, there are reductions in marginal costs and potentially increasing economies of scale, which does not immediately reflect market power.

Finally, the dataset constructed for this study using ORBIS data allows the connection of firm-level markup data with datasets in intangibles and innovation to test directly the markup-innovation or intangibles link. Since ORBIS provides an identifier per firm, our markup estimates can be linked to ORBIS data on intangible assets, patents, and innovation indexes from the ORBIS Intellectual Property databases, including social media data. For instance, in the next chapter, we successfully measured innovation and intangible investment intensity with social media posts of ORBIS-covered firms.

Still, the chapter has its limitations. We do not yet connect estimated markups, output elasticities, or expenditure revenue shares with data that could explain the trends in a formal econometric setting. It also does not decompose the change between De Loecker and our preferred specification. Beyond the plot visuals, it would be helpful to provide the magnitude of each component that makes markups different, especially the differences between constant parameters and rolling windows estimations. We expect to improve this aspect in future versions of this paper. Furthermore, the Forlani et al. (2023)'s method could also be subject to the criticism of Doraszelski and Jaumandreu (2021), that the error term is not separately identified and will be included in the markup estimates. We can argue that by taking the demand heterogeneity out of the residual, fewer unknowns

are left in the error term, but still, removing more unknowns must be a priority for urgent contributions on this front.

2.2 Conceptual Framework

What are superstar firms, and why do they matter? Despite lacking a unified definition, it's applied broadly to individuals, firms, regions, sectors, and more, emphasizing concentration of output, skewed income or profit distributions, large rewards at the top (Rosen, 1981), or exhibiting a significantly greater share of income than peers, pulling away further over time (Manyika et al., 2018). The firms most commonly acknowledged as superstars are those of the type of the FAANGs (Facebook, Amazon, Apple, Netflix, and Google) or, as more recently called, the MATANAs (Microsoft, Apple, Tesla, Amazon, and Nvidia), depending on who is at the top of the distribution.

The market power literature has popularized the concept of superstars. Since market power centers on firms with the power to set prices and monopolize markets, this stream of literature has focused on "exposing" superstars' extraordinary growth, emphasizing extreme profitability outcomes and their potential cannibalization effect and displacement of other firms (Autor et al., 2017; Andrews et al., 2015).

For instance, discussing superstars in the framework of the "winner takes all" dynamics of some competition models is popular. In these models, two players compete for a prize, exerting efforts based on their positions relative to competitors. This often results in a few firms gaining a significant market share, thanks to critical differentiators and low substitution elasticity between superstars and smaller firms, as in Rosen (1981) model. Particularly evident in digital-ICT sectors, more prominent, mature firms in high-tech industries dominate, acquiring smaller innovative firms and impacting business dynamics and country welfare negatively by diverting resources and discouraging healthy competition (De Loecker et al., 2018; Calligaris et al., 2018; Andrews et al., 2016a; Criscuolo and Martin, 2019; Autor et al., 2017).

Regardless of the primarily negative connotation of superstars, the recent advent of micro-founded research to investigate market power has improved the analysis toolkit to explore more in-depth their emergence, prevalence and consequences for economies worldwide (De Loecker et al., 2018; De Loecker and Eeckhout, 2018; Hall, 2018; Eslava and Haltiwanger, 2019, 2018). Novel structured models and empirical strategies have enabled researchers to obtain better insights into increasing gaps between heterogeneous firms and the emergence of powerful groups. Similarly, new micro-data sources provided the possibility of estimating markups, profit shares, costs and investments at the firm level and getting a more realistic measure to study the forces behind Superstar firms, which is essential for assessing the reasons for higher or lower markups, and consequently

identify market power (Syverson, 2019).

2.2.1 Market Power and Superstar Firms

Market power is a firm's influence on the price of its products. Estimating markups, a traditional approach in industrial organization literature, is a direct measure of market power. Markups represent the price-to-marginal cost ratio at the firm's profit-maximizing output, influenced by the shape of the firm's residual demand curve. Steeper demand curves indicate more significant margins and greater market power, with policymakers concerned about the implications for market competition and consumer surplus (Syverson, 2019).

Identifying superstars often involves recognizing firms with consistently or increasingly higher markups than their peers. However, estimating markups is challenging due to scarce data on critical variables in firms' competitive dynamics and relevant markets. Obtaining information on prices, quantities, and factors of production is often impossible. Hence, market power, and therefore the identification of Superstars, had been relegated to macro measures such as market concentration, combined market share of major firms, number of competitors, profit rates, and market entry costs.

The widely used concentration measure is the Herfindahl-Hirschman Index (HHI), calculated as the sum of squares of market shares across firms in relevant markets. While widely used, concentration has limitations in identifying heterogeneity across firms, geographies, and sectors, restricting its ability to infer the presence of market power or superstar firms.

First, although concentration measures are straightforward, they may not capture exceptional market power cases. They rely on revenue and output or cost information or are unable to identify cases of monopolistic competition across levels (e.g., national vs local concentration). Even with low concentration in atomistic monopolies, their inelastic residual demand curves indicate significant market power, which these measures may miss (Syverson, 2019)[64].

Second, and perhaps the most problematic aspect of concentration measures, is that they can conflate outcomes with causes. Concentration can be a "market outcome" and not a cause– that is, it arises as a result of the competitive interactions of firms rather than determining the competitive interaction. Moreover, its relationship with competition might not always be as inverse as expected; higher concentration might indicate more competition.

For instance, in Cournot's competition with homogeneous products, concentration correlates positively with market power. However, a higher concentration can coexist

with higher competition in models allowing for product variety substitutes, indicating less market power (Syverson, 2019). Increased competition can increase profits as firms compensate for entry costs, even though margins may rise. Models suggest that greater substitution or competition can decrease welfare (Melitz and Ottaviano, 2008). The increasing availability of differentiated products (e.g., different car brands) challenges the concentration-market power relationship[65].

2.2.2 Micro-founded research on market power

Micro-founded research has addressed some of macro research's most significant obstacles to making inferences on market power. However, it is not without a cost. There are two micro-founded approaches to estimating market power via markup estimation. First, the demand-side approach estimates markups by determining the profit-maximizing price-marginal cost through the slope of the reverse demand curve. This is perhaps the ideal method since it directly estimates the demand curve. Still, this method is challenging due to data requirements on prices and quantities, assumptions on consumer behavior, and the imposition of a competition model between firms. It has been applied selectively to specific industries such as cars or beer industries (Berry et al., 1995; Koujianou Goldberg and Hellerstein, 2012).

This is why the supply side or production approach, popularized by Robert Hall, has become favored. It relies on firm output and input cost data, utilizing the cost minimization process and reducing data requirements and assumptions. Hall (1988) showed that under cost minimization, for any variable input (e.g., freely adjustable by firms at any time, as opposed to quasifixed inputs as many forms of capital), a firm's markup will equal the ratio of two values: the elasticity of output to that variable input, and the share of revenues the input is paid. The numerator needs to be recovered from the production function estimation, and the denominator can be observed directly in accounting data. This approach, further developed by De Loecker and Warzynski (2012), has become widely used for inferring market power, offering flexibility for nonconstant returns and general technologies (De Loecker et al., 2018; Hall, 2018).

An essential contribution of De Loecker and Warzynski (2012)'s work is that it addressed issues in Hall's methodology by introducing a control function to address unobserved factors and firm-level heterogeneity that bias markups when estimating the production function, such as productivity growth. They also incorporated less flexible inputs and recovered markups at the firm level instead of industry only. However, problematic characteristics remain, which we aim to address in this study partially.

2.2.3 Stylized facts in market power

In recent decades, reported market power rises in developed economies have garnered significant attention globally, materializing in multiple academic and policy research documents De Loecker and Eeckhout (2018); De Loecker et al. (2018, 2020). Studies can be grouped into three categories. The first, adopting revenue-based concentration as a sign of market power, focuses on increased concentration measures such as the Herfindahl-Hirschman index (HHI), CR4, and CR20[66]. Evidence suggests rising industry concentration in developed countries, with the U.S. seeing notable increases in various sectors (Autor et al., 2017, 2020).

In the U.S., the share of sales by the top 4 firms increased significantly in various sectors between 1982 and 2012: around 5% in manufacturing, 4.4% in service industries, 15.0% in retail industries, and 2.1% in the wholesale sector (Autor et al., 2017, 2020). In Europe, concentration increases are comparatively lower. However, European Commission-monitored mergers antitrust market analyses, which define markets more accurately, reveal even higher concentration increases, up to 10 times higher than traditional firm or industry-level measures (Affeldt et al., 2021)[67]. See also Furman (2015); Autor et al. (2017); Barkai (2020); Grullon et al. (2019); Philippon and Gutierrez (2017); De Loecker et al. (2020); Syverson (2019); Hall (2018); Diez et al. (2018); Akcigit and Ates (2018); Eggertsson et al. (2021); Philippon and Gutierrez (2017); Christopher J. Nekarda (2020)

The second group emphasizes the most direct measure of market power: markups. Studies like those by De Loecker et al. show substantial markup increases, indicating heightened market power globally, with a markup increase from 1.15 in 1980 to 1.6 in 2016 (Loecker and Eeckhout, 2021)[68]. Only in the U.S. markups increased 60% from 1.2 in 1955 to 1.61 in 2016; in Europe, the estimated increase was 15% between 1980 and 2016. Aligned with these findings, reports indicate excessive profits, not just for covering fixed costs. Profit rates increase from one to eight percent in the U.S. for the same period as for markups (De Loecker et al., 2020). See also Philippon and Gutierrez (2017); Eggertsson et al. (2021); Barkai (2020); Akcigit and Ates (2019).[69]

The third group of studies examines macroeconomic trends linked to rising markups or increased concentration. For example, the decline in the labor market share of GDP in the U.S., especially post-2000s, authors suggest a decline of up to two-thirds of the world's GDP (Dao et al., 2017). Multiple reasons apart from market power have been proposed to explain declined labor share[70]. Studies explaining labor share in relation to superstars have mixed views. One the one hand, some authors relate declined labor share to less stringent anti-trust enforcement that allows superstars to exercise market power (Karabarbounis and Neiman, 2014; Akcigit and Ates, 2018; Philippon and Gutierrez, 2017). Autor, Dorn, Katz, Patterson, and Van Reenen (2020). On the other hand, lower

labor share, at least in the U.S. and other industrialized countries, is explained by output reallocation triggered by more globalization (more competition) and technological change, favouring more productive firms that often have smaller labor shares.

Studies linking macro trends and market power also point to a decline in the investment share in physical capital, contrary to its expected increase due to a drop in labor share[71]. Philippon and Gutierrez (2017) emphasises a substantial reduction in firms' investment share out of revenue, from 20% (1962-2001) to 10% (2002-2015) in the U.S., in contrast to Europe, where investment levels have been more in line with profitability measures and Tobins Q (Philippon and Gutierrez, 2017; Dottling, Gutierrez, and Philippon, 2018).

This research strand also explores the decline in physical capital investment, linked to the rise of intangible capital (e.g., software, business models), and its impact on increased market power, driven by characteristics like scalability and stronger intellectual property protections that provide a competitive advantage to intangible-intensive firms, resulting in adverse effects on overall investment (Crouzet and Eberly, 2018, 2019; Zhang et al., 2019; Zheng et al., 2022; de Ridder, 2020; Altomonte et al., 2020; Ayyagari et al., 2022). For instance, Gutiérrez and Philippon (2017b), shows that market power, measured by HHI concentration indexes, combined with intangible investments and potential short-term management, significantly diminishes investment rates in the U.S. The study indicates that approximately 30% of the decline in investment rates is linked to the dominance of intangibles, displacing traditional capital investments.

Finally, there is a substantial body of work suggesting market power as a reason for declining dynamism, encompassing reduced worker reallocation rates (Decker et al., 2014), a more even firm-growth distribution Decker et al. (2016c), and diminished firm entry rates (Karahan et al., 2019).

However, despite some consensus on rising market power in advanced economies, ambiguities persist. Not all countries experience this trend; for instance, the Netherlands shows stable markups from 2006 to 2016 (van Heuvelen et al., 2021). Furthermore, different indicators of increasing market power, such as concentration, profits, or markups, can contradict each other theoretically and empirically (Syverson, 2019; Gutiérrez and Philippon, 2017a; Hwang et al., 2022). The dynamics behind rising markups also vary; Spain's increase is driven by small, unproductive firms (García-Perea et al., 2021), while in the U.S. and China, markups surge due to larger, high-markup superstar firms (Autor et al., 2020; Baqaee and Farhi, 2020b; Zheng et al., 2022). Additionally, there is a growing trend of studies exposing serious methodological challenges in De Loecker and coauthors' benchmark methodology to estimate markups. See Raval (2023b,a); Bond et al. (2020); De Ridder et al. (2021); Loecker and Eeckhout (2021).

2.2.3.1 Superstar firms: The good Superstars' emergence and dominance have been said to emerge more as a product of globalization and technological forces and not so much of an anti-competitive nature per se. For instance, Autor et al. (2020) attributes the decline in labor share, especially in the U.S. and OECD countries, to an environment favoring the most productive firms, where heightened product market competition in the middle of globalization and technological forces favors the Superstars, raising concentration, facilitating a winner-takes-all scenario.

When competition intensifies, Autor et al. (2020) says that even though markups decrease and labor share declines for the average firm due to competitive pressure, the reallocation effect that shifts economic activity toward the most productive firms is more substantial. These highly productive firms produce more and typically have lower labor shares, often larger firms. As a result, the overall economic output increases, widening the gap relative to labor and reducing the aggregate labor share. Furthermore, the authors note that in the U.S., as found by De Loecker et al. (2020), aggregate weighted (by firms' sales)-average markups are rising due to changes in market shares, driven by the concentration and markups of the largest firms, which tend to have the highest markups. This suggests that greater competition can lead to higher markups, unlike what is typically expected.

Akerman et al. (2022) also proposes that globalization and technological advancements, like the Internet, have intensified product market competition and empowered consumers with better access to price information, increasing their price sensitivity. With heightened competition, the cost of entry rises, favoring more resourceful and productive firms. This reallocation of sales and value-added from less productive to more productive firms, often with smaller labor share initially, drives an increase in market share for the most productive firms and higher concentration.

Assessing specific U.S. and European trends challenges the negative view of superstars. Autor et al. (2020) discover similar trends in labor share and concentration between the U.S. and the supposedly more anti-trust stringent Europe, contradicting Philippon and Gutierrez (2017)'s view and findings. Ayyagari et al. (2022) also finds that positive shocks to competition (e.g. the increase in Chinese imports to the U.S.) do not affect superstar firms' capital[72], output, or markups, questioning the role of monopoly power.

This has led to the discovery that several factors may contribute to concentration without implying market entrenchment. First, strong network effects, like those of FAANGs, may be disrupted as technology advances and dependency on specific hardware decreases, destabilizing leadership positions (Tucker, 2018). Second, falling prices of ICT technologies and intangible capital benefit firms with high fixed costs, fostering efficiency gains rather than pricing power (Bessen, 2017). Third, rapid declines in quality-adjusted equipment prices, capital deepening, and increased productivity, especially in

ICT, naturally lead to higher concentration (Acemoglu et al., 2018; Karabarbounis and Neiman, 2014)[73]. The prevalence of robots and declining capital prices relative to labor, when the elasticity of substitution is greater than 1, is a potential reason (Acemoglu et al., 2018; Karabarbounis and Neiman, 2014). Factors such as those contribute to labor-capital substitution and declining labor share without market entrenchment.

Thus, the capacity of investments in productivity-enhancing and technology-related assets plays a significant role in the emergence of superstars without entrenchment(Crouzet and Eberly, 2018; Tambe et al., 2020). Superstar firms are typically intangible-intensive, especially in sectors like healthcare and high-tech. In such sectors, investments in supply chains and new business models amplify the impact of technological investments, naturally leading to efficiency gains that reflect later in increased markups. For instance, Crouzet and Eberly (2018); Tambe et al. (2020) demonstrated that healthcare and high-tech exhibited higher concentration driven by productivity gains from intangibles. In contrast, in consumer sectors (e.g., wholesale and retail trade and agriculture), it was more likely due to market power. (Autor et al., 2020) also finds a higher concentration exclusively on more productive and innovative sectors; they find a positive between concentration, productivity growth, and industry patent intensity in the U.S. Additionally, Tambe et al. (2020) illustrates superstar's digital capital predicts their productivity well.

Finally, reduced aggregate investment could be either because of measurement problems or benevolent reasons, not because of unwilling superstars. Brynjolfsson et al. (2017) and Corrado et al. (2021) explain physical capital investment may appear to be declining because there is an implementation lag of breakthrough technologies such as artificial intelligence (AI). These technologies are still in the diffusion stage. They require acquiring and installing complementary physical and additional intangible assets, such as supercomputers, innovative business models, and skills, which takes time.

2.2.3.2 Superstar Firms: The bad Although Autor et al. (2020, 2017)'s model suggests innovation and rapid technological progress as the primary reason behind superstars, and a challenge arises when the expected efficiency gains from reallocation between firms do not translate into overall productivity improvements. Aggregate productivity growth in the U.S. and Europe has been sluggish, notably below long-term averages (de Ridder, 2020)[74]. Thus, as positive as it looks, the superstar effect based on reallocation presents challenges. There are some factors potentially muting positive outcomes.

First, superstar firms might contribute modest productivity gains to aggregate productivity growth. de Ridder (2020) shows that after initially boosting productivity, the expenses in innovative investments (e.g., corporate R&D) were followed by a significant

decline in productivity growth in the U.S. and France.

Second, specific environments can counteract the benefits of enhanced productivity. Incumbents that are more likely to innovate than entrants or capable of erecting barriers to entry are problematic. For example, Bessen (2022) shows how big tech firms, through proprietary software protection, have hampered the innovation of smaller firms.

Also, when there are knowledge and technology diffusion obstacles from frontier to laggard firms, it widens productivity gaps and hinders catch-up. Andrews et al. (2015) highlights that national frontier firms[75], once they overcome country-specific factors to adopt innovation from global frontier firms, protect their competitive advantages, making it more difficult for national laggard firms to catch up, increasing productivity dispersion (Decker et al., 2017)[76]. Akcigit and Ates (2019) also suggests that reduced corporate taxes, increased R&D subsidies, increased regulations, and heavier intellectual property use have further limited knowledge diffusion.

Third, Superstar firms can naturally thrive in an increasingly intangible economy. By 2013, major developed countries spent £1.10 on intangible assets for every £1 invested in tangible assets. Intangible investments (e.g., software, business models, organizational capital, and intellectual property) provide considerable advantages in marginal cost reduction and efficiency gains compared to tangible capital, increasing economies of scale[77] (Haskel and Westlake, 2018).

For example, de Ridder (2020) estimates markups using De Loecker et al. (2020)'s method to demonstrate that intangibles, particularly software, lower marginal costs and increase fixed costs. The scalability of intangibles allows superstars to reuse assets at zero marginal cost. Software-intensive firms, for example, leverage intangibles like Enterprise Resource Planning and automated business processes to reduce marginal sales costs. In addition, firms producing software benefit from minimal reproduction costs for additional units, shifting from marginal to fixed costs.

The extent of the advantage varies based on individual firm characteristics, such as varying capacities to deploy intangibles (e.g., managerial abilities). Thus, incumbent intangible firms capable of leveraging these advantages can lower prices through cost reduction while increasing revenue. Subsequently, potential competitors with superior products might be deterred from entering the market, unable to match the low prices offered by these cash-flow-rich companies[78]. Positive liquidity shocks increase intangible investment, enabling higher markups. Altomonte et al. (2020) explores liquidity constraints on intangible capital in French firms (2004-2014). This suggests financial constraints for startups, as intangibles are not widely accepted as collateral, hindering their ability to accumulate such assets.

Furthermore, superstars are more likely to be the ultimate beneficiaries of a fall in the price of ICT investment because they can acquire more ICT and exploit the benefits of their combination with intangibles (Autor et al., 2020; Haskel and Westlake, 2018)[79]. For example, this ICT-intangible combination has helped Amazon to become one of the most successful companies in the world. Amazon leverages its cloud platform for data storage and processing, enhancing product recommendations and customer service, ultimately building brand recognition and loyalty.

The multiplier power of intangibles is evident in studies adjusting markup calculations with intangible capital. In this context, aggregate markups are lower than initially predicted by De Loecker et al. (2020)[80], but it is higher for intangible-intensive. In the top 10% of firms' markup distribution, industries with high intangible intensity exhibit higher markups than those with low intangible intensity industries (de Ridder, 2020).

This implies that highly intangible industries could exercise market power if conditions allow. For instance, preeminent audit firms like Deloitte, PricewaterhouseCoopers (PwC), Ernst & Young, and KPMG, the top firms of the U.S. MSA industry market, exhibit collective market power since their fees are higher than average audit firms. They have a privileged competitive position since audit production is highly intangible, client-specific, complex, and cash-flow risky (Zhang, Ye, Simunic, and Chu, 2019).

Hence, intangible investments in the considerations for market power and superstars can present a double sword. Increasing markups drive firms to invest in outperforming competitors. However, excessive markup power eventually hampers overall investment due to diminishing returns, following the concept that escaping competition initially incentivizes innovation. However, overly high markup power curtails investment and innovation rewards (Zheng, Zhang, and Zhang, 2022; Aghion, Bloom, Blundell, Griffith, and Howitt, 2005)[81].

2.2.3.3 The challenge Given the above discussion on the good and the bad about Superstars, this paper's main contribution is to provide more evidence on some of the issues that prolong the ambiguities of the debate. As of today, one of the biggest obstacles is confirming whether aggregate markups are rising and, if so, which companies are contributing. Likewise, in this chapter, we explore at least partially that if markups are rising, it might be due more to technological change and globalization rather than market power per se. As section 2.3.1.1 exposes, we apply recently developed methods that address some of De Loecker and Warzynski (2012); De Loecker et al. (2020)'s methodology's most controversial assumptions to estimate markups and some additional techniques that, at least partially, tackle measurement issues that could erroneously lead to believe aggregate markups are rising when they are not because rather those apparent markups rise are emerging from increasing returns to scale that emerge from technological

change.

For instance, a measurement error by omission of intangibles hampers the conclusions on market power via markup estimations. First, suppose the investment has shifted to intangibles. Thus, once omitted, when testing the quantitative impacts of variables that push investment down, they will not impact effectively measured investment; the responsiveness to standard shocks in investment declines.

Second, when using De Loecker et al. (2020)'s methodology, the omission of intangibles biases markup estimations upwards since it impacts the denominator and numerators of the markups formula (Section 2.3.1). Equation 42 shows a firm's markup as the ratio between the output elasticity of the variable input and a firm's revenues spent on flexible inputs (e.g., cost of materials for production, labor, and electricity, among others.), so defining what is understood as variable costs is determinant.

Intangible assets can be considered fixed costs rather than variable because they represent a sunk cost (Haskel and Westlake, 2018). Intangible investments, such as expenditures in marketing, once done, cannot be recovered and contribute to a firm's capital stock (e.g., brand value). Thus, they should be considered a fixed or capital expense. However, in available firm-level datasets, they are usually considered an operating expense and therefore added as flexible input costs (Ayyagari et al., 2022). Thus, the denominator would be lower if it were possible to deduce them from the variable costs correctly. Holding the numerator constant, markups could be higher.

Likewise, intangibles could boost the denominator if intangibles inflate variable costs (Traina, 2018). For example, part of the variable input expenses besides labor or materials is all incurred costs getting the product to the consumer. Some of them are variable and should be added to the denominator. Hence, markups could remain flat or go down. Yet, available data on expenditures on flexible inputs do not capture these costs in intangibles, and markup levels vary throughout estimations.

The numerator in the markup equation can also change because of intangibles. Under certain assumptions, the output elasticity is equivalent to the labor share, which has declined as described above. Highly intangible industries demand a more skilled labor force, which decreases its demand for unskilled workers, which reduces the elasticity of the output of labor. Thus, depending on which effect is dominant markup levels can be highly variable with this methodology.

Furthermore, causality is another issue. Is it intangibles generating market power or vice-versa? Intangibles can flatten the marginal cost curve since they can propitiate increasing economies of scale (e.g., reduced marginal costs due to reduced necessary inputs for production) and exacerbate network effects. The problem with increasing

economies of scale and network effects is that both can signal or lead to market power. Accordingly, understanding whether intangibles came before market power or vice versa requires methods to tackle this simultaneity problem[82].

What is certain, however, is that evidence on intangible assets' implications for market power not only invites authorities to be alert (See CMA (2022)) but also, perhaps more importantly, to look for estimations methods that contribute to gaining some certainties. This is why, in this chapter, we aim at improving markup estimation methods while giving a more prominent role to technological changes propitiating the emergence of Superstars instead of preemptively blaming the phenomenon on anti-competitive practices[83].

2.3 Methodology

We follow two different micro-founded approaches to estimate markups. First, we follow the work of De Loecker and Warzynski (2012); De Loecker et al. (2018); De Loecker and Eeckhout (2018); De Loecker et al. (2020) and second, we follow Forlani, Martin, Mion, and Muûls (2016); Forlani, Martin, Mion, and Muûls (2023), who improve on some limitations of the De Loecker style model. This second approach has two variations depending on the assumed structure of the production technologies (e.g., Cobb-Douglas or Translog).

We estimate markup models for around 131 countries across different industry sectors and a separate analysis for high-tech sectors. For the latter, we used a high-tech classification system of industrial codes recently created by the OECD, called the J-tag taxonomy (Inaba and Squicciarini, 2017). This classification proposes a definition of Information and Communication Technologies (ICT) based on the technology classes used by the International Patent Classification (IPC) system. Based on the knowledge of patent experts, the strategy aligns OECD definitions of ICT at the industry level (4-digit industry codes) with those encompassed by types of patents.

These estimation methods are based on the supply-side approach. Thus, to make inferences on market power, the method finds a proxy for the firm's price-marginal-cost markup relying on the cost minimization process of the firm and assuming a production technology (production function) across economic sectors. The methods rely on yearly accounting information at the firm and establishment levels. We aim to compare the results between the different methods to demonstrate markups are not increasing as much as widely acclaimed. The following section describes the benchmark model, De Loecker and Warzynski (2012); De Loecker et al. (2020) and econometric strategies, providing a structure to explain the application of an alternative model.

2.3.1 Model 1: De Loecker et al.

Following, De Loecker and Warzynski (2012) we assume a Cobb-Douglas production function of the style:

$$Q_{it} = A_{it} L_{it}^{\alpha_L} M_{it}^{\alpha_M} K_{it}^{\gamma - \alpha_L - \alpha_M} \tag{34}$$

Where Q_{it} is the output, $A_{it}$is hicks-neutral technological change, $L_{it}$is labor capital, $K_{it}$is capital, and $M_{it}$is the cost of materials for transforming inputs into outputs is our proxy for the flexible input. We assume that capital and labor face adjustment costs. This function is continuous and twice differentiable. Assuming cost minimization, producers face the following Lagrangian function:

$$\mathcal{L} = P_L L_{it} + P_M M_{it} + P_K Kit - \lambda_{it} \left(Q_{it}(.) - \overline{Q_{it}} \right) \tag{35}$$

Where $P_L, P_K$and $P_M$are the prices of labor, capital, and variable input, respectively. The first order condition with respect to each of the inputs is equal to:

$$\frac{\partial \mathcal{L}}{\partial X_{it}} = P^{X_{it}} - \lambda_{it} \frac{\partial Q_{it}(.)}{\partial X_{it}} = 0 \tag{36}$$

Where $X_{it} = \{L_{it}, K_{it}, M_{it}\}$. From the first order condition with respect to the level of output $\frac{\partial \mathcal{L}}{\partial Q_{it}}$, we obtain that the marginal cost of production is equal to λ_{it}. From 36 we get that in relation to the variable input:

$$P^{M_{it}} = \lambda_{it} \frac{\partial Q_{it}(.)}{\partial M_{it}} \tag{37}$$

Rearrangingλ_{it} and multiplying both sides by $\frac{M_{it}}{Q_{it}}$we get the expression that:

$$\frac{P^{M_{it}}}{\lambda_{it}} \frac{M_{it}}{Q_{it}} = \frac{\partial Q_{it}(.)}{\partial M_{it}} \frac{M_{it}}{Q_{it}} \tag{38}$$

Therefore, cost minimization implies that the optimal demand for the flexible input M is when the elasticity of output with respect to M (right hand of the equation above) is equal to the fraction between expenditures in the flexible input M and marginal cost multiplied by total output Q. Assuming the markup is simply prices over marginal costs, we express markups:

$$\mu_{it} = \frac{P^{Q}it}{\lambda_{it}} \tag{39}$$

Where $_{P^{Q}it}$ is the price of a product or a price index for the multi-product company i. With equations 38 and 39 we find that markups can be expressed in terms of:

$$\mu_{it} \frac{P^{M_{it}}}{P^{Q}it} \frac{M_{it}}{Q_{it}} = \frac{\partial Q_{it}(.)}{\partial M_{it}} \frac{M_{it}}{Q_{it}} \tag{40}$$

Therefore, the share of expenditures in the flexible input M in total sales equals the markup multiplied by the elasticity of the flexible input concerning output. The left-hand side of the equation is denominated s_{Mit} and the expression after the markup is denominated α_M in this form:

$$\mu_{it} s_{Mit} = \alpha_M \tag{41}$$

Hence, markups can be identified as the fraction between the input elasticity and expenditure shares. The first term is the one that needs to be estimated econometrically from the production function, and the second one is directly observed from the accounting data for each firm. Notice that a standard technology is assumed. Model 1 then refers to the results in terms of the expression:

$$\mu_{it} = \frac{\hat{\alpha}_{Mit}}{s_{Mit}} \tag{42}$$

The strategy for estimating an unbiased alpha depends on expressing the production function in terms of available data. Therefore, we express equation 34 in terms of the natural logarithm as

$$q_{it} = a_{it} + \alpha_L l_{it} + \alpha_M m_{it} + (\gamma - \alpha_L - \alpha_M) k_{it} \tag{43}$$

Now, since we rarely measure what we want in the previous equation above, we must relate the equation to the data. Starting with the left hand side. The output comes from microdata. There is no information available at this stage on the number of units of output, but there is information on total revenues. Therefore we define:

$$r_{it} = q_{it} + p_{it} \tag{44}$$

Where $r_{it}$is the natural logarithm of operating revenues expressed as the sum of the natural logarithm of the number of units of product $q_{it}$and the natural logarithm of the price of that product p_{it}. Now, for the right-hand side, we have most of the data except, of course for the exogenous technological change or shifter a_{it} and other unobserved variables. Thus, on the right-hand side, we have, labor l_{it}as the cost of employees in firm i at year t, capital $k_{it}$is the value of the capital stock of firm i at year t, and cost of materials $m_{it} = r_{it} - v_{it}$, where $v_{it}$is the value added. All these are in natural logarithms. Subsequently, we can assume that the log-linear revenue equation able to be estimated follows the form:

$$r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta k k_{it} + \underbrace{p_{it} + a_{it} + d_{it}}_{\varepsilon_{it}}$$

$$r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + \varepsilon_{it} \tag{45}$$

Where ε_{it} is the error term[84]. Since revenues are not double deflated, the prices go to this error term[85] along with other omitted variables that could not be captured with the data. This includes the technological shifter a_{it}, pricesp_{it} and a demand shock d_{it}that captures unobserved consumer preferences that also impact output decisions.Therefore, $\varepsilon_{it} = p_{it} + a_{it} + d_{it}$ and revenues can be affected by random shocks to consumer demand shocks. α_X is the parameter that captures the output elasticity concerning input X.

De Loecker and Warzynski (2012); De Loecker et al. (2018) papers follow Levinsohn and Petrin (2003, hereafter LP) methodology in which they assume that materials serve as a good proxy for Productivity. This is why the method is called a control function approach since it will address the endogeneity issue of the inputs with unobserved Productivity by introducing this proxy as a control in the production function estimation.

Estimating coefficient β_M consists of a 2-stage GMM instrumental variables approach following Ackerberg et al. (2015). Taking equation 45, they observe logged output r_{it} and assume that it is given by $r_{it} = lnQ_{it} + \epsilon_{it}$, where r_{it} is the natural logarithm of a firm's realized sales, epsilon is unanticipated shocks to production and shocks ϵ_{it} are i.i.d including measurement error. Firms do not observe epsilon when making optimal input decisions.

In the first stage, r_{it} is regressed on the 3rd-degree polynomial of k, l, m, and control variables. Following De Loecker et al. (2019), firm sales shares are included to control for markups (at the 3-digit industry level). In our section with multi-country data, we include country-fixed effects. The first stage is aimed at removing the measurement error and unanticipated shocks. In the second stage, we estimate output elasticities with respect to variable inputs and the state variable k by fitting an AR(1) process for productivity to the data (via GMM). Thus we can identify α_M and compute the firm-level markup.[86]

The first stage depends on the assumptions that (1) materials can be expressed as a function $m_{it} = \phi_M(a_{it}, d_{it}, k_{it})$, of capital k_{it}, the productivity shock a_{it} and d_{it} that can be considered a vector of additional variables that can affect input demand choice (e.g., a product demand shock), and (2) strict monotonicity of this input demand function, strictly increasing in a_{it}. The latter property allows to invert the demand equation of materials and obtain a proxy of productivity, $a_{it} = {h_t}^{-1}(k_{it}, m_{it}, d_{it})$ in the production function estimation. The productivity shock is expressed as an unknown function of observables. By substituting this into the production function, equation 45, we obtain:

$$r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + {h_t}^{-1}(k_{it}, m_{it}, d_{it}) + \varepsilon_{it} \tag{46}$$

Since the unknown function of observables can only be treated non-parametrically, the first three parameters are not identifiable and are subsumed to a function Φ

$$r_{it} = \Phi(l_{it}, m_{it}, k_{it}, d_{it}) + \varepsilon_{it} \tag{47}$$

This equation can be estimated non-parametrically, and an estimate $\hat{\Phi}$ for the function ϕ and ε can be obtained. $\hat{\Phi}$ is then plugged into the estimations of the second stage. Thus the expected value of the equation 47 leads to

$$\Phi(l_{it}, \hat{m_{it}}, k_{it}, d_{it}) = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + a_{it} \tag{48}$$

where $E[\varepsilon_{it} \mid I_t] = 0$, and a_{it} is an estimate of productivity per function ${h_t}^{-1}(k_{it}, m_{it}, d_{it})$.

The second stage identifies the production function by imposing an AR(1) process on the productivity and using the estimate of equation 47 to identify the parameters. The AR(1) assumption is necessary to address the endogeneity issue of outputs and inputs being affected by productivity simultaneously, biasing a least squares estimation. The methodology achieves that by plugging in the residuals of the first stage estimation as an estimate of productivity into the log revenue equation.[87]

Thus, by replacing an AR(1) process

$$a_{it} = g(a_{it-1}) + \nu_{a,it}$$

into equation 46 (or 47), we get

$$r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + g(a_{it-1}) + \nu_{a,it} + \varepsilon_{it}$$

and since we know from the estimates in equation 48 we know that

$$\Phi\left(l_{it}, \hat{m_{it}}, k_{it}, d_{it}\right) - \beta_L l_{it} - \beta_m m_{it} - \beta_k k_{it} = a_{it}$$

This equation is expressed in lagged terms, and replacing it into the revenue equation again, we get:

$$\begin{aligned} r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + g(\Phi(l_{it-1}, \hat{m_{it-1}}, k_{it-1}, d_{it-1}) \\ - \beta_L l_{it-1} - \beta_m m_{it-1} - \beta_k k_{it-1}) + \nu_{a,it} + \varepsilon_{it} \end{aligned} \tag{49}$$

So the second stage moment condition to estimate β_m along with the other production function parameters is:

$$\begin{aligned} E\left[\nu_{a,it} + \varepsilon_{it} \mid I_{t-1}\right] &= E\left[r_{it} - \beta_L l_{it} - \beta_m m_{it} - \beta_k k_{it} \mid I_{t-1}\right] \\ &- \left[g\left(\Phi\left(l_{it-1}, \hat{m_{it-1}}, k_{it-1}, d_{it-1}\right) - \beta_L l_{it-1} - \beta_m m_{it-1} - \beta_k k_{it-1}\right) \mid I_{t-1}\right] \\ &= 0 \end{aligned} \tag{50}$$

In summary, to replicate these authors' exercise, we assume that our control function for the productivity shock considers that materials and other semi-flexible inputs can be captured via proxy by semi-flexible inputs with lagged values, capital, and an error term. Consequently, the estimation is based on these four moments:

$$E\left[\begin{pmatrix} k_{it} \\ l_{it-1} \\ m_{it-1} \\ \varepsilon_{it-1} \end{pmatrix} \nu_{\epsilon it}\right] = 0$$

$$\beta_m = \hat{\alpha_M}$$

$$y_{it} = m_{it} - p_{it} + \ln\left(\mu_{it}\right) + f_t + \xi_{it}$$

This model De Loecker and Warzynski (2012); De Loecker et al. (2018, 2020), imposes two restrictions. First, a scalar Hicksneutral productivity term, and second, common technologies or which is the same, standard technology parameters (elasticities) by establishing a Cobb-Douglas production function across the set of producers within the sector. The latter implies that alpha parameters are constant across firms in a sector. Then, markups only vary across sectors and firms with changes in their shares of expenditures in their variable input in total revenue.

2.3.1.1 Main critiques to De Loecker et al. method A growing body of work highlights estimation and identification issues of De Loecker et al. (2020)'s methodology, implying that the estimated firm-level markups are unobservable and biased, higher or lower.

Two of the most prominent critiques emerge from the method's assumption that firms are price takers. The first critique appears from using the revenue to measure output.

The technique relies on the production function estimation approach, requiring output data. However, usually, there is only revenue data, which creates an omitted variable bias because prices go to the residual when estimating the production function. These prices tend to be correlated with lagged values of the variable input in the presence of market power. See Ridder et al. (2022).

The production function estimations using revenue data, under conditions of market power of firms, result in estimated output elasticities being biased downwards since using revenue elasticities concerning the variable input reflects the price-setting behavior of such firms that cannot be controlled for a priori. The critical assumption of this production function methodology is that there is perfect competition in the cost minimization problem. However, price-setter firms reduce prices when raising output, leading to understating actual output elasticities. Furthermore, limited firm-level pricing data seriously restricts the ability to control these price movements. Thus, when there is market power, there is no way to control it, and the estimated elasticities are biased. Indeed, using revenues instead of outputs can bias the elasticity estimates downwards so severely that the estimated markups are uninformative about true markups. This happens because the estimated markup, the revenue elasticity ratio over the flexible input share, can be precisely equal to one (Bond, Hashemi, Kaplan, and Zoch, 2021; Klette and Griliches, 1996), which is uninformative.

The second critique emerges precisely from the omission of markups in the first stage of the estimation procedure. Doraszelski and Jaumandreu (2019, 2021) observe that output can only be purged of measurement errors and shocks in the first stage of the procedure if the researcher can observe markups in the first place. Since the latter is the point of the exercise, the Hall 1986 method on which De-Loecker et al. base its estimations is not feasible.

Furthermore, there could be an identification problem in the production function estimation when oversimplifying the firm's technology, like using a Cobb-Douglas. When firms are price setters, meaning that firms have market power, the elasticities estimated under the profit maximization assumption can not be estimated non-parametrically in cases of non-linear productivity dynamics, and firms face heterogeneous demand schedules, with unobserved variation in a demand shifter.

This latter critique, however, is a little less problematic, according to other authors. Calvino et al. (2018) suggested that although not ideal, assuming a Cobb-Douglas or Translog production function, each has strengths and weaknesses when used to estimate markups. For the Cobb-Douglas production function, no variation in output elasticities exists across firms within the same industry. Consequently, variation in markups over time and producers is driven by revenue share. Instead, the Translog production function

retrieves firm-level output elasticity estimates; consequently, markup variation is given by firm-level variation in revenue shares and output elasticities. In the case of a Cobb-Douglas, the output elasticity is joint to all firms of a given sector; in the case of a translog, the concerning derivatives of intermediates are firm-specific. A firm-specific value for the output elasticity implies that each firm will likely combine production inputs differently. However, the values obtained with the Cobb-Douglas are simpler to estimate and, therefore, generally considered more stable in the literature than those obtained through the Translog.

Finally, Yeh et al. (2022) and Morlacco (2020) also note that the De Loecker et al. methodology leads to biased markups when firms have market power in the market of the variable input (monopsony).

However, according to De Ridder et al. (2021), although the estimated markups under this methodology may not be entirely informative about the level of the true markups because of the biases in the elasticities, it can provide well-estimated results about their dispersion, both in the cross-section and over time. Moreover, markups estimated under this methodology still reflect the same sign and order of magnitude to correlations with profitability measures, and trends are also well estimated with revenue data.

To argue this, De Ridder et al. (2021) test on firm-level French data and simulations that although the methodology does not control for market power, production functions estimated with revenue data produce a bias in elasticity which is equal to the average of the demand elasticity across firms with the same production function. In some oligopoly models, these elasticities are the inverse of the firm-level markup. When firms have heterogenous markups, the bias may cancel out the average but not the variation in firm-level markups. Thus, revenue-estimated markups correlate positively with true markups (or those that should have been estimated with proper output data).

Regardless, estimated markups using revenue data can be higher or lower than the true average markups. When the authors contrast estimations made with accurate output data, they find that revenue-estimated markups tend to be higher than true average markups.

There are other critiques related to the type of cost data De Loecker et al. (2018); De Loecker and Eeckhout (2018); De Loecker et al. (2020) use. Traina (2018) criticizes De Loecker et al. (2018); De Loecker and Eeckhout (2018) for using Costs of Goods Sold (COGS) as a measure of the variable input. "Cost of Goods Sold" (COGS) is directly reported by the firm's financial statement of their data for the U.S. study and decides to measure variable input with it. It includes all expenses directly attributable to the production of the goods sold by the firm and comprises materials and intermediate inputs, labor costs, and energy. Following the prior work on intangible investment measurements

such as Peters and Taylor (2017), Traina (2018) tests what would happen with the markup estimations using Hall's methodology but considering as flexible inputs not only COGS but also selling, general, and administrative (SG&A) spending. The authors argue that such a cost has substantially increased over recent decades. Peters and Taylor (2017) argues that a fraction of SGA can be considered replacement costs of assets such as organizational capital, including human capital, brand, customer relationships, and distribution systems. Therefore, those could be considered a proxy for fixed costs that add to COGS as variable costs and markups decrease. However, as discussed in the section 2.2.3.3, the contrary could occur with a different assumption regarding intangibles and their relationship with variable costs, rendering De Loecker's benchmark markup equation unreliable.

We attempt to address some but not all of these issues with the new model. De Loecker et al.'s work, has been one of the most influential, and the recent versions of their model have been intended to correct for associated endogeneity issues. However, despite controlling for the endogeneity associated with the productivity shock, there is not yet a successful attempt to control for endogeneity driven by prices, so estimates for $\hat{\alpha_m}$ remain inconsistent. Prices go to the error term, which is then affected simultaneously by productivity shocks, demand shocks, capital, materials, and labor behavior. Yet, labor and materials are a function of those same variables. In other words, the authors are using as instruments the same variables they should want to instrument for $l_{it} = \phi_l(a_{it}, d_{it}, k_{it})$. and $p_{it} = \phi_p(a_{it}, d_{it}, k_{it})$..

2.3.2 Model 2: New

Most approaches will recognize that the error term in equation 45 likely correlates with the explanatory variables. Hence, they suggest instrumental variables or control function approaches to deal with the matter. The problem with this is that,

$$\epsilon_{it} = p_{it} + a_{it}$$

i.e., the price variable is entirely relegated to the shock term. However, under the maintained assumptions

$$p_{it} = \phi_P(a_{it}, d_{it}, k_{it})$$

i.e., prices are a function of the productivity and demand shock and the quasi-fixed factor capital. Similarly

$$m_{it} = \phi_M(a_{it}, d_{it}, k_{it})$$

$$l_{it} = \phi_L(a_{it}, d_{it}, k_{it})$$

i.e., the flexible production factors are functions of the same variables. This implies that we cannot find any instruments for this problem as no driver of materials and labor would also be correlated with the error term ϵ_{it}. Similarly, if there is a control function that would control for all relevant variation, it could not separately be identified from the production function. What's worse, if we try alternative timing assumptions (e.g. a la Ackerberg et al.) that could potentially make the problem tractable e.g., we could suggest that prices are like capital fixed in the short-run this will destroy the validity of equation 42; i.e., it would change the short-run profit maximization problem so that we could no longer back out valid estimates of markups from equation 42.

Hence, we propose applying the Forlani et al. (2023) method to U.S. and European data as an alternative. The methodological framework develops in a simple CobbDouglas context with log-linear demand for simplicity. However, all issues carry over to more complex frameworks with more flexible production and demand structures. Accordingly, we consider the production function expressed in De Loecker's models, equation 34.

There are two fundamental assumptions in this application[88]. First, the demand function is invertible. So consider a residual demand function of the form:

$$Q_{it} = \Lambda_{it}^{1-\eta_{it}} P_{it}^{-\eta_{it}} \tag{51}$$

Where η_{it} is a firm and time-specific elasticity of demand, we assume for the sake of simplicity in the present paper that $ln\left(\Lambda_{it}\right) = d_{it}$. The latter indicates demand shocks or the appeal of a product to a customer, a consumer's preference for a particular product in a monopolistic competition model[89].

The second assumption is that we can simplify the regression only to require parameters related to the quasi-fixed production factors described in equation 52. See appendix O for a complete method description.

$$LHS_{it} \equiv \frac{r_{it} - s_{Mit}\left(m_{it} - k_{it}\right) - s_{Lit}\left(l_{it} - k_{it}\right)}{s_{Mit}} = \frac{\gamma}{\alpha_M} k_{it} + \frac{1}{\alpha_M}\omega_{it} \tag{52}$$

Therefore, we can obtain an estimate of $\frac{\gamma}{\alpha_M}$ from equation 52 subject to assumptions about the error term, which we discuss in the next section. This estimate is going to be equivalent to the coefficient of materials (or elasticity of the flexible input) in equation 45; $\frac{\gamma}{\alpha_M} \equiv \beta$. The advantage of this equation is that in contrast to the linear approximation of revenue, as presented in equation O.81 from Appendix O, is that all unobservable variables to the econometrician are removed. For instance, markups. On the left-hand side (LHS). Besides the decomposed shock and its random error, the transformed equation of log revenue contains only observables on the left. From this, we can get an estimate

of markups defined as:

$$\frac{\widehat{\mu_{it}}}{\gamma} = \frac{1}{\widehat{\left(\frac{\gamma}{\alpha_M}\right)} s_{Mit}} \tag{53}$$

Note that we are only able to identify markups over the returns to scale parameter γ. However, this is the right measure to understand if there are any excess markups. We would not expect firms to operate if $\gamma < \mu_{it}$ as they would make negative profits, corresponding to an inability to recover possible fixed costs. Also, note that the residual comprises the technical TFP and demand shock, $\omega_{it} = (a_{it} + d_{it})$. Again, this is the relevant measure of possible welfare impacts. For simplicity, we refer to this estimator in equation 53 as the μ-ω production function estimator.

This model addresses three problems simultaneously with productivity estimation methods. Namely, demand shocks, markups, and Total Factor Productivity (TFP) heterogeneity (Forlani et al., 2016, 2023). The difference between Model 2 and Model 3, available in section 2.3.3, is the underlying technology; Model 2 assumes a CobbDouglas production function, whereas Model 3 assumes a more flexible technology with a translog. The rest of the assumptions are shared between these two models.

These models improve on De Loecker and Warzynski (2012); De Loecker et al. (2018, 2020) models because they deviate from the perfect competition usual assumption that, as discussed, is problematic. In contrast, these specifications make explicit assumptions regarding market structure and propose a monopolistic competition approach. In particular, the alternative model uses the DixitStiglitz demand system, a strategy that provides a tractable method of introducing imperfect competition and increasing returns to scale into the model.

This approach is critical because, as discussed most recently in Syverson (2019), markup behavior can result from changes of an increase in the share of pure profits and or increasing returns to scale. Similar evidence to the discussed above, research has found how production technologies with increasing returns generate natural monopolies given that as a consequence of such returns, large production entities are more productive than smaller ones (see Van Ark and Monnikhof, 1996, Van Biesebroeck, 2005 cited in Forlani et al. (2016)). Monopolistic competitive markets are characterized by many producers who each enjoy some market power (i.e., thanks to each producing a variety of the same product). Nonetheless, such market power is given precisely because of free entry (a proxy for competition) so that profit opportunities are limited. These model types perfectly fit the case of superstar firms because they can be the product of natural monopolies, and the De Loecker style models are not capturing it.

As in previous studies, another valuable characteristic of these models is that they can

separate demand from productivity shocks without imposing zero correlation. In doing so, the alternative model can isolate such shocks from markup shocks, which can affect productivity and cause endogeneity problems in production function estimations. Thus, it is possible to estimate markups independently, limiting the biases.

2.3.3 Model 3: New, more flexible model

Also following Forlani et al. (2023), Model 3 tracks the same assumptions as Model 2 but is more general because it uses a translog production function as the technology across firms. It also assumes a more general demand function that frees the model from imposed structures to consumer demand. The variables are also expressed in terms of differences, averages, and their levels, as the previous two models did. See appendix P for details on the equations. Equivalent to equation 52 in Model 2, the econometric expression that can be estimated is:

$$LHS_{it}\frac{\bar{\alpha}_{Mit}}{\gamma} - \Delta k_{it} = \frac{1}{\gamma}\Delta\omega_{it} \tag{54}$$

Where

$$LHS_{it} = \frac{\Delta r_{it} - \frac{\gamma}{\bar{\mu}_{it}} + \bar{s}_{Mit}\left(\Delta m_{it} - \Delta k_{it}\right) + \bar{s}_{Lit}\left(\Delta l_{it} - \Delta k_{it}\right)}{\bar{s}_{Mit}}$$

To estimate equation 54, the model assumes as well a stochastic process AR(1) for the residual omega, equivalent to $\omega_{it} = \rho\Delta\omega_{it-1} + \eta_{it}$. Thus, equation 54 can be rewritten in the form of

$$\hat{\eta}_{it} = LHS_{it}\frac{\bar{\alpha}_{Mit}}{\gamma} - \Delta k_{it} - \frac{\rho}{\gamma}\left[LHS_{it-1}\frac{\bar{\alpha}_{Mit-1}}{\gamma} - \Delta k_{it-1}\right] \tag{55}$$

To identify the relevant parameters, this equation is also solved with GMM. The moment conditions correspond to the first derivatives of the residual η with respect to each of its parameters, namely $\frac{\rho}{\gamma}, \frac{\alpha_M}{\gamma}, \frac{\alpha_{KM}}{\gamma}, \frac{\alpha_{LM}}{\gamma}, \frac{\alpha_{MM}}{\gamma}$. Once found these parameters, markups can be identified in the form of:

$$\frac{\mu_{it}}{\gamma} = \frac{\hat{\alpha}_{Mit}}{\bar{s}_{Mit}} \tag{56}$$

This model, in contrast to Model 2, allows variation in markups at the firm level different from the one provided merely by the changes to expenditures shares in the flexible input. So μ varies by α (output elasticity) and by the s (expenditure share), not only on the latter as in model 2.

It has been found that markups' dispersion under this technology may be higher than true dispersion (De Ridder et al., 2021). Also, such imposition might be slightly restrictive when the true technology is more complex. However, markups based on this functional

form are robust to biases in production function estimates up to a sector-fixed effect. Both production functions' estimation has strengths and weaknesses. For the Cobb-Douglas, one assumes no variation of output elasticities across firms within the same industries; consequently, variation in markups over time and producers are driven by that of material costs shares on sales. The translog production function allows retrieval of a firm-level output elasticity, which leads to markup variation associated with firm-level variation in both sales and output elasticities. Assuming a firm-specific value of the output elasticity implies that each firm is likely to combine production inputs differently. However, different authors have highlighted that the Cobb-Douglas specification is much simpler to estimate. Therefore, it is generally considered much more stable in the literature than those obtained with the translog, as described in Model 3 Calvino et al. (2018). For this reason, we focus on estimating Models 1 and 2, assuming a cobb-douglas production function.

2.3.4 Implementation: Model 2 (New)

For comparison, we will implement a simple version of both the μ-ω and the more conventional estimator suggested in De Loecker et al. model's 42. The μ-ω estimator is fairly agnostic regarding the stochastic structure of the decomposed shock ω (includes a and d) and various timing assumptions as long as we can assume that there is at least one quasi-fixed factor. As with Model (1), the method requires making some assumptions on the error term omega. To keep things as simple as possible initially, we assume that both ϵ_{it} and ω_{it} evolve as AR(1) processes:

$$\omega_{it} = \rho_{\omega}\omega_{it-1} + \nu_{\omega it} \tag{57}$$

$$\epsilon_{it} = \rho_{\epsilon}\epsilon_{it-1} + \nu_{\epsilon it} \tag{58}$$

The above allows to identify equation O.83 via OLS, assuming an exogenous k_{it} and using,

$$E\left[\begin{pmatrix} k_{it} \\ \omega_{it-1} \end{pmatrix} \nu_{\omega it}\right] = 0$$

We build on these AR assumptions, but before we use a convenient expression of lagged omega ω_{it-1} based on its lagged components[90], a lagged version of equation 52 and the log-linear approximation of revenue from equation O.80.

$$\begin{aligned} LHS_{it-1} &= \frac{\gamma}{\alpha_M}k_{it-1} + \frac{1}{\alpha_M}\omega_{it-1} \\ &= \frac{\gamma}{\alpha_M}k_{it-1} + \frac{1}{\alpha_M}\left(a_{it-1} + r_{it-1}\frac{\alpha_M}{s_{Mit-1}} - q_{it-1}\right) \end{aligned} \tag{59}$$

and reorganizing for a_{it-1} then we have:

$$a_{it-1} = \alpha_M LHS_{it-1} - \gamma k_{it-1} - r_{it-1}\frac{\alpha_M}{s_{Mit-1}} + q_{it-1} \tag{60}$$

So we replace a_{it-1} equation and equation 57 into equation 52 so obtain the expression for estimation:

$$LHS_{it} = \frac{\gamma}{\alpha_M}k_{it} + \rho_\omega LHS_{it-1} - \frac{\rho_\omega \gamma}{\alpha_M}k_{it-1} + \frac{\nu_{\omega_{it}}}{\alpha_M} \tag{61}$$

By re-writing equation 61 as a linear regression

$$LHS_{it} = \beta k_{it} + b_2 LHS_{it-1} + b_3 k_{it-1} + \epsilon \tag{62}$$

where $\frac{\gamma}{\alpha_M} \equiv \beta$, $b_2 = \rho_\omega$, $b_3 = \frac{\rho_\omega \gamma}{\alpha_M}$ and $\epsilon = \frac{\nu_{\omega_{it}}}{\alpha_M}$. Since we assume that labor is also a variable input, as materials, both labor and materials are endogenous, and so the assumption of $\mathrm{E}(\nu_{\omega_{it}} l_{it}) = \mathrm{E}(\nu_{\omega_{it}} m_{it}) = 0$ does not hold, then we cannot estimate this equation to obtain our parameter of interest β via OLS, β is endogenous. Thus, we instrument with lags of order 2, LHS_{it-2}, and with lags of order 1 and two of capital, revenue, and materials. We include country and year dummies to perform the estimations separately by two-digit industries for the U.S. and the European region.

Notice that we cannot obtain the value of parameter γ directly, which will prevent us from limiting our markup estimates to the scaled version as in equation 53. First, we do not count with quantity information at the firm level in the data. Thus, contrary to Forlani et al. (2023) where in a second stage, they estimate the quantity equation 43, in combination with the parameters obtained from the first stage (the estimation of equation 62), we cannot obtain γ. Second, notice that in principle, one could look to exploit the parameters' constraints in the estimation of equation 62, to obtain an estimate of α_M directly to recover the parameters fully, without the need to estimate in a second stage, the quantity equation, as in Forlani et al. (2023). However, as it is seen from the restrictions, this is not feasible.

In summary, this new model builds upon De Loecker's approach, which uses a proxy variable approach and allows for the underlying presence of demand heterogeneity, d_{it}. The contrast with De Loecker's is that this new model is so explicit in its assumptions about demand heterogeneity (called product appeal) that, besides allowing us to measure it, it uses it to recover an additional equation not used in conventional methods to estimate markups. The new model, besides using equation 43, the quantity equation of the production function, uses equation O.81, the log revenue equation because it can be written in terms of observables and the heterogeneity allowed in the model parameters. This allows the model to land a simultaneous system of equations that although in this model, we only use the first stage, serves for a complete system when the quantity data

at the firm level is available.

Furthermore, the alternative method takes advantage of two insights. First, the AR(1) processes assumptions imposed on the critical parameters in light of previous methods, but here, specific to the quantity TFP a and the demand component (product appeal), d. Second, the markup equation resulting from the profit maximization process from equation 42. These elements result in only needing to estimate parameters of an equation with linear methods such as OLS (for revenue equations, when there are no endogenous factors) or IV (for quantity equations), in contrast to potential non-stable, joint-non-lineal equations as De Loecker and other authors had done previously.

We estimate the IV regression of equation 62 for the European data in the first differences. The LHS to be estimated is described below. Lower letters represent the natural logarithm of the variables:

$$DLHS_{it} = \frac{Dr_{it} - \bar{s}_{Mit}\left(Dm_{it} - Dk_{it}\right) - \bar{s}_{Lit}\left(Dl_{it} - Dk_{it}\right)}{\bar{s}_{Mit}} \tag{63}$$

Where,

$$\begin{aligned} Dr_{it} &= r_{it} - r_{it-1} \\ Dk_{it} &= k_{it} - k_{it-1} \\ Dm_{it} &= m_{it} - m_{it-1} \\ Dl_{it} &= l_{it} - l_{it-1} \\ \bar{s}_{Mit} &= 0.5\left(s_{Mit} + s_{Mit-1}\right) \\ \bar{s}_{Lit} &= 0.5\left(s_{Lit} + s_{Lit-1}\right) \end{aligned}$$

For the U.S. estimations of this model, the LHS is defined as below since we do not have different materials from labor costs, just the bundle of costs COGS. "Cost of Goods Sold" (COGS) is an item directly reported by the financial statement of the firm they use to measure the variable input. It includes all expenses directly attributable to the production of the goods sold by the firm and comprises materials and intermediate inputs, labor costs, and energy. Then the regression equivalent to 61 for the U.S. is:

$$LHS_{it} = \frac{\gamma}{\alpha_{COGS}} k_{it} + \rho_{\omega} LHS_{it-1} - \frac{\rho_{\omega}\gamma}{\alpha_{COGS}} k_{it-1} + \frac{\nu_{\omega_{it}}}{\alpha_{COGS}} \tag{64}$$

where $LHS_{it} = \frac{r_{it} - \frac{\alpha_{COGS}}{\mu_{it}}\left(cogs_{it} - k_{it}\right)}{s_{COGSit}}$, and in first differences:

$$DLHS_{it} = \frac{\gamma}{\alpha_{COGS}} Dk_{it} + \frac{1}{\alpha_{COGS}} \omega_{it} \tag{65}$$

Where

$$
\begin{aligned}
DLHS_{it} &= \frac{Dr_{it} - \bar{s}_{COGSit}\left(Dcogs_{it} - Dk_{it}\right)}{\bar{s}_{COGSit}} \\
Dr_{it} &= r_{it} - r_{it-1} \\
Dk_{it} &= k_{it} - k_{it-1} \\
Dcogs_{it} &= cogs_{it} - cogs_{it-1} \\
\bar{s}_{COGSit} &= 0.5\left(s_{COGSit} + s_{COGSit-1}\right)
\end{aligned}
$$

COGS share is $s_{COGSit} = \frac{COGS_{it}}{\exp(r_{it})}$. We estimate a 2-stage IV regression to control for endogeneity.

2.3.5 Implementation: Model 1 (De Loecker et al.)

Alternatively, we regress equation 45 directly following the 2-stage GMM procedure as Ackerberg et al. (2015), following De Loecker, using lagged values of the flexible factors as well as contemporaneous values of capital stocks as instruments:

$$
E\left[\begin{pmatrix} k_{it} \\ l_{it-1} \\ m_{it-1} \\ \epsilon_{it-1} \end{pmatrix} \nu_{\epsilon it}\right] = 0
$$

This distills the essence of the approach from much of the existing literature. Note that given the model assumptions, this is inconsistent as (flexible) prices remain in the error term. Given this, we expect a positive correlation with the flexible factors and the error term leading to an upward bias in the estimate of β_M (and β_L), which we find in the results below. This would mean we overestimate the market power degree in 53.

2.4 Data

We used COMPUSTAT data for the U.S. and ORBIS data for 26 European Union countries plus the U.K. Both databases are often used for empirical analysis to compute markups at the firm level; we use both to contrast the two regions.

For the U.S. estimations, we use the bundle of variable inputs, not the individual inputs, so the bundle includes materials and labor costs as in De Loecker et al. (2020). To compute the markup for Model 1 as in equation 42, we use as estimated output elasticity the coefficient resulting from the estimations considering revenue data (Gross Output) as the dependent variable and applying the De Loecker method assuming the bundle COGS variable, as the flexible input[91]. "Cost of Goods Sold" (COGS) is an item directly

reported by the financial statement of the firm they use to measure the variable input. It includes all expenses directly attributable to the production of the goods sold by the firm and comprises materials and intermediate inputs, labor costs, and energy. Subsequently, we adapt the LHS as in equation 65. For estimations for Europe, we use the separate shares of materials and labor costs over sales to compute the LHS, also in differences.

Compustat is a database of financial, statistical, and market information on active and inactive global companies worldwide. We consider only U.S. companies for this study. We base our estimations for De Loecker et al.'s methods (Model 1) on the STATA codes and routines shared by the authors at De Loecker et al. (2020, 2018). We apply the same assumptions imposed by these authors to obtain the final sample, to confirm whether we obtain the exact estimates as them, and therefore test for the same sample, our alternative method. See Appendix L. After we confirm we obtain the exact estimates as in De Loecker et al. (2020), we add necessary assumptions. Most importantly, to eliminate extreme materials share values, an assumption that De Loecker et al. (2020, 2018) was just partially accounted for in their paper. See Appendix M for those additional assumptions.

Following De Loecker et al., we drop firms with negative sales, COGS values, or SGA expenses. We also trim 1% extremes of the sample based on the sales-to-cogs ratio within each year (all sectors). Later, we trim the 1% extremes of cost shares defined as $\frac{COGS}{COGS+capitalexpenses}$ and as $\frac{COGS}{COGS+capitalexpenses+SGAexpenses}$. Capital expenses are those the authors estimate using an estimated user cost multiplied by the value of capital.

Summary statistics about the final sample are provided in Table 9. The final sample of De Loecker et al. includes 248,358 observations, of which 21,393 are firms with between 1 and 62 years of financial data and an average of 11.6. Capital equals the property plant and equipment gross variable in Compustat (PPEGT). Notice that only 247,082 observations have non-missing capital. However, the authors assigned estimated markups to all firms regardless of whether they had non-missing capital or not. All firms with a markup can be included in the national or sector-level aggregate calculations by having a non-missing COGS share and an elasticity estimated for their corresponding sector and year.

TABLE 9
Summary Statistics U.S.: De Loecker et al. sample

	Mean	Median	SD	p10th	p90th	Min	Max	N
Sales	1,933.44	148.03	10,250.74	6.52	3,095.13	0.00	461,824.69	248,358
COGS	1,293.19	89.23	7,609.35	3.51	2,026.82	0.00	407,563.59	248,358
Capital	1,467.13	58.62	9,559.33	2.25	1,958.79	0.00	456,525.69	247,082

TABLE 10
U.S. COGS share: Original De Loecker et al.

	Mean	Median	SD	p10th	p90th	Min	Max	N
COGS Share in Sales (Adj.)	0.62	0.65	0.19	0.34	0.84	0.04	1.00	241,619
COGS Share in Sales (Raw)	0.68	0.66	0.83	0.34	0.86	0.04	80.59	248,358

TABLE 11
U.S.: De Loecker et al. sample after additional assumptions

	Mean	Median	SD	p10th	p90th	Min	Max	N
Sales	1,983.84	158.50	10,387.20	7.90	3,199.63	0.00	461,824.69	241,619
COGS	1,324.40	94.66	7,710.11	3.98	2,089.02	0.00	407,563.59	241,619
Capital	1,494.05	61.12	9,669.73	2.42	2,010.10	0.00	456,525.69	240,508

Note: Summary statistics for U.S. listed firms from 1955 to 2016. Data are obtained from COMPUSTAT (balance sheet and income statement variables). Nominal values are deflated using US GDP and are expressed in thousands of 2010 USD. The dataset contains 21,393 unique firms across 23 sectors at the two-digit level, according to NAICS. Values are expressed in thousands of dollars.

De Loecker et al. (2020) includes in their sample firms with COGS share of up to 800% of sales. Although those observations represent less than 1% of the sample, to control for still extreme outliers in the estimation of markups, we assume that the COGS share is not greater than one and we eliminate approximately 6,736 observations (firm x year) as shown in Table 10. See Table Q.34 in the appendix to see the cogs share distribution of the original sample. We estimate markups for a slightly smaller sample for both models so, plots might slightly change. However, we present estimates with their original sample for comparison purposes.

ORBIS is a database commercially provided by the Bureau van Dijk and contains standardized financial statements for over 300 million companies worldwide (Bureau van Dijk, 2007). For ORBIS, we restricted the sample following the recommendation of former OECD analyses using this data. The sampling strategies and reporting threshold in ORBIS vary across the underlying data sources for each country over time, so concerns are relative to the degree of representatives at the country level. ORBIS contains mainly information from European companies relative to other regions; as companies tend to be larger, the proportion of firms reporting for Europe and other countries is more similar. Therefore, we limited the sample to only firms displaying, on average, at least 20 employees throughout the analysis Calligaris et al. (2018).

We also exclude firms that do not report an industry code. The ORBIS results, unless reported otherwise, rely on the sample of firms over the period 1982-2016, where we eliminate companies with a reported cost of materials plus the cost of labor to operating revenue ratios in the 99 and bottom one percent. This database permits 1) to estimate labor share more accurately than De Loecker et al. (2018) because they only have the labor share for a few firm-year pairs and had minimal information on total wage bills. 2) to have another type of proxy for the flexible input in the markup estimations, we use the

difference between operating revenue and value-added, given that the cost of materials has many missing observations.

There are different challenges with ORBIS. Using accounting data is generally problematic since these were not built to measure markups, and there is excellent measurement error. But most importantly, it is that it is challenging to capture variable costs adequately. The cost of materials variables used in this paper is similar to COGS (Cost of goods sold), the proxy for variable input used by De Loecker style models. These reflect costs regarding the purchase and transformation of input into a product. These items are thought to be in a more significant part variable, but some might arguably be fixed. The same companies can report more flexible inputs in other accounts, such as in the SG&A (selling, general, administrative expenses), so they are not captured when COGS-like accounts are used. Using accounts similar to SG&A is not a very practical solution either because these accounts include mostly fixed costs, but again; companies could report fixed costs that increase with the size of operations so that they are variable costs. Thus, the classification of expenses can vary by sector and affect the estimations [92].

The final sample used for markup estimations for European Union countries, including the UK in ORBIS, contains 7,933,787 observations, of which 1'393,980 distinct firms in 24 countries between 1 and 23 years of financial data and an average of 5.69 years.

The final sample used for markup estimations for European Union countries, including the UK in ORBIS, contains 7,933,787 observations, of which 1'393,980 distinct firms in 24 countries between 1 and 23 years of financial data and an average of 5.69 years. We cover the period 1994-2016. We deflate the relevant variables for their corresponding purchasing power parity-adjusted country-industry-sector-year-specific deflators. Gross output or sales are deflated by the Gross Output Deflator, cost of materials by the Intermediate Input Deflator, cost of employees by CPI (Consumers Price Index), and investment and capital by Gross Fixed Capital Formation Deflator, all 2005 Prices [93].

The analysis focuses on mining, manufacturing, non-financial market services (excluding real estate), and utilities sectors according to NACE Rev. 2 for 59 2-digit sectors. We impose a threshold of 500 observations per country for the entire period. Because of this, Greece, Lithuania, and Malta are dropped from the sample. Croatia is also excluded due to the lack of industry-specific deflators.

ORBIS financial data reports firm-level information on different versions of consolidated and unconsolidated financial accounts. We focus on unconsolidated accounts for all firms unless the firm is a stand-alone one, meaning that we choose the consolidated accounts for those firms that do not have an ultimate owner.

For the benchmark specification of production function estimation, 5-year rolling

estimations, we set a minimum threshold of several observations, given the number of instruments required in the De Loecker model to converge during the GMM procedure. So, 1) at least 1000 observations in each industry and 2) at least 50 observations per period-industry sector cell: a period is equivalent to a 5-year rolling window between 1998 and 2020. After these assumptions, we preserve all 2-digit sectors. See in Table 12 the summary statistics.

TABLE 12
Summary Statistics Europe: De Loecker et al. sample after add. assumptions

	Mean	Median	SD	p10th	p90th	Min	Max	N
Sales	43,012.06	4,277.34	661017.28	122.51	48,961.41	0.29	346,868,352.00	7,305,235
Cost of Employees	6,618.98	971.28	99,534.44	22.05	7,926.90	0.00	41,158,916.00	7,305,235
Cost of Materials	32,345.53	2,521.12	518817.45	59.87	36,251.25	0.04	239,160,656.00	7,305,235
Capital	10,213.00	329.50	249397.09	2.90	7,515.27	0.00	135,913,440.00	7,305,235

Note: Summary statistics for European firms, including the UK, from 1994 to 2016. Data are obtained from ORBIS (balance sheet and income statement variables). Nominal values are deflated using variable-specific deflators and are expressed in thousands of 2015 EUR. The dataset contains 21,393 unique firms across 23 sectors at the two-digit level, according to NAICS. Values are expressed in thousands of euros. After imposing a minimum threshold of 500 observations per country for the full period, Greece, Lithuania, and Malta are dropped. Croatia is also excluded since it does not count with deflator values at the industry level.

2.5 Results

This section describes the production function and markup estimation results on the U.S., EU, and UK ORBIS data. We start by reviewing two versions of estimated revenue elasticities concerning materials for Models 1 and 2. First, allowing for production technologies to be different across sectors (estimating separately for each sector) and second, allowing simultaneous changes in production technologies across sectors and time.

We then compare the levels and dispersion of the aggregate markups from the various specifications. We continue by examining the variable input and trends of elasticities, the markup components, to understand the trend. Finally, we look at how the aggregate unweighted and weighted markups for the two regions differ to understand the role of company size as suggested by Autor et al. (2017).

There are good reasons to consider that technology varies across productive sectors across the economy and consequently has time-sector varying elasticities. Retail, for example, differs substantially from how highly specialized industries operate. For instance, how Walmart produces differs from medical device companies. Even more important is to acknowledge that technology is time-varying. Technology has likely been evolving in two decades of data. The latter is significant for estimating markups since systematic technological change implies a time-varying output elasticity that misleads results without proper consideration. For instance, if we only rely on sector-varying parameters, we could be underestimating the true α if that true parameter was increasing, which would mistakenly lead to overestimated markups. Recent work has implemented the

strategy of rolling windows between 3 and 10 years to allow for these properties of technological change (See Chiavari and Goraya (2021); Jo (2020); Baqaee and Farhi (2020b,a). Furthermore, allowing changes over time allows us to consider whether a particular sector has become more labor-intensive than other inputs. Consequently, it is possible to control for potential structural changes to returns to scale, as it is one of our main interests.

We consider two types of estimations of elasticities to compute markups. First, our benchmark specification, the *Rolling* version of the estimates α_{st}^{V}, varies across sectors and time. Since panel data is required, we use 5-year rolling windows so that the elasticity estimates for the year t are based on financial data in years t-4,t-3,t−2, t-1, and t. Thus, for time variation, the regressions were estimated in looping regressions of windows of five years for each industry.

Second, when the elasticity varies only across sectors, α_{s}^{V}, we refer to the *No−rolling* estimations. No−rolling refers to a production function estimation with constant coefficients for each sector. For European countries (ORBIS sample), estimations were made for the 2-digit NACE Rev. 2 industry classification sectors. For the U.S., estimations were made for 2-digit NAICS sectors.[94]

2.5.1 Production function estimates

The measure of markups in the equation 42 is the ratio of output elasticity α and shares of the variable input in sales $\frac{PQ}{P_M M}$. Since the share is directly obtained from the firm's financial data, we need to estimate elasticities for the two models. The benchmark specification uses sector and time-specific estimated elasticities to capture the technology changes across sectors and time.

For model 2 (New), we obtain a coefficient we call $\beta_{st} = \frac{\hat{\gamma}}{\alpha_{M,st}}$ correspondent to the coefficient obtained for the variable Dk in the regression. Thus, we consider that the estimated alpha is equivalent to the inverse of the beta, in the form $\alpha_{\hat{M},st} = \frac{1}{\beta_{st}}$.

We obtain the estimated alpha coefficient directly for model 1 (De Loecker et al.), so $\alpha_{\hat{M},st} = \alpha_{\hat{M},st}$. In table 13, we present summary statistics of the estimated elasticities (weighted by sales) and the variable input shares for the U.S. We generally observe that Model 1's elasticities are higher than Model 2's, as expected. Our preferred specification, **New** row, yields an average output elasticity of 0.77, while **De Loecker et al.** is 0.84.

TABLE 13
U.S. Output Elasticities (Weighted by sales)

	Mean	Median	SD	p10th	p90th	Min	Max	N
De Loecker et al. (No Rol)	0.84	0.84	0.01	0.84	0.85	0.83	0.89	203,861
De Loecker et al.	0.84	0.84	0.02	0.82	0.87	0.82	0.89	203,861
New (No Rol)	0.77	0.76	0.01	0.76	0.77	0.76	0.84	203,861
New	0.77	0.77	0.03	0.73	0.81	0.73	0.90	203,861
Variable input share (cogs/sales) Weighted	0.67	0.66	0.03	0.65	0.73	0.64	0.75	203,861

Note: The Table presents summary statistics of the estimated output elasticities concerning the variable input (materials) for the four specifications from the estimations of cobb-Douglas production functions. No Rol: No-rolling regressions
Title without No Rol: No rolling regressions with heterogeneous elasticities within two-digit NACE codes.
The mean output elasticities presented in this table (Table 13) seem relatively the same between the "rolling" and "no rolling" versions in De Loecker's estimations because we estimate the output elasticities for De Loecker's "no rolling" version of the estimates by sector but constant over time. In contrast, the plots in Figure 30 with the green, long dashed lines present the markups for the "no rolling" version of the De Loecker using the constant elasticity for all years and sectors of 0.85 as they did in their paper, De Loecker et al. (2020). Put differently, we consider "no rolling" in this table as elasticities that change over sector but not across time, while in their paper they just assumed a constant elasticity value. To be consistent with this, we plotted as well the "no rolling" version of the "new" model, assuming a constant elasticity exactly the same as De Loecker's, this is why the green, long dashed lines look the same between the panel (a) and (b) of Figures 30 and 31.

2.5.2 Aggregate markups

We next compute markups using the estimated elasticities. In the remaining analysis, we focus on aggregate markups in levels. To treat for outliers, we trim an extra 0.1% of the bottom and top of the estimated markup distribution, but we focus on the non-trimmed sample for comparison, Which leaves 203,861 firm*year observations.

2.5.2.1 Levels and distribution The average aggregate markup is computed as a weighted average, as defined in:

$$\mu_t = \sum_i w_{it}\mu_{it} \tag{66}$$

Where w_{it} refers to the weight of each firm's sales (revenues) in total sales for the year in the whole sample, $w_{it} = \frac{R_{it}}{R_t}$ and $R_t = \sum_i R_{it}$. It is essential to highlight that in the original paper of De Loecker, they consider all sales of the firms in the sample, regardless of whether they contributed to the estimation of the elasticities or had a non-missing value for COGS shares. For instance, firms that potentially had the materials variable missing were considered to add to the total sales. Summary statistics are presented in Table 14. We account for weighting markups to control the influence of size by sales (or employees, in case the weights account for this).

To confirm we obtain a baseline to compare the results of the proposed New specification,

we reproduce De Loecker's estimations and graphs. Thus, we estimate the production function using the methodology applied to their sample. Later, we compute the markups. We use their code in Stata published along with (De Loecker et al., 2020) to estimate Model 2, De Loecker et al.[95]

TABLE 14
U.S. Markups (Weighted by sales)

	Mean	Median	SD	p10th	p90th	Min	Max	N
AGG MARKUP (PF De Loecker et al.)	1.39	1.42	0.09	1.25	1.49	1.19	1.58	203,861
AGG MARKUP (PF De Loecker et al. (no rolling))	1.39	1.42	0.11	1.21	1.51	1.17	1.57	203,861
AGG MARKUP (PF New)	1.27	1.29	0.07	1.17	1.36	1.14	1.40	203,861
AGG MARKUP (PF New (no rolling))	1.26	1.29	0.10	1.09	1.36	1.06	1.43	203,861

Note: It estimated production-function coefficients for the two models' different specifications. It was considered an essential first stage in which a third-order expansion in the production-function inputs and time-fixed effects were included.

TABLE 15
EU Markups (Weighted by sales)

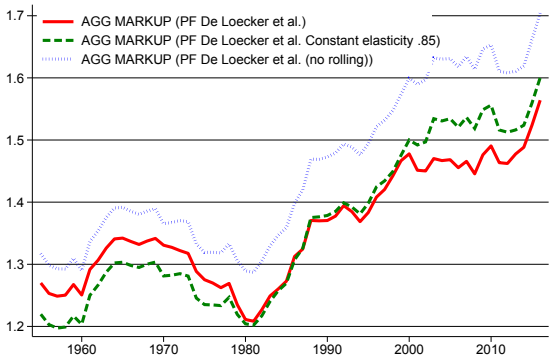
	Mean	Median	SD	p10th	p90th	Min	Max	N
AGG MARKUP (PF New)	1.07	1.07	0.02	1.05	1.09	1.03	1.11	7,305,235
AGG MARKUP (PF New (No Rolling))	1.06	1.06	0.02	1.04	1.08	1.03	1.11	7,305,235
AGG MARKUP (PF De Loecker)	1.25	1.26	0.02	1.22	1.28	1.21	1.28	7,305,235
AGG MARKUP (PF De Loecker (No Rolling))	1.26	1.26	0.02	1.23	1.28	1.22	1.29	7,305,235

Note: Estimated production-function coefficients for the two models' different specifications. It considers a basic first stage in which a third-order expansion in the production-function inputs and time-fixed effects were included.

Figure 29 reports the resulting aggregate markups weighted by sales, as in De Loecker et al. (2020). Axis Y represents the markup estimator, and axis X, represents the years. The benchmark specification and the rolling estimations of the average markup are presented as a solid line. The dashed lines represent the estimated markups for the economy's assumed constant elasticity of 0.85.

The figure corresponds precisely to the trends and levels as in Figure II (A), page 576 of their paper, so we can confirm we correctly estimate elasticities and share of the variable input in revenue applying their methodology.

FIGURE 29
Average U.S. Markup (weighted by sales). De Loecker et al.



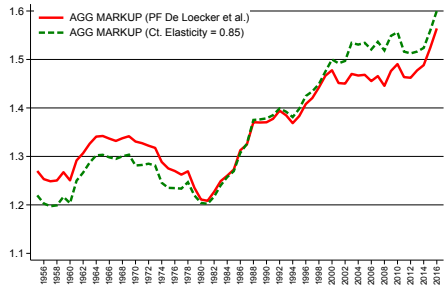
Note: Original De Loecker et al. (2020)'s replicated results using the original sample as described in Table 11, Rolling vs. no rolling estimations

Thus, for the sample with additional assumptions, figures 30 and 31 present the evolution of the average weighted (by sales) markup estimators for models 1 and 2. For the U.S., Figure 30a reports model 1 results. We observe similar trends and levels with Figure 30, which represents that despite limiting the sample to COGS shares below one, the results remain similar. Figure 30b reports the results for model 2, and it is noticeable that although trends are similar, the levels are lower. As more recent studies highlight (see De Ridder et al. (2021)), using the 2-stage GMM ACF methodology might lead to biased markups upwards (or downwards). The IV-GMM estimate delivers a biased estimate of the actual value of α if prices and lagged variable inputs are correlated, and the relationship of the correlation can be positive or negative; so using revenue rather than quantity data to measure output creates an omitted variable bias because the revenue production function has priced in the residual. Thus, we cannot trust the levels from figure 30a. However, we no longer need to deal with this issue as in model 2, we can see how upward De Loecker's estimates can get for the case of the U.S. Appendix R, figure R.23 presents the non-rolling estimations for both models.

Figure 31a presents model 1's European results. The patterns are similar to evidence found for French data in De Ridder et al. (2021) and other sources of European data in De Loecker and Eeckhout (2018) for the same analysis period. Although there is an increase, the levels of the markups tend to be lower than those estimated for the U.S. Table 15 presents summary results for Europe. Appendix R, figure R.24 presents the non-rolling estimations for both models.

FIGURE 30
Average U.S. Markup (weighted by sales) - Adjusted Sample (1955 - 2016)

(a) De Loecker et al.



(b) New

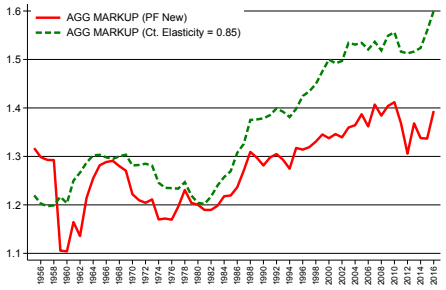
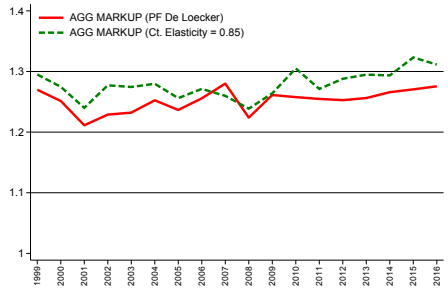
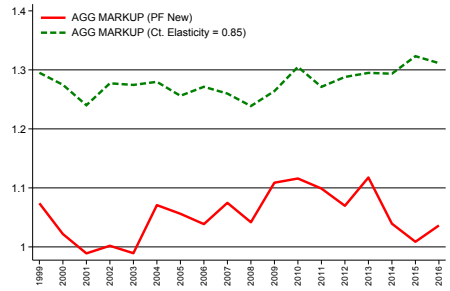


FIGURE 31
Average European Markup (weighted by sales) - (1998 - 2016)

(a) De Loecker et al.



(b) New



Note: The figures present the aggregated (weighted by sales) markup for the adjusted sample, contrasting the "rolling" (time-changing elasticity version, green long-dashed lines) and "non-rolling" (or constant elasticity version, red continuous lines) of estimations for both models. All figures contrast the rolling version of the markups for each model (red, solid lines) with a markup assuming a constant output elasticity for all sectors and years of 0.85 (green, dashed lines). Panels (a) represent model 1. Panels (b) present the aggregate markups for Model 2.
The mean output elasticities presented in this table (Table 13) seem relatively the same between the "rolling" and "no rolling" versions in De Loecker's estimations because we estimate the output elasticities for De Loecker's "no rolling" version of the estimates by sector but constant over time. In contrast, the plots in Figure 30 with the green, long dashed lines present the markups for the "no rolling" version of the De Loecker using the constant elasticity for all years and sectors of 0.85 as they did in their paper, De Loecker et al. (2020). Put differently, we consider "no rolling" in this table as elasticities that change over sector but not across time, while in their paper they just assumed a constant elasticity value. To be consistent with this, we plotted as well the "no rolling" version of the "new" model, assuming a constant elasticity exactly the same as De Loecker's, this is why the green, long dashed lines look the same between the panel (a) and (b) of Figures 30 and 31.

Figure 32 shows the difference between weighted and unweighted markups for the U.S. The black (dashed) line is the unweighted average markup over time. It is calculated by adding up the markups of each firm in the data set and then dividing by the number of firms. The red (solid) line is the weighted average markup. It counts the markups by firms with a larger market share more heavily than those of smaller ones.

If the increase in markups were driven by the dominance of large, very profitable firms, we would expect the red (solid) line to be substantially above the black. That is, we would not expect small independent businesses to be increasing their markups. One could think instead that smaller firms face more vigorous competition and are pushed to slash prices. However, small independent businesses vastly outnumber large firms. They monopolize the unweighted average.

Yet, the unweighted average is where the most significant increase is observed. This chart tells the story of fierce competition among large multinational firms but ever-increasing market power by small businesses in the U.S. Figure 33 shows similar results for Europe.

FIGURE 32
Average U.S. Markup (Weighted vs. Unweighted) - Adjusted Sample

(a) De Loecker et al. (b) New

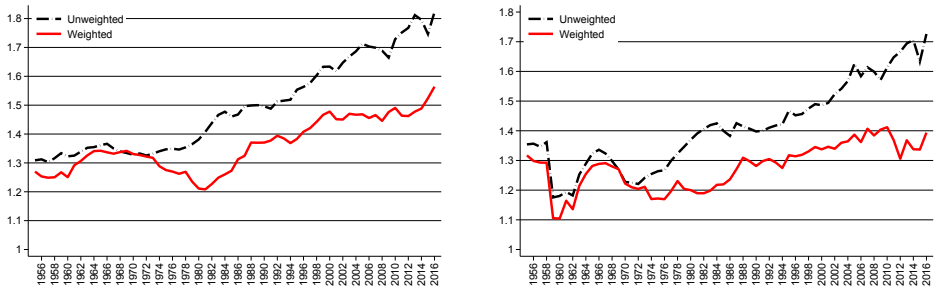
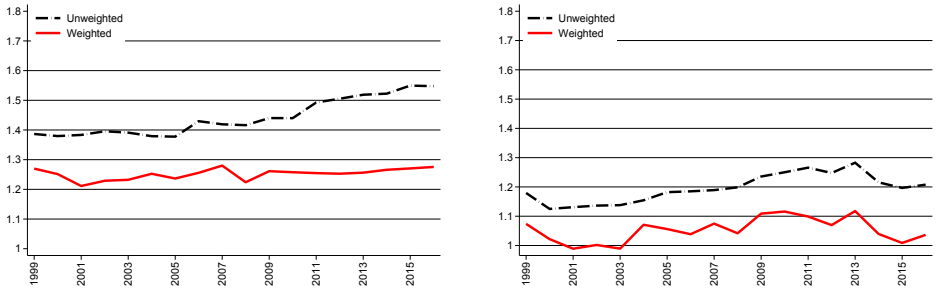


FIGURE 33
Average European Markup (Weighted vs. Unweighted)

(a) De Loecker et al. (b) New

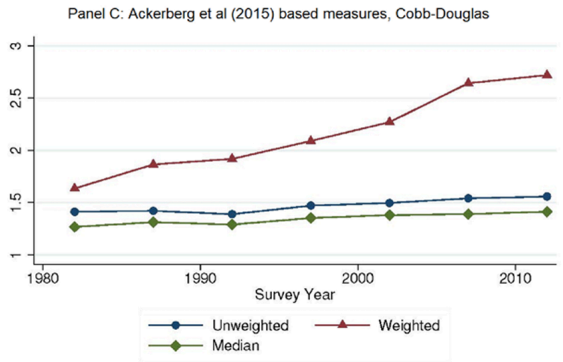


Note: The Figure contrasts both models' weighted average markups (red, solid lines) and unweighted average markups (black, dashed lines).

While Autor et al. (2020) shows sales-weighted markup to lie above the simple average markup, in contrast to the above, implying a positive association between markups and firm size, De Loecker et al. (2020) shows the opposite to be true (See Figure 34). This contradiction may be explained by the different data sources used in the two papers: The listed Compustat firms in the data used by De Loecker et al. (2020) are likely to represent only the most prominent firms of the Economic Census used by Autor et al. (2020). The two results may, therefore, be reconciled if the negative association between markups and

firm size found in De Loecker et al. (2020) is only a feature of the right tail of the firm distribution but is swamped out by a positive association between markups and firm size in the larger economy, as documented by Autor et al. (2020).

FIGURE 34
U.S. Weighted vs. Unweighted markups: Autor et al. 2020



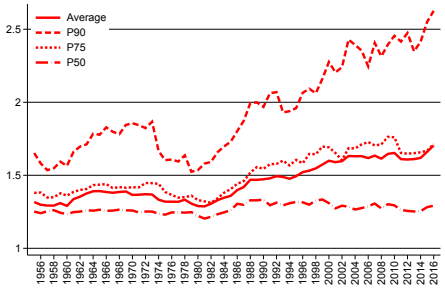
Note: Plot extracted from Autor et al. (2020), figure X, p. 68. Sales-weighted and unweighted markup trends.

In Figures 35 and 36, we present the percentiles of average weighted (by sales) markup estimators for each model. The hashed lines represent the estimated markups for a given distribution percentile to indicate the evolution over time. The average aggregate markup is the continuous line. We follow De Loecker's strategy of ranking the firms by markup every year, and to obtain the percentiles, we weigh each firm by its market share in the entire sample for the year. Doing this helps the percentiles to be directly comparable with the share-weighted markup. Since the ranking is updated each year, firms at the top may differ each year, showing the persistence of the markup process.

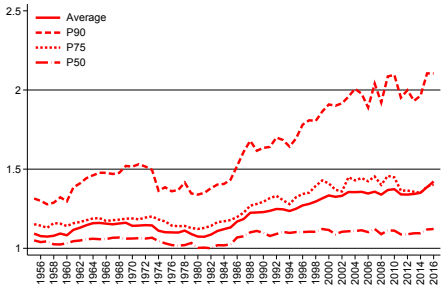
The U.S. and Europe present similar results for both models. They show an increase emerging almost entirely from the firms at the top half of the distribution of markups, meaning that smaller firms with the highest markups are the ones increasing the markups.

FIGURE 35
Average markup percentiles (weighted by sales). COMPUSTAT (U.S. only)

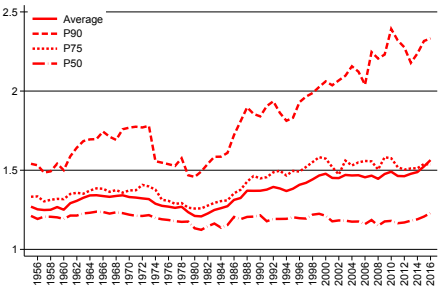
(a) De Loecker et al. (No Rolling)



(b) New (No Rolling)



(c) De Loecker et al. (Rolling)



(d) New (Rolling)

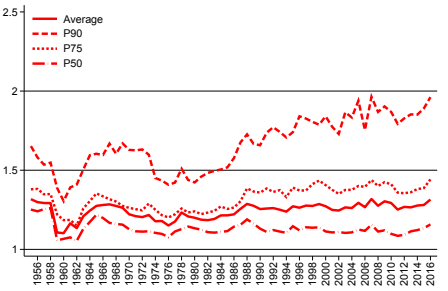
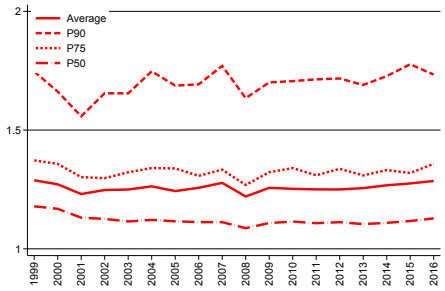
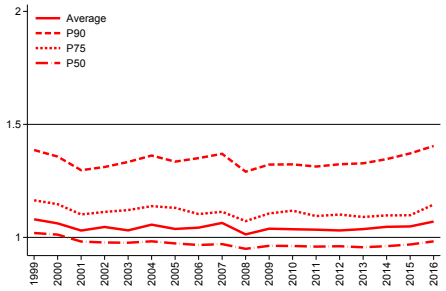


FIGURE 36
Average markup percentiles (weighted by sales). ORBIS (Europe only)

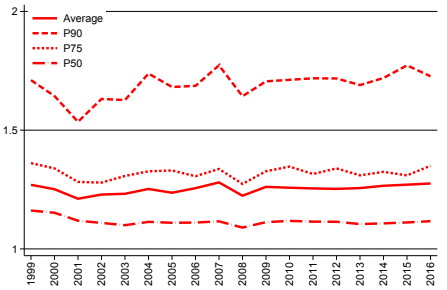
(a) De Loecker et al. (No Rolling)



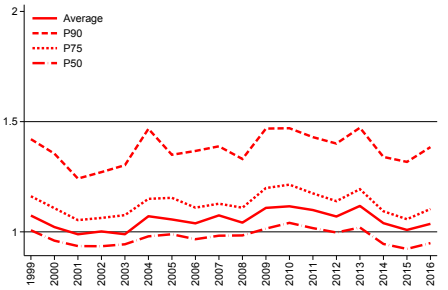
(b) New (No Rolling)



(c) De Loecker et al. (Rolling)



(d) New (Rolling)



2.5.2.2 Components of markups To understand the reasons behind the differences between Models 1 and 2, we observe the trends of the components of markups. Figure

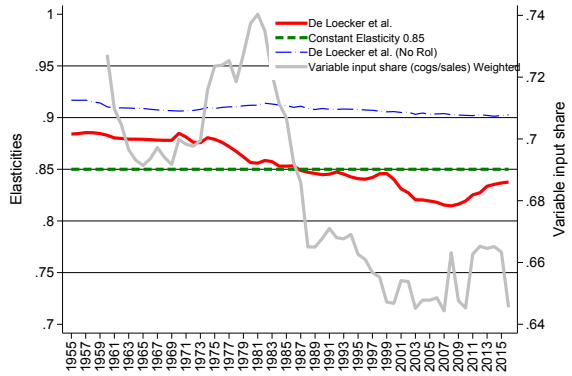
37 presents the variable input shares and the α estimators (output elasticities) for the models. Figure 37a red, hashed line shows that the (*Rolling*) version of the output elasticity in the U.S. is decreasing over time in the De Loecker method but that the magnitude of the decline is less than the one observed for the variable input share (grey, thick line). Thus, the steep increase in markups is driven by a more significant decline of the denominator relative to the decrease in the numerator. The green (long hash) line represents constant elasticities (*No-Rolling*) whose variation is only attributed to the sectors, and the declining trend for elasticities is not even there.

In contrast, Figure 37b shows that with the new model, the decline of the estimated elasticities is much more pronounced and seems similar to the decrease in variable input shares. Therefore, the markup increase is less noticeable as the numerator and denominator fall at similar rates. Thus, we infer that the returns to scale or the technologies have changed over time rather than the difference between the revenues and costs.

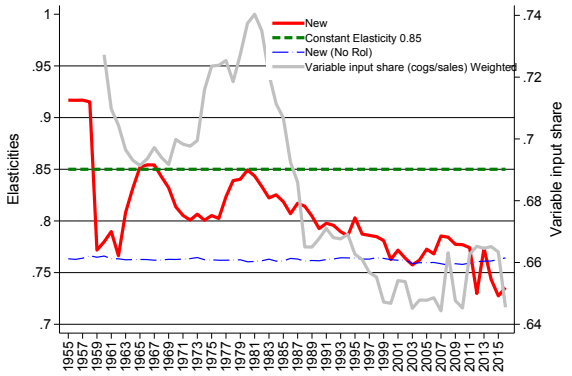
The latter aligns with what has been found about declining labor shares. Notice that in Model 2, figure 37b, the decline in the labor share is slightly less than the decline in variable input shares. Paul (2020) shows how the fall of the labor share has been the greatest for unskilled labor, but for skilled labor has increased, which aligns with the findings where the type of labor most demanded by a rising intangible economy has been skilled labor. Thus, skilled labor is the labor type getting the most of labor's income share. Markups still increasing in these circumstances means that the decline in labor share is slightly less than the decline in costs, which we precisely observe in model 2. Velasquez (2023) also shows how estimated labor shares have declined over time, using the version of the time-varying markup.

FIGURE 37
Elasticities and Variable Input Shares. (U.S.: 1955 - 2016, EU: 1999 - 2016)

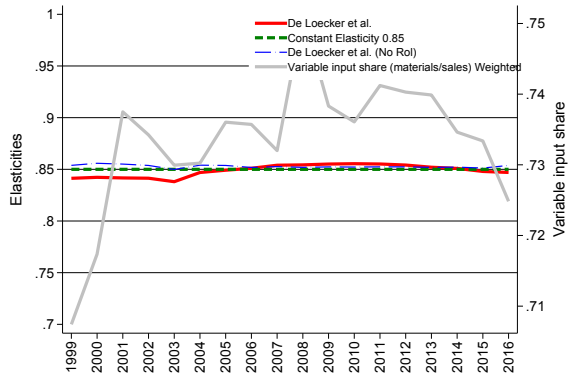
(a) De Loecker et al.(U.S.)



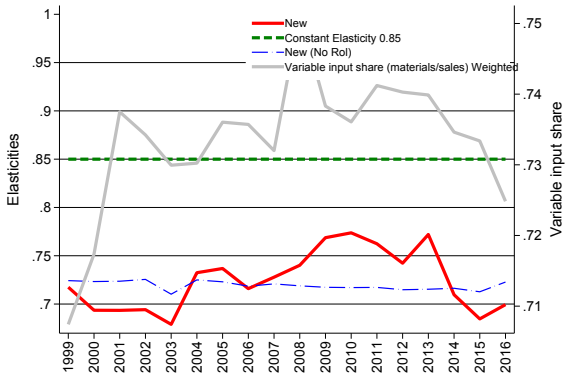
(b) New (U.S.)



(c) De Loecker et al. (Europe)



(d) New (Europe)



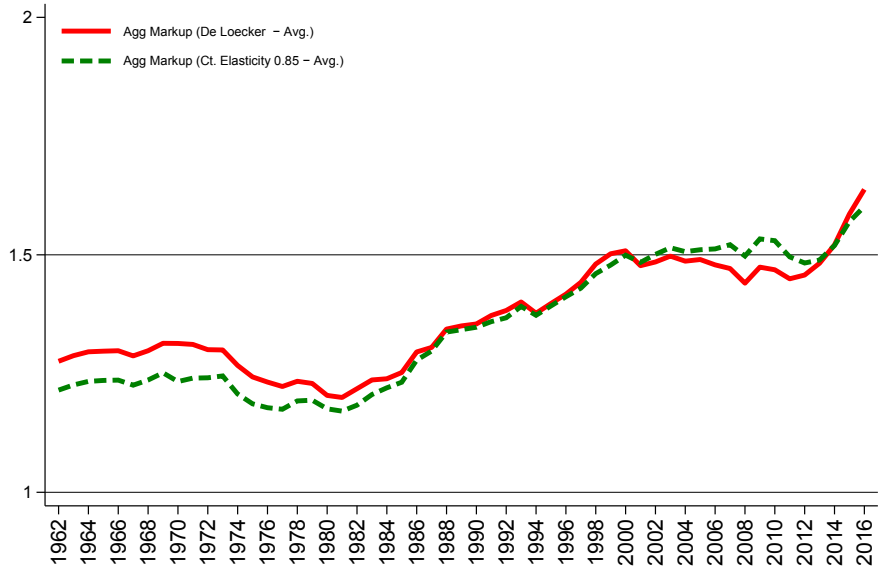
Note: Estimated output elasticities to variable input: $\alpha_{\hat{M},st}^{\text{model}_1} = \alpha_{\hat{M},st}$ and $\alpha_{\hat{M},st}^{\text{model}_2} = \frac{1}{\beta_{st}}$. Rolling corresponds to estimators resulting from the rolling regressions.

2.5.3 Sectors

This section presents aggregate markups by sector for both regions of study. Panels (a) are model 1 estimations, and panels (b) are model 2 estimations, contrasting a rolling version of the aggregate markup (red, solid line) with a constant version only varying by the share of materials costs (or cogs) in revenue (green, dashed line),

FIGURE 38
U.S. Manufacturing

(a) Average Markups (De Loecker)



(b) Average Markups (New)

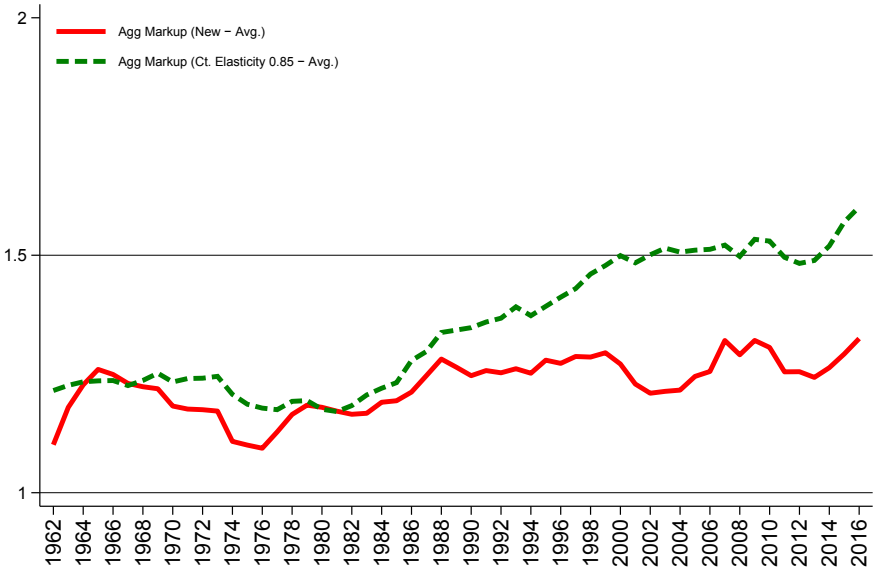
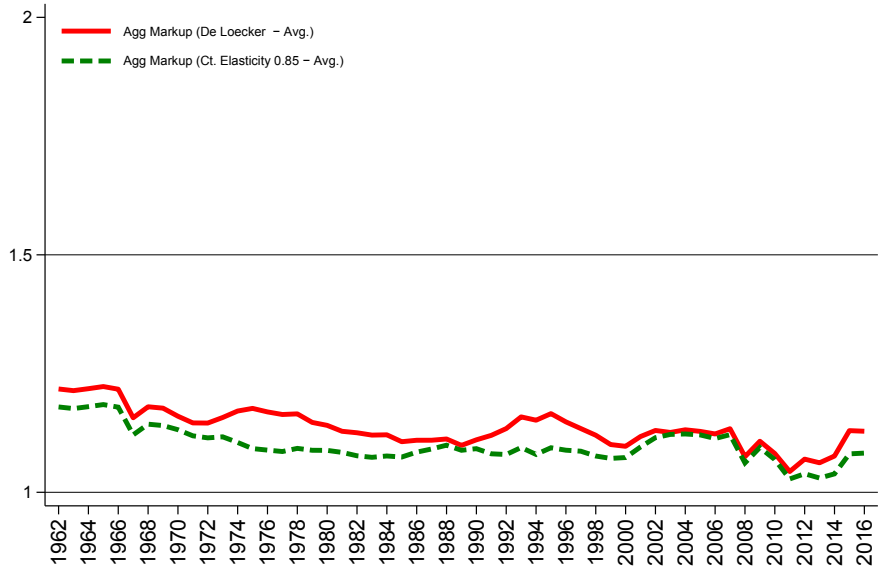


FIGURE 39
U.S. Wholesale

(a) Average Markups (De Loecker)



(b) Average Markups (New)

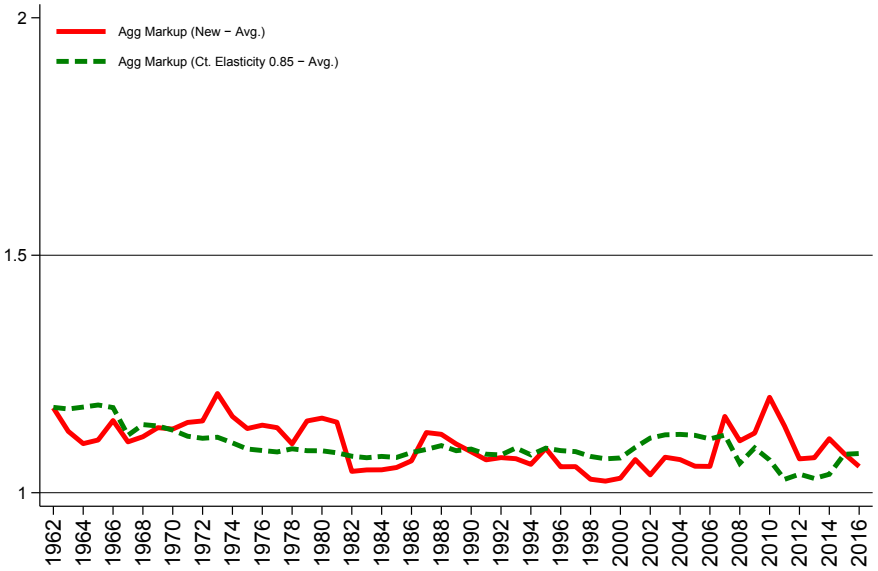
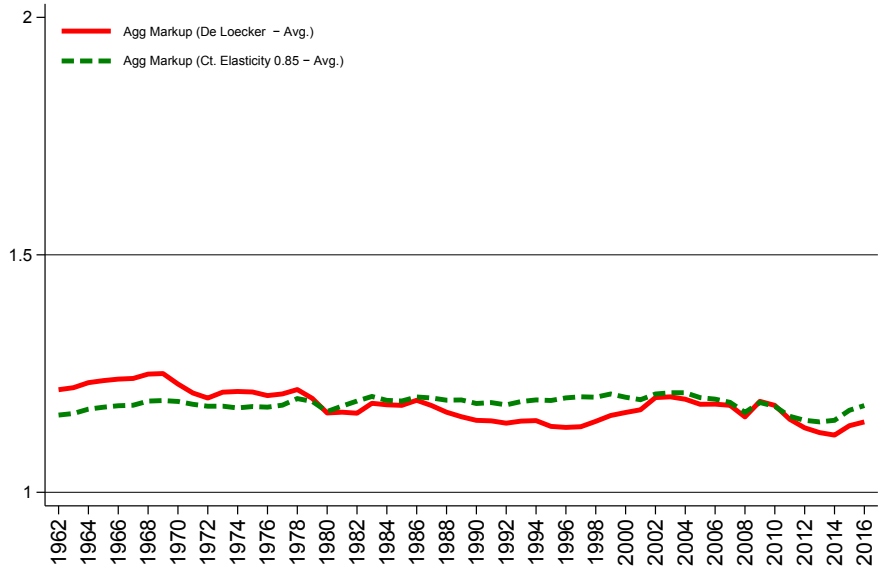


FIGURE 40
U.S. Retail

(a) Average Markups (De Loecker)



(b) Average Markups (New)

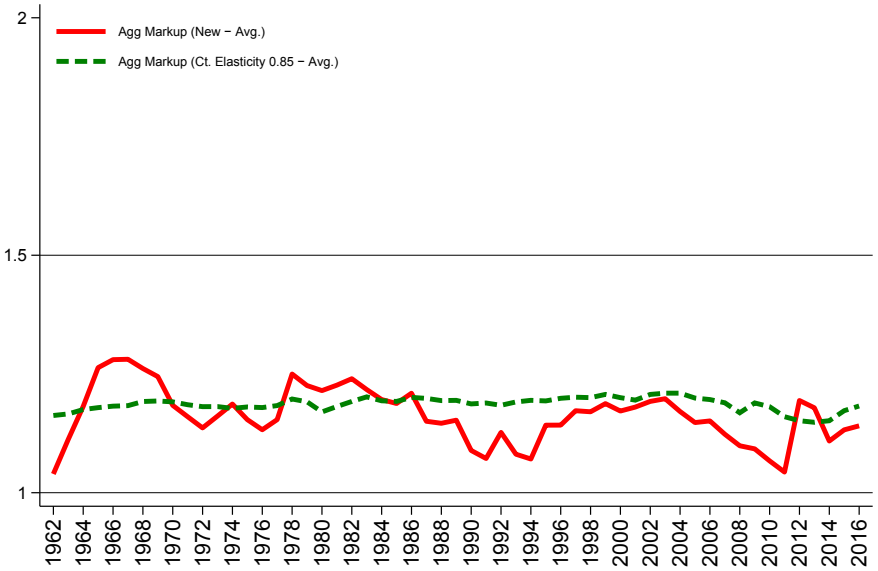
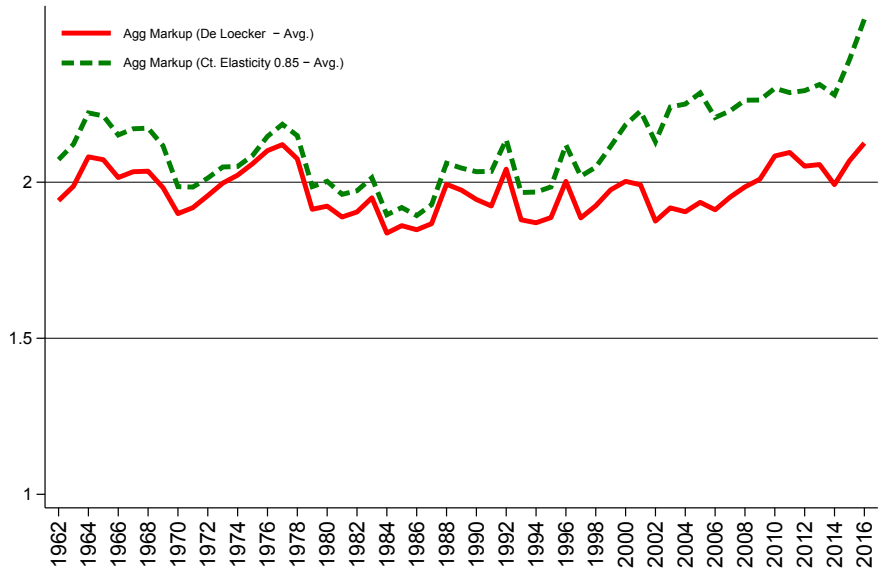


FIGURE 41
U.S. Information

(a) Average Markups (De Loecker)



(b) Average Markups (New)

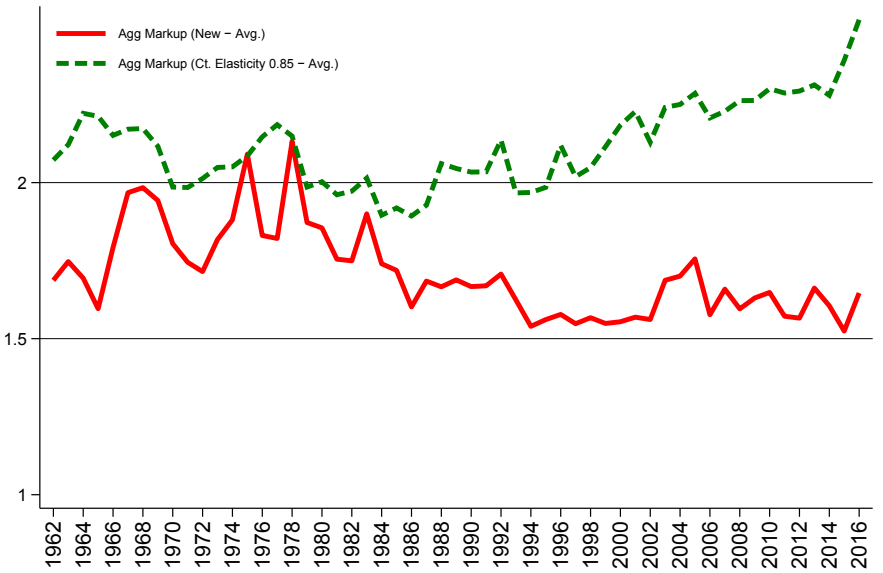
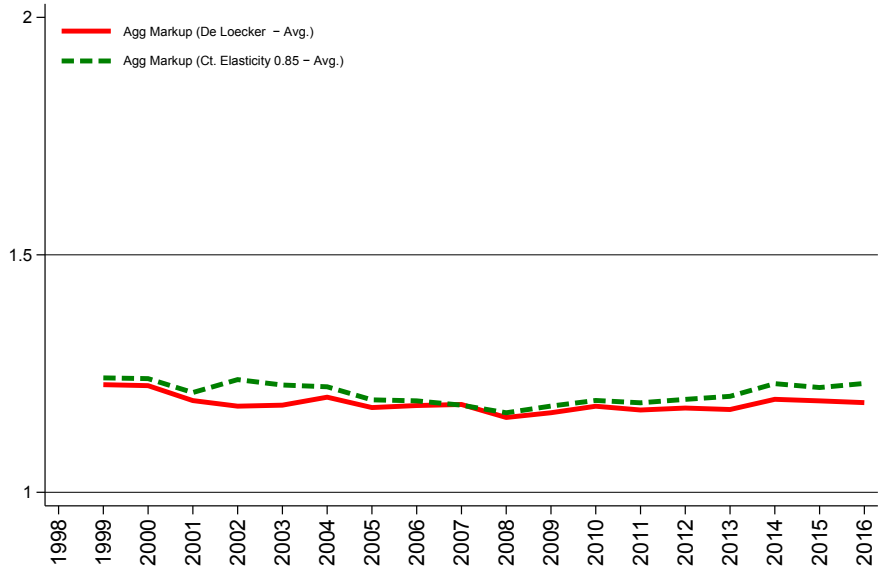


FIGURE 42
Europe Manufacturing

(a) Average Markups (De Loecker)



(b) Average Markups (New)

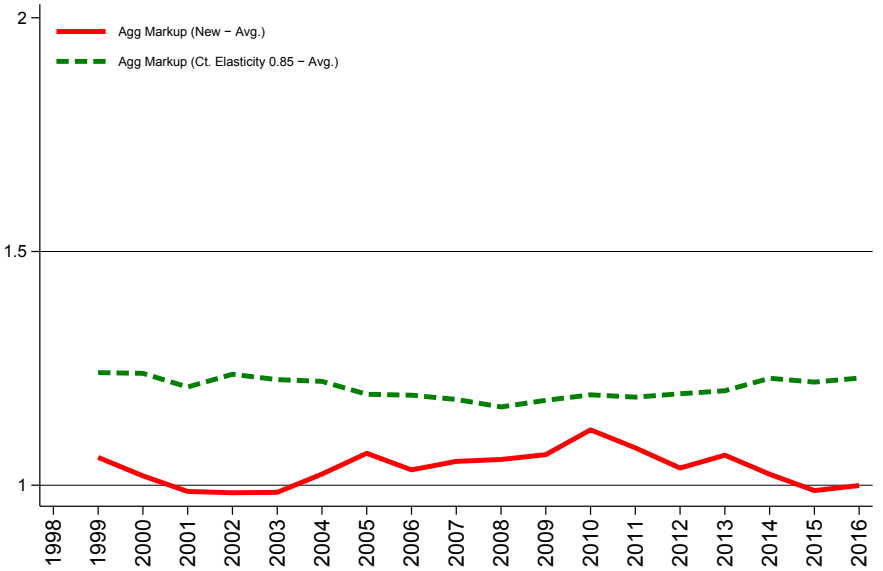
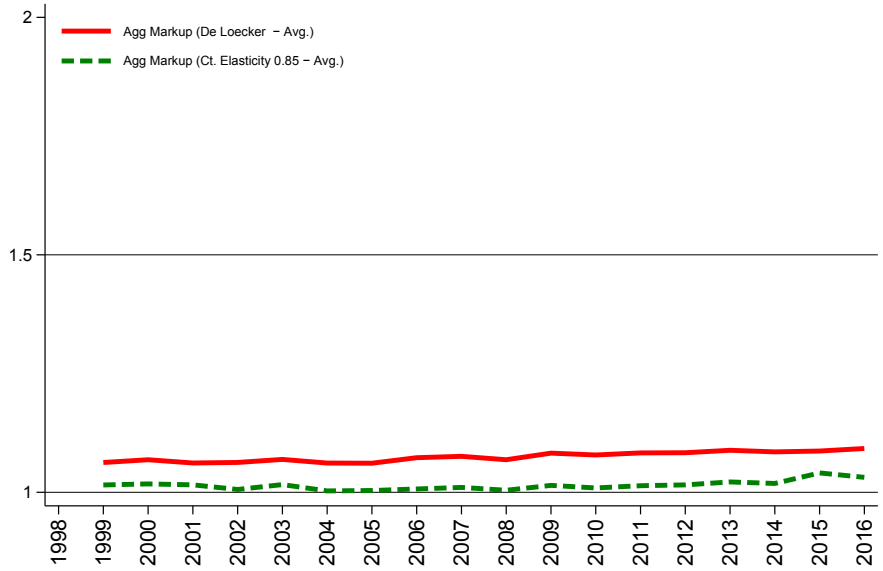


FIGURE 43
Europe Wholesale

(a) Average Markups (De Loecker)



(b) Average Markups (New)

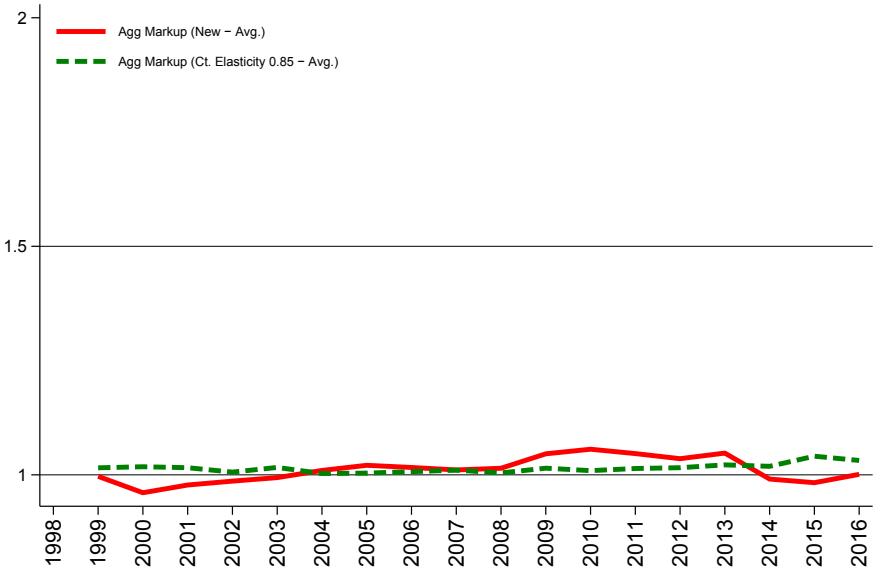
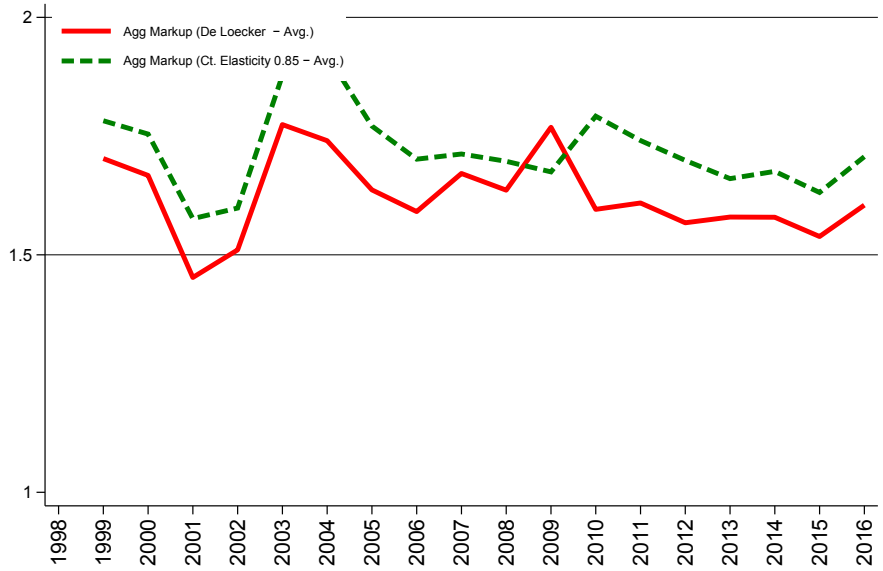


FIGURE 44
Europe Information

(a) Average Markups (De Loecker)



(b) Average Markups (New)

2.5.4 Percentiles by sector

FIGURE 45
U.S. Average Markups for Information by percentile

(a) De Loecker et al. (Rolling) (b) New (Rolling)

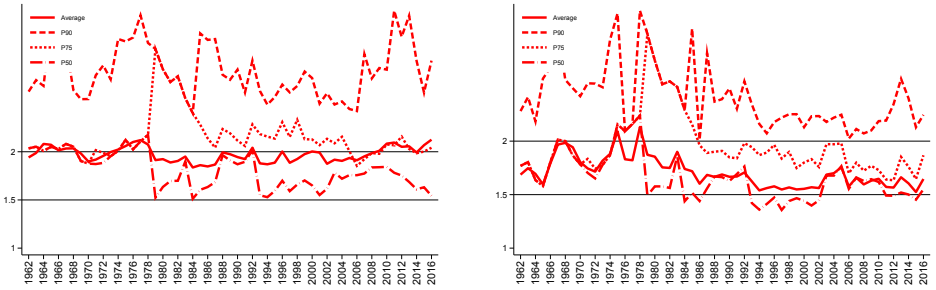


FIGURE 46
U.S. Average Markups for Manufacturing by percentile

(a) De Loecker et al. (Rolling) (b) New (Rolling)

FIGURE 47
European Average Markups for Information by percentile

(a) De Loecker et al. (Rolling) (b) New (Rolling)

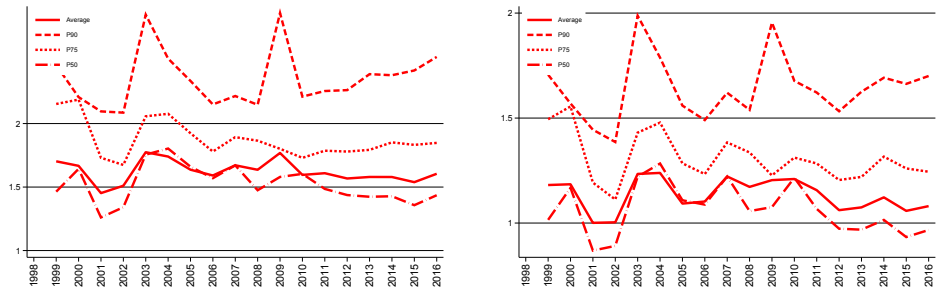
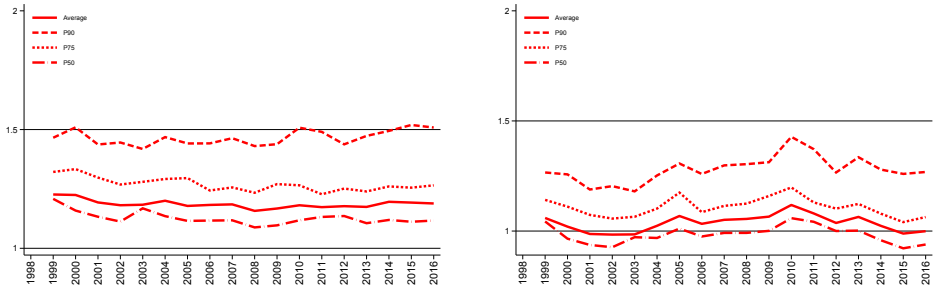


FIGURE 48
European Average Markups for Manufacturing by percentile

(a) De Loecker et al. (Rolling) (b) New (Rolling)

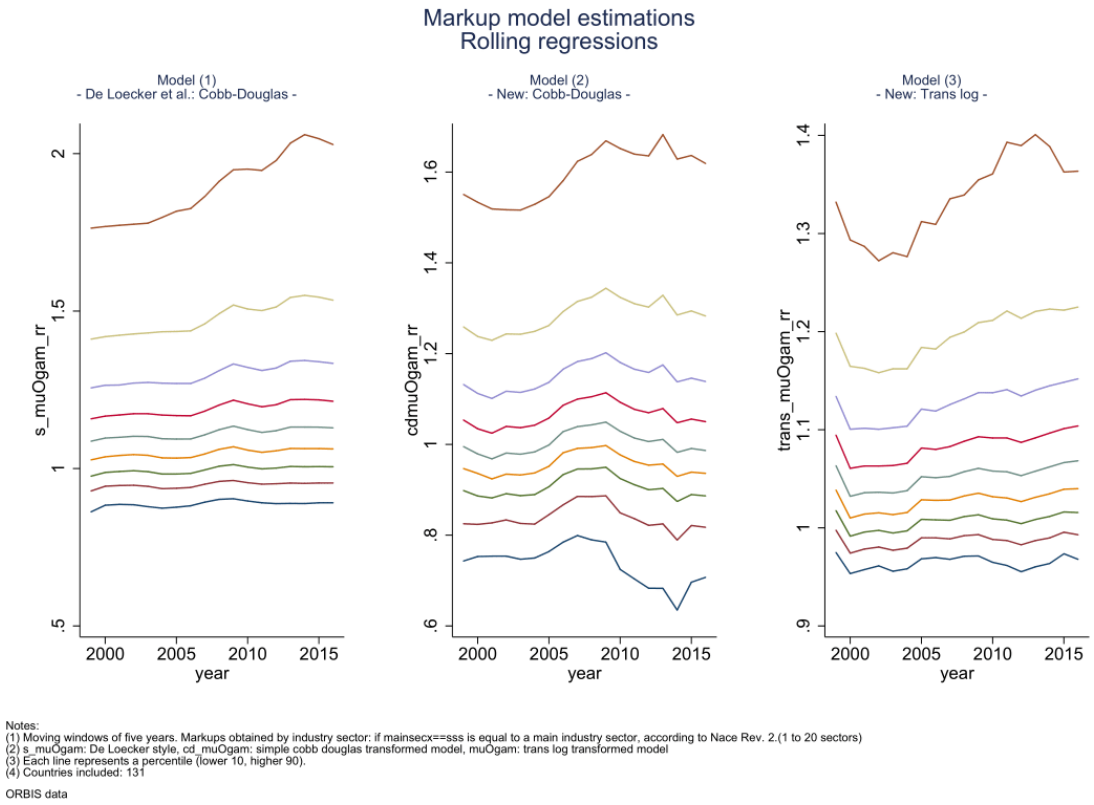


2.5.5 Global

This section explores a global sample of countries from ORBIS. Besides the EU countries and the U.S., it includes countries such as Canada, Japan, and South Korea, as long as it counts 50 firms' × year observations within each industry sector. Regressions were estimated for 1-digit NACE Rev. 2 industry classification sector and high-tech sectors following the J-tag ICT classification taxonomy from the OECD. See appendix Q for descriptive statistics.

Figure 49 presents weighted markup estimators for models 1, 2, and 3 (as described for the translog production function described in the appendix 2.3.3). Axis Y represents the markup estimator and axis X years. Each line represents the estimated markups for a given percentile of the distribution. The bottom line indicates the evolution in time of the 10th percentile of the distribution of estimated markups. The top line indicates the markups for the 90th percentile of the distribution in each given year. Each line in between, therefore, represents the different percentiles in between. See appendix U for markup estimations by sector. As with the U.S. and European cases, we observe lower average markups for the new model.

FIGURE 49
All sectors. Rolling within the sector



Notes:
(1) Moving windows of five years. Markups obtained by industry sector: if mainsecx==sss is equal to a main industry sector, according to Nace Rev. 2.(1 to 20 sectors)
(2) s_muOgam: De Loecker style, cd_muOgam: simple cobb douglas transformed model, muOgam: trans log transformed model
(3) Each line represents a percentile (lower 10, higher 90).
(4) Countries included: 131

ORBIS data

2.6 Conclusion

Recent evidence suggests market power has increased in the past two decades, more noticeable in industrialized economies. Several authors find that what seemed to be a purely microeconomic phenomenon of specific industries and markets has now extended to the overall economy, and it has been allegedly driven primarily by an environment of lax anti-trust regulation (Philippon and Gutierrez, 2017). The implications of increased average levels of market power across the board are significant since they can degrade key metrics in economy-wide well-being, including investment, innovation, total output, and income distribution. According to some authors, the emergence of a few "superstar firms", or firms at the extreme top of the income distribution, could be a symptom of the problem (Syverson, 2019; De Loecker et al., 2018; De Loecker and Eeckhout, 2018; Manyika et al., 2018; De Loecker et al., 2020; Loecker and Eeckhout, 2021).

Nonetheless, the evidence on market power seems problematic mainly because the methodologies used to estimate it are questionable. Either because it is assumed that rising concentration means increasing market power, which is not always the case under certain conditions, or because policymakers and academia overlook significant deficiencies and accept the results proclaimed by the predominant methodology of De Loecker and

Warzynski (2012); De Loecker et al. (2018, 2020) to measure firm-level markups (price over marginal costs). This methodology proxies markups by estimating output elasticities of a flexible input (e.g., materials or labor) and dividing it by the share of revenues on that flexible input.

For this reason, in this chapter, we proposed an empirical strategy that could address some of these issues and provide more enlightening evidence to this debate for academics and policymakers. To compute firm-level markups, we applied alternative markup estimation approaches that introduce more realistic assumptions of monopolistic competition, such as those used by Forlani et al. (2016, 2023) and computed markups assuming technological changes yearly (we estimated output elasticities for every year using 5-year rolling windows). Notice precisely that the most prominent critique of De Loecker's work was the lack of consideration of changing elasticities over time, as unaddressed endogeneities in the estimation procedures emerged from assuming the unrealistic assumption of perfect competition.

In particular, we make two types of comparisons. First, we contrast the De Loecker method ("no-rolling" or constant output elasticities version of model 1) with the Forlani et al. (2023) ("no-rolling" version of model 2). Second, we contrast each model's "rolling" or time-changing output elasticities version and "no-rolling".

In summary, unlike De Loecker's work predictions, we found that markups maintained levels of prior decades in recent years (see Figures 30 and 31). Indeed, we found that they have been declining in specific industry sectors. We also show that their markup levels are lower by simply reestimating De Loecker with changing technologies over time (contrasting model 1 "rolling" with the "no-rolling" version). Yet, markups are even lower when comparing models 1 and 2 "rolling" trends.

After presenting critical evidence in prior literature documenting the supposed increase in market power, the chapter focused on exposing, via discussion, the contrast between the concerning and not-so-concerning aspects of market power and the Superstar phenomenon.

On the one hand, the positive side is that Superstars may result from a complete economic model transformation. The recent combination of the power of new ICT technologies and the increasing returns to scale and reduction of marginal costs propitiated by intangible assets such as organizational capital, business models, big data, and software, among others, may create productivity efficiencies that increase concentration.

On the other hand, intangibles can be a good thing for efficiency, but concentrated in specific hands like cash-rich Superstars could be detrimental and lead to situations where these firms could exercise market power. For instance, Superstars' intangibles could

generate so much in net income that these firms could acquire any younger and smaller firms to eliminate competition without obstacles if not regulated. Superstars' intangible intensity could allow them to reduce their costs so much that Superstars' prices are unbeatable by competitors. These scenarios may require efforts from policy to make the acquisition or production of intangibles more accessible to the broader population of firms so their benefits can be enjoyed not only by a few.

This discussion points out the central premise of the chapter, which is that despite the risks of the second scenario, we contribute by providing evidence for the first one so we can think of Superstars in a more positive light. Accordingly, market power could be a result but not necessarily a cause of current trends. Put differently, we show evidence of improvements in production technologies (output elasticities have changed over time), so it is likely that decreasing marginal costs and increasing economies of scale have been behind rising concentration. Still, more importantly, that market power is not necessarily the main protagonist, as markups tend to remain stable or decline.

Decreasing marginal costs means producing each additional unit of a product or service is lower. When lower marginal costs, firms can reduce prices without sacrificing profitability. Lower marginal costs encourage firms to compete by offering lower prices to consumers, which puts downward pressure on markups. Increasing economies of scale also suggest that as firms produce more, their average costs decrease. This means that larger firms can produce at a lower cost per unit than smaller firms. Larger firms can use this cost advantage to offer products at competitive prices, again leading to lower markups.

Furthermore, decreasing marginal costs and increasing economies of scale often lead to increased competition in the market. Firms are incentivized to compete on price, which further reduces markups. Consumers benefit from lower prices as firms vie for their business. Moreover, consumers may expect and demand lower prices as they become accustomed to lower prices due to cost efficiencies and competition. Firms offering more affordable products or services are more likely to attract customers[96].

In general, we find various results that correspond with this narrative. First, especially in the U.S., the smaller firms with the highest markups are the ones with increasing markups (See Figures 32 and 33). This result tells the story of fierce competition among prominent Superstars and relatively higher markups for smaller firms. Although the following is speculative, it could be one explanation for this trend. The U.S. data is from the manufacturing sector, and recent evidence suggests that manufacturing's intangible assets intensity has increased over time in recent years. In 1980, the U.S. had a healthy group of "High Street" with many thriving small and medium-sized businesses engaged in robust competition. Additionally, local manufacturers participated in a nationwide marketplace with limited market power.

Over the subsequent decades, a transformation took place. This transformation involved the emergence of massive retailers and intricate global supply chains boosted by the efficiencies of intangible assets, like innovative business models and organizational capital, resulting in even more competitive pricing. The competition became so intense that many typical High Street businesses could not keep pace. Manufacturers, already contending with slim profit margins in the national market, were entirely displaced by global suppliers.

However, despite this intense competition, not all small businesses succumbed, and U.S. manufacturing endured. The surviving businesses shared a common trait - they were immune to being undercut by the large retailers and global suppliers. They thrived by offering specialized retail experiences and niche products or serving otherwise underserved markets.

Second, the average markups of firms in the top percentile of the distribution (p90) are relatively stable in the U.S. and even more so in Europe in the most recent decades. The rolling version of the alternative model application shows no increasing gap between that group's markups and those in lower percentiles, like firms at the 75th and the median (see Figures 35 and 36). This indicates that Superstars, in terms or markups, are not necessarily a robust result in prior literature.

Third, the observed trends in aggregate markups are explained by what is found for estimated output elasticities and shares of revenue on flexible inputs. The alternative model estimations show that the decline of the estimated elasticities is much more pronounced than that obtained with the De Loecker method, and it seems similar to the decrease in variable input shares. Therefore, the markup increase is less noticeable in the alternative specification as the numerator and denominator fall at similar rates.

Declining output elasticities may tell the story of diminishing returns of inputs, such as labor or materials, that ultimately affect aggregate labor share. It may also reveal the story of a more intangible economy that demands more skilled than unskilled labor. Given the prevalence of unskilled labor in economies, aggregate labor share and overall labor seem to present diminishing returns for production. Again, a decreasing elasticity also may reveal improvements in production efficiency. Firms may have become more efficient in utilizing these inputs thanks to intangibles, which allow them to produce more output with the same or fewer resources. This could be a positive sign of operational improvements and cost savings. Furthermore, they are potentially revealing technological advancements. Declining elasticity may result from adopting new technologies or production methods, as these advancements often lead to increased output with the same or fewer inputs, reducing the overall elasticity of these inputs.

Finally, the results by industry sectors show that U.S. and European manufacturing

and information exhibit markups that have remained stable over time or, indeed, have declined. Considering prior evidence showing that the intensity of the intangible asset of these sectors has increased in recent years, see Corrado et al. (2022); these results are pretty fascinating. These two industry sectors presented the most noticeable differences between the "no rolling" (green, long dashed lines in plots) and "rolling" (red, continuous lines) versions of estimated markups (see Figures 38 and 42).

Again, one would expect that if the combination of ICT technologies and intangibles has had a role, then markups should not been increasing as much in sectors that are characteristically more intangible, like the information sector, or where recent technological advances mixed with intangibles have taken central stage as in manufacturing. It has been identified that increasing intangibles in manufacturing subsidiaries has facilitated their absorption of the headquarters' technology transfer and enhanced the productivity of the local core activities (Szalavetz, 2017).

Also, R&D investments can be even more effective for the high-tech manufacturing sector than services (Felix Roth and Rammer, 2023). Artificial intelligence (AI) has also become the ultimate tool in manufacturing to improve efficiency and productivity. Generative and data-centric AI or synthetic data have helped automatize tasks to solve many manufacturing operations challenges. For instance, production facilities can determine the likelihood of future failures in operational machinery, allowing for optimizing production schedules and preventative maintenance and repairs to be scheduled in advance (Stackpole, 2023).

The results and discussion leave essential lessons and opportunities for future research. First, the resulting datasets of this chapter's exercise can be easily linked with other firm-level data sources to more formally test the connection between intangibles and markups. In particular, we could connect estimated markups for European and other non-US countries using ORBIS financial data with the innovation and intangible investment indicators obtained in the third chapter of this thesis. The third chapter's indicators of intangibles can allow researchers to explore how non-price factors (e.g., advertising, branding, and customer loyalty) can influence firm-level markups instead of market power.

Second, future research must access more detailed, fine-grained firm-level pricing and cost data. This may require collaboration with data providers, government agencies, or companies willing to share proprietary data. This includes quality and product differentiation information, which can affect pricing strategies. Quality-adjusted price indices and product-level data can help control for these factors. Using microdata and gathering longitudinal can help capture these and more heterogeneity across firms and industries and track markup changes.

Fourth, improve the robustness of the results by incorporating more formally the

economic properties of "intangibles" that can lead to misleading markup estimates—for instance, the influence of intangibleness sunkness or scalability as Haskel and Westlake (2018). Understand how firms allocate their costs and assess their implications for pricing. Examine how global supply chains and international competition affect firm-level markups; globalization and offshoring may influence firms' pricing strategies. Exploring sectoral and regional markup variations remains crucial, as they may exhibit distinct pricing dynamics.

In general, it is urgent to keep improving and proposing alternative markup measures and methodologies. Recognize that markup measurement can be context-specific, and different approaches may be appropriate for different industries or settings. That is why it is also crucial to conduct extensive robustness checks to test the sensitivity of results to various modeling and measurement choices. Precisely, a significant lesson emerging from this study is that it is necessary to promote transparency in data and methodologies. This encourages replication studies to verify findings and enhance the robustness of results, again for better policymaking in very controversial topics such as this one.

3 Twitnovation and Aboutnovation[2]

3.1 Introduction

Innovation, R&D, and investment in intangible assets such as brands, management practices, and certain types of corporate culture have been identified as important drivers of economic growth (Haskel and Westlake, 2018; Bloom and Van Reenen, 2007; Bloom et al., 2012; Guiso et al., 2015). However, comprehensive measurement of these remains elusive, in particular at the level of individual firms. Here, we explore if data derived from social media can help overcome some of the main obstacles to obtaining empirical indicators that help understand these drivers of success.

There are at least two reasons why this should be the case. First, social media could reveal critical drivers of innovation. According to innovation management research, creativity is the ability to generate new and original ideas, and as such, is the foundation of innovation, which is the process of turning new ideas into successful products, services, or processes. Social media can reveal firms' creativity or ability to generate new and original ideas through their unique communication or interaction styles with others exposed in their posts. Such interaction can also transform into successful methods to find novel ideas for their products or solutions to their problems (Amabile, 1988; Csikszentmihalyi, 1996; Vaisnore and Petraite, 2011; Chesbrough, 2003). Furthermore, social media is a crucial tool for firms to announce their new products or the adoption of new practices so that it can serve as a direct barometer of a firm's innovation level (Belso-Martinez et al., 2013).

Second, social media activity by firms is in itself an intangible investment. Especially for consumer-facing firms, social media can be an essential tool for marketing, public relations, and customer service. Companies use social media platforms to build and strengthen their brand presence online, increasing their brand equity and, therefore, their company value. Moreover, social media provides a means of direct interaction with customers, contributing to crucial firms' intangible assets such as customer engagement and loyalty (Haskel and Westlake, 2018). Furthermore, firms will provide information in social media posts indicative of other intangible investments, like their networking or engagement with peers, employees, and other stakeholders, indicating levels of intangibles in social capital (Nahapiet and Ghoshal, 1998).

The challenge will be to distill meaningful information about such intangible investments and sources of innovation from the noise of unstructured social media posts. In this paper, we start this quest by developing several different measures. First, we look at

[2]This chapter is adapted from the co-authored work "Twitnovation: Can Social Media explain Firm Performance in Latin America?" sponsored by the Inter-American Development Bank, coauthored by me, with Eugenie Dugoua and Ralf Martin.

what we have dubbed "Twitnovation": an indicator of novelty in social media posts, as it signals whether a new word or combination of strings has been mentioned[97] by a firm for the first time; we use tweets posted "*by*" a firm. Our working hypothesis is that genuine innovation is a subset of Twitnovation, as posts by a firm may reveal its creativity.

We explore this idea by relating Twitnovation and conventional indicators of firm performance, such as labor productivity. Why would twitnovation matter for productivity? Our indicators may signal firms' attunement to novel things. They might reveal not necessarily that they are creating new tech, yet they are one of the first firms talking about it or are being creative in the way they express themselves, and all of that is a good thing. For example, talking for the first time about "working from home", "smart devices", or "API" says something good about what they are as a business.

Second, in this paper, we develop several indicators based on dictionaries of relevant n-grams, i.e., word combinations associated with technology adoption or practices related to intangibles or corporate culture.

We also develop measures under the concept of "aboutnovation" to get some indication of the level of investments of a firm in the intangible of customer engagement. To build aboutnovation indicators, we primarily use tweets *"about"* the firm (e.g., the number of times a firm is mentioned in other posts by any user). However, we occasionally combine the "by" and "about" tweets.

We use firm-level data primarily from the Bureau van Dijk's (BvD) Orbis database financial dataset for Latin America and the United Kingdom. In Latin America, because it provides a unique opportunity, the Interamerican Development Bank (IDB) has helped a series of countries develop dedicated innovation surveys. These surveys contain information on many intangibles measures one might be interested in for future research where we would like to validate such indicators with traditional innovation data sources. Of course, such surveys are typically infrequent and for small samples of firms, so besides validating, these proposed measures are intended to be complimentary. When available, these surveys can provide training data for social media-based indicators. Hence, the current study offers preparation and background for such an exercise.

We also focus on the United Kingdom to contrast emerging economies with developed countries and to take advantage of the accessibility of better financial data. The UK has better coverage in ORBIS and social media use. Twitter, in particular, has greater penetration in companies' marketing strategies. This could reveal a different perspective on geographical and development level differences.

Besides differentiating indicators according to what they might be measuring (e.g., innovations, customer engagement), it is helpful to categorize them into three groups:

semantic indicators, meta-dictionaries, and context-specific dictionaries.

Our Twitnovation and Aboutnovation measures- explained in detail in the following sections - are examples of semantic measures. They only use the structure of the body of the text, e.g., when and where words occur within a corpus. This approach leads to indicators that are not context-specific and automatically change over time and across different areas or sectors. Such indicators are also independent of language and should have more external validity.

Meta dictionaries share a certain generality with semantic indicators: we class our glossary of intangible and new technology terms - detailed below - in this category. These dictionaries collect words that transcendent across areas of application and over time. Examples include words such as "research" or "reputation" to capture firms' intangible intensity. While they are more general, they would not be independent of language.

Our primary objective in using context-specific dictionaries is to find terms relating to technologies and intangibles in what firms post about. These could include words like "Bluetooth" or "artificial intelligence". Our premise is that if a firm is a product innovator, it is more likely that this firm is informing people about it (i.e., words such as "smartphone" or a "COVID-19" vaccine) and doing it years before nobody would have known. Some of these will be pretty general. However, in many cases, they will be specific to a particular sector. They are also likely highly unstable over time. Almost by definition, whatever innovative technology today will not be innovative tomorrow. But, of course, this depends on the time resolution and interval. In many applications, we will look at time windows over which a given technology is still diffusing and has not been superseded by even newer technology. Consequently, relevant n-grams will stay stable over any given data horizon.

Linking social media accounts to individual firms is not trivial. Standard business data repositories, such as the ORBIS database, have started to collect social media information. However, such data is often incomplete or wrong. Larger firms, such as multinationals, will have many social media channels. It is not necessarily clear which account is the most relevant, especially when relating information to unconsolidated balance sheet data from a specific location, such as a typical Latin American country. Indeed, we might not necessarily be identifying the most relevant social media channel for the matter at hand. There might well be a social media account of a Colombian subsidiary of a Chinese company; however, existing mappings only include the leading social media channel of the Chinese headquarters. The same occurs for the U.K. data.

Second, many firms do not engage with social media at all, or they might not engage with the platforms that allow easy access to relevant data. For instance, Twitter is much more open and accessible for researchers than, for example, Facebook or LinkedIn.

Third, even conventional balance sheet data is somewhat patchy in some countries we considered.

Fourth, although Twitter is relatively open in providing researchers access to their data back-end, data transfer restrictions mean data collection has to take a long time.

Some of these issues might be sorted with further research or simply by time. It is plausible to assume that an ever more significant fraction of firms will have some social media presence, and access to the underlying data for research should improve. For now, we end up with a relatively small sample of firms for Latin America, which is one of the main weaknesses of this chapter. However, we have substantially more for the U.K., where we could undertake a joint analysis of performance variables and social media-derived indicators. Also, the exact number varies depending on the particular performance variable concerned.

That notwithstanding, we can establish significant relations between a measure of labor productivity and Twitnovation. We also find that word-based intangible and technology indicators correlate meaningfully with labor productivity within country $\times$ sector cells. We also find some relationships with variables related to customer engagement. However, a noteworthy result here is that a simple count of mentions of a firm by the community at large - a measure we have dubbed the pinnacle of our *Aboutnovation* measures, provides for a surprisingly good indicator of firm performance and that such an indicator is more important than any measure of how a firm responds to such mentions with the sample at the moment. This result is based on our samples, the small for Latin America and the larger sample for the U.K. Our goal is to overcome data access restrictions of the Twitter API in future research and build up a more comprehensive sample to study this and perform an out-of-sample analysis. However, we are aware this will take more time.

We also find that they are jointly relevant; i.e., each broad type of indicator captures different aspects of the variation of labor productivity. Overall, we can explain about 20% of the within-country $\times$ sector cell variation in labor productivity. We do not find much evidence for a systematic relationship between those measures and innovative activity measured via patents. Thus, this could be a consequence of the fact that firms in Latin America are not as active in this sphere as elsewhere.

In summary, for Latin America, we found that from the broad group of measures, the technology indicators explain the bulk of variation (about 15 percent). In contrast, intangible indicators describe three and Twitnovation on its own about 2 percent of the within-cell variation. Even though the latter number is small, it is remarkable that such a simple indicator derived from social media contains such a strong signal. Also, while technology n-grams explain a comparatively more significant fraction, we expect such measures to perform less well in an out-of-sample exercise. Indeed, we would expect

Twitnovation to perform just as well as it is an indicator that doesn't require any training set to be computed. Explore this will be the subject of future research.

For the UK, the analysis primarily focused on exploring the relationship between labor productivity and tweets *about* the firm. We strikingly found the same magnitude and level of significance that we found for Latin America. For each 1% increase in the number of mentions of the average firm in Latin America, labor productivity increases by 0.0807%. In the UK, labor productivity increases by 0.0837%.

Therefore, this chapter's main contribution is to provide alternative measures for complex and hard-to-measure concepts crucial for determining firms' success, such as innovation and intangible investments. Although there have been substantial efforts in prior literature to capture them, such as through the analysis of innovation, management, and intangible assets surveys, the lack of frequency, the costs, and the rapidly changing technologies, among multiple other difficulties, make our proposed measures an attractive alternative, or at least, a valid complementary tool. Social media indicators could provide faster, cheaper, and more reliable indicators for firms and policymakers.

The chapter has its notable limitations, however. Besides the small sample for Latin America, it does not directly address the "selection bias" issue, which can be problematic. Our primary analyses, at least for Twitnovation, only include firms with a Twitter account that effectively tweet. This may lead to potentially skewed or incomplete results. We do show that there is an apparent difference between the firms with Twitter accounts and those who do not. Firms with Twitter accounts tend to be larger regarding revenue and employees. This limits our generalizability to the broader population of firms, reducing our findings' external validity and making it difficult to conclude the impact of tweeting on labor productivity for all firms.

Also, we might incur an omitted variable bias. Firms that choose to tweet may have specific characteristics or strategies that differ from non-tweeting firms. We may omit crucial variables that could influence labor productivity by not considering the non-tweeting firms. Finally, even if we find a correlation between Twitnovation and higher productivity among firms that tweet, it does not necessarily imply causation. There may be confounding factors or selection issues at play.

To address these issues, in the future analysis, we suggest considering a more comprehensive analysis that accounts for the firms that do not tweet. One way to do this is by using matching techniques, instrumental variables, or propensity score matching to create a more balanced sample that includes both tweeting and non-tweeting firms.

3.2 Conceptual Framework

The source of innovation and economic growth of modern firms goes beyond tangible assets like machines and equipment (Haskel and Westlake, 2018; Bloom and van Reenen, 2007). Intangibles are knowledge-based assets that are difficult to touch or measure but have recently played a significant role (Haskel, 2020). Businesses invested mainly in physical assets like railroads and vehicles from the nineteenth century until about 25 years ago. Yet, in the last 25, investment in intangible assets has soared. Recent McKinsey Global Institute (MGI) research found that, by 2019, intangibles accounted for 40 percent of all investment in the United States and ten European economies, up 29 percent from 1995. The intangible investment appears to have surged again in 2020 as digitalization accelerated in response to the COVID-19 pandemic (Hazan et al., 2021).

Interestingly, we do not refer only to things like R&D or patent investments. Although these intangibles are a significant source of growth, they are typically highly skewed as only a few firms tend to account for most investments in such assets. Thus, they do not fully explain the widening gap between firms' equity market and accounting valuations based on physical assets.

On the contrary, evidence has shown that innovative and high-growth companies invest in many intangibles beyond the scope of R&D. For instance, software tools, attributed design, and strategies to improve branding, business practices, or managing after-sales services (Jonke, 2012).

Acquiring intangibles and knowledge assets can increase innovation capacity, resulting in more profitability and productivity. For instance, R&D acts as direct innovation input. Other intangibles, such as training or business models, provide firms with other knowledge assets that, although not direct innovation inputs, facilitate the different phases throughout which the innovation can eventually occur (Montresor and Vezzani, 2014).

For instance, early management and marketing literature identified brands as strategic capital (See Farquhar (1989); Aaker (1991)). Influential work on accounting highlighted brand names as critical drivers of financial outcomes in many innovative companies (Lev, 2001). Furthermore, firms have recognized for a long time in their business reports, customer relations, and organizational capital as valuable assets (Bontis and Fitz-enz, 2002).

Furthermore, under the theory of Corrado et al. (2009); Haskel and Westlake (2018)'s work on intangibles, assets like branding, design, marketing strategies, and organizational know-how are part of the *Economic competencies*, the core intangible assets that provide a firm with a competitive advantage, as they set a firm apart and contribute to its success.

However, the non-physical nature of intangibles (direct or indirect inputs for innovation), and even more so, these more modern intangibles, or those beyond R&D or patents, are very complex to measure. Consequently, the measurement of innovation drivers is even more difficult. As a result, measuring intangible investments' magnitude has remained elusive, and even more so, their relationship with innovation and performance.

The reasons for such difficulties vary. For example, the rapid pace of newly emerging tech and its augmented complexity makes it hard to keep up with new forms of intangibles associated with such new developments. Also, intangibles are not always seen as an asset because of the difficulties of valuing them. It is easy to assign a value to intangibles based on the price at which they can be purchased on the market (e.g., a Microsoft software program). Still, intangibles generated internally (e.g., employee-created software) are usually not included in firms' balance sheets. Instead, the resources used to make that internal possible are considered costs with amortization or as part of other inputs or expenses. Subsequently, liquidating these intangibles is cumbersome relative to a physical asset. Moreover, public market data on the previous valuations or purchases of intangibles is unavailable, making it challenging to find comparable values and, therefore, make valuations by proxy.

At the macro level, there has been substantially more progress, at least partially, by applying methodologies to capitalize costs and consider them investments rather than amortizable costs in developed countries (see Corrado et al. (2009)). Incorporating these methods into national accounts has moved slowly, and it remains incredibly obscure in developing countries, where national accounting systems are outdated, poorly standardized, and lack uniformity within and across industries, preventing the creation of fully comparable data between countries.

However, at the firm level, measurements are more deficient. Incorporating intangibles systematically into company financial accounts has not materialized significantly because they are conceptually inconsistent (See CPA Ontario 2021 cited in Corrado et al. (2021)). Furthermore, the information generated through business surveys like innovation surveys is often incomplete or of poor quality. Questionnaires omit questions about intangibles or are exclusively focused on innovation, ignoring its potential drivers. Also, when questions on intangibles are asked, surveyors do not establish standards for framing them, which leads to potential biases and a lack of comparability. Intangibles can be a fuzzy concept, and surveyors do not explain clearly to businesses what intangible assets mean or how intangibles can be interpreted in the context of their own companies. In addition, innovation surveys, in general, are not applied with enough frequency (maximum conducted yearly), so innovation measurements cannot reflect the actual pace of technological progress (Martin and Baybutt, 2022).

Accordingly, measuring innovation alone has been a challenge in itself. Although there is more expertise on the matter via standard innovation surveys in developed countries, the task is even more constrained by the lack of comparable and reliable data in developing countries. As intangibles, operationalizing the definition of innovation is challenging. Survey respondents are called to make subjective judgments because of this. Ultimately, what constitutes innovation may be considered differently across different contexts or geographies (Cirera and Muzi, 2020).

This study attempts to overcome some of these problems by proposing alternative measurements of firms' innovation and intangible capital investments that can predict a firm's performance. We do so by exploiting big data from Twitter, or as currently called **X**, a social networking and microblogging service on which users post and interact with messages known as "tweets". Tweets were initially restricted to 140 characters, but on November 7, 2017, this limit was doubled for non-CJK languages. Registered users can post, like, and retweet tweets, but unregistered users can only read them. Users access Twitter through the website's interface, standalone apps for mobile devices, desktop applications, or by SMS.

Specifically, this study aims to contribute to the related literature by answering these research questions: (i) Does social media data provide valid measures of investment in intangibles and innovation of firms? (ii) What is the contribution of measured intangibles and innovation to the economic performance of firms in Latin America and the U.K.? In particular, we propose the measure of "Twitnovation" to measure the novelty in tweets as a proxy for innovation, and "aboutnovation", which measures one of the crucial intangibles for enhancing firms' innovation capacity and performance, such as customer engagement.

3.3 Data

3.3.1 Latin America

We use data on Latin American firms from Orbis, a Bureau van Dijk (BvD) database containing financial variables and social media account links. We also use data from innovation surveys rolled out in Peru, Argentina, Colombia, and Paraguay, so we primarily focus on firms in these countries, covering the period of January 2018 to June 2022.

We collect Twitter handles associated with firms in three ways. First, the Orbis database contains a social media field that sometimes provides the firm's Twitter handle. Each Bvd ID is associated with only one Twitter handle in our sample. However, the same Twitter handle can be associated with multiple BvD IDs because Orbis includes many

local branches of large firms.

Second, in Google, we search for the name of a firm followed by "Twitter" and the country's name (e.g., Colombia). When firms have a Twitter account, Google search typically successfully shows the correct Twitter account as the top result. Third, the country-specific innovation surveys sometimes provide a website URL for the firms. We then scrape social media information from the firm's website.

We use the Twitter Timeline API to collect all the tweets posted on the firms' Twitter accounts. The API allows us to retrieve all tweets since creating the firm's account. In total, we obtained 22,769,478 tweets from 1724 handles. The average handle has about 13,200 tweets, 7,300 of which are labeled as being in Spanish in the Twitter database.[98]

Tweet-talkative firms vary greatly: some handles only have one tweet, while the most active account has 954,534 tweets (in Spanish). The most active handles are: claroteayuda, perfilcom, eluniversalctg, claroperu, americatv_peru.

Table 16 summarizes the number of firms and Twitter handles collected for each country. The first row shows the total number of firms in the Orbis database (unique BvD ID). Most do not have Twitter accounts. Even when Orbis provides a Twitter handle, the handle may not be found on the Twitter API (which we call a "bad" handle), or the handle is located, but there are no tweets in the account.

Table 17 illustrates the differences between our sample and the rest of Orbis firms. Firms with tweets are overall much larger than the average Orbis firm, both in terms of the number of employees, turnover, and total assets. They also generally have a higher labor productivity, proxied by the turnover ratio on the number of employees, except in Peru.

TABLE 16
Handles - Sample Composition

	Columbia	Peru	Argentina	Paraguay
Nbr of Orbis Firms	1283605	98239	800858	54868
Nbr of Orbis Firms with Handle	2011	253	361	35
Nbr of Orbis Firms with Tweets	1568	187	291	26
Nbr of Handles	1462	214	315	32
Nbr of Handles with Tweets	1143	159	256	24
Nbr of Bad Handles	235	32	43	6
Nbr of Additional Handles	71	132	0	0

Note: Additional handles are collected by scraping the results of a Google search or an innovation survey.

We also use Orbis to obtain sector information (2-digit NAICS code). Table 18 shows that the top ten sectors include, among others, Wholesale Trade, Manufacturing, and Technical Services.

TABLE 17
Comparing Mean Financials of Handle Sample with All Orbis Firms

	Columbia		Peru		Argentina		Paraguay	
	All Orbis	With Tweets	All Orbis	With Tweets	All Orbis	With Tweets	All Orbis	With Tweets
Nbr Employees	5	454	39	1,247	18	248	7	94
Turnover	809,351	107,944,836	24,417,275	191,918,440	185,045	258,880,651	60,580	20,866,365
Total Assets	1,053,685	289,140,590	550,948,729	732,657,758	278,966,299	1,793,737,179	58,898,298	161,550,592
Turnover / Employees	152,694	237,841	623,451	153,877	10,495	1,044,776	8,906	222,629

Note: Values are the corresponding sample's average number of employees, turnover, or total assets.

TABLE 18
Top 10 Sectors (2-digit NAICS) (in %)

	All	Columbia	Peru	Argentina	Paraguay
Wholesale Trade	20	22	25	19	15
Manufacturing	19	21	20	22	30
Technical Services	12	12	13	22	NA
Finance/Insurance	7	7	9	11	20
Retail Trade	7	8	4	7	10
Health Care	6	8	2	4	5
Construction	6	8	5	2	NA
Support Services	5	6	5	6	5
Transportation/Warehousing	5	5	7	4	15
Real Estate	4	3	9	3	NA

Note: The table shows the percentage of firms in selected sectors. Sector information was missing for 35% of the firms in our sample. The percentages are calculated relative to the sub-sample of firms with sector information.

3.3.2 United Kingdom

We also use data on British firms from Orbis, the Bureau van Dijk (BvD) database. We collected 145,652 Twitter handles connected to 670,918 different British BvD IDs. From this sample, we focused only on Twitter handles associated with firms (denoted by BvD ids) with non-missing industry sectors at the 2-digit level (NACE Rev 2.) and labor productivity from 2017 to 2021. Only 44,683 thousand out of the 145,652 fulfill this requirement. We also eliminate firms with negative values of the number of employees or assets. See Table 19.

TABLE 19
Summary Financials UK:
All UK Orbis Firms with a Handle and Non-Missing Turnover or Employees

	Sum	Mean	SD	Min	Max	N
Nbr Employees	13,372,922	299.3	4,620.0	1.0	561,766.3	44,683
Turnover	3,686,613	82.5	1,717.7	-0.2	272,797.3	44,683
Total Assets	17,891,925	481.8	18,640.6	0.0	2,326,009.8	37,136
Turnover / Employees	38,171	0.9	104.0	-0.2	21,953.8	44,683

Note: Values are for the full sample available in ORBIS for the UK that complies with the non-missing criteria for labor productivity. It presents summary statistics for the number of employees, turnover, or total assets. Values of total assets, turnover and turnover per employee are expressed in millions USD.

Each ORBIS ID is uniquely associated with a Twitter handle using the same method as before. We also focused on using Twitter Academic Research's Timeline API to collect tweets posted *about* the firms, so we searched for all tweets that contained *@[company's Twitter Handle]* in the body of the tweet between 2017 and 2021. The API also allows us

to retrieve all tweets posted by the firms since creating the firm's account, which is the primary data source of the Twitnovation measures. However, as explained towards the end of the results section 3.4, the aboutnovation measure tends to have more explanatory power, so the primary use of the UK data is for the aboutnovation analysis. In the appendix, we show some estimations of Twitnovation for the UK for guidance. See appendix H.

To avoid exacerbating the Twitter API's monthly quota quickly and overcome the limitations faced in collecting the Latin-American data, we randomly selected 10,000 handles out of the 44,683 thousand with non-missing labor productivity for the same sample. In contrast to the data for Latin America, we cut off collecting mentions once the number of mentions (tweets about) per firm exceeded 200,000 for the sample period. Tweets retrieval using the API is designed to start collecting from more recent to older tweets, so firms that reach the 200,000 cut-offs have a higher frequency of mentions in the most recent years, and the excluded tweets may be more from 2017 and 2018. We randomized the order of querying of the 10,000 handles. From this exercise until the moment of submitting this thesis, we queried a total of 7,148 handles, of which 6,289 Twitter handles have been found and for which we obtained 33,985,864 tweets *about*. The average handle has about 15,300 tweets; 30,648,026 tweets about out of the total 33,985,864 are labeled as being written in English. Table 20 shows the differences in financials between the Handles found and not found on Twitter, as mentioned.

TABLE 20
Mean Financials of Handle Sample with All Orbis Firms: UK

	Mean	N
0		
Nbr Employees	151.3	1,542
Turnover	27.7	1,542
Total Assets	191.4	1,216
Turnover / Employees	0.3	1,542
1		
Nbr Employees	400.5	6,206
Turnover	206.5	6,206
Total Assets	1,461.0	5,622
Turnover / Employees	4.5	6,206
Total		
Nbr Employees	350.9	7,748
Turnover	170.9	7,748
Total Assets	1,235.2	6,838
Turnover / Employees	3.6	7,748

Note: Values are for the queried sample of UK ORBIS firms out of the firms in Table 19. It presents summary statistics for the number of employees, turnover, or total assets by group, where 0 represents Handles not found (or not mentioned firms) and 1 represents Handles found (or mentioned firms). Values of total assets, turnover, and turnover per employee are expressed in millions USD.

We highlight that when studying the relationships between the different built social media indicators and between such indicators and productivity, we limit the analysis to 2018 to 2021 so the study periods are comparable. Otherwise, if analyses are isolated, the sample

period can vary. The sample being more recent, covering 2018 to mid-2022, is due to the use of Twitter in the region. The firm's penetration of this tool usage has been slower in emerging countries. Therefore, the data is more limited for 2017. Furthermore, we will explain some methodological implications for the construction of the Twox indicator that we will explain in section 3.4.1.1. In the case of the UK, we take 2017 and limit it to 2021 to have full years of information.

3.4 Social Media based indicators

3.4.1 Twitnovation

The idea of Twitnovation is to measure novelty in tweets. We suggest that firms' innovation could be closely related or a subset of Twitnovation. For instance, a firm introducing a new product or practice could be expected to discuss this in social media posts. It could also reflect a corporate culture close to current affairs and developments. For instance, firms offering comments on new societal and technological trends (and the resulting new terms) could also be expected to be more agile in innovation in their core business or adopt new technologies and approaches.

We measure tweet innovation by tracking when new words appear and how often they are mentioned. We argue that a word's novelty decreases with the frequency with which it has been mentioned in the past and with the time since it was first said. Hence, our benchmark for what is new will be derived based on the existing body of text and the flow of time. Consequently, the indicator automatically adjusts as time moves on. It can be defined on any body of text without the need to define dictionaries from any other sources of knowledge.

We define the Twitnovation Index, or Twox, of word w in tweet i published at time t as follows:

$$Twox_{wt} = \delta^{Q_{w,t}},$$

Where $Q_{w,t}$ is the number of times word w has been mentioned before tweet i was posted at time t. δ is a depreciation factor between 0 and 1: novelty wears off faster when δ is lower. In what follows, we set δ to values between 0.8 and 0.999. For instance, novelty decreases by 50% after about 700 mentions in the latter case.

At any particular time t, $Q_{w,t}$ will be more significant if we consider a larger sample of tweets. In particular, at the beginning of Twitter, e.g., in 2012, few firms had ever posted tweets, and therefore, most words' $Q_{w,t}$ were small, leading to high values of Twox. After enough tweets have been posted, however, we should expect that the most common words would have appeared, and variations in Twox can more truthfully capture the appearance of new ideas on Twitter.

To calculate Twox, we use all the tweets posted by the corporate handles we collected. To help us grow the vocabulary size in our corpus, we also use other Spanish tweets posted by users that tagged in their tweets the corporate handles we study. Coming up with a novel sequence of characters may be easy, but it is more difficult to find a novel sequence that is subsequently widely adopted. We define a measure of Twitnovation *success* as follows:

$$TwoxSuccess_{wt} = C_w^{Total} \times Twox_{wt}$$

where C_w^{Total} is the total number of times word w appeared *subsequently* to its first appearance in the corpus. Importantly, this measure of Twitnovation success is also useful to filter out new words that do not get adopted by other users (e.g., perhaps just a typo). There are two potential issues with $TwoxSuccess_{wt}$. Firstly, right-censoring: we can only know how often a word is repeated until our sample ends. Hence, by construction, words arriving later for the first time will be less "Twoxsuccessful". Secondly, Twoxssuccess doesn't lend itself as a leading indicator as we have to wait - theoretically - until the end of time or - practically - until the end of the sample to compute it. Hence, along similar lines, we also compute

$$TwoxSuccess_{wt}^{\omega} = C_w^{\omega} \times Twox_{wt} \tag{67}$$

where C_w^{ω} represents the number of appearances of a given word within a fixed time window of ω after (and not including) its first emergence. Note that if we define ω at length shorter than the width of a time step in the analysis undertaken, then we can use $TwoxSuccess_{wt}^{\omega}$ as a leading indicator, e.g., if we are conducting analysis at the weekly level but set ω to one day.

Twox and TwoxSuccess are defined at the level of words and time points. For most analyses, we need to aggregate this to the level of firms j and periods τ. We do this in the following way: we define the Twitnovation count of firm j in a week t as the sum of the Twitnovation Index of all the words posted by firm j in week τ. Formally:

$$TwoxCount_{j\tau} = \sum_{w \in W_{j\tau}} Twox_{w,t \in W_{j\tau}},$$

where $W_{j\tau}$ is the set of words posted by firm j in period τ. Intuitively, this measure can be interpreted as a word count where the words are novelty-weighted. Similarly, we define the Twitnovation Success of firm j in week t as the sum of the Twitnovation Success of all the words posted by firm j in week τ.

Finally, we calculate the Twitnovation Share of firm j in week τ as the average

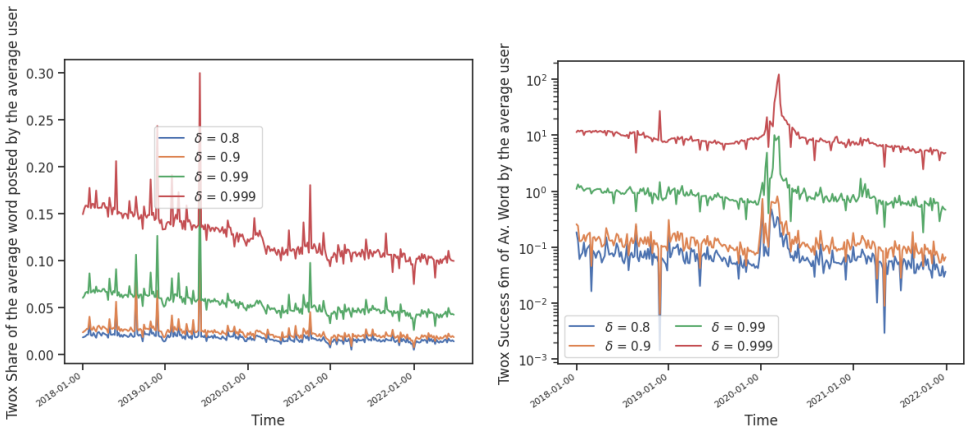
Twitnovation Index of the words posted firm j in week τ:

$$TwoxShare_{j\tau} = \frac{\sum_{w \in W_{j\tau}} Twox_w}{||W_{j\tau}||},$$

where $||W_{jt}||$ the number of words posted by firm j in week t. We can interpret this measure as the share of words with a Twitnovation Index of one.

3.4.1.1 Twitnovation Trends Here we look at average twitnovativeness over time and many other descriptive statistics. First, Figure 50b illustrates how the Twox Share and Twox Success vary over time. We focus on the time window from January 2018 to June 2022. We could, in theory, go further back in time; however, our twox measure requires an extensive enough dictionary of references to infer what is or not novel. Also, similar trends are observed over time for the UK, as shown in Figure H.18 in Appendix H.

FIGURE 50
Trend Over Time for the Average Firm:
Latin America



(a) Twox Share (b) Twox Success (6-month window)

We note that, unsurprisingly, the Twox Share decreases over time. The Twox Share is also tinier for smaller values of δ. We note several peaks throughout the period. In particular, the most prominent Twox Success peak corresponds to the beginning of the COVID-19 pandemic in 2020. Figure 51b shows that the most successful twitnovative word in 2020 was "covid". In 2022, many of the most successful twitnovative words refer to entertainment events such as new movie releases or concerts (e.g., "batmanconsigue", "dance_in_cinemascompra"). Table G.15 in Appendix G provides several examples of twitnovative tweets in different sectors in 2022 for Latin America. The type of twitnovative tweets differs by sector, so it is important to control for industry sector effects in our regression analysis later on. For instance, the Information sector shows

a prominent telecommunications firm in Peru, *Claroperu*, announcing new mobile plan promos. In contrast, *Empresaspolar*, a prominent manufacturing firm announces the launch of a new product called "toddy pudding". Figure H.19 in Appendix H shows what Twitnovate words look like for the UK.

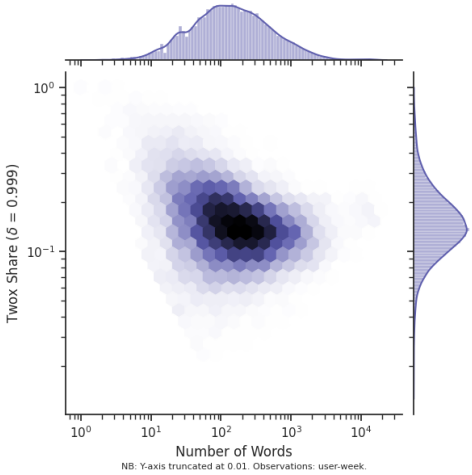
FIGURE 51
Examples of Twitnovative Words: Latin America

(a) 2020 (b) 2022

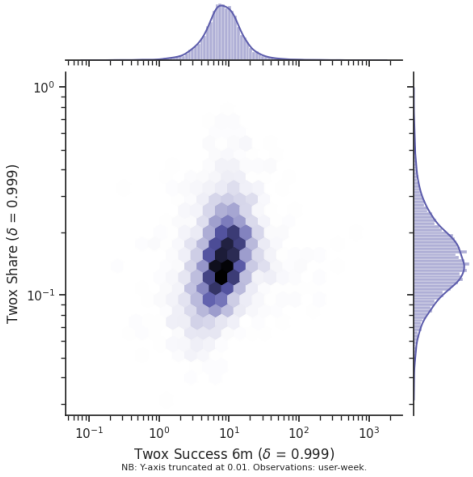
Note: The wordclouds show words with a Twox value of 1 in the corresponding year, meaning that the words first appeared that year. The size of the word on the cloud is proportional to the word's Twox Success (within 6 months). δ was set to 0.999.

Figure 52 illustrates the variation when the data is aggregated at the firm-week level. In Figure 52a, we note that when firms post more words in a given week, they also tend to have a lower value of Twox Share. In other words, it does not seem to be the case that by tweeting a lot, one can be more likely to hit upon a novelty. On the other hand, figure 52b shows that weeks with higher Twox share also tend to have a higher value of Twox Success.

FIGURE 52
Variation at the Firm-Week Level: Latin America



(a) Twox Share and Number of Words



(b) Twox Share and Twox Success

Note: δ was set to 0.999. For Twox Success, we used a 6-month window. The bell-shaped figures represent the frequency distribution of each X and Y variable; the vertical axis distribution corresponds to axis Y, and the horizontal one corresponds to axis X.

3.4.2 N-Gram Based Measures of Technologies and Intangibles

We use a list of keywords to identify tweets related to technology and intangible assets. Specifically, we construct variables at the firm-week level that capture the number of times particular keywords are mentioned. Appendix F shows the full list of terms used. Webuilt the list based. on prior work by Corrado et al. (2005); Haskel and Westlake (2018); Office for National Statistics et al. (2016); Bloom et al. (2020). Given our context, we include both the English and Spanish versions of each term. Technologies include artificial intelligence, 3D printing, Bluetooth, cloud, driverless, electric cars, fracking, LEDs, machine vision, monoclonal antibody, optical fiber, solar, and virtual reality. For intangibles, we use keywords related to training, software, reputation, R&D, design, and organizational capital.

Table 21 below explores correlations between the Twox measures and the technology and intangibles bigrams. We note that technologies related to artificial intelligence, the cloud, or touch screens correlate with higher Twox Share and Twox Success values. Intangibles related to design and organizational characteristics also correlate with Twox Share, although the coefficients for Twox Success are not significant.

TABLE 21
Correlations between Twox and Tech Bigrams
- Latin America -

	(1) Twox Share	(2) Twox Success
Nbr of Words (log)	0.155***	0.316***
	(0.036)	(0.033)
rfid	0.500***	0.162
	(0.182)	(0.173)
ai_long	-0.070*	-0.142***
	(0.039)	(0.039)
ai_short	0.137***	0.150***
	(0.047)	(0.047)
threedprinting	0.047	0.041
	(0.066)	(0.090)
solar	-0.038	-0.150
	(0.110)	(0.110)
oled	-0.039	-0.095*
	(0.060)	(0.052)
virtual_reality	0.034	-0.039
	(0.050)	(0.049)
driverless	0.007	-0.180
	(0.127)	(0.174)
smartphone	-0.204*	-0.178*
	(0.111)	(0.106)
disk_drive	0.013	0.138
	(0.116)	(0.127)
cloud	0.185**	0.283***
	(0.084)	(0.073)
smartcard	0.009	-0.123
	(0.139)	(0.350)
touch_screen	0.292***	0.224**
	(0.095)	(0.093)
machine_vision	0.000	0.000
	(0.000)	(0.000)
training	-0.052	-0.067
	(0.044)	(0.044)
software	0.004	-0.071
	(0.059)	(0.050)
research	-0.113	-0.131*
	(0.082)	(0.073)
design	0.560***	0.148
	(0.147)	(0.138)
reputation	0.022	0.036
	(0.055)	(0.055)
organizational	0.124**	-0.078
	(0.060)	(0.081)
Constant	-3.257***	0.103
	(0.192)	(0.173)
Tech Bigrams	Binary	Binary
Week FEs	X	X
Sector FEs	X	X
Observations	140,670	140,663
Pseudo-R2	0.030	0.101

Linear regression. Standard errors in parentheses
clustered at the user and week level.
Outcome variable: log of Twox Share or Twox Success with delta = 0.999

Note: Each cell represents the correlation of each firm-week value of the respective Twox variable and the appearance in that firm-week of the respective Tech-bigram one or more times (e.g., *rifd* is a dummy variable = 1 if the technology was mentioned by the firm in that particular week 1 or more times). ai_long represents a richer dictionary of terms related to AI than ai_short.

3.4.3 Measuring Customer Service

So far, we worked with social media data *by* firms. However, in addition, we can look at data *about* firms. There could be several reasons why somebody mentions a firm in a Tweet, e.g., this could be the CEO or some other official representative of the company tweeting. It could be non-official employees of the firm. It could be some other kind of stakeholder, e.g., an environmental group concerned or impressed by the firm's actions. It could also be actual or potential customers of the firm issuing praise or complaints.

There are several approaches one could pursue with this setup. For example, one could construct various indicators of closeness or distance that might indicate different issues. Or, firms might try to greenwash their image in social media posts, whereas civic society might draw an image of a firm that violates the environment. For now, we focus on the existence of a response to such posts by the firm and their speediness in that.

To do so, we use the Twitter API to collect tweets that mention the handle of firms in our sample, i.e., any tweet that contains the symbol "@" followed by a firm handle. For example, `"I love @subway"`. Some firms receive a very high number of mentions, and the Twitter API imposes a monthly quota on the number of collected tweets. Hence, to collect "about" tweets, we focus on a narrower time window spanning 2017 to 2021.

3.4.3.1 Rate of response To develop proxies of customer engagement, we calculate how often and when firms respond to the tweets that mention them. Consider a tweet p that mentions a firm j (once or more times) at a time t. Firm j can choose to address or not address a mention and do so with varying speed, i.e., they might respond after some time γ_p to a mention posted at time t_p. The firm can address that tweet by replying, retweeting, or quotation. Accordingly, we define the indicator function I for each tweet (mention) p of firm j as follows:

$$I_{jph} = \begin{cases} 0, & \text{if firm}j \text{ does not address tweet } p \text{ posted at time } t_p \text{ by time } t_p + h \\ 1, & \text{if firm}j \text{ addresses tweet } p \text{ via a reply, a retweet, or a quote by time } h + t_p \end{cases} \tag{68}$$

The total number of distinct tweets mentioning firm j in some period t is defined as P_{jt}. We denote $AddressedMentions_{jht}$ the number of mentions made during a given period t, which is addressed within a time window h. Formally:

$$AddressedMentions_{jht} = \sum_{p=1}^{P_{jt}} I_{jph} \tag{69}$$

Our main indicator of customer responsiveness will be the response rate, a measure

that captures the percentage of addressed mentions. Specifically, we define the Rate of Response, RR_{jht}, as the percentage of mentions received in period t that firm j addresses within a delay h. We compute RR_{jht} for $h \in \{hour, week, month, year\}$. Formally:

$$RR_{jht} = \frac{AddressedMentions_{jht} \times 100}{P_{jt}} \tag{70}$$

3.4.3.2 Time to reply We compute a second indicator of customer responsiveness, similar to the response rate defined above. Still, each response is weighted by the inverse of the firm's time to respond. The idea is that the less time the firm takes to respond, the more efficient we can consider it.

To clarify, consider a firm j that takes γ_p time (e.g., in hours) to address (by reply, retweet, or quote) a post p that mentions it. We define a weighted version of the rate of response as follows:

$$I_{jp} = \begin{cases} 0, & \text{if firm} j \text{ does not address tweet } p \text{ posted at time } t_p \\ \frac{1}{\gamma_p}, & \text{if firm} j \text{ addresses tweet } p \text{ via a reply, a retweet or a quote} \end{cases} \tag{71}$$

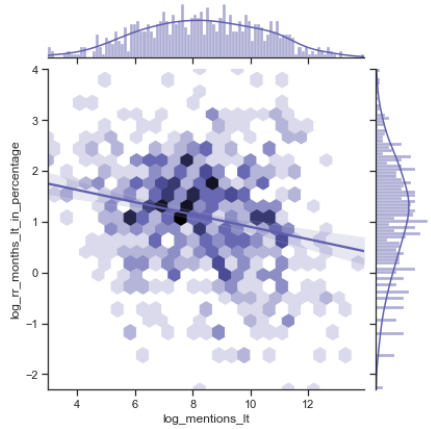
$$RR_{jt}^{\text{weighted}} = \frac{\sum_{p=1}^{P_{jt}} I_{jp}}{P_{jt}} \tag{72}$$

Note that, unlike RR_{jht}, the weighted measure allows differentiating between posts that were never addressed (at least over the sample period) from those that were addressed late.

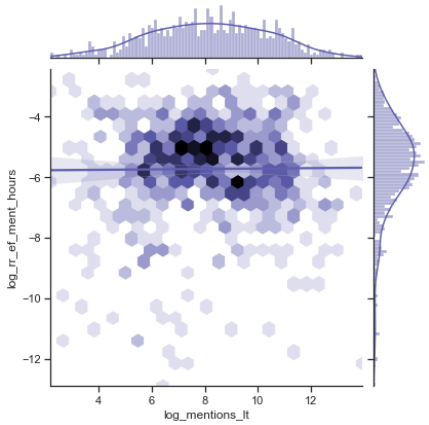
3.4.3.3 Customer service descriptive statistics For the Latin American sample, we collected over 29 million tweets *about* 1205 firms (and posted by almost four different millions of Twitter accounts). Only a small percentage of those mentions are ever addressed (2.6%). See Table J.18 in Appendix.

We observe a negative and significant correlation between the firm-level number of total mentions P_j and the response rate, potentially meaning that the more mentions a firm has, the greater the difficulties in coping with the required demand in responses (See Figure 53a). Yet, when correlating P_j with the weighted response rate, there is no significant relationship (see Figure 53b). Figure J.20 in the Appendix also shows a negative correlation between the time taken to reply (in minutes) and P_j.

FIGURE 53
Firm-Level Response Rate: Latin America



(a) Within One Month Response Rate



(b) Weighted Response Rate

Note: The x-axis shows the total number of mentions (in log) received by firms between 2017 and 2021. Observations are at the firm (user) level. In panel b, the response time calculating the weighted rate is in hours.

In table 22, we regress the twox success measure against the customer engagement measures at the firm level controlled by industry and the number of words posted by the firms. The number of mentions and the response rates are significant for twox success.

We observe a negative correlation between Twox and the response rate. This result might indicate that the more responsive firms get to be with customers, they do not learn much from that interaction that could be reflected in finding new solutions to problems or product ideas. However, we would need to check more deeply what types of interactions occur.

Also, we find a positive correlation between the number of mentions and twox. We could interpret this as that widely commented-about firms could take valuable information from those mentions for novel ideas. However, responding to them might not be as effective.

Notice that when we add n-grams of intangible assets, some turn to be significant to explain twox success. For example, "design" investments are negatively correlated with Twox.

The following is purely hypothetical, but one could think of many reasons why this could happen. For example, a firm overprioritizing design can lead to complacency, where the firm relies solely on design for competitive advantage, neglecting other innovative areas. Also, a myopic focus on design narrows the firm's scope, potentially overlooking crucial aspects of innovation like technology and business model improvements. Moreover, excessive publicizing of design makes the firm vulnerable to imitation by competitors, potentially eroding market differentiation. Finally, overemphasis on design may overshadow the importance of fostering a holistic culture of innovation, which includes creative thinking and risk-taking in various aspects of the business.

Table G.16 in the Appendix shows similar specifications for the regression but uses the Twox share as the dependent variable, but coefficients are not significant.

TABLE 22
Correlation between Twox Success, Customer Engagement and Bigrams - Latin America -

	(1) Twox	(2) Twox	(3) Twox	(4) Twox	(5) Twox	(6) Twox	(7) Twox
ln(1 + Number of words)	0.000	0.000	-0.000	-0.004	-0.004	-0.003	-0.004
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)
ln(1 + Number of mentions)	0.001***	0.001***	0.001***	0.003**	0.003**		
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)		
ln(1 + Rate of Response perc.)	-0.002***	-0.003***	-0.003**	-0.006*	-0.006*		
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)		
ln(1 + Rate of Response (weighted))		0.041	0.051	0.111	0.126		
		(0.054)	(0.055)	(0.206)	(0.208)		
training			0.000	-0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
software			0.000**	0.001	0.001	0.000	0.000
			(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
research			0.000	0.000	0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
design			-0.003**	-0.017	-0.017	-0.031**	-0.031**
			(0.001)	(0.013)	(0.013)	(0.014)	(0.013)
reputation			-0.000	-0.000	-0.000	0.001	0.001
			(0.000)	(0.002)	(0.002)	(0.002)	(0.001)
organizational			0.001**	0.002*	0.002	0.000	0.001
			(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
ln(Number of Employees)				0.001	0.002	0.002*	0.001
				(0.001)	(0.002)	(0.001)	(0.002)
ln(Number of Employees) sq.					-0.000		0.000
					(0.000)		(0.000)
Bigrams	None	None	All	All	All	All	All
Full Sample	Yes	Yes	Yes	Yes	Yes	Mentioned	Mentioned
Country FEs X Sectors FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	824	824	824	259	259	265	265
R2 within	0.057	0.057	0.070	0.193	0.194	0.067	0.067
F-test TechBigrams	No	No	3.873	23.652	12.575	No	No

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: Mean Twox .9 Success 1 hour

Note: The table presents regressions at the firm level, including industry-fixed effects. The number of employees is at the yearly level per firm.
Rate of Response Perc.: RR_{jht} defined in tweets addressed within a month.
Rate of response (weighted): $RR_{jt}^{\text{weighted}}$ where time γ is defined in hours to reply.
Each cell referring to an intangible (e.g., training) represents the correlation of each firm-week value of the respective Twox variable (Twox Success) and the appearance in that firm's history (between 2017 and 2021) of the respective intangible-bigram one or more times (e.g., *training* is a dummy variable = 1 if the firm mentioned any word from the dictionary associated to training in the year to which that week belongs to.).

3.4.4 The relation between social media indicators and firm performance

In this section, we explore how our measures of Twitnovation, customer engagement, and bigrams correlate with firm performance. Here, our primary outcome variable is labor productivity, proxied by the turnover ratio over the number of employees.[99]

Table 23 reports regression results of the following main specification:

$$LaborProd_j = \beta TW_j + \alpha_{sc} + \epsilon_j \tag{73}$$

where TW_j is a vector of social media derived indicators for firm j and α_{sc} are (2-digit) sector $\times$ country fixed effects.

In Column 1, we consider various Twitnovation measures along with a control for the number of words used by the firm. We note that, as a whole, these are significantly correlated with labor productivity and explain about 2 percent of the variation. That said, the only significant indicator is Twitnovation success within the hour. Note that the average word only receives 0.004 subsequent mentions within its first hour of appearing (Std. Dev. 0.014). This so low because most words are never mentioned. Hence, the coefficient of 21.759 would imply that a 1 standard deviation increase is associated with a 31% increase in labor productivity.

Column 2 focuses on one Twitnovation measure only. This reduces the coefficient size somewhat but it remains of a similar order of magnitude and significance. Overall, this explains still around 1.7 percent of the within variation.

In Column 3, we add the set of intangible n-gram variables. We note that, as a group, they capture about 3% of the within-variation of labor productivity. That said, the only individual coefficient that emerges as significant is for "reputation". To illustrate what reputation captures in this context, let's take some examples. @unicordoba_col, a university, tweeted " *en habla sobre certificacion iso que recibio unicordoba* ". It is announcing the acquisition of a new international quality certification from ISO, the International Organization for Standardization. Improvements in service quality have been shown to positively correlate with productivity, as it usually leads to inputs being more efficiently utilized.

As an other example, @aminaperu, a company, also announced they obtained an ISO certification. They tweeted the following: "*obtuvimos la certificacion iso luego de haber aprobado con exito y sin observaciones la auditoria a los procesos del sistema integrado de gestion sig ademas renovamos las certificaciones iso e iso*". Firms also use Twitter to announce their attendance of trade fairs and exhibitions such as @lagacetasa: "*lumbrales acoge este fin de semana su feria comercial y multisectorial*".

TABLE 23
Regressions of Labor Productivity
- Latin America -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)
ln(1 + Number of words)	0.102	0.080	0.115*	0.113	0.101
	(0.075)	(0.070)	(0.067)	(0.074)	(0.075)
Mean Twox .9 Success 1 hour	21.764***	13.798**			13.985**
	(7.351)	(5.547)			(6.179)
Mean Twox .9 Success 30 days	-1.167				
	(0.849)				
Mean Twox .9	1.438				
	(1.458)				
training			-0.003		0.001
			(0.005)		(0.003)
software			0.000		-0.093**
			(0.002)		(0.046)
research			-0.007		0.035
			(0.019)		(0.029)
design			-1.406		-0.949
			(0.962)		(2.322)
reputation			0.659*		0.405
			(0.356)		(0.279)
organizational			-0.022		0.036
			(0.129)		(0.129)
rfid				0.010	0.030
				(0.071)	(0.095)
ai_long				-0.020*	-0.018
				(0.011)	(0.011)
ai_short				0.034*	0.036*
				(0.019)	(0.021)
lcds				-0.598***	-0.313
				(0.217)	(0.281)
threedprinting				0.043	0.026
				(0.041)	(0.059)
solar				0.019*	0.021**
				(0.011)	(0.010)
oled				0.000	-0.011
				(0.004)	(0.012)
virtual_reality				-0.076	-0.096
				(0.195)	(0.200)
gps				0.018	-0.032
				(0.042)	(0.051)
driverless				-0.580*	-0.378
				(0.303)	(1.016)
smartphone				0.005	0.007
				(0.007)	(0.007)
fracking				-0.071	-0.096
				(0.050)	(0.066)
disk_drive				0.105	0.099
				(0.194)	(0.212)
cloud				-0.006	0.143
				(0.040)	(0.087)
electric_cars				0.012	0.004
				(0.015)	(0.014)
smartcard				2.726***	2.829***
				(0.745)	(0.761)
bluetooth				0.332***	0.339***
				(0.107)	(0.112)
optical_fiber				-0.092	-0.065
				(0.130)	(0.128)
touch_screen				-0.273	-0.209
				(0.259)	(0.268)
Country X Sectors FEs	X	X	X	X	X
Observations	311	311	311	311	311
R2 Within	0.028	0.022	0.036	0.149	0.194

Linear regression.Standard errors in parentheses

clustered at the user level.

Outcome variable: ln(LP = Revenues/Number of Employees)

Column 4 considers our set of technical terms, which explain about 15% of the variation in productivity as a group. Column 5 considers all variables combined. This leads to a within-R2 of 20%. Crucially, our Twitnovation measure is still highly significant and, in terms of magnitude, is hardly affected at all. It would, therefore, seem that this rather crude and simple-to-calculate measure has quite some explanatory power distinct from a rather large set of n-grams. That said, both sets of n-grams remain statistically significant; however, in the case of the intangible n-grams, that reduces to a 10% significance level for the joint significance of all variables (un-reported result).

We repeat the exercise for (log) patent counts as outcome (see Table I.17 in the Appendix)[100]. Our indicators seem to have very little explanatory power here. Even in the column that includes all indicators, we only account for less than 10% of the within variation. One issue, however, might be that firms in Latin America generally patent very little. Out of our sample of over 1000 firms where we can identify Twitter handles we have only 41 with at least one patent.[101]

Columns 1 to 4 in Table 24 report regression results when we include our customer engagement variables in addition to Twox and bigrams. The sample size is much smaller because the Twitter API quotas limited the number of tweets *about* firms that we could collect. Hence, when we include our customer engagement variables (number of mentions and response rate), the sample is smaller than when we only use indicators based on tweets from firms' timelines.

Column 1 includes only the customer engagement variables as detailed in section 3.4.3. Interestingly, only the log number of mentions the firm receives correlates well with productivity. The correlation remains significant even when controlling for our complete set of other indicators in Column 3. It appears that, by and large, it also accounts for the variation previously captured by our Twitnovation measures, as the Twitnovation coefficient is now smaller and no longer significant. What could be behind this?

In Columns 4 and 5, we examine if it is a simple size effect by including the number of employees (logged) as an additional control variable, i.e., more prominent firms might be more often tweeted *about* and be more productive. However, there is no suggestion that this is the case.

TABLE 24
Regressions of Labor Productivity on Customer Engagement
- Latin America -

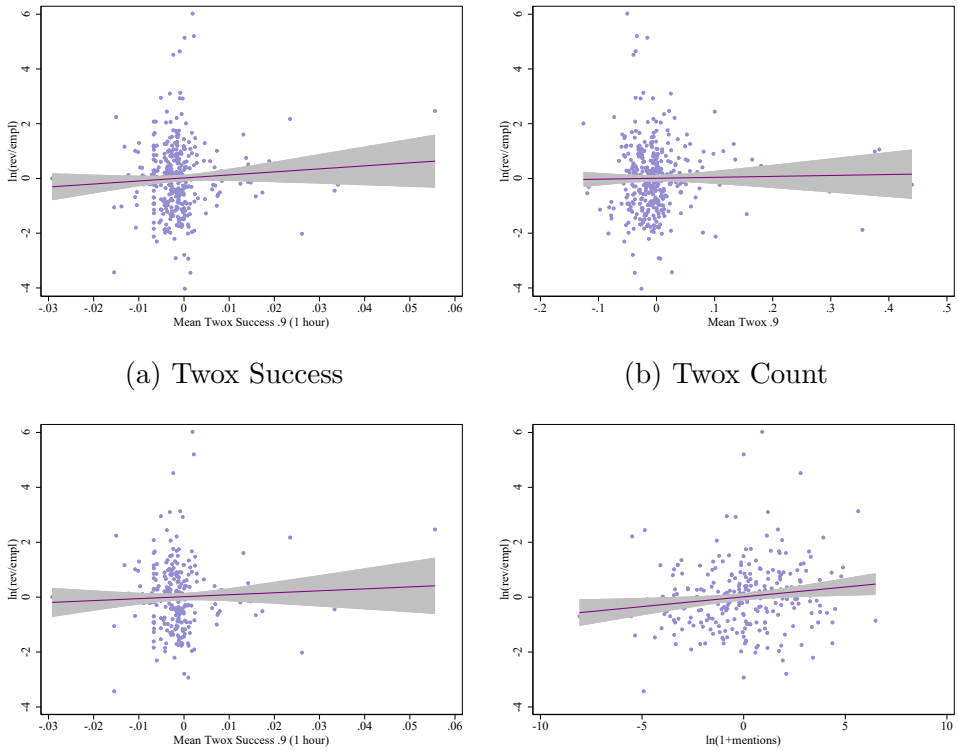
	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)	(6) ln(Rev/Empl)
ln(1 + Number of mentions)	0.135***		0.101**	0.101**	0.115**	0.130***
	(0.044)		(0.046)	(0.046)	(0.047)	(0.043)
Rate of Response perc.	0.028		-0.037	-0.037	-0.042	0.024
	(0.041)		(0.068)	(0.068)	(0.071)	(0.042)
Rate of Response (Weighted)	-22.504		22.632	22.632	26.012	-14.611
	(27.765)		(38.924)	(38.924)	(40.797)	(27.368)
ln(1 + Number of words)	0.024	-0.115	-0.177*	-0.177*	-0.123	0.058
	(0.071)	(0.093)	(0.096)	(0.096)	(0.096)	(0.069)
Mean Twox .9 Success 1 hour		12.737*	9.221	9.221	10.771*	10.022*
		(6.562)	(6.253)	(6.253)	(6.203)	(5.711)
ln(Number of Employees)					-0.143**	-0.157***
					(0.056)	(0.059)
Bigrams	None	All	All	All	All	None
Full Sample	X	Mentioned	X	X	X	X
Country FEs X Sectors FEs	X	X	X	X	X	X
Observations	215	215	215	215	215	215
R2 within	0.077	0.311	0.336	0.336	0.369	0.137

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)

We also illustrate these relationships in Figure 54. Panel a) reports a scatter plot of labor productivity on Twitnovation success (Within transformations of sector-country cells in each). The figure illustrates a positive relationship. However, the fit is not super tight. It gets even worse in panel (b), where we report simply Twitnovation shares. By contrast, Panel (d) reports a scatter plot for productivity vs. the log of mentions. This is a remarkably close fit. For completeness, Panel (c) reports the scatter plot for Twitnovation Success for the smaller sample where we have customer engagement data.

Hence, it would appear that out of all our variables considered, the number of times a firm is talked about – which we propose to call "Aboutnovation" – has the most power in accounting for productivity variation.

FIGURE 54
Regressions of Labor Productivity on Customer Engagement
- Latin America -



(a) Twox Success

(b) Twox Count

(c) Twox Success for mentioned firms

(d) Mentions

Note: The Figure reports scatter plots of within transformations - within (2 digit) sector×country cells - for labor productivity (revenue over employment) vs various Tweet-based indicators, i.e. the difference between a particular value and the sector× country cell average. Panel (a) reports on Twox Success (within 1 hour). Panel (b) reports Twox count. Panel (c) reports Twox Success for the sample for which we have customer engagement data. Panel (d) reports the (log) of the mentions counts. The graphs report all observations for except firms with extreme values (outliers) these include: sancorok,herbalife ,pirelli ,exxonmobile, bavaria_sa.

3.5 Aboutnovation

We further explore the mentions of companies and implement two methods to classify them. First, by reference type according to the Twitter API outputs. Second, by the sentiment of the mention, as determined by sentiment analysis. The process of analyzing text to determine if the emotional tone of the message is positive, negative, or neutral.

3.5.1 Aboutnovation Language

We consider that not all mentions are necessarily equal and that the context in which they are made matters. To explore this, we describe below some differences in the language of the tweets.

The top three languages of mentions are Spanish, English, and Portuguese for the Latin

American sample. In the UK, the top language is English, followed by tweets in an unidentified language, usually tweets including only media links (See tables J.24 and J.25 in the Appendix).

Language is critical to identify the nature of aboutnovation. First, it distinguishes the type of customers being reached by firms. For instance, whether companies are reaching a global base of customers beyond a local base. The primary language of the countries included in the Latin American sample is Spanish. Finding a substantial portion of mentions in English can signal multiple potential issues. Perhaps we are capturing a multinational that shares the same handle with all its subsidiaries worldwide, the company is genuinely reaching an English-speaking customer base, or it could describe something about the language of employees and other stakeholders that talk about the company, among other multiple reasons.

Language can also be a limitation in our analysis. It is rare to find machine learning classifiers for Twitter in general, but it is even rarer to find them for non-English languages. The degree of development and massification of these in Spanish and other languages is lower than that of English tweets. Furthermore, to the best of our knowledge, analyzing mentions of companies is still a nascent practice, if any, which adds to the language issue. Hence, in future research, we aim to develop our original classifier of companies "by" and "about" tweets or social media outputs.

Given these limitations, we develop regression analysis that first considers the totality of mentions in numbers (without discriminating by language), but later on, for sentiment analysis, we make language distinctions. For the Latin American data, we focus on Spanish tweets; for the UK data, we focus on English tweets. Although classifiers are limited in non-English languages, in particular for Twitter classification, we use a classifier trained on similar data sources to Twitter for the analysis in Spanish. We explain further in the corresponding section.

3.5.2 Aboutnovation Types

The first classification method takes advantage of the features describing a tweet provided by the API. First, we distinguish by *reference type* and by *type of retweet*.

The Twitter API allows identifying the mention type through the *referenced.tweets.type* expansion.[102]. This feature indicates the type of relationship between a Tweet and the Tweet returned in the response (the tweet (mention) we found by searching @[company user handle]). There are three reference types. A mention can have reference type, retweeted (this Tweet is a Retweet), quoted (a Retweet with a comment, also known as Quoted Tweet), or replied_to (this Tweet is a reply). We consider a 4th category that we call "independent", which refers to tweets (mentions), not referring to any other tweet, so

it has a null value in the reference type variable. For the UK, we combine "independent" and "replied to" into the same independent tweets (mentions) category.

We explain the differences with some examples. A mention of Company A of the type "retweeted" means that someone reposted a tweet in which Company A is mentioned. The tweet is re-posting another Tweet that includes the handle of our company of interest in the form of **@handle**. As explained below, a retweet can be direct or indirect. Direct retweets are re-posts of Tweets by users (or the company itself) of a company's tweet. An indirect retweet is a re-post of someone else's Tweet that mentioned the company; it was not directly a company's tweet.

A mention (tweet) of Company A of the type "quoted" means that the mention of the company can occur only in the comment added to the quoted tweet since we do not retrieve the referenced tweet in the quote. A mention (tweet) of Company A of the type "replied to" means that the tweet is a response to a company's tweet or someone else's Tweet mentioning the company. A mention (tweet) of Company A of the type "independent" means that the tweet is an original text mentioning a company and is not referring to anyone else's tweet.

Tables 25 and 26 present different properties of the mentions by reference type in Latin America and the UK. In both regions, retweets are the primary reference type of mentions. Replies and independent tweets follow. Quoted tweets are the least common.

In Latin America, Independent mentions tend to have more likes than any other mention type, with an average of 6 likes per tweet. Likes are represented by a small heart below each Tweet, and users can click on it to show appreciation for a Tweet. Mentions that are retweeted are the most retweeted, with around 746 retweets. Notice that the number of likes and retweets in the last two columns in the table refer to the actual mention (or tweet) that we count; it is the appreciation of the tweet that we consider a mention. For the UK, we combine replies and independent tweets in the same category. When considered together, the average of likes per reference type is higher for the Quoted tweets.

TABLE 25
Mentions by type - Latin America -

Type of Mention	N.Mentions	(%)	Avg.Likes	Avg.Retweets	N.Likes	N.Retweets
Total	29,305,576	100	1.5	414.2	44,664,086	**
Replied	7,461,924	25.5	1.3	0.2	9,865,474	1,575,338
Retweeted	16,265,647	55.5	0	745.5	0	**
Quoted	538,276	1.8	4.7	1.5	2,508,233	796,402
Independent	5,039,729	17.2	6.4	1.9	32,290,379	9,485,277

Note: The table contains the number of mentions (tweets) of a company by the reference type. It includes mentions with a valid date for years between 2017 and 2021.
** refers to unavailable values for that category.

TABLE 26
Mentions by type - UK -

Type of Mention	N.Mentions	(%)	Avg.Likes	Avg.Retweets	N.Likes	N.Retweets
Total	33,985,864	100.0	1.6	1,105.0	53,950,768	**
Retweeted	17,472,191	51.4	0.0	2,148.6	0	**
Independent	15,859,037	46.7	3.2	0.7	50,826,341	11,325,995
Quoted	654,636	1.9	4.8	1.3	3,124,427	820,557

Note: The table contains the number of mentions (tweets) of a company by the reference type. It includes mentions with a valid date for years between 2017 and 2021.
** refers to unavailable values for that category.

Tables 27 and 28 distinguish further between indirect and direct retweets. An "indirect retweet" is a re-post of a tweet that mentioned a company i, but it was not the company the original's tweet issuer. For instance, **@playmobiluk** is mentioned via an "indirect retweet" through the tweet: *"RT @glowormfest: COMPETITION TIME In partnership with our Headline Sponsors @PlaymobilUK, we're giving a lucky Glowormer a chance to w…"*. In this case, **@glowormfest** tweeted about **@playmobiluk**, and it was the tweet that was re-posted. A "direct retweet" refers to someone else but the company or the company retweeting a company i's tweet; for instance, **foxsportasia** was mentioned via a re-post of the following company's tweet: *"RT @FOXSportsAsia: This is embarrassing for ! The Timor-Leste football team were left stranded at the airport, and later taken to th…"*.

Tables 27 and 28 present in the category "rest" all tweets that do not fall in any indirect or direct retweet categories. For the reference type "retweeted', all mentions under that category should be, in principle, indirect or direct retweets. However, as seen in the tables, 6,640 and 1,026 tweets (mentions) fall out of the indirect-direct subcategory of retweets. This occurs because we include withheld tweets in our collection of tweets. According to countries' policies, Twitter can withhold access to the tweet depending on the rules or policies established by a country regarding content, preventing us from retrieving the original text. For instance, mention of **@sports_england** was withheld, and the only text retrieved was *This Tweet from @VGAGymnastics has been withheld in: United Kingdom.*, so we know **@sports_england** was mentioned but do not know how.[103].

TABLE 27
Mentions by type and subtypes - Latin America -

Reference type	N.Indirect RT	N.Direct RT	Rest	Total
Retweeted	10,679,220	5,579,787	6,640	16,265,647
Replied_to	117	51	7,461,756	7,461,924
Independent	12,713	15,381	5,011,635	5,039,729
Quoted	566	228	537,482	538,276

TABLE 28
Mentions by type and subtypes - UK -

Reference type	N.Indirect RT	N.Direct RT	Rest	Total
Retweeted	8,736,573	8,734,592	1,026	17,472,191
Independent	20,502	26,596	15,811,939	15,859,037
Quoted	1,036	6,506	647,094	654,636

Note:
N. Indirect RT: Number of Indirect Retweets of the company
N. Direct RT: Number of Direct Retweets of a company's Tweet
Rest: All tweets that are not re-posting someone else's tweet.

There are interactions between reference types and indirect-direct retweets. For example, direct or indirect retweets can have the "replied to" reference, indicating that the tweet is a re-posting of a company's tweet or someone else's Tweet mentioning the company, that was a reply to someone.

A retweet can also be a re-post of a Tweet that is not referencing anybody's tweet, so its reference type is "independent". In our sample, that tweet would be a re-posting of a tweet mentioning a company that originally was not referencing any other tweet. In essence, it is a retweet of an original tweet.[104] Retweets with "retweeted" reference indicate re-posts of Tweets that were already a retweeted tweet. The mention of the company occurred in the original retweet that is being retweeted. Appendix J.2 contains examples of tweets under each category for both regions. Tables J.19 for English tweets in Latin America, Table J.20 for Spanish tweets in Latin America and Table J.21 for the U.K.

3.5.3 Aboutnovation sentiment

As the Latin American sample comes from predominantly Spanish-speaker countries, and a substantial portion of that region's sample of mentions is in Spanish, we estimate sentiment for tweets in that language. For reference, we also assess the sentiment of English tweets in Latin America using VADER Sentiment Analysis. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool attuned explicitly to sentiments expressed in social media and works well on texts from other domains (Hutto, 2020).

Table 29 presents the classification of the Spanish tweets in three categories: positive, neutral, and negative, using 0.9 as the probability selected for the threshold. For this analysis, we use *sentiment-spanish*, a Python library that uses convolutional neural networks to predict the sentiment of Spanish sentences (Bello, 2021). The author trains the model using over 800,000 reviews of users of the pages *eltenedor*, *decathlon*, *tripadvisor*, *filmaffinity* and *ebay*.[105]. All these reviews were extracted using web scraping with the project opinion-reviews-scraper[106]. We consider this library suits the nature of the

"about" tweets in Spanish and constitutes a good source of training data because it is likely that reviewers on those websites are comparable to the type of users or customers in our sample. Such reviewers not only reveal digital engagement but also that they write their opinions regarding a product or a service, actions that are comparable to a user posting their comments on Twitter about a company.

Table 30 presents sentiment classification results of English language mentions in the same categories for the U.K. We use Google Big Query, the machine learning algorithms of distant supervision and training methods as suggested by Go et al. (2009) for classifying the sentiment of Twitter messages. The authors use emoticons as noisy elements to label the training sample of tweets. Emoticons are pictorial representations of a facial expression using characters—usually punctuation marks, numbers, and letters. For instance, :) represents a happy face. In alignment with prior methods that use users reviews signals, such as stars to rank restaurants, their method relies on the "distant" approach of labeling the training data with noisy signals such as emoticons. The method allows completing the labeling task faster than doing it by hand as is traditional in supervised machine learning. We chose the threshold for the accuracy of the method to be above 90% when trained with emoticon data.

The sentiment results indicate that Spanish tweets are more negative than English tweets. However, this could be attributed to the differences in the classification methodology. In principle, we could use the emoticons technique for the Spanish tweets but we also want to explore different approaches to classify tweets about companies, so this is also part of the exercise.

TABLE 29
Sentiment of mentions in Spanish - Latin America (Bello (2021) methodology)

	N Tweets	Pct.
NEGATIVE	6,747,452	77.9
NEUTRAL	1,367,489	15.8
POSITIVE	548,698	6.3
TOTAL	8,663,639	100.0

TABLE 30
Sentiment of mentions in English - UK (Go et al. (2009) - GBH methodology)

	N Tweets	Pct.
NEGATIVE	2,291,500	6.7
NEUTRAL	23,082,616	67.9
POSITIVE	8,612,281	25.3
TOTAL	33,986,397	100.0

Note: Pct.: Percentage. N Tweets: Number of Tweets (mentions).

Figure 55 shows examples of aboutnovation words by positive and negative sentiment categories for two companies, Selfridges and Tableau, present in the UK sample. It is characteristic of both companies that the words associated with negative sentiment are related to negative experiences with customer service or the product or service offered by the company.

FIGURE 55
Examples of Aboutnovation Words: Positive and Negative Sentiment

(a) Selfridges: Positive

(b) Selfridges: Negative

(c) Tableau: Positive

(d) Tableau: Negative

Note: The wordclouds show words corresponding to their frequency in company mentions from 2017 to 2022. The minimum frequency of the word has to be five and a maximum of 100. The plot was based on the top 1000 more frequent words per company x sentiment. The maximum number of mentions per company considered was 200,000.

Notice that sentiment is unequal across types of references. Table 31 shows sentiment scores discriminated by language and reference type, using Vader for English tweets and Bello, 2021 for Spanish tweets in Latin American. Table 32 shows sentiment scores using

Go et al. 2009., for the U.K.

Sentiment scores are a metric for measuring customer sentiment, revealing the emotional intensity in the text, and they have a range depending on the technique. Primarily, VADER sentiment analysis relies on a dictionary that maps lexical features to emotion intensities, the sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. A lexical feature is any element used to communicate via text. For example, in a tweet, there are words and emoticons like ":-)", acronyms like "LOL", and slang like "meh". VADER sentiment analysis is made to capture such colloquialisms and it maps them to intensity values. In Vader, emotion intensity or sentiment score is measured on a scale from -4 to +4, where -4 is the most negative and +4 is the most positive. The midpoint 0 represents a neutral sentiment. Sample entries in the dictionary are "horrible" and "okay," which get mapped to -2.5 and 0.9, respectively. In addition, the emoticons "/-:" and "0:-3" get mapped to -1.3 and 1.5 (Hutto, 2020; Hutto and Gilbert, 2014). The technique for the Spanish tweets outputs sentiment scores in the range of 0 and 1. The sentiment score in this case is a probability of a string variable text of being "positive". Low probabilities mean that the text is negative (numbers close to 0), high probabilities (numbers close to 1) mean that the text is positive (Bello, 2021).

For the U.K. and Latin America, English "quoted" mentions tend to have a more positive sentiment, as the sentiment score tends to be higher than the rest of the reference types. Retweeted and independent tweets follow. In contrast, "retweeted" are the reference type with the highest sentiment score in Spanish tweets for Latin America. "Quoted", "independent", and "replied to" reference types follow, in that order.

TABLE 31
Average sentiment score of mentions by reference type and language:
Latin America
- Vader (Hutto, 2020; Hutto and Gilbert, 2014) and Sentiment-Spanish (Bello, 2021) -

Language	Reference type	%	N. mentions	Sentiment EN	Sentiment SP
en	Quoted	0	62,931	0.309	**
en	None	20	789,928	0.284	**
en	Replied_-to	20	833,311	0.228	**
en	Retweeted	60	2,091,136	0.302	**
en	Total	100	3,777,306	0.282	**
es	Quoted	0	129,290	**	0.148
es	None	20	1,371,096	**	0.145
es	Replied_-to	20	2,092,037	**	0.136
es	Retweeted	60	5,071,216	**	0.186
es	Total	100	8,663,639	**	0.167

Note: The "None" category represents independent tweets.
It was not possible to estimate sentiment for all mentions in English or Spanish; primarily because those tweets after clean-up didn't have any remaining meaningful words or emoticons. Hutto (2020) was used for English Tweets, and Bello (2021) for Spanish tweets.

TABLE 32
Average sentiment score of mentions by reference type and language:
United Kingdom
- Go et al. (2009) - GBH methodology -

Language	Reference type	%	N. mentions	Sentiment EN
en	Quoted	1.7	534,668	0.449
en	Independent	44.8	13,727,550	0.346
en	Retweeted	53.5	16,385,808	0.329
en	Total	100.0	30,648,026	0.340

3.5.4 Aboutnovation Types and Labor Productivity

This section reports regression results of the following specification:

$$LaborProd_j = \beta TW_j + \alpha_m + \epsilon_j \tag{74}$$

Where TW_j is a vector of social media-derived indicators for firm j and $\alpha_m = \alpha_{sc}$ are (2-digit) sector $\times$ country fixed effects for Latin America regressions and $\alpha_m = \alpha_s$ are (2-digit) sector fixed effects for UK regressions.

Table 33 presents the results for Latin America. Column 3 presents a positive and significant coefficient for mentions after including controls such as the number of words and the company size. Column 4 introduces the squared value of company size. Column 5 replaces logged mentions with the mentions' reference types. Quoted and retweeted

mentions are combined in the same category. Interestingly, the logged number of replied_-to mentions coefficient is positive and highly significant. For every 1% increase in this type of mention, labor productivity increases by about 0.26%.

TABLE 33
Regressions of Labor Productivity on Customer Engagement by type
- Latin America -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)	(6) ln(Rev/Empl)
ln(1 + Number of mentions)	0.107**	0.106**	0.120***	0.121***		
	(0.042)	(0.042)	(0.044)	(0.045)		
ln(1 + Number of words)		0.002	0.027	0.028	0.046	0.147**
		(0.064)	(0.063)	(0.064)	(0.065)	(0.071)
ln(Number of Employees)			-0.139**	-0.105	-0.113	-0.161
			(0.060)	(0.166)	(0.166)	(0.155)
ln(Number of Employees) sq.				-0.003	-0.008	0.003
				(0.015)	(0.015)	(0.014)
ln(1+independent)					-0.159	
					(0.147)	
ln(1 + quoted or retweeted)					0.014	
					(0.081)	
ln(1 + replied to)					0.262**	
					(0.115)	
ln(1+direct ret.)						-0.070
						(0.048)
ln(1+indirect ret.)						0.688
						(0.633)
ln(1+rest of types)						-0.586
						(0.597)
Bigrams	None	None	None	None	None	None
Full Sample	X	X	X	X	X	X
Country FEs X Sectors FEs	X	X	X	X	X	X
Observations	215	215	215	215	215	215
R2 within	0.046	0.046	0.083	0.083	0.139	0.071
R2	0.629	0.629	0.643	0.643	0.661	0.634

Notes: Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)
ln(1 + x): ln(1+ number of mentions of type x)
Independent: N. of mentions (tweets) of the independent type
Quoted: N. of mentions (tweets) of the quoted type
Retweeted: N. of mentions (tweets) of the retweeted type
Direct ret: N. of retweets of a company's tweet'
Indirect ret: N. of retweets of a user's tweet mentioning a company'
Note: ln(Number of employees) sq.: ln(Number of employees) × ln(Number of employees)

In Column 6, we replace the logged number of mentions instead by the type of mention in direct retweets, indirect retweets, and the rest. None of these types are significant. These results indicate that just the most engaging tweets directly replying to a tweet are more significant for a firm's revenue-to-employee ratio.

The importance of a retweet versus a reply can vary depending on the context and

marketing goals of the companies involved. For some companies, the primary aim may be to reach a larger audience and spread information (favoring retweets), while others may prioritize meaningful engagement and discussion (favoring replies). Ultimately, the significance of retweets versus replies to labor productivity can be many, and at this point of the analysis, we can only speculate. However, we discuss some reasons why this could be happening.

A positive and significant impact of replies can indicate high customer engagement and relationship-building capabilities. A high number of replies suggests active engagement and conversation between the users mentioning the company, the company, and its customers or audience. Replies often indicate direct interaction, personalized responses, and a willingness to address specific inquiries or feedback. This level of engagement can lead to stronger customer relationships, improved customer satisfaction, and loyalty. Positive customer interactions can increase sales, repeat business, and customer advocacy, positively impacting the company's turnover-to-labor ratio.

Replies to tweets mentioning the company or to a company's tweets often involve feedback, suggestions, or ideas from customers or the general audience. Listening and responding to this feedback can provide valuable insights into customer preferences, needs, and pain points. By incorporating customer feedback, the company can improve its products, services, and processes, enhancing overall efficiency and productivity. This customer-centric approach can positively influence the turnover-to-employees ratio.

More replies may also indicate active participation from employees in engaging with customers or addressing their concerns. This involvement signifies a collaborative and engaged workforce that takes ownership of customer interactions. Employees actively contributing to conversations through replies can foster positive brand perception and nurture customer relationships. Enhanced employee engagement and collaboration can positively impact job satisfaction and labor productivity.

The lack of significant impact from retweets about the company suggests that companies value direct communication and personalized interactions. While retweets can help increase the company's visibility and reach a wider audience, they may not necessarily foster deeper engagement or generate valuable insights. By prioritizing replies and focusing on direct communication, the company emphasizes the value of individualized attention and active engagement, which can lead to stronger customer relationships and improved labor productivity.

Table 34 shows similar results for the UK. Column 1 exhibits a regression of the logged mentions against labor productivity; the coefficient is positive, highly significant, and close in magnitude to the coefficient found in Latin America. The coefficient remains positive and significant in all specifications, with some variations in magnitude after

controls. Column 2 includes the number of words used in the mentions (tweets) after clean-up (e.g., eliminating stopwords), and Column 3 also includes the number of words before cleaning. In both columns, the impact of "aboutnovation" is slightly lower but remains significant. Column 6 includes a dummy variable indicating whether the firm is mentioned or not. Table J.29 presents some variations in the Appendix.

In contrast to the regressions for Latin America, we include both mentioned and not mentioned firms (but queried) in the sample. The dummy coefficient is positive and significant, indicating that just the fact of being mentioned matters. Column 7 reveals that direct retweets hurt labor productivity, whereas anything but retweets positively impacts productivity.

TABLE 34
Regressions of Labor Productivity on Customer Engagement
- United Kingdom -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)	(6) ln(Rev/Empl)	(7) ln(Rev/Empl)
ln(1+Number of Mentions)	0.084***	0.050***	0.032**	0.052***	0.057***	0.128***	
	(0.004)	(0.017)	(0.016)	(0.017)	(0.017)	(0.027)	
ln(1+Number of Words C.)		0.027**		0.030**	0.026**	-0.046*	-0.060***
		(0.012)		(0.012)	(0.013)	(0.025)	(0.020)
ln(1+Number of Words)			0.041***				
			(0.011)				
ln(Number of Employees)				-0.036***	0.006	0.005	0.012
				(0.008)	(0.023)	(0.023)	(0.023)
ln(Number of Employees)^2					-0.005**	-0.005**	-0.006***
					(0.002)	(0.002)	(0.002)
Mentioned (1: Yes)						0.229***	0.187***
						(0.065)	(0.068)
ln(1+direct ret.)							-0.123***
							(0.015)
ln(1+indirect ret.)							-0.014
							(0.009)
ln(1+rest of types)							0.290***
							(0.023)
Full Sample	X	X	X	X	X	X	X
Sectors FEs	X	X	X	X	X	X	X
Observations	7748	7748	7748	7748	7748	7748	7748
R2 within	.0743	.0748	.0755	.0782	.0789	.0801	.0992
R2	0.259	0.260	0.260	0.263	0.263	0.264	0.279

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)
Note: Words C.: Words resulting from the count after cleaning the text of the mention (tweet)
Words: Words resulting from the count of the raw text of the mention (tweet).
ln(1 + x): ln(1+ number of mentions of type x)
Independent: N. of mentions (tweets) of the independent type
Quoted: N. of mentions (tweets) of the quoted type
Retweeted: N. of mentions (tweets) of the retweeted type
Direct ret: N. of retweets of a company's tweet'
Indirect ret: N. of retweets of a user's tweet mentioning a company
Rest of types = N. of mentions - Direct ret. - Indirect ret.

These results align with our earlier findings for Latin America. Retweeting a company's tweet without adding original content or commentary may not provide much value in fostering meaningful discussions or engagement. It could be seen as a passive action,

indicating a lack of active involvement or critical thinking. On the other hand, replying to a tweet or mentioning the company with an original tweet shows a proactive effort to contribute to the conversation, share insights, or provide additional context. This active engagement can stimulate discussions that might be more useful for the company to enhance labor productivity.

Furthermore, retweeting alone does not require much effort or thought. It is a simple act of sharing existing content without adding any personal touch or contribution. This can lead to shallow engagement with the company's message, resulting in limited interaction with helpful information for labor productivity. Conversely, replying to a tweet or mentioning the company with an original tweet demonstrates a deeper level of engagement. It encourages dialogue, prompts further conversation, and allows for the exchange of ideas, potentially leading to more meaningful outcomes and improved productivity.

Similarly, retweeting a company's tweet does not provide an opportunity to provide feedback or share suggestions directly. Feedback is vital for companies to identify areas for improvement, address concerns, and enhance their products or services. By replying or mentioning the company with an original tweet, individuals can offer valuable feedback, express opinions, or provide suggestions, which can help the company identify opportunities for growth, innovation, and increased productivity.

This intuition aligns with our findings assessing the sentiment of the mentions. Table 35 shows that positive mentions have a negative association with labor productivity, while negative tweets have a positive association. This might indicate that firms that receive negative feedback can use this information to improve their productivity. In contrast, firms that receive positive feedback do not have the opportunity to learn and improve from customers' feedback. A proxy for how influencing is a particular mention is the number of likes to that mention. Table J.28 in the Appendix shows that when including the average likes of the mentions, the value of the coefficient of mentions (column 2) remains positive but lower (than in column 1), suggesting that it could be even more important how much others support such a mention, regardless of whether is favourable or not.

We also illustrate these relationships in Figure 56. Panel a) reports a scatter plot of labor productivity on logged value of one plus the number of mentions for our sample in the UK (Within transformations of sector-country cells in each). The figure illustrates a positive relationship. This is a remarkably close fit. It also fits nicely in panel (b), where we report only mentioned firms.

For completeness, Figure 57 reports the scatter plots for the logged mentions against labor productivity for the Latin American and UK samples. The color intensity indicates the company's position in the labor productivity distribution of the total sample for each

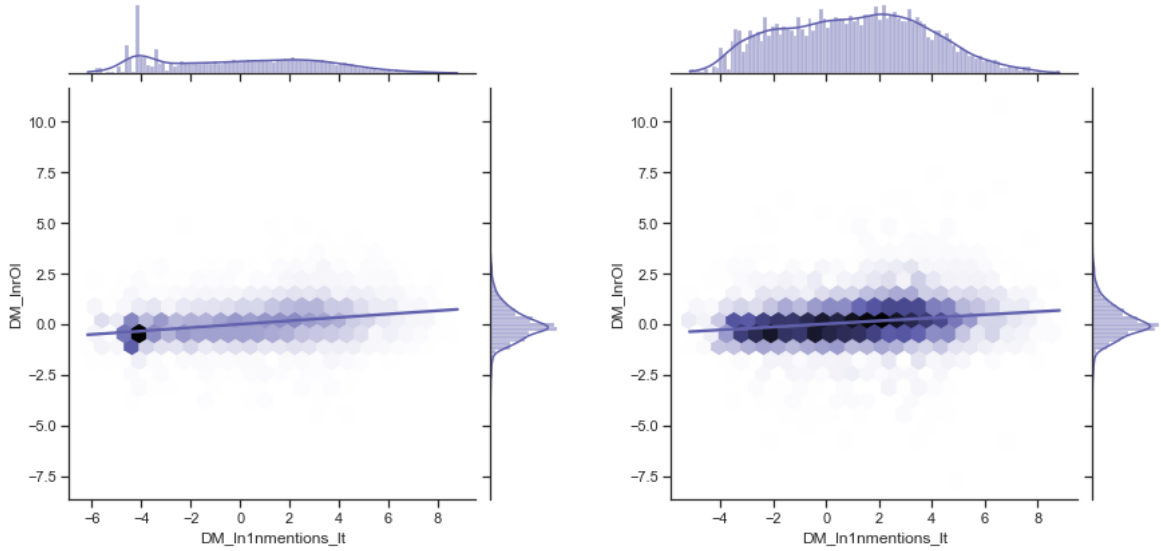
region. Green represents the top quantiles, and purple represents the bottom quantiles. The size of the dot represents relative frequencies. The relationship between the variables is positive for both samples, and the coefficients are close in magnitude.

TABLE 35
Regressions of Labor Productivity on Customer Engagement
- United Kingdom -
- Sentiment (GBH Methodology) -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)	(6) ln(Rev/Empl)
ln(1+Number of Mentions)	0.084***	0.050***	0.057***	0.128***		
	(0.004)	(0.017)	(0.017)	(0.027)		
ln(1+Number of Words C.)		0.027**	0.026**	-0.046*	0.023*	0.007
		(0.012)	(0.013)	(0.025)	(0.012)	(0.023)
ln(Number of Employees)			0.006	0.005	0.013	0.013
			(0.023)	(0.023)	(0.023)	(0.023)
ln(Number of Employees)^2			-0.005**	-0.005**	-0.007***	-0.007***
			(0.002)	(0.002)	(0.002)	(0.002)
Mentioned (1: Yes)				0.229***		0.063
				(0.065)		(0.069)
ln(1+pos. GBH)					-0.123***	-0.117***
					(0.017)	(0.018)
ln(1+neu. GBH)					0.135***	0.145***
					(0.023)	(0.026)
ln(1+neg. GBH)					0.049***	0.050***
					(0.015)	(0.015)
Full Sample	X	X	X	X	X	X
Sectors FEs	X	X	X	X	X	X
Observations	7748	7748	7748	7748	7748	7748
R2 within	.0743	.0748	.0789	.0801	.0879	.088
R2	0.259	0.260	0.263	0.264	0.270	0.270

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)
Note: Words C.: Words resulting from the count after cleaning the text of the mention (tweet)
Words: Words resulting from the count of the raw text of the mention (tweet).
ln(1 + x): ln(1+ number of tweets with sentiment type x)
Pos: Number of positive tweets for the firm i
Neg: Number of negative tweets for the firm i
Neu: Number of neutral tweets for the firm i
The sentiment was calculated using Big Query ML based on Go et al. (2009) methodology (GBH) to classify the sentiment of Tweets.

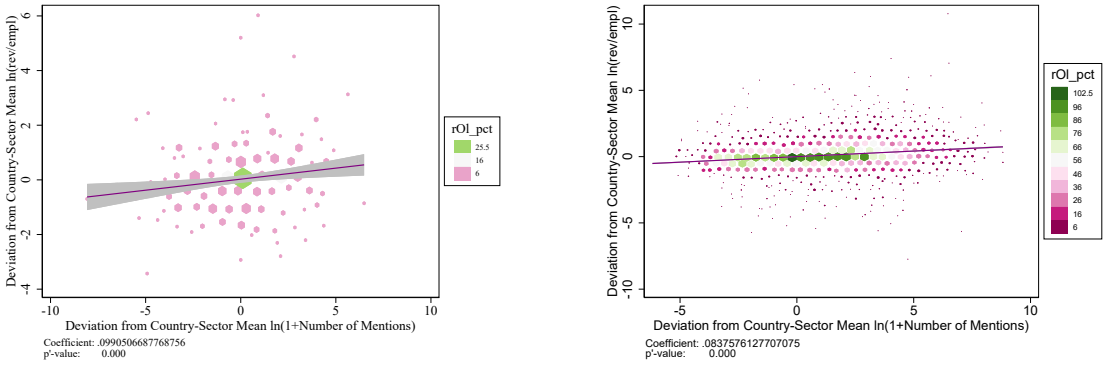
FIGURE 56
Regressions of Labor Productivity on Customer Engagement (Mentions)
United Kingdom



(a) All (b) Only mentioned

Note: The Figure reports scatter plots of within transformations - within (2-digits) sector×country cells - for labor productivity (revenue over employment) vs natural logarithm of one plus the number of mentions; i.e. the difference between a particular value and the sector× cell average. Panel (a) reports on all UK firms. Panel (b) reports only mentioned UK firms. Confidence intervals are reported in the shaded area at 95%.
Greener tones represent firms with higher labor productivity quantiles for the full sample *times* period, and pink tones represent lower quantiles. The size of the color represents the frequency of the firms,

FIGURE 57
Regressions of Labor Productivity on Customer Engagement (Mentions)
Latin America vs. United Kingdom



(a) Latin America (b) United Kingdom

Note: The Figure reports scatter plots of within transformations - within (2-digits) sector×country cells - for labor productivity (revenue over employment) vs natural logarithm of one plus the number of mentions; i.e., the difference between a particular value and the sector× cell average. Panel (a) reports on Latin American companies. Panel (b) reports on United Kingdom companies.

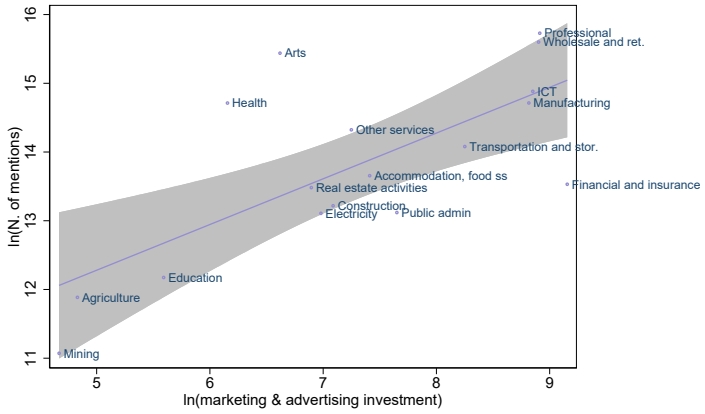
3.5.5 Aboutnovation as a Proxy for Intangible Investment

The aim of building alternative or proxy indicators of innovation or intangible assets is to find reliable solutions to the myriad of problems that traditional sources entail. However, it is important to find some validation that these indicators correlate somehow with such traditional sources. For this reason, in this subsection, we explore such a correlation.

The EUKLEMS – INTANProd database provides detailed data on intangible investments for 27 EU Member States, the US, Japan, and the United Kingdom, across 40 industries (although coverage may vary over time and across countries), 23 industry aggregates, over the period 1995-2020. The database includes information on critical variables for studying productivity, including output, intermediate inputs, gross value added, employment, compensation of employees, and iThis dataset contrastsoss tangible and intangible assets. We use this dataset to contrast our aboutnovation measures with marketing and advertising investments. Does aboutnovation reveal accurate patterns of companies' investment in social media management?

We argue that the type of intangibles we capture with aboutnovation variables is that that reveals how effectively companies engage with customers. We suggest that if a firm has successfully increased such engagement, it is due in part to its efforts in investment in marketing and advertising. Figure 58 reveals a positive, significant correlation between the KLEMS data on advertising investment and the number of mentions for each industry sector.[107]

FIGURE 58
Correlation between investment data and Aboutnovation
- United Kingdom -



Note: The y-axis shows firms' total number of mentions in the UK (expressed as natural logarithm) by industry sector between 2017 and 2021. The x-axis plots the total intangibles investments (expressed as natural logarithm) corresponding to the total investment in advertising and marketing from 2015 to 2017, according to EUKLEMS, by industry sector. The correlation is at the sector level according to the EU-KLEMS category of NACE Rev. 2 codes classification https://euklems.eu/archive-history/query-archive/ and https://euklems-intanprod-llee.luiss.it/query-data/. The previous EU KLEMS database up to the 2017 release (revised July 2018) can be accessed via http://www.euklems.net. The EUKLEMS – Release 2019 is available via Archive and History at https://euklems.eu/. The new EU KLEMS – Release 2021 is available at https://euklems-intanprod-llee.luiss.it/.

Professional services, wholesale trade, ICT, manufacturing, and Finance are at the top right corner of the plot. ICT, finance, and professional services are expected to have a substantial presence in social media. They invest substantially to do so, more than sectors such as mining or agriculture.

Manufacturing firms often produce tangible consumer products that individuals widely use and consume. These products, such as electronics, automobiles, or consumer goods, directly impact people's daily lives. As a result, people are more likely to engage in conversations, share experiences, and express opinions about these products on social media platforms like Twitter. In contrast, agricultural or mining firms may produce goods that are less directly connected to consumers' everyday experiences, resulting in relatively fewer conversations or mentions on social media.

Manufacturing firms often invest heavily in marketing and branding initiatives to promote their products and engage with their target audience. These efforts can include social media campaigns, influencer collaborations, and interactive content, encouraging people to talk about and share their products on platforms like Twitter. Agricultural or mining firms, on the other hand, may have less consumer-focused marketing strategies, as their target audience is often business-to-business (B2B) rather than individual consumers.

Furthermore, certain manufacturing products, such as smartphones, gaming consoles, or fashion items, tend to have a broader appeal and are often associated with trends or popular culture. People are more likely to discuss or share their experiences with these products on social media due to their popularity and the desire to be part of ongoing conversations. While crucial for various industries, agricultural or mining products may not have the same widespread popularity or trendiness among the general public, leading to fewer tweets about them. For these reasons, companies in agriculture or mining do not see the value in investing in engaging with customers via social media.

Likewise, social media platforms like Twitter are more commonly used in urban areas and by individuals with internet access. Manufacturing firms often produce products that cater to urban consumers, who are more likely to have access to the internet and engage on social media platforms. Conversely, agricultural or mining firms are typically associated with rural or resource-intensive areas, where internet penetration may be lower, resulting in fewer conversations or mentions on social media.

Finally, certain manufacturing products can become part of people's lifestyles or personal identities, leading to more discussions and mentions on social media. For example, people may tweet about their favorite fashion brands, tech gadgets, or fitness equipment to express their personal preferences or showcase their lifestyle choices. Agricultural or mining products may have less of a direct impact on personal identity or lifestyle, which can contribute to fewer tweets or discussions related to these sectors.

It's important to note that these reasons are general observations, and there can be exceptions and variations depending on specific industries, products, or regions. Additionally, the prominence of social media platforms and the patterns of online conversations can evolve, influencing the dynamics of discussions about different industries over time.

3.6 Conclusion

We have developed a range of indicators based on social media posts for a sample of firms in the U.K. and Latin America. Our primary objective was to provide more readily available but valid data sources to capture complex and hard-to-measure drivers of success, such as innovation and intangible investments.

We considered two types of tweets, those "by" and "about" firms for both regions, but we primarily focused on the "about" tweets for the UK. We also explored different indicators that a priori could be expected to be proxies for innovation, investment in various forms of intangibles, corporate culture, and technology adoption. We experimented with different conceptual approaches. In particular, we attempted to develop indicators independent of specific contexts such as the industry sector – they are only based on the structure found in the text, something we have dubbed as "Semantic" –time or language while also working with different kinds of dictionaries. In particular, we developed a measure called "Twitnovation" which derives an index of novelty from social media posts.

By geometrically discounting how much a particular word has been used, we calculate an index called Twox for every word. Once a word first emerges in the whole corpus of Twitter posts we have, this index has a number of one, and it is elevated to the power of Q, or how many times that word has been mentioned by a particular point in time. The more it has been mentioned the word counts less as something very new. The lower the depreciation factor considered, the faster it takes for innovations to wear off.

We also have Twox success. Perhaps there are a lot of words that people say but it never catches on so this is not really a great new thing but if it does, then it sends sort of a signal of innovation diffusion, or the concept of how new ideas are spread and adopted by others. Thus, we argue that by multiplying Twox by the number of times the word or string has been mentioned after it was mentioned the first time, we would avoid potential mistakes or typos, since if they were, they could not have become that popular and survive in time.

We can compute these indicators in many specifications. Like, how many times a word has ever been said up to the present time or using specific time windows. For example, how much this has been mentioned from 1 hour of its first emergence, or in a year.

We also derived measures called "Aboutnovation" which derives an index of customer engagement and constitutes a proxy of such an important intangible asset for firms. For instance, we account for the number of times a firm has been mentioned, which indicates awareness about the company. Also, we include the rate of response, which indicates whether a Twitter user who mentioned the firm was ever addressed in any shape or form. Like via reply, re-tweet or quotes. This would be expected to indicate a deeper level of engagement of firms with their customers.

We find strong evidence that social media posts from firms have information that captures the drivers of the success or failure of a firm. We account for 20 percent of the variation in labor productivity within sector × country cells. Indeed, Twitnovation – a straightforward measure based on the novelty of words in tweets – performs remarkably well. However, Aboutnovation – a simple count of the number of times a firm has been tweeted about - seems to be an even stronger indicator of a firm's performance, accounting for up to 63 percent of the overall variation.

This is not entirely surprising in some ways: most celebrities, influencers, and thought leaders are also the ones who are most likely being talked about. And the same is possibly true for firms. However, some might have had entirely different priors. For instance, we might have suspected that a large part of tweeting about firms has to do with customers complaining about some product or service, so that could expect a negative relationship.

The beauty of working with tweets is that there is almost always more information to explore what is going on. For this reason, we explored which type of mentions drive this result.

In further research, we would explore the type of customer tweeting about firms. Is it, for example, customers celebrating a firm's products, or is it about complaints that the firm manages to address successfully? Perhaps having a lot of overly critical customers is what makes a firm perform well. Or is the result driven by something entirely different? Maybe a firm's employees are mainly responsible for Aboutnovation. What makes a firm productive is having employees passionate about their work and tweeting about it on social media.

Developing social media data for this purpose is a work in progress and, at this stage, faces various weaknesses. Besides the small sample for Latin America, there is a selection issue resulting from focusing primarily on firms that tweet, especially for the Twitnovationvation measure. Also, Twitnovation, as a proxy of novelty in tweets, could be criticized as an accurate measure of firm innovation. For instance, we do not include in this analysis Twox Success "weighted" by how many firms are tweeting the new word at the same time, not just after the word was mentioned the first time. Twox Success as it is now may be not enough to filter out typos, and truly valuable information.

However, this early exercise would suggest that it can be fruitful. Immediate next steps include sampling further data, in particular, via the Twitter API to sample "about" tweets. A critical exercise will explore how the various measures perform out of the sample. Furthermore, having a complete balance sheet and performance data would be highly valuable. In some cases, it might be possible to derive this from official data sources rather than rely on commercial resources such as Orbis.

Yet, more importantly, we want to improve our Twinotvation index by employing ChatGPT more effectively rather than simply computing sentiment categories. ChatGPT is a model trained to follow the instructions in a prompt and provide a detailed response using artificial intelligence (OpenAI, 2023). For instance, we can ask Chat GPT if a tweet is considered innovative (according to a definition of innovation that we give it) for a particular period in time, and return a response on a scale (e.g., from 0 to 5, high vs. low, etc.). We could either make prompts at the tweet level or at the firm-year level corpus of tweets (i.e. combining all tweets by the company in that year as a single document). We have to be mindful of tweet types. Original tweets by a firm can reflect the firm's innovation capacity, retweets probably do not. In such a way, we can use a much richer corpus of text to find words or combinations of words truly novel and representative of the firm's actual innovation level.

We also look forward to relating Twitnovation measures with marketing and advertising expenditures, as with Aboutnovation. Finding significant relationships could reveal that efforts in marketing investments effectively result in more innovation for the firm and that Twox measures nicely track firms' successful efforts to attract valuable content and interactions.

Marketing spending may drive innovation in firms through three primary channels. First, it promotes customer-centricity by gathering feedback and market research, informing innovative product development. Second, it builds brand recognition and resources, with revenue from marketing supporting innovation investment. Additionally, cross-functional collaboration, often encouraged by marketing, fosters an innovation culture. Finally, marketing drives market growth and product adoption by differentiating products, promoting their adoption, and supporting innovation during market expansion. These pillars highlight how marketing expenditure sparks innovation in firms.

Such a measure could allow for testing different marketing expenditure scenarios depending on expected outcomes in innovation, and budgeting for intangibles of this sort could be more systematic for companies.

4 Conclusions

This thesis argues that the observed co-existing macro trends over the last two to three decades in developed countries, including the emergence of superstar firms that outpace their peers in growth, rising measured markups, decreased firm entry, reduced labor and investment shares, sluggish productivity growth, and heightened market concentration, are not primarily driven by the exercise of market power through anticompetitive practices.

I suggest that these trends are more likely to emerge from a complete economic model transformation. The rising prevalence of intangible investments, globalization, and the ICT revolution transformed how firms can reach dominant positions as superstar firms.

This thesis partially addresses this hypothesis using novel empirical strategies, data sources, and theoretical propositions to demonstrate that declining business dynamics and supposedly increased market power are tightly related to the rise of the intangible economy.

Through the first chapter, ***Business Dynamics and Intangible Investments***, I aimed to answer whether the emergence of intangible assets can be behind declining business dynamics in the U.K. I created an original longitudinal database using businesses' official register data for the 1998-2018 period that would allow me to connect business dynamics indicators with data on intangible intensity. The empirical strategy allowed us to identify firms' groups behind the slowing dynamics trends like high-growth young SMEs and mature and large firms in most ICT and intangible sectors losing steam in employment participation and growth.

The chapter's major weakness is not connecting the business dynamics with intangible data at the firm level in the empirical exercises. However, the stylized facts, regression analysis, and theoretical model looked to supply this deficiency with intra-group analysis. Put differently, the chapter attempted to identify which groups of firms were behind the trends. Prior evidence has identified that aggregate trends are explained by what is happening to particular groups of firms. In the chapter, the proposed theoretical model inspired by the work of Hopenhayn (1992); Hopenhayn and Rogerson (1993) and Bartelsman et al. (2013) was also used to conduct comparative statics analyses on these groups.

We compared how dynamics would be altered for firms of different sizes and employment growth rates, indicating that the more intangible intensity a sector or a firm is, the less elastic it is to respond to shocks via employment expansion or contraction. Primarily, despite a remarkable business dynamics recovery that set the UK apart from other countries after the great recession of 2009, we found the UK shared with other nations

the fact that high-growth firms have become, on average, fewer, smaller, and older, particularly in sectors such as information, content and media, and in some high-ICT-manufacturing sectors.

A crucial aspect we wanted to argue was that the economic properties' of intangibles have played a substantial role in firms' employment or entry and exit decisions. Although the chapter does not propose direct measures to test this, the discussion, stylized facts, and regression findings suggest that intangible properties could be valid mechanisms for what is observed in the trends. In particular, a rising intangible economy may trigger labor frictions, as it tends to demand more skilled types of labor, and to incentive substitute more unskilled labor when shocks occur. Given the prevalence of skilled relative to unskilled labor, it is logical that aggregate labor shares may suffer as a product of more intangibles.

Is it that bad that high-growth firms in terms of employment are fewer or becoming smaller and older? I would say not really. Is it bad that mature and older firms, often the ones employing most people, increasingly employ fewer people? Yes, perhaps. However, the solution is not to demonize these patterns but rather to establish policies that deal with the implications of intangibles properly; instead, academics and policymakers should be more concerned with understanding the phenomenon and creating better metrics to target specific firms for extra support. For instance, providing more financing options to very high-intangible young SMEs, or ensuring that more people are qualified with relevant skills in a more intangible economy.

The U.K. case is interesting to analyze since additional labor-related frictions could impede keeping up with the rising intangible investments, like excessive regulation of highly skilled immigrants and insufficient investment in a skilled and creative workforce. This could be related to the decrease in intangible investments experienced in recent years (out of our sample period).

Exactly, beyond the discussion, the chapter contributed to improving data-based evidence on this controversial topic by implementing a well-regarded indicators standardization routine (the OECD's business dynamics program, DynEmp3) that could facilitate comparative studies. It also facilitated creating a replicable list of assumptions to build a longitudinal database using business register data in the UK so researchers in the future do not have to be limited to the U.K. Office for National Statistics official but not necessarily timely data releases. Furthermore, the theoretical model proposed is a practical application of prior authors' work that enlightens the discussion, and puts in the jargon of the debate the word "elasticity" which is a more comprehensive and flexible concept to convey how new trends, or shocks, in technology and assets, make macro trends to "bounce".

The second chapter, ***Market Power and Technology***, examined whether markups, the preferred market power signal, have increased in recent decades using data from the U.S., Europe, and other countries.

By applying an alternative method to the predominant of De Loecker and Warzynski (2012); De Loecker et al. (2020) to estimate firm-level markups, we showed that markups have not changed their trends much in recent years. We found that the numerator and the denominator of the ratio used to compute markups have decreased at similar rates in recent years. The output elasticities have declined at the same rate as the share of expenditures in flexible production inputs.

We argued that what has reduced elasticities has been the arrival of increasing returns to scale and reducing marginal costs is a product of combining the rising of the intangible economy, ICT progress, and globalization. We show empirical evidence that production technologies have changed in recent decades to prove this. Consequently, markups have remained relatively stable or even declined in certain sectors, such as information and manufacturing, which have become more intangible in recent years.

Besides this main finding, the chapter contributes in other ways. It raises awareness about the flaws of a methodology that multiple policy and academic outlets accept as a fact. It applies methodologies that respond more to the needs and realities of recent economic trends, with assumptions beyond perfect competition. Eventually, it leaves a ready-to-use database in ORBIS able to be connected with any other variables on innovation and tangible investments from ORBIS itself and the complimentary database of ORBIS IP.

The last point suggests an interesting venue to explore in future research. To explore more directly if intangibles impact the behavior of markups, this could be a fruitful empirical journey. ORBIS IP is a database used to access and analyze information related to intellectual property (IP) and innovation for various firms. It provides data on patents, trademarks, trade secrets, and other forms of IP registered by companies. ORBIS IP can be used to measure intangible assets and innovation of firms by identifying the IP assets a company owns. By analyzing the historical IP filings and trends of a company, researchers can gauge its innovation and R&D efforts over time in relationship with the patterns in markups.

ORBIS IP also permits the comparison of IP portfolios between companies, enabling researchers to study the relative innovation capabilities of firms in the same industry. Also, by combining IP data with that from Zephyr, another dataset owned by the same data company and linkable to ORBIS, researchers can analyze markups, considering mergers and acquisitions. This would also allow us to dig deeper into potential IP disputes or patent infringement lawsuits, to explore further the economic property of "spillovers" of intangible assets.

Demonstrating that markup trends are related to technological changes and not primarily for anti-competitive practices per se is key for making the controversial debate less aggressive towards a particular group of firms and, instead, for making it more constructive to redirect policy efforts to supply the needs where is relevant. Furthermore, emphasizes that methodologies should be addressed with a microscope before making impactful policy decisions.

The third and final chapter, ***Twitnovation and Aboutnovation: Social Media Indicators to Measure Innovation and Intangible Investments***, proposed novel empirical techniques to measure the hard-to-measure but essential drivers of firms' success, such as intangible assets and innovation. We took advantage of higher-frequency and readily available data, such as social media from and about companies like Twitter. We build two types of measures, semantic measures, which are not necessarily context-specific but can extract meaningful information on how a text is structured. For instance, our Twitnovation (*twox*) measure, or the primary Aboutnovation measure, proxies the level of investment in the intangible investment of customer engagement by counting the number of times other Twitter users have mentioned a firm.

We contrast those with dictionary-based measures. We are interested in particular technologies or intangible investment types, and we wanted to identify the words that relate to those in the tweets; for instance, a firm talking about AI, R&D, Bluetooth, or Research is indicative of the level of "investment" or at least of awareness of that firm about these assets, which ultimately says something good about the business.

In contrast to the two first chapters, this one was more concerned with contributing to the literature with alternative approaches to improve the measurements of such concepts, which ultimately is one of the most meaningful obstacles preventing consensus today on many of the topics discussed in the thesis.

In alignment with the suggested connection between markup estimates and ORBIS IP data, the resulting dataset of this chapter could be connected with markup data. The primary source of Twitter handles or users so we could know the firm's social media handlers to collect their posts was extracted from ORBIS. This would allow us to explore the links suggested in this thesis even more formally. Indeed, entry and exit firms' information could be inferred using "ownership" information from ORBIS data. This would allow the creation of an original dataset that could simultaneously test the business-dynamics-markups-intangibles link, which this thesis partially treats in each chapter.

In summary, this thesis' main contributions were empirical, applying recent methodological approaches, creative data sources, and strategies to link it. The thesis also contributed with practical, theoretical propositions based on prior work. Building on prior theoretical models in economics is crucial for leveraging the cumulative knowledge

of the field. It promotes efficiency and empirical relevance and helps refine and test the robustness of existing models, ultimately fostering intellectual progress and innovation.

The common ground of all three chapters centres around the properties of intangible assets under the theoretical framework of the 4S as called by Haskel and Westlake (2018). For instance, intangibles like software can be highly scalable. Unlike physical products, the software can be replicated and distributed without high additional cost. Once the initial development is done, creating copies of the software for additional users or clients is inexpensive and doesn't require extra resources. Thus, superstars who own or develop proprietary software, for example, can sell it at no additional cost. At the same time, the network effects play their part via established user networks and higher switching costs, leading to organic dominance without inherent anticompetitiveness.

Similarly, more intangibles can be a rich innovation source, as demonstrated in the last chapter. "Aboutnovation" not only positively impacted firms' labor productivity but also seemed to have a positive relationship with our innovation proxies of "Twinotvation". Intangibles can trigger innovation even via the most indirect ways one might think of, such as via effective marketing or customer engagement, which reflects later in better performance.

Like scalability, intangibles' economic properties of spillovers, synergies, and sunkness not only reduce marginal costs and enhance economies of scale but, more generally, entirely alter the incentives of players in the market that in one or another way may tend to organically favor highly intangible, productive, cash-rich companies, like the Superstars. Thus, this thesis suggests to look at these issues from a more laudable point of view.

Superstar firms' productivity and innovative prowess should not be chastised by policy but rather seen as assets that can strengthen economies. These firms have the potential to drive innovation and boost global competitiveness without harming consumers with higher prices. Their economies of scale result in cost efficiencies and can lead to job creation if labor markets can be professionalized and adapted to technological progress demands. Moreover, superstar firms' success often spills into related sectors, fostering generalized innovation and progress. They excel in managing intangible assets crucial in the modern economy, such as intellectual property and data, which can teach laggard firms how to do it. They also attract investment, benefiting smaller companies within their ecosystem. While regulation is essential to ensure fair competition and protect consumers, policymakers should seek sounder empirical methodologies and data-based evidence to inform their views and decisions. Better evidence could instead allow for understanding the best ways to harness Superstar firms' potential to encourage economic growth, innovation, and responsible business practices rather than preemptively and extremely condemning their success.

References

Aaker, D. A. (1991). *Managing Brand Equity*. New York: The Free Press.

Acemoglu, D., U. Akcigit, H. Alp, N. Bloom, and W. Kerr (2018). Innovation, Reallocation, and Growth . *American Economic Review* (November).

Acemoglu, D. and D. Autor (2011). Chapter 12 - skills, tasks and technologies: Implications for employment and earnings. Volume 4 of *Handbook of Labor Economics*, pp. 1043–1171. Elsevier.

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification Properties of Recent Production Function Estimators. *Econometrica 83*(6), 2411–2451.

Affeldt, P., T. Duso, K. Gugler, and J. Piechucka (2021). Market Concentration in Europe: Evidence from Antitrust Markets. *SSRN Electronic Journal*.

Aghion, P. and U. Akcigit (2015). Innovation and Growth: The Schumpeterian Perspective.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: an Inverted-U Relationship. *The Quarterly Journal of Economics 120*(2), 701–728.

Akcigit, U. and S. Ates (2018). What Happened to the U . S . Business Dynamism ? Diagnosis of Ten Facts through the Growth Theory.

Akcigit, U. and S. Ates (2019). What Happened to U.S. Business Dynamism?

Akerman, A., E. Leuven, and M. Mogstad (2022, January). Information frictions, internet, and the relationship between distance and trade. *American Economic Journal: Applied Economics 14*(1), 133–63.

Altomonte, C., D. Favoino, M. Morlacco, and T. Sonno (2020). Markups , Intangible Capital , and Heterogeneous Financial Frictions. *Working Paper 822390*(822390), 1–62.

Amabile, T. M. (1988). A model of creativity and innovation in organizations. *10*, 123–167.

Andrews, D., C. Criscuolo, and P. N. Gal (2015). Frontier Firms, Technology Diffusion and Public Policy. *OECD: The Future of Productivity: Main Background Papers.*.

Andrews, D., C. Criscuolo, and P. N. Gal (2016a). The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy. *OECD Publishing 5*(November), 1–76.

Andrews, D., C. Criscuolo, and P. N. Gal (2016b). The Global Productivity Slowdown, Technology Divergence and Public Policy: A Firm Level Perspective. *Brookings Institution Hutchins Center Working Paper* (September), 1–50.

Anyadike-Danes, M., K. Bonner, M. Hart, and C. Mason (2009). Measuring business growth: high growth firms and their contribution to employment in the UK. *London: NESTA* (October), 1–52.

Araujo, G. and J. Hambur (2018). Which Firms Get Credit? Evidence from Firm-level Data?

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2017). The Fall of the Labor Share and the Rise of Superstar Firms.

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2020). The Fall of the Labor Share and the Rise of the Superstar Firms. *The Quarterly Journal of Economics* (February), 645–709.

Ayyagari, M., A. Demirguc-kunt, and V. Maksimovic (2022). The Rise of Star Firms: Intangible Capital and Competition.

Azoulay, P., B. F. Jones, J. D. Kim, and J. Miranda (2020). Age and High-Growth Entrepreneurship. *American Economic Review: Insights 2*(1), 65–82.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. John Wiley and Sons, Ltd.

Bank of England (2022). Financial Stability Report, December 2022. Technical report, Bank of England, London.

Baqaee, D. R. and E. Farhi (2020a). Online Appendix to Productivity and Misallocation in General Equilibrium. *Quarterly Journal of Economics 135*(1), 105–163.

Baqaee, D. R. and E. Farhi (2020b). Productivity and Misallocation in General Equilibrium. *Quarterly Journal of Economics 135*(1), 105–163.

Barkai, S. (2016). Declining Labor and Capital Shares.

Barkai, S. (2020). Declining Labor and Capital Shares. *Journal of Finance 75*(5), 2421–2463.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013). Cross-country differences in productivity: The role of allocation and selection. *American Economic Review 103*(1), 305–334.

Bartelsman, E. J., P. A. Gautier, and J. De Wind (2016). Employment Protection, Technology Choice, and Worker Allocation. *International Economic Review 57*(3), 787–826.

Bello, H. (2021). sentiment-analysis-spanish.

Belso-Martinez, J. A., F. X. Molina-Morales, and F. Mas-Verdu (2013). Combining effects of internal resources, entrepreneur characteristics and kis on new firms. *Journal of Business Research 66*(10), 2079–2089. Strategic Thinking in Marketing Strategic Management in Latin America Corporate Social Responsibility and Irresponsibility Managing Global Innovation and Knowledge.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Bessen, J. (2017). Information Technology and Industry Concentration. *Boston University School of Law Law & Economics Paper 1*(4), 21–37.

Bessen, J. (2022). *The New Goliaths How Corporations Use Software to Dominate Industries, Kill Innovation, and Undermine Regulation*. Yale University Press.

Bijnens, G. and J. Konings (2017). Where has the Belgian Business Dynamism gone? The decline in high growth (small) firms.

Bijnens, G. and J. Konings (2018a). Declining Business Dynamism.

Bijnens, G. and J. Konings (2018b). *Declining business dynamism in Belgium*. Number November. Small Business Economics.

Bloom, N., C. Genakos, R. Sadun, and J. Reenen (2012, 02). Management practices across firms and countries. *Academy of Management Perspectives 26*.

Bloom, N., T. A. Hassan, A. Kalyani, J. Lerner, and A. Tahoun (2020). The Geography of New Technologies. *Institute for New Economic Thinking Working Paper Series*, 1–45.

Bloom, N. and J. Van Reenen (2007, 11). Measuring and Explaining Management Practices Across Firms and Countries*. *The Quarterly Journal of Economics 122*(4), 1351–1408.

Bloom, N. and J. van Reenen (2007, 11). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics 122*.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2020, April). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. Working Paper 27002, National Bureau of Economic Research.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics 121*, 1–14.

Bontis, N. and J. Fitz-enz (2002). Intellectual capital roi: A causal map of human capital antecedents and consequents. *Journal of Intellectual Capital 3*(3), 223–247.

Bravo-Biosca, A. (2016a). Firm growth dynamics across countries: Evidence from a new database.

Bravo-Biosca, A. (2016b). Firm growth dynamics across countries: Evidence from a new database.

Bravo-Biosca, A., C. Criscuolo, and C. Menon (2016). What drives the dynamics of business growth? *Economic Policy*, pp. 703–742.

Brynjolfsson, E., D. Rock, and C. Syverson (2017). Artificial intellegience and the modern productivity paradox.

Bureau van Dijk (2007). Orbis Internet User Guide.

Buss, T. F. (2002, feb). Emerging High-Growth Firms and Economic Development Policy. *Economic Development Quarterly 16*(1), 17–19.

Byrne, D. M. *FEDS Notes 2015-06-02* (2015-06-02).

Caballero, R. and M. L. Hammour (2005). The Cost of Recessions Revisited: A Reverse-Liquidationist View . *Review of Economic Studies 72*(2), 313–341.

Cabral, L. (2018). Standing on the Shoulders of Dwarfs: Dominant Firms and Innovation Incentives.

CalChamber. California Chamber of Commerce. (2022). Joint Liability. Legislature Shifts Workplace Labor Enforcement to Private Sector by Expanding Joint Liability.

Calligaris, S., C. Criscuolo, and L. Marcolin (2018). Mark-ups in the digital era.

Calvino, F. and C. Criscuolo (2019). Business Dynamics and digitalization. *OECD, Technology and Industry Policy Papers* (62), 59.

Calvino, F., C. Criscuolo, and C. Menon (2015). The Future of Productivity: Main Background Papers. Cross-country evidence on start-up dynamics.

Calvino, F., C. Criscuolo, and C. Menon (2018). The employment dynamics Stata program dynemp3. Technical Report January.

Calvino, F., C. Criscuolo, and R. Verlhac (2020). Declining business dynamism: Structural and policy determinants. *OECD Publishing November*(94).

Chesbrough, H. W. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston, MA: Harvard Business Press.

Chiavari, A. and S. Goraya (2021). The Rise of Intangible Capital and the Macroeconomic Implications. Technical report.

Christopher J. Nekarda, V. A. R. (2020, 12). The cyclical behavior of the price-cost markup. *Journal of Money, Credit and Banking 52*, 319–353.

Cirera, X. and S. Muzi (2020). Measuring innovation using firm-level surveys: Evidence from developing countries . *Research Policy 49*(3), S0048733319302306.

CMA (2022). The State of UK Competition.

Competition and Markets Authority (2022). The state of UK Competition.

Corrado, C., C. Criscuolo, J. Haskel, A. Himbert, and C. Jona-Lasinio (2021). New Evidence on Intangibles, Diffusion and Productivity. *OECD Science, Technology and Industry Working Papers 10*.

Corrado, C., J. Haltiwanger, and D. Sichel (2005). *Measuring Capital and Technology: An Expanded Framework*. Number August. University of Chicago Press.

Corrado, C., J. Haskel, and C. Jona-Lasinio (2017). Knowledge Spillovers, ICT and Productivity Growth† . *Oxford Bulletin of Economics and Statistics 79*(August 2017), 592–618.

Corrado, C., J. Haskel, and C. Jona-Lasinio (2019). Productivity Growth, Capital Reallocation and the Financial Crisis: Evidence from Europe and the US. *Journal of Macroeconomics 61*(March 2018), 103120.

Corrado, C., J. Haskel, and C. Jona-Lasinio (2021). Artificial intelligence and productivity: An intangible assets approach. *Oxford Review of Economic Policy 37*(3), 435–458.

Corrado, C., J. Haskel, C. Jona-Lasinio, and M. Iommi (2016). Intangible investment in the EU and US before and since the Great Recession and its contribution to productivity growth. Technical report.

Corrado, C., J. Haskel, C. Jona-lasinio, and M. Iommi (2022). Intangible Capital and Modern Economies. *Journal of Economic Perspectives 36*(3), 3–28.

Corrado, C., J. Haskel, C. Jona-lasinio, and Rincon (2019). Notes on Productivity leaders, laggards and intangibles.

Corrado, C., C. R. Hulten, and D. Sichel (2009). Intangible capital and U.S. Economic Growth. *Review of Income and Wealth 55*(3).

Criscuolo, C. and R. Martin (2019). How to measure superstar firms ? [Work in progress].

Crouzet, N. and J. Eberly (2018). Intangibles, Investment, and Efficiency. *AEA Papers and Proceedings 108*, 426–431.

Crouzet, N. and J. C. Eberly (2019). Understanding Weak Capital Investment. *NBER Working Paper May*(25869), 1–65.

Csikszentmihalyi, M. (1996). *Creativity: flow and the psychology of discovery and invention*. HarperCollins Publishers.

Dao, M., M. Das, Z. Koczan, and W. LianPublication (2017, 7). Why is labor receiving a smaller share of global income? theory and empirical evidence. *Working Paper No. 2017/169 169*.

Davis, S. J. and J. Haltiwanger (2014). Labor Market Fluidity and Economic Performance.

De Loecker, J. and J. Eeckhout (2018). Global Market Power.

De Loecker, J., J. Eeckhout, and G. Unger (2018). The Rise of Market Power.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The Rise of Market Power and the Macroeconomic Implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J. and F. Warzynski (2012, May). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

de Ridder, M. (2020). Market Power and Innovation in the Intangible Economy. *LSE Research Online Documents on Economics* (September).

De Ridder, M., B. Grassi, and G. Morzenti (2021). The Hitchhiker's Guide to Markup Estimation. *Working Paper*.

Decker, R., J. Haltiwanger, J. Jarmin, and J. Miranda (2019). Changes Business Dynamism and Productivity: Shocks vs. Responsiveness. In *Micro-data perspectives on the productivity puzzle: evidence for policy*, London.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014, September). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives 28*(3), 3–24.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2017). The Role of Entrepreneurship in US Job Creation and Economic Dynamism. *The Journal of Economic Perspectives 28*(3), 3–24.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016a). Declining Business Dynamism : Implications for Productivity ? (September), 1–69.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016b, may). Declining business dynamism: What we know and the way forward. In *American Economic Review*, Volume 106, pp. 203–207.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016c). Where has all the skewness gone? The decline in high-growth (young) firms in the U.S. *European Economic Review 86*, 4–23.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2020). Changing Business Dynamism and Productivity: Shocks versus Responsiveness. *American Economic Review 110*(12), 3952–3990.

Demsetz, H. (1973). Industry Structure, Market Rivalry, and Public Policy. *The Journal of Law Economics 16*(1), 1–9.

Diez, F., D. Leigh, and S. Tambunlertchai (2018). Global Market Power and its Macroeconomic Implications. *IMF Working Papers 18*(137), 1.

Doraszelski, U. and J. Jaumandreu (2019). Using Cost Minimization to Estimate Markups.

Doraszelski, U. and J. Jaumandreu (2021). Reexamining the De Loecker & Warzynski (2012) Method for Estimating Markups. *Ssrn* (2012), 1–28.

Dottling, R., G. Gutierrez, and T. Philippon (2018). Is There an Investment Gap in Advanced Economies? If So, Why? *SSRN Electronic Journal*, 1–44.

Eberly, J., J. Haskel, and P. Mizen (2021). "POTENTIAL CAPITAL", WORKING FROM HOME, AND ECONOMIC RESILIENCE. *Working Paper 29431*, 7–12.

Eggertsson, G. B., J. A. Robbins, and E. G. Wold (2021). Kaldor and piketty's facts: The rise of monopoly power in the united states. *Journal of Monetary Economics 124*, S19–S38. The Real Interest Rate and the MarginalProduct of Capital in the XXIst CenturyOctober 15-16, 2020.

Eisfeldt, A. L. and D. Papanikolaou (2014). The Value and Ownership of Intangible Capital. *American Economic Review 104*(5), 189–94.

Eslava, M. and J. Haltiwanger (2018). The Life-Cycle Growth of Plants: The Role of Productivity, Demand and Distortions . In *Ssrn*, pp. 1–53.

Eslava, M. and J. Haltiwanger (2019). Web Appendix for : The Life-cycle Growth of Plants : The Role of Productivity , Demand and Wedges . Appendix A : measuring exact price indices.

Faggio, G., K. Salvanes, and J. Van Reenen (2007). The Evolution of Inequality in Productivity and Wages: Panel Data Evidence .

Farquhar, P. H. (1989). Managing brand equity. *Marketing Research 1*, 24–33.

Felix Roth, A. S. and C. Rammer (2023). The role of intangibles in firm-level productivity – evidence from germany. *Industry and Innovation 30*(2), 263–285.

for National Statistics, O. (2006). Business Structure Database User Guide. Version 1. Technical Report May.

Forlani, E., R. Martin, G. Mion, and M. Muûls (2016). Unraveling Firms : Demand , Productivity and Markups Heterogeneity Emanuele Forlani.

Forlani, E., R. Martin, G. Mion, and M. Muûls (2023). Unraveling Firms: Demand, Productivity and Markups Heterogeneity. *The Economic Journal* (uead031).

Fukao, K., T. Miyagawa, K. Mukai, Y. Shinoda, and K. Tonogi (2009, September). Intangible investment in japan: measurement and contribution to economic growth. *The Review of Income and Wealth 55*(3), 717–736.

Furman, J. (2015). Business Investment in the United States: Facts, Explanations, Puzzles, and Policies. *Council of Economic Advisers. Remarks at Progressive Policy Institute*, 1–21.

García-Perea, P., A. Lacuesta, and P. Roldan-Blanco (2021). Markups and cost structure: Small spanish firms during the great recession. *Journal of Economic Behavior Organization 192*, 137–158.

Go, A., R. Bhayani, and L. Huang (2009, 11). Twitter sentiment classification using distant supervision. *University of Standford Press 1.*

Goldschlag, N. and J. Miranda (2020). Business Dynamics Statistics of High Tech Industries. *Journal of Economics and Management Strategy 29*(1), 3–30.

Grullon, G., Y. Larkin, and R. Michaely (2019). Are US industries becoming more concentrated? *Review of Finance 23*(4), 697–743.

Guiso, L., P. Sapienza, and L. Zingales (2015). The value of corporate culture. *Journal of Financial Economics 117*(1), 60–76. NBER Conference on the Causes and Consequences of Corporate Culture.

Gutiérrez, G. and T. Philippon (2017a). Declining Competition and Investment in the U.S.

Gutiérrez, G. and T. Philippon (2017b). Investmentless Growth: An Empirical Investigation. In *Brookings Papers on Economic Activity*, pp. 89–169.

Gutiérrez, G. and T. Philippon (2018). How EU Markets Became More Competitive Than US Markets: A Study of Institutional Drift.

Hahn, J., J. Hausman, and G. Kuersteiner (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. *Journal of Econometrics 140*(2), 574–617.

Hall, R. E. (1988). The Relation between Price and Marginal Cost in U. S. Industry. *Journal of Political Economy 96*(5), 921–947.

Hall, R. E. (2018). New Evidence on the Markup of Prices over Marginal Costs and the Role of Mega-Firms in the US Economy.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013). Who creates jobs? Small versus large versus young. *The review of economics and statistics XCV*(2), 347–360.

Hart, D. (2008). The Politics of 'Entrepreneurial' Economic Development Policy in the U.S. States. *Review of Policy Research 25*(2), 141–68.

Haskel, J. (2020). Monetary policy in the intangible economy - speech by Jonathan Haskel. Number February, London, pp. 1–5.

Haskel, J., G. Awano, M. Franklin, and Z. Kastrinaki (2010, 07). Measuring investment in intangible assets in the uk: Results from a new survey. *Economic and Labour Market Review 4*, 66–71.

Haskel, J. and R. Martin (2002). The UK Manufacturing Productivity Spread.

Haskel, J. and S. Westlake (2018). *Capitalism Without Capital: The rise of the intangible economy*. Princeton, New Jersey: Princeton University Press.

Hazan, E., S. Smit, J. Woetzel, B. Cvetanovski, M. Krishnan, B. Gregg, J. Perrey, and K. Hjartar (2021). Getting tangible about intangibles. *McKinsey Global Institute* (June), 40.

Hopenhayn, H., J. Neira, and R. Singhania (2022, 7). From population growth to firm demographics: Implications for concentration, entrepreneurship and the labor share. *Journal of Money, Credit and Banking 90*, 1879–1914.

Hopenhayn, H. and R. Rogerson (1993). Job Turnover and Policy Evaluation: A General Equilibrium Analysis. *Journal of Political Economy 101*(5), 915–938.

Hopenhayn, H. A. . (1992). Entry , Exit , and firm Dynamics in Long Run Equilibrium. *Econometrica 60*(5), 1127–1150.

Huang, R. and J. R. Ritter (2009). Testing theories of capital structure and estimating the speed of adjustment. *The Journal of Financial and Quantitative Analysis 44*(2), 237–271.

Hutto, C. (2020). vadersentiment.

Hutto, C. and E. Gilbert (2014, 06). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *ighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014. 1.*

Hwang, K., A. Savagar, and J. Kariel (2022). Market power in the UK.

Inaba, T. and M. Squicciarini (2017). ICT: A new taxonomy based on the international patent classification.

Intellectual Property Office (2022). Intangible Assets Intensity of Bottom Growth Firms: A UK Perspective. Technical Report January.

Jo, A. (2020). The Elasticity of Substitution between Clean and Dirty Energy with Technological Bias. *ETH working paper* (October).

Jonke, R. (2012). Managing after-sales services.

Karabarbounis, L. and B. Neiman (2014). The Global Decline of the Labor Share Get access Arrow. *The Quarterly Journal of Economics 129*(1), 61–103.

Karahan, F., B. Pugsley, and A. Şahin (2019, May). Demographic origins of the startup deficit. Working Paper 25874, National Bureau of Economic Research.

Klette, T. and Z. Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics 11*(4), 343–61.

Koujianou Goldberg, P. and R. Hellerstein (2012, 03). A Structural Approach to Identifying the Sources of Local Currency Price Stability. *The Review of Economic Studies 80*(1), 175–210.

Lev, B. (2001, 01). Baruch lev, intangibles: Management, measurement and reporting, bookings institution press, 2001.

Li, B., C. Liu, and S. T. Sun (2021). Do corporate income tax cuts decrease labor share? regression discontinuity evidence from china. *Journal of Development Economics 150*, 102624.

Loecker, J. D. and J. Eeckhout (2021). Global Market Power. pp. 1–34.

Manyika, J., S. Ramaswamy, J. Bughin, J. Woetzel, M. Birshan, and Z. Nagpal (2018). Superstars: the Dynamics of Firms, Sectors, and Cities Leading The Global Economy. *Mckinsey&Company Discussion Paper* (October).

Martin, J. and C. Baybutt (2022, August). The 'f words': Why surveying businesses about intangibles is so hard (escoe dp 2022-20). Technical Report ESCoE DP 2022-20, Economic Statistics Centre of Excellence (ESCoE).

Mason, G., C. Robinson, and C. Rosazza Bondibene (2016). Sources of Labour Productivity Growth at Sector Level in Britain, 1998-2007: a Firm-Level Analysis.

Melitz, M. J. and G. I. Ottaviano (2008). Market Size, Trade, and Productivity. *Review of Economic Studies 75*, 295–316.

Montresor, S. and A. Vezzani (2014). On the rd giants' shoulders: Do fdi help to stand on them? JRC Working Papers on Corporate RD and Innovation 2014-01, Joint Research Centre (Seville site).

Morlacco, M. (2020). Market Power in Input Markets: Theory and Evidence from French Manufacturing. *Working Paper* (October), 1–58.

Nahapiet, J. and S. Ghoshal (1998, 04). Social capital, intellectual capital, and the organizational advantage. *The Academy of Management Review 23*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*(6), 1417–26.

OECD (2021). Bridging the gap in the financing of intangibles to support productivity: Background paper.

Office for National Statistics (2017). UK business; activity, size and location: 2017. Statistical bulletin. Technical report, Office for National Statistics, London.

Office for National Statistics (2018). Business demography , UK : 2017. pp. 1–11.

Office for National Statistics (2021). UK business; activity, size and location: 2021. Statistical bulletin. Technical report, Office for National Statistics, London.

Office for National Statistics (2022). Investment in intangible assets in the UK: 2020. ONS website, article. Technical report, Office for National Statistics, London.

Office for National Statistics (2023a). Labour market overview, UK: April 2023. ONS website, statistical bulletin. Technical report, Office for National Statistics, London.

Office for National Statistics (2023b). UK Innovation Survey (UKIS) 2023. Available at: 10.5255/UKDA-SN-8556-2. Accessed on: October 1 of 2023.

Office for National Statistics, Imperial College London, and UK Data Service. (2016). Investment in Intangible Assets Survey, 2009-2010: Secure Access. [data collection]. 3rd Edition.

OMahony, M. (2020). Intangible Capital, Productivity, and Labor Markets.

OpenAI (2023). ChatGPT: A Large-Scale Generative Model for Natural Language Communication. https://chat.openai.com. Accessed: October 20, 2023.

Organisation for Economics Co-operation and Development (2020). High-Growth SMEs and Employment. Technical report.

Paul, S. (2020). Understanding the global decline in the labor income share. *IZA World of Labor*, 1–9.

Peters, R. H. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics 123*(2), 251–272.

Philippon, T. and G. Gutierrez (2017). Investmentless Growth: An Empirical Investigation. *Brookings Papers on Economic Activity Fall*.

Piketty, T. (2014). *Capital in the Twenty-First Century*. The Belnakp Press of Harvard University Press.

Prashar, N. and M. Hart (2019). Job Creation and Destruction in the UK 1998-2018. pp. 1–11.

Raval, D. (2023a). A flexible cost share approach to markup estimation. *Economics Letters 230*, 111262.

Raval, D. (2023b, 01). Testing the Production Approach to Markup Estimation. *The Review of Economic Studies 90*(5), 2592–2611.

Reshef, A. and G. Santoni (2023). Are your labor shares set in beijing? the view through the lens of global value chains. *European Economic Review 155*, 104459.

Ridder, M. D., B. Grassi, and G. Morzenti (2022). The Hitchhiker's Guide to Markup Estimation. *CEPR Press Discussion Paper*.

Rockhopperhr Law Firm (2023). Employment law and HR glossary.

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review 71*(5), 845–858.

Schoar, A. (2010). The Divide between Subsistence and Transformational Entrepreneurship . *Innovation Policy and the Economy 10*(1), 57–81.

Schumpeter, J. (1942). *Capitalism, Socialism and Democracy.* (2003 ed.). London: Routledge London and New York.

Shane, S. (2010). Why Encouraging More People to Become Entrepreneurs Is Bad Public Policy . *Small Business Economics 33*(2), 141–49.

Stackpole, B. (2023, June). Ideas made to matter, artificial intelligence: For ai in manufacturing, start with data.

Suleyman, G. and Y. Ozkara (2021). *Intangible capital meets skilled labor: The Implications for U.S. Business Dynamism*.

Syverson, C. (2019). Macroeconomics and Market Power: Facts, Potential Explanations and Open Questions. *Brookings Working Paper* (January).

Szalavetz, A. (2017, 03). Intangible investments at multinational companies' manufacturing subsidiaries: do they promote innovation-based upgrading? *Equilibrium 12*, 63.

Tambe, P., L. Hitt, D. Rock, and E. Brynjolfsson (2020, December). Digital Capital and Superstar Firms. NBER Working Papers 28285, National Bureau of Economic Research, Inc.

Traina, J. (2018). Is Aggregate Market Power Increasing? Production Trends Using Financial Statements. *Ssrn* (February), 1–21.

Tucker, C. (2018). Network Effects and Market Power: What Have We Learned in the Last Decade? *Antitrust Spring*, 72–29.

UK Statistics Authority (2020). Assessment of compliance with the Code of Practice for Official Statistics: UK Business Demography Statistics. Technical Report October.

Vaisnore, A. and M. Petraite (2011, 11). Customer involvement into open innovation processes: a conceptual model. *Social Sciences 73*.

van Heuvelen, G. H., L. Bettendorf, and G. Meijerink (2021). Markups in a dual labour market: The case of the netherlands. *International Journal of Industrial Organization 77*, 102762.

Van Reenen, J. (2018). Increasing Differences Between Firms: Market Power and the Macro-Economy.

Velasquez, A. (2023). Production Technology, Market Power, and the Decline of the Labor Share.

Yeh, C., C. Macaluso, and B. Hershbein (2022). Monopsony in the US Labor Market. *American Economic Review 112*(7), 2099–2138.

Zhang, P., M. Ye, D. Simunic, and L. Chu (2019). Market Power and Competition in Audit Markets. *Hawai'i Accounting Research Conference (HARC) University of Hawai'i*.

Zheng, B., M. Zhang, and X. Zhang (2022). The rise of market power and firms' investment: Evidence from China. *Accounting and Finance 62*(5), 4807–4830.

Notes

[1]This chapter is co-authored with Ralf Martin, Imperial College London, and LSE.

[2]This chapter is adapted from the co-authored work "Twitnovation: Can Social Media explain Firm Performance in Latin America?" sponsored by the Inter-American Development Bank, coauthored by me, with Eugenie Dugoua and Ralf Martin.

[3]describes the portion of a company's increased output that cannot be explained by increased capital or labor inputs and thus is considered a measure of operational efficiency

[4](e.g. models with non-homogenous goods permit co-existing high concentration with high competition)

[5]"The DynEmp project presents evidence on the employment dynamics of start-ups and incumbents across more than 20 OECD and non-OECD economies since 1998." The U.K. has only updated DynEmp3 data up to 2011, officially performed by the OECD, but I could update this thesis up to 2018. See https://www.oecd.org/sti/dynemp.htm for more information.

[6]A young firm is considered less than or equal to 6 years old, following OECD definitions in DynEmp. A mature firm is older than six years.

[7]A small firm has 249 employees or less, following OECD definitions in DynEmp. A large firm refers to a firm with 250 employees or more

[8]A high-growth firm is an incumbent or exiting firm with over ten employees and with an annual employment growth rate above 20 percent over three consecutive years, OECD definitions in DynEmp

[9]A high-tech firm is one reporting an industry code that classifies within the list of high-tech sector as considered in DynEmp, version 3, by the OECD.

[10]Labor productivity was computed using ORBIS -Bureau van Dijk financial database of firms

[11]extracted from the EU-KLEMS dataset at the industry level, for growth and productivity research

[12]defined as the standard deviation of a time series of the gap between a firm's and its industry's average productivity. This pattern has been observed as stronger or weaker, depending on the type of productivity measure (i.e., TFP or Revenue Labor Productivity). Rising productivity dispersion is characteristic of developed countries such as the U.S., where Total Factor Productivity based on shares' (TFPS) within-industry dispersion increased from about 0.46 in the 1980s to 0.51 in the 2000s (Decker et al., 2020).

[13]For example, rising concentration, in alignment with the tainted premise about superstars; that a smaller number of large firms now dominate many industries, which can make it more difficult for new businesses to enter the market and compete. Another one is regulatory burdens, where governments have enacted several regulations to protect consumers and workers over the years. While these regulations are often necessary, they can make starting and running a business more costly and time-consuming. Also, it could be the rise of the gig economy. More and more people now work as freelancers or contractors rather than full-time employees of traditional companies. This can make it more difficult for businesses to scale up and grow. It may also be related to demographics. The populations of many developed economies are aging, which can lead to a decline in the number of people starting new businesses.

[14]A high-growth firm is an incumbent or exiting firm with over ten employees and with an annual employment growth rate above 20 percent over three consecutive years, per OECD definitions in software routine DynEmp. We may also refer to a high-growth firm as one at the top 90th percentile of the employment growth distribution in a particular year.

[15]The INTAN-Invest database is a benchmark dataset for intangible investment. It was created by the INTAN-Invest project, funded by the European Commission's Seventh Framework Programme (FP7).

[16]"The DynEmp project presents evidence on the employment dynamics of start-ups and incumbents across more than 20 OECD and non-OECD economies since 1998." See https://www.oecd.org/sti/dynemp.htm for more information.(Calvino et al., 2018)

[17]The U.K. Business Structure Database is a snapshot of the Inter-Departmental Business Register and Office for National Statistics (ONS) business survey data for research use, taking complete account of changes in ownership and restructuring of businesses.

[18]A young firm is considered less than or equal to 6 years old, following OECD definitions in DynEmp. A mature firm is older than six years.

[19]A small firm has 249 employees or less, following OECD definitions in DynEmp. A large firm refers to a firm with 250 employees or more

[20]A high-growth firm is an incumbent or exiting firm with over ten employees and with an annual employment growth rate above 20 percent over three consecutive years, OECD definitions in DynEmp

[21]A high-tech firm is one reporting an industry code that classifies within the list of the high-tech sector as considered in DynEmp, version 3, by the OECD.

[22]declining dispersion differentials 90-10 and 90-50 of employment growth rates responding with a 2-period lag of delay

[23]The IDBR reports three different types of units of analysis (Figure 3). Administrative, statistical, and reporting units (also called observation units). The administrative units are those reported as VAT traders or PAYE employers. The reporting units hold the mailing address for the business and are the units for which registered companies report their survey data to ONS. However, the BSD's primary focus and, thereby, primary structure, is the statistical units.

[24]Firm group refers to the group to which a particular firm belongs in a particular reference period. A firm cannot belong to the same group in the same reference period (i.e., the same year). These distinct group categories are explained in detail in subsections 2.1.4.

[25]EU KLEMS is an industry level, growth, and productivity research project. EU KLEMS stands for EU level analysis of capital (K), labor (L), energy (E), materials (M), and service (S) inputs.

[26]As defined by a UK law firm: "RESPONDENT: This is the legal term for the employer in employment tribunal proceedings. In discrimination claims, claimants may sue individuals and a company, meaning that individuals may also be respondents to a claim.".Rockhopperhr Law Firm (2023). Also, see "there are two ways in which one employer can be held liable for the actions of another employer in employment law: if the employer exercises sufficient control over the second employer's workers, or if it is statutorily liable. The Industrial Wage Orders define an employer as any person who directly or indirectly, or through an agent or any other person, employs or exercises control over the wages, hours, or working conditions of any person." CalChamber. California Chamber of Commerce. (2022)

[27]The share of young firms is lower and the dispersion of young firms is higher relative to mature firms.

[28]Switching from good-producing sectors such as manufacturing to more volatile, high-dispersion sectors such as Retail Trade and Services.

[29]They measure the dispersion of idiosyncratic productivity shocks with a proxy, the Total Factor Productivity differentials between plants of the same firm and between firms. Differentials of TFP would incentivize companies to reallocate resources. The variable captures the firm-level realization of productivity shocks.

[30]For instance, the U.S. electronics manufacturing sector transitioned to special-purpose equipment manufacturing after 2001, leading to lower demand and less responsiveness to productivity shocks (Byrne, Byrne).

[31]Corrado et al. (2017) find a significant presence of spillovers by finding a positive difference between the estimated marginal product of a factor and the marginal product implied by the factor remuneration in a perfect competition context

[32]—university degrees and equivalents—

[33]https://www.bbvaopenmind.com/en/articles/intangible-capital-productivity-and-labor-markets/

[34]After controlling for relevant factors firm size, capital or labor inputs, dispersion of such measures, and fixed effects, they signal the quality of labor as the mediator effect. Employment-based firm size only affects the relationship between intangibles and productivity dispersion when one refers to industries with low-skill-intensive sectors. High-skills-intensive sectors require the size to scale up and keep up with the demands imposed for having more skills-demanding intangibles—evidence of increasing within-industry productivity dispersion along with increasing investment in IT capital intensity. In general, industries that experience the most growth also experienced the most increases in IT-related goods investments, signalling a significant tie between IT and greater productivity variance (Faggio et al., 2007)

[35]Although workers from home were still low (no more than 5% of the total workforce), the trend increased during the pandemic.

[36]A transition from labor-intensive (a more flexible input) to intangible-intensive economies (an input with more adjustment and fixed costs).

[37]They estimate the elasticity of the Average Revenue Product (ARP) for both forms of capital, intangible and tangible, to idiosyncratic productivity shocks. They find that the elasticity is much higher for tangible capital, meaning that the ARP of intangibles is much more volatile than the ARP of tangible capital, leading to more extraordinary misallocation. To prove this, they use a structural framework of firm dynamics (quantitative framework of firm and investment dynamics with adjustment costs - particularly fixed adjustment costs) using the empirical techniques of the industrial organization literature to show that the change in production from a flexible input -labor- to input with high adjustment costs - intangible capital - can explain a sizeable fraction of some of the significant trends documented for the U.S. economy, such as rising average firm size, dispersion of Total Factor Productivity revenue, sales-weighted profit rates and increasing industry concentration.

[38]The program also allows setting up the data in the panel for local units or plants, per the definitions established above. The local units or plants are also called establishments in the DynEmp3 codes

[39]"The STAN database is a comprehensive tool for analyzing industrial performance at a relatively detailed level of activity across countries. It includes annual measures of output, value-added and its components, labor input, investment, and capital stock from 1970 onwards, which allow users to construct a wide range of indicators to focus on areas such as productivity growth, competitiveness, and general structural change." (OECD 2016).

[40]The Pay As You Earn (PAYE) system is a method of paying income tax and national insurance contributions.

[41]The U.K. Innovation Survey collects data from organisations designed to capture advances in innovative activity. This includes innovations in production, new goods and services, and markets in which businesses operate.

[42]The Investment in Intangible Asset (IIA) Survey was launched by ONS in October 2009. It is a unique survey of firms in the U.K., drawn from the business register to represent the market sector of the economy. The survey measures firms' investment in six categories of intangible assets: employer-funded training, software, research and development (R&D), reputation and branding, design, and business process improvement.

[43]www.intaninvest.net

[44]www.spintan.net

[45]When we run the estimations, some of the dependent variables coming from the business dynamics estimators may have one or two years less available by construction; depending on the way DynEmp3 calculated them. Thus, there might be fewer observations in an estimation set. The dispersion differential variables, for instance, have one observation less, starting in the t+1 year instead of t.

[46]See appendix A.1 for further details on these differences.

[47]The OECD-Eurostat Manual on Business Demography statistics defines high-grow enterprises as "enterprises with an average annualized growth greater than 20% a year, over a 3-year period, and with 10 or more employees at the beginning of the observation period". Growth is measured by turnover or employment, with the turnover definition giving higher numbers than the employment definition.

Gazelles are the subset of high-growth enterprises which are up to five years old. In the European Union, the Commission implementing regulation (EU) No 439/2014 has lowered the recommended growth threshold to "greater than 10% per annum, over a three-year period".

[48]The evolution of knowledge-intensive services (KIS) and basic services in the last 30 years in the UK has been characterized by a shift towards KIS and a decline in basic services. In 1992, KIS accounted for 28% of the services sector in the UK, while basic services accounted for 72%. By 2022, KIS accounted for 39% of the services sector, while basic services accounted for 61%. This means that KIS have grown by 39% over the past 30 years, while basic services have declined by 15%.

[49]Or also $E\left[x^s_{c,t}\vartheta^s_{c,m}\right] = 0$ for all t and m. If $E\left[x^s_{c,t}\vartheta^s_{c,m}\right] \neq 0$ for $m < t$ but $E\left[x^s_{c,t}\vartheta^s_{c,m}\right] = 0$ for all $m \geq t$, the variable is said to be predetermined. Intuitively, suppose the error term at time t has some feedback on the subsequent realizations of xit. In that case, xit is a predetermined variable

[50]just in some contexts with also use a proxy for **Demand**, the sector's aggregate real value added

[51]Those results suggest that more intangibles are associated with fewer contributions of Young SMEs to employment.

[52]The DynEmp3 program recreates the following steps. First, it calculates the employment growth rate between t and $t-1$ for each firm i in the sample, denominated as $e_{i,t}$. Second, it ranks each employment growth rate $e_{i,t}$, from the lowest to the highest and classifies each firm's employment growth rate in five intervals of growth distribution. Firms in the bottom 10% of the distribution of employment growth rate ($p = 1$), firms between the 11th and 25th percentile ($p = 2$), firms between the 26th and 50th percentile ($p = 3$), firms between the 51st and 75th percentile ($p = 2$), firms between the 76th and 90th percentile ($p = 2$), and firms in the top 10% of the distribution ($p = 6$). Third, it estimates the weighted-average employment growth rate $\bar{e}_{c,t}$ by cell class c at time t.

[53]Corrado, Criscuolo, Haskel, Himbert, and Jona-Lasinio (2021) divides the services sectors by high and low labor skill intensity. They run separate regressions of these sectors' 90-10 productivity dispersion differential as a function of their share of intangible investments, and an interaction term of the intangibles share and size. This size variable is labor dispersion between firms growing at the top and the bottom in employment. They show that the coefficient for the interaction term is only positive and significant in the low-skilled services sectors regression. For the high-skilled sectors' regression, this is not observed. Thus, the greater the labor dispersion, the greater the productivity dispersion but only in low-skilled sectors. In other words, labor dispersion only matters in industries that are not still so much in need of complementary skilled workers for their intangible assets. Furthermore, when they do the same regression but for the manufacturing industry only (low and high-skilled sectors combined), they find that the coefficient for the interaction is still positive and significant but lower than that for the total services sectors.

[54]Again, highly intangible sectors are declining their intensity while maintaining their reduced responsive behavior to labor dispersion. To exemplify why this could be the case, let's notice what has been shown by prior studies on intangibles and productivity dispersion across industry sectors.

[55]Despite young SMEs being, on average, contributing the most to net job creation throughout the period, the patterns are different over time and across sectors. After a period of stability of low contribution, Young SMEs from manufacturing slightly increased their contribution towards the end of the sample period.

Services young SMEs' contribution, which had been noticeably higher than manufacturing, quickly recovered after plummeting during the financial crisis. However, such recovery seemed to have stopped in years close to the Brexit vote in 2016. Interestingly, mature manufacturing SMEs have turned to destroying fewer jobs in recent years.

In contrast, in the service sector, this age×size group fluctuated around the same contribution values during the period. Furthermore, larger firms contributed more to net job creation or destroyed fewer jobs. Primarily in the manufacturing sector (section 1.6.1.2).

[56]The evolution of knowledge-intensive services (KIS) and basic services in the last 30 years in the UK has been characterized by a shift towards KIS and a decline in basic services. In 1992, KIS accounted for 28% of the services sector in the UK, while basic services accounted for 72%. By 2022, KIS accounted

for 39% of the services sector, while basic services accounted for 61%. This means that KIS have grown by 39% over the past 30 years, while basic services have declined by 15%.

[57]Another issue is that DynEmp3 uses slightly different definitions of entering, incumbent, and exiting firms, among other variables relative to the ONS publications. This alters further estimated variables such as entry, exit, and job reallocation estimated rates. For instance, in DynEmp3, an entry or a "birth" is considered a firm that was not in year $t-1$ but appears in year t (as seen in Figure 2); whereas in the ONS publication, an entry is considered a firm that was not in $t-1$ or $t-2$; which makes their birth counts much more restrictive (Office of National Statistics 2019). Those reports are considered a death if it is not active for more than one period, so if the firm is absent only between $t-2$ and t, then it is not considered a birth in this case. Also, in DynEmp3, an exit or a "death" is considered a firm in year t, but it does not appear in year $t+1$ (Figure 2). Conversely, the ONS defines the death of a business when the firm is not on the active file in year t and is no longer present in the active file in $t+1$ and $t+2$. This is also related to the disadvantages of using a concatenated version of the cross-sectional files of the BSD to form a longitudinal micro-dataset and use it as input for DynEmp3. For example, the BSD includes variables for each enterprise, such as birth and death, indicating these demographic events. However, these dates are not entirely accurate.

[58]The main findings include (a) that the U.K.'s business dynamics exhibit similar behavior to other countries, with a secular decline over the past two decades with an accelerated decline after the 2000s; (b) In contrast to other cases, however, there was a recovery after the 2009 financial crisis. (c) The improvement after the crisis halted again after the 2016 Brexit vote; (d) High-Growth firms' numbers have slightly declined over time, as has their growth pace and average size, but their average median age has increased. Also, large mature firms employ fewer people, while young and big firms have recovered their share of total employment and net job creation roles in recent years. Once connecting these trends to intangible investment trends, the principal results indicate that rising intangible intensity may be leading to a) declining dispersion differentials 90-10 and 90-50 of employment growth rates, responding with a 2-period lag of delay, b) declining entry rates (less number of businesses entering the market and creating employment), c) increasing exit rates in terms of employment (no statistically significant), d) declining job reallocation, and e) declining number of high-growth young and small firms.

[59]The implications of increased average levels of market power across the board are significant since they can degrade key metrics in economy-wide well-being, including investment, innovation, total output, aggregate productivity, and income distribution. It allows firms to set higher prices, impacting consumer costs and reducing purchasing power. Higher markup firms might decrease factors' demand, affecting production inputs like labor. This, in turn, can decrease workers' income share, potentially stifling innovation.

[60] In economic papers on production function estimations, the "control function approach" is a statistical technique used to address unobserved factors or endogeneity in production models by introducing control variables that help disentangle the effect of interest from other potentially confounding variables, enhancing the precision and reliability of production function estimates. It allows researchers to isolate the impact of specific factors on production output while controlling for unobservable influences.

[61]Orbis is the world's most powerful comparable data resource on private companies. It has information on over 425million companies.

[62]Compustat is a comprehensive market and corporate financial database published by Standard and Poor's, covering thousands of companies worldwide, with info dating as far back as 1950.

[63]We use the command "rolling" in Stata, which allows using the "panel" option, so Stata treats each firm as a separate entity and applies the rolling window calculations separately for each entity within their respective panels. We also use the "one-step" option to specify that the operation we want to perform is a one-step operation. This means that the result of the operation will be calculated based on the data within the rolling window, and then the window will move forward one step (a year in this case) at a time. The result for each step will be stored in the output dataset.

[64]An associated index to concentration is the Lerner index. It describes the relationship between elasticity and price margins for a profit-maximizing firm; however, it relies on prices and cost information, which is often unavailable and therefore measured with revenues and can be subject to measurement error

[65]These theoretical implications have been explored in a substantial empirical body of work. See, for example, Syverson 2004a 2004b cited in Syverson (2019) find that increases in the easiness for consumers to substitute products force out least efficient producers and increase skewness in the size in the distribution. Goldmanis et al. 2010 cited in Syverson (2019), demonstrate that search cost reductions (a sign of greater sustainability or competition) shift away from an activity from smaller, higher producers and toward larger, lower-cost producers. As substitution/competition increases (.. falls), firm profits increase

[66]CR4 is the fraction of total sales that is accrued by the four largest firms in an industry, CR20 is the fraction of sales accrued by the 20 largest firms

[67]These analyses address measurement issues by examining product/geographic antitrust markets affected by the merger official studies on legal cases by the European and therefore define markets more accurately. A significant criticism of academic studies is that market shares are measured at the national industry level and revenues at the firm level because economic markets and product level revenues are generally not observed directly in the data. Indeed, both the "numerator" (product level sales) and the "denominator" (sum of sales in the relevant product and geographic market) might not – and generally do not – reflect economically sensible definitions (Affeldt et al., 2021).

[68]Loecker and Eeckhout (2021) on 134 countries and 70,000 firms

[69]The estimation of markups is often paired with the estimation of profitability measures. Higher markups alone may not imply increased economic profits per se if firms also incur higher fixed costs to produce goods and services. A rise in markups could reflect the increase in those fixed costs (i.e. just enough to recover the cost of capital) so that firms with markups above one are not necessarily making excess profits. The evidence of excessive profits, however, is ample, arguably as the result of less competition.

[70]For example, the growing importance of international trade, offshoring or foreign sourcing of inputs (particularly with China)(Reshef and Santoni, 2023), a decrease in corporate tax rates (Li et al., 2021), a slowdown in population growth Hopenhayn et al. (2022), and a suggested shift in social norms and labor market institutions leading to lower labor shares Piketty (2014).

[71]Robbins and Getz Wold 2008 also demonstrate that a declining natural interest rate aligns with increasing markups.

[72]Superstar firms are defined as ROIC in the top 10% of the distribution.

[73]Karabarbounis and Neiman (2014) hypothesize that the cost of capital relative to labor has fallen, driven by rapid declines in quality-adjusted equipment prices, especially of ICT, which could lower the labor share if the capital-labor elasticity of substitution is greater than 1

[74]Less than 0.5% between 2005 and 2018 way below the long-term average of 1.3% and Europe's since the 1970s, with very vapid periods of small recoveries until the present. In countries such as France and the UK, productivity growth flatlined

[75]Frontier firms are the most productive firms in each 2digit industry code in 23 countries.

[76]Andrews, Criscuolo, and Gal (2015) find evidence between 2001 and 2013 of increasing gaps in productivity across sectors between productivity frontier (most productive firms in each 2digit industry code in 23 countries) and laggard companies. In particular, they distinguish between global and national frontier and laggard firms, emphasizing the role of technological diffusion between these two groups. Global frontier firms are characterized by being larger, more profitable, younger, and more likely to patent and be part of a multinational conglomerate. The authors explain that national laggard firms in countries only catch up to the global frontier level once the leader/frontier national firm catches up with the global frontier. This is possible only when the national leader manages to overcome country-specific factors to adapt the innovation or technology, so laggard national firms have an easier job. Patterns like these partly explain the increased productivity dispersion in the United States (also shown by Decker et al. (2017). The explanation for this pattern of the rise of the "best versus the rest" dynamics is that precisely leaders have turned better to protect their competitive advantages, contributing to a slowdown in aggregate productivity. Similarly, Andrews et al. (2015, 2016) find that across the OECD economies, the gap between the average productivity levels of frontier firms (top 5 percent of firms with the highest

productivity) and the laggard ones has been widening, suggesting the rise of "best versus the rest" dynamics.

[77]scalability, Spillover, sunk ness, and synergies are characteristics identified by Haskel and Westlake (2018).

[78]De Loecker et al. (2018) also has similar findings when including overhead costs (as a proxy for intangibles) as a factor in production function estimations to estimate firm-level markups

[79]Autor et al. (2020) find that technological progress that reduces the relative cost of capital (measured as the ratio of capital to value added) hurts labor share. Still, it is not all that explains the fall in labor share. The superstar firm effect (between-firms reallocation of output) remains significant, including capital deepening. These ideas arguably co-exist with Autor et al. (2020) superstar mechanism because to test this hypothesis, the authors introduce the capital to added-value ratio variable to regressions of labor share against concentration, finding that although such variable as expected has significant and adverse effects on labor share; the concentration coefficient remains negative and about the same value. This is also consistent with rising size-weighted aggregate markups, as De Loecker et al. find, at least in the U.S., since larger firms have higher markups in their findings.

[80]1.31 using the cost shares approach and 1.21 using the production function estimation approach

[81]Zheng et al. (2022) examined Chinese firms from 2000 to 2019, and found evidence of an inverted U-shaped relationship between markups and investment.

[82]The reasons that could be causing increasing skewness in concentration and productivity simultaneously can be divided into demand or supply causes. On the demand side, if companies' margins are increasing, something should have flattened the firm's residual demand curve. Maybe consumers are improving their search abilities to find cheap and good quality products. Intangible assets can explain those shifts in demand. For example, better reputation, marketing strategies, or even software applications can help consumers do this.

[83]Authorities need to establish appropriate antitrust measures to control features of these assets that can eventually lead to market power, and subsequently negatively impact resource allocation, labor share and incentives to innovate, among others. Higher markups are worth investigating since they signal something of competition intensity. Perhaps in that market, consumers face higher search costs. Hence, they are unaware of alternatives, or a firm's products are so differentiated that others cannot expand on that particular market segment. However, the contrary can be true, and intangibles can reduce search costs.

[84]

$$r_{it} = \beta_L l_{it} + \beta_m m_{it} + \beta_k k_{it} + a_{it} + \varepsilon_{it} \tag{75}$$

[85]$r_{it} - p_{it} = q_{it} + (p_{it} - p_{it})$

[86]Following De Loecker and Warzynski (2012), the estimation of markups at the firm level includes a correction to the value of sales. Sales are adjusted by the measurement error obtained in the first stage of the estimation. Since we do not directly observe the correct expenditure share for input M (we only observe estimated output), which is given by

$$\ln y_{it} \exp(\epsilon_{it})$$

, we take advantage of the estimated value for the error term. We correct the input share such as the final estimate of the markup is:

$$\mu_{it} = \frac{\alpha_M}{S_{it} \exp(\epsilon_{it})}$$

.

This correction is necessary because it removes variation in expenditure shares that comes from variation in output that is not correlated with the function with unknown functional form in the first stage that takes into account the variation in the three inputs and the instruments; in other words, it removes output variation not correlated to variables impacting input demand including input prices,

productivity, technology parameters, and market characteristics, such as elasticity of demand and income levels (De Loecker and Warzynski, 2012).

However, we will not include this correction in our estimations since in their De Loecker et al. (2020) paper they present markups estimations without this correction, and we want to compare them with the most recent paper estimations.

[87]They do it in lags since the AR(1) is introduced in this manner, so the moment condition in the second stage is:

$$\begin{aligned} & E\left[\xi_{it}+\varepsilon_{it} \mid I_{it-1}\right] \\ & =E\left[r_{it}-\beta_k k_{it}-\beta_m m_{it}-\beta_l l_{it}\right. \\ & \left.-g\left(\tilde{\Phi}_{t-1}\left(k_{it-1}, l_{it-1}, m_{it-1}\right)-\beta_k k_{it-1}-\beta_l l_{it-1}\right) \mid I_{it-1}\right]=0 \end{aligned}$$

where I refers to the information set. However, this suffers from the functional dependence problem highlighted by Ackerberg et al., where the proxy variable is the same free variable. By construction

$$a_{it}=\rho_a a_{it-1}+\nu_{a,it}$$

[88]Model 2 assumes a Cobb Douglas production function as in equation 34, but Model 3 assumes a Translog

[89]Those shocks are not the focus of this model.

[90]Remember that the full model version decomposes ω between a demand heterogeneity shock d and a quantity TFP shock, each evolving as linear stochastic Markov processes.

[91]There are good reasons to prefer gross output estimations over value-added, as suggested in Levinhson and Petrin (2003). Gross Output Functions allow to have multiple first-order conditions to recover the markup (i.e. labor and or materials and test overidentifying restrictions. Furthermore, avoid having to impose extra assumptions of assuming that a fixed proportion of materials is used for producing as a unit of output (as Leontief technology). Moreover, if one decides to do so, the computation of markups would be based on the coefficient of a much less flexible input, such as labor. Labor is exposed to frictions such as hiring and firing costs, which ultimately create a bigger wedge between the marginal product of the input and input price, other than the markup. However, in order to be able to identify all variable factors of production, to make explicit that all input prices of flexible inputs vary across firms and are serially correlated. This assumption is necessary for the identification of the coefficient of materials.

[92] Challenges with the UK data: for entry and exit mostly: "Whilst conceptually straightforward, in reality, entry and exit may not be easily defined. This may partly be due to threshold rules that data sets are constructed under. Firms may fall in and out of the population if their turnover or employment is close to the threshold. Moreover, as with the Annual Respondents Database (ARD) in Britain, smaller firms may be sampled from the population, complicating death and birth information (see below). Thus, the data requirements for dynamic decomposition are often greater than for static decomposition." cite Sources of labor Productivity Growth at Sector Level in Britain, 1998-2007: a Firm-Level Analysis

[93]We use the deflators compiled by the OECD from different sources, including

[94]Once we compute the output elasticities for both models, we also apply certain assumptions to them when needed. We make sure all elasticities are under one for Models 1 and 2; in particular, for Model 2, we make sure that coefficient Dk per equation 65 is positive and greater than one for each of the 5-year window * sector cells. See the Appendix for a complete list of assumptions.

In order to make the estimations the most comparable possible between the U.S. and European samples, we do not trim markup estimates as other authors. For example, OECD (2018) adjusts markups equal to one for those estimated as right under 1 (e.g., 0.95 is made 1), or by trimming by sector top percentiles of the markup distribution. De Loecker's public codes for their De Loecker et al. (2020) paper do not trim extreme markups unless they refer to the plots depicting trends of some percentiles of the markup distribution. They eliminate markup values above 3.5 and below 0.7. However, more and more recurrent practice on papers on this topic is to eliminate extreme markup values by trimming 0.1% of the estimated markups by year, or a combination of the two, as in Ridder et al. (2022).

[95]The authors share the set of codes plotting graphs and tables of the markups but not directly their computation of output elasticities resulting from the 2-stage GMM ACF procedure.

[96]It's important to note that various factors, including industry dynamics, market structure, and consumer behavior, can influence the relationship between markups, marginal costs, and economies of scale. While decreasing markups may be a general trend when costs decrease and economies of scale increase, individual firms and industries may experience variations in this trend based on their specific circumstances.

[97]We are relying on data from Twitter, but the idea could be extended to other platforms.

[98]We drop non-Spanish tweets for our analysis.

[99]As discussed earlier, any other balance sheet-based measure would lead to substantially smaller data due to an incomplete data situation.

[100]For simplicity we use the $ln(1 + PatentCount)$ transformation.

[101]We obtained patent information from Orbis IP.

[102]https://developer.twitter.com/en/docs/twitter-api/expansions

[103]See https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country. "Many countries, including the United States, have laws that may apply to Tweets and/or Twitter account content. In our continuing effort to make our services available to people everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to withhold access to certain content in a particular country occasionally." Twitter, 2023.

[104]Notice that our lookup consisted of finding the (tweet id) in the *referenced_tweet_id* of the complete timelines of the firms AND finding the author of that tweet; the one that corresponded to the author that firms were referring to when posting their tweet. In other words, we matched both variables, the tweet id and the author id of the tweet. This means we include only direct references to the customers, which can only occur by replying or quoting a tweet, not a retweet. If we were to match all tweet ids regardless of the authors of the tweet, we would have found the retweets as well.

[105]*Tripadvisor* is an online travel information and booking website with reviews, photographs, and informative forums about various hotels and restaurants. *El Tenedor* is a Tripadvisor company focused on food reviews and recommendations for Spanish-speaking customers. Source: Tripadvisor. (2023, May 19). In Wikipedia. https://en.wikipedia.org/wiki/Tripadvisor*Decathlon* is a French sporting goods retailer with significant brand recognition in Spain and Latin America and a robust e-commerce presence. Source: Decathlon. (2023, May 12). In Wikipedia. https://en.wikipedia.org/wiki/Decathlon.*FilmAffinity* is a movie recommendations website created in Spain. In 2016, was reported to list more than 125,000 movies and series and had 556,000 reviews written by its users. Source: FilmAffinity. (2023, April 23). In Wikipedia. https://en.wikipedia.org/wiki/FilmAffinity

[106]Review Scraper API, https://datashake.com/review-scraper-api

[107]Mentions can also reveal a firm's innovation degree. For instance, a firm mentioned along with words like "new" as part of the text reveals that the company's name is somehow associated with a new pattern. Perhaps it is not only the case that a customer is aware of a new product or launch, but it is aware of any innovative behavior of the firm. Appendix J.2, Tables J.22 and J.23, show some examples of these mentions.

Appendices

Appendix A UK data comparability

A.1 BSD and comparability with UK official demography reports

DynEmp3's business dynamics indicators slightly differ from those published in the U.K.'s Office of National Statistics (ONS) publications, such as the UK Business Demography annual bulletin; for statistics, see Table A.1.

TABLE A.1
Differences in definitions

Definitions	Business Structure Database	UK Business Demography (ONS Publication)	UK business: activity, size and location (ONS Publication)	OECD- DynEmp3
Collection of data	Snapshot of IDBR in March	Extract from IDBR in September each year, but takes a look at information over the reference period to define active status	March each year, but it is only a snapshot, so if it was active or inactive in a prior later/month it does not take it into account. "The data held in the BSD is a separate publication from the UK Business Demography, and it is called UK business; activity, size and location."	Program requires to add the month in which the snapshot is taken (March, in the case of UK).
Active firms	The definition of the population of active enterprises is those enterprises with turnover and/or persons employed greater than zero and at least one administrative unit linked to the enterprise.	"The starting point for demography is the concept of a population of active businesses in a reference year (t). These are defined as businesses that had either turnover or employment at any time during the reference period. Births and deaths are then identified by comparing active populations for different years." September, but the enterprise could have been active in any point of the year (with turnover or employment different than zero). It was active in 2017 if either had employment/turnover>0 in October 2016 or in March of 2017 for example.	These are defined as businesses that had either turnover or employment at a particular point in time (snapshot). The main administrative sources for the IDBR are VAT trader and PAYE employer information passed to the ONS by HM Revenue & Customs under the Value Added Tax Act 1994 for VAT traders and the Finance Act 1969 for PAYE employers; details of incorporated businesses are also passed to ONS by Companies House.	Active firms for entry rate= Number of Incumbents + number of entrants in year t Active firms for exit rate= Number of Incumbents + number of exiting firms in year t
Birth (of an enterprise)	The Birth of an enterprise is the bringing into being of an enterprise where no enterprise existed before and no other enterprises are involved. Since the enterprises is an organisational unit producing goods or services, a birth amounts to the creation of a combination of production factors. Births and deaths will result in the creation and deletion of identity numbers, as many other events.	"A birth is identified as a business that was present in year t, but did not exist in year t-1 or t-2. Births are identified by making comparison of annual active population files and identifying those present in the latest file, but not the two previous ones." This concept differs from their other definition of "employer births", which has to comply with one of two conditions to qualify it as a birth by growth (employees someone for the first time). New business registrations are referred to as business births.	It does not report	An enterprise is not in t-1 but it is in t, but it enters with positive employment
Death (of an enterprise)	The Death of an enterprise is the opposite to the above. The death of an enterprise amounts to the dissolution of a combination of production factors.	A death is defined as a business that was on the active file in year t, but was no longer present in the active file in t+1 and t+2. In order to provide an early estimate of deaths, an adjustment has been made to the latest two years deaths to allow for reactivations. These figures are provisional and subject to revision. Businesses that have ceased to trade are referred to as business deaths.	It does not report	An enterprise was in t-1 but it is not longer in t, but it leaves taking positive employment with it
Entry rate	It does not report	Birth rate is calculated using the number of births as a proportion of the active enterprises.	It does not report	Birth rate is calculated using the number of births (entrants) as a proportion of the active enterprises as defined above (incumbents + entrants).
Exit rate	It does not report	Death rate is calculated using the number of deaths as a proportion of the active enterprises.	It does not report	Death rate is calculated using the number of deaths (exiting firms) as a proportion of the active enterprises as defined above (incumbents + exiting firms).

Source: Author based on ONS publications.

There are two core reasons. First, the BSD is not the dataset used by the ONS to build its demographics bulletin. Indeed, the ONS builds its extract from the IDBR's file of active firms to calculate the different business demographics indicators. The data in the Business Structure Database –BSD- is used for a separate publication called UK business; activity, size, and location[108]. Second and directly related to the first, the ONS publication applies assumptions somewhat different from the ones applied to the BSD. A

significant difference between the Business Demographics publication and the BSD is the definition of active firms. This concept is essential to estimate enterprises' entry and exit rates every year in any business dynamics analysis. Generally, entry rates are equivalent to the number of births divided by the number of active firms in a reference period. Exit rates are equivalent to the number of deaths divided by the number of active firms in a reference period.

The *ONS business demographics* bulletin defines active firms as businesses that had either turnover or employment at any point during the reference period, for instance, a year.[109] This bulletin also takes an annual snapshot of the IDBR in September, but they can review the activity status of the firm at any point in time in the reference period. In contrast, the BSD defines active firms as those firms with positive turnover[110] or employment at just the point in time they take the snapshot at in March[111]. Hence, the total stock of active businesses is higher in the demographics bulletin than in the BSD (and therefore, higher than in the activity bulletin), leading to a BSD-based analysis to have potentially fewer active firms.

To illustrate this with an example, see Figure A.1. Let us assume that activity is measured only concerning employment with disregard for turnover levels, ignoring links to local units. Then, we can access four sources to gather information about the level of employment for the year 2018 of firm x. If we had direct access to the IDBR (source No. 1) since the register is more or less continuously updated, we would have an excellent idea about the levels of employment throughout the year so we could make an educated assessment of whether this an active or an inactive firm in 2018.

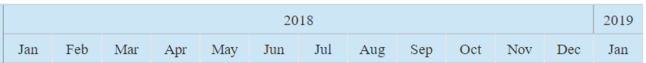
If we only had access to the BSD (source No. 2), we could only access the snapshot in March to assess the activity status of firm x. If the firm does not report employment that month, we would say that the firm was inactive in 2018. If we are part of the ONS (source No. 3), we could review the IDBR and check whether the firm is active in 2018 by reviewing employment levels throughout the year. If a firm reports employment activity at any point during this year, even if we only take a snapshot of this firm's data in September, we can know with more certainty whether the firm is active for 2018.

Let us also consider the application of DynEmp3 as another source (source No. 4). In this case, the program permits setting the employment variable as recorded as a point in time in a reference period, in this case, year t. The DynEmp3 user is only required to enter the number of the month in which the snapshot is taken. Thus, we consider firm x as an active firm according to the employment information provided in the snapshot for March in the BSD. If employment is not reported and it is missing, as happened with source No. 2, that firm is classified as inactive and further computations needing that type of observation ignore that firm's information.

All in all, the different timelines in the collection of data affect the overall definition of active firms and consequently, the different business dynamics indicators that depend on them.

FIGURE A.1
UK Business demographics' data collection timeline differences

	Data source / variable	Collection method	Start	End	Days
	Employment 2018 ⊖		1/1/18	31/12/18	259
1	Inter-Departmental Business Register	Live business register: Updated frequently (data are provided from HMCE weekly, monthly or quarterly)	1/1/18	31/12/18	259
2	Business Structure Database	Annual snapshot of the IDBR at a point in time	14/3/18	14/3/18	1
3	UK Business Demography (ONS Publication)	Annual snapshot of the IDBR at a point in time but recognizes activity at any point during the reference year	14/9/18	14/9/18	1
4	DynEmp3 (OECD's business demography program)	Annual snapshot of the business demographics dataset selected at a point in time	14/3/18	14/3/18	1



Notes: Author illustration based on Office for National Statistics (2006) and Hart and Prashar (2019).

Another issue is that DynEmp3 uses slightly different definitions of entering, incumbent, and exiting firms, among other variables relative to the ONS publications. This alters further estimated variables such as entry, exit, and job reallocation estimated rates. For instance, in DynEmp3, an entry or a "birth" is considered a firm that was not in year $t-1$ but appears in year t (as seen in Figure 2); whereas in the ONS publication, an entry is considered a firm that was not in $t-1$ or $t-2$; which makes their birth counts much more restrictive (Office of National Statistics 2019). Those reports are considered a death if it is not active for more than one period, so if the firm is absent only between $t-2$ and t, then it is not considered a birth in this case. Also, in DynEmp3, an exit or a "death" is considered a firm in year t, but it does not appear in year $t+1$ (Figure 2). Conversely, the ONS defines the death of a business when the firm is not on the active file in year t and is no longer present in the active file in $t+1$ and $t+2$.

This is also related to the disadvantages of using a concatenated version of the cross-sectional files of the BSD to form a longitudinal micro-dataset and use it as input for DynEmp3. For example, the BSD includes variables for each enterprise, such as the birth date and death date, indicating these demographic events. However, these dates are not entirely accurate.

In the BSD, a firm i at year t is flagged as dead when certain conditions apply. If a firm does not appear in the "active firms" file of a particular year in the IDBR, that firm will be considered "inactive", and a death date is reported that year. Yet, this situation can occur multiple times and not consistently over the sample period (from 1998 to 2018). For instance, the firm can appear and disappear from the active files intermittently, for 1, 2, or more consecutive years each time, so that the same firm can have multiple death dates throughout the whole period. In this study's sample, at least 12 percent of the 8 million firms (reporting units at the enterprise level) have between 2 and 17 different death years reported.

Figure A.2 illustrates this issue. The BSD has an active variable, which signals whether the firm is found in the IDBR files of active firms at year t. The IDBR has two types of files, one with the active and the other with the inactive firms' files. active is equal to 1 if the firm is active that year or 0 otherwise. The figure presents the activity history of 8 fictional firms in the BSD throughout the analysis.

FIGURE A.2
Activity history of fictional sample firms

Company id / activity status (1=Yes, 0=No)	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Activity history in 22 years of data (BSD's 22 yearly cross-sectional files)*
1	1	1	1	1	1	1	0	0	0	0	0	0	.	.	.	.	.	.	.	.	.	.	111111000000.....
2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	0	0	.	.	.	.	.....000....
3	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	.	.	.	.	.	.	1	111111111111100.....1
4	1	1	1	1	1	1	0	0	0	0	0	1	1	1	0	0	0	1	1	1	1	1	1111110000011100011111
5	1	1	1	1	1	1	1	0	0	0	0	0	.	.	.	.	.	.	.	.	.	.	111111100000.....
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	11111111111111111111
7	1	1	1	0	0	.	.	.	.	.	.	.	.	.	.	1	1	1	1	1	1	1	11100.....1111111
8	.	.	.	.	.	.	.	.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	00000000000000

Notes: Author's calculations based on the BSD. Office for National Statistics (2019) These are not real data, but an example for explanation purposes.

Firm (No. 1) presents an apparent clear life cycle since it has a continuous active status over a few years and a consistent inactive status. To check whether the firm is still active after 2002 is relatively straightforward. Each missing dot indicates that the firm is not found in that year's cross-sectional file of active firms in the IDBR. Thus, one could say that the firm died in 2008. The entire BSD presence of Firm (No. 1) is illustrated in Figure A.3.

FIGURE A.3
Fictional Firm No. 1

firm id	year	sic 03	sic 07	sic92	sic92-2digits	sic 3-dig (approx.)	change to SIC Class. Version	employment	turnover	live_lu	live_ru	live_vat	live_paye	birth date	death date	activity (BSD)	activity (ONS)	activity (IDBR)	activity (DynEmp3)	birth (DynEmp3)	death (DynEmp3)
1	1997			37200	37	372	2007	1	89	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	1998			37200	37	372	2007	1	86	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	1999			37200	37	372	2007	1	80	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	2000			37200	37	372	2007	1	71	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	2001			37200	37	372	2007	1	66	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	2002			37200	37	372	2007	1	69	0	1	0	1	1981	.	1	1	1	1	1981	2002
1	2003			37200	37	372	2007	1	74	0	1	0	1	1981	.	1	1	1	.	.	.
1	2004			37200	37	372	2007	1	74	0	0	0	0	1981	2002	0	1	0	.	.	.
1	2005			37200	37	372	2007	1	74	0	0	0	0	1981	2002	0	1	0	.	.	.
1	2006	####			.	372	2007	1	74	0	0	0	0	1981	2002	0	1	0	.	.	.
1	2007	####	####		37	383	2007	1	74	0	0	0	0	1981	2002	0	1	0	.	.	.
1	2008	####	####		37	383	2007	1	74	0	0	0	0	1981	2002	0	1	0	.	.	.
1	2009	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2010	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2011	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2012	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2013	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2014	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2015	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2016	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.
1	2017	.	.		.	.	.	.	.	.	.	.	.	1981	2002	.	.	.	.	.	.

Notes: • Light shadowed area: Dropped observations • Dark shadowed area: Input panel for DynEmp3 • These are not real data, but an example for explanation purposes

In contrast, firm (No. 2) has a very irregular activity status, and its death date is unclear. This firm appears only in three cross-sectional files of active firms for 2012, 2013, and 2014. Yet, the variable active is equal to 0 for each of those years. However, this firm may report non-missing values for employment and turnover for all those supposedly "inactive" years, as seen in Figure A.4. The whole history of the firm (No. 2) for the sample period is presented in the latter figure.

FIGURE A.4
DynEmp3 input for Fictional Firm No. 2

firm id	year	employment	turnover	live_lu	live_ru	live_vat	live_paye	birth	death	activity (BSD)	activity (ONS)	activity (IDBR)	activity (DynEmp3)	birth (DynEmp3)	death (DynEmp3)
2	1997	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	1998	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	1999	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2000	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2001	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2002	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2003	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2004	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2005	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2006	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2007	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2008	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2009	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2010	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2011	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2012	19	1147	0	0	0	0	1973	2011	0	1	0	1	1973	2012
2	2013	19	1147	0	0	0	0	1973	2012	0	1	0	1	1973	2012
2	2014	19	1147	0	0	0	0	1973	2012	0	1	0	1	1973	2012
2	2015	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2016	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2017	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012
2	2018	.	.	.	.	.	.	.	.	.	.	.	.	1973	2012

Notes: • These are not real data, but an example for explanation purposes

Most of these problems are associated with the intrinsic nature of the IDBR itself. Although this is a 'live register' (it is updated more or less continuously and the data is then picked up by the BSD every March) there are lags in the data. The IDBR is a set of March snapshots and not a conventional time series, so it does not necessarily record data that report activity levels for March, they are data as of March. Moreover, the March data can refer to a range of time periods over the previous years. Also, the IDBR has lagged due to the administrative sources (VAT/PAYE), which means it is possible that a business that is continuing to trade can appear to cease on the IDBR. Delays can occur mainly for two reasons.

First, suppose an old VAT scheme is de-registered and there is a delay in creating and/or matching the new VAT scheme. In that case, it can leave the enterprise without a live administrative source,, automatically flagging it as a death. Second, VAT-based units whose turnover drops to zero are flagged as dead in the IDBR but reborn when a non-zero turnover is reported again later.

To account for these irregularities, the ONS publication adjusts its death counts yearly for a number of firm "reactivations". The Eurostat and OECD's business demography manual recommends waiting two years after the reference period to allow for reactivations of firms before death figures are calculated. The ONS, to publish official death count figures every year, has decided to publish unadjusted counts every year but update every year's figure after two years. For instance, in 2011, the ONS published that around 244 thousand firms had died, but in 2013 it realized that around 14 thousand had reactivated. Therefore, the settled count of deaths for 2011 is 230 thousand.

A reactivation mainly occurs due to lags in the administrative sources (VAT/PAYE), which means it is possible that a business that is continuing to trade can appear to cease on the IDBR. Delays can occur mainly for two reasons. First, suppose an old VAT scheme is de-registered, and there is a delay in creating and/or matching the new VAT scheme. In

that case, it can leave the enterprise without a live administrative source, automatically flagging it as a death. Second, VAT-based units whose turnover drops to zero are flagged as a death in the IDBR but are reborn when a non-zero turnover is reported again in a later period.

The ONS also considers the concept of a "new birth". If a company's situation complies with either two continuity conditions out of a set of three, the birth of a completely different company is considered. For example, if a firm changes its controlling legal unit (condition one) and changes its principal activity (condition two), but the central location of the old company remains the same (condition three), a new firm emerges, and the old firm dies; thus it re-enters the register as a different one. The BSD applies the same sort of 2-year "correction" to the death date of a company every year. If a company has not been found in IDBR's active files of the past 2 consecutive years, the death date remains the same as reported for the first time (year) that it did not appear in the active file. However, when the firm appears and disappears multiple times in the register, as explained before, a new dead date is generated. Thus, despite the 2-year correction, the death date variable can have multiple death dates reported for the same firm. An example of the latter is firm (No. 8) in Figure A.5.

FIGURE A.5
DynEmp3 input for Fictional Firm No. 8

firm id	year	employment	turnover	live_lu	live_ru	live_vat	live_paye	birth date	death date	activity (BSD)	activity (ONS)	activity (IDBR)	activity (DynEmp3)	birth (DynEmp3)	death (DynEmp3)
8	1997	.	.	.	.	.	.	.	.	.			.	.	.
8	1998	.	.	.	.	.	.	.	.	.			.	.	.
8	1999	.	.	.	.	.	.	.	.	.			.	.	.
8	2000	.	.	.	.	.	.	.	.	.			.	.	.
8	2001	.	.	.	.	.	.	.	.	.			.	.	.
8	2002	.	.	.	.	.	.	.	.	.			.	.	.
8	2003	.	.	.	.	.	.	.	.	.			.	.	.
8	2004	.	.	.	.	.	.	.	.	.			.	.	.
8	2005	.	.	.	.	.	.	.	.	.			.	.	.
8	2006	5	200	0	0	0	0	2005	2005	0	1	0	.	2005	2017
8	2007	5	200	0	0	0	0	2005	2005	0	1	0	.	2005	2017
8	2008	5	200	0	0	0	0	2005	2005	0	1	0	.	2005	2017
8	2009	8	189	0	0	0	0	2005	2005	0	1	0	.	2005	2017
8	2010	8	340	0	0	0	0	2005	2009	0	1	0	.	2005	2017
8	2011	5	390	0	0	0	0	2005	2010	0	1	0	.	2005	2017
8	2012	5	212	0	0	0	0	2005	2011	0	1	0	.	2005	2017
8	2013	5	218	0	0	0	0	2005	2012	0	1	0	.	2005	2017
8	2014	5	216	0	0	0	0	2005	2013	0	1	0	.	2005	2017
8	2015	5	230	0	0	0	0	2005	2014	0	1	0	.	2005	2017
8	2016	5	233	0	0	0	0	2005	2015	0	1	0	.	2005	2017
8	2017	5	250	0	0	0	0	2005	2016	0	1	0		2005	2017
8	2018	5	300	0	0	0	0	2005	2017	0	1	0	.	2005	2017

Notes: These are not real data, but an example for explanation purposes

Also, in the BSD, a company might have changed its core business (change of industry code) in some years, but this would not be recorded as the death of the old or the birth of a new firm in the particular year of such events. Additionally, the BSD does not include entry into the population of firms due to mergers, breakups, split-offs, or other restructuring.

A.2 Assumptions applied to BSD to run DynEmp3

We assume the following assumptions to consider some of the above issues with the BSD and improve the quality of the estimated indicators, first, regarding the death of a firm. We assume that the error is consistent across firms and define a death year for any enterprise as the maximum death year reported for that company by the BSD. Once the maximum year of death is selected, we drop that firm's yearly observations beyond that death year. In that way, the last year a firm appears on the panel is the assumed exit year as required by DynEmp3.

A.2.0.0.1 Death of a firm This first assumption disregards any intermittent inactive periods of the firm. This is another form to apply the concept of reactivations but in a broader manner. I consider that regardless of the company's disappearing from the BSD cross-sectional files for 1, 2, or more years intermittently throughout the period, the death year of the firm shall be the maximum death year reported for that firm. In that case, if a death year is never reported in any of the cross-sectional files of that firm in the BSD, that firm is considered to have remained active for the period. To illustrate this, see again Figure A.5. Columns birth (DynEmp3) and death (DynEmp3) are the assumed birth and death years, respectively, for fictional firm 8. I consider the death of the firm to occur in the year 2017 because this is the maximum death year reported. In cases of firms such as firm 4 in Figure A.3, when the firm has intermittently shown consecutive inactive periods, it does not matter whether the firm was inactive for 2 years only or more; the actual firm death will occur the final death year reported. Thus, firm 4 is considered to be active and not have died during the period of analysis.

DynEmp3, on top of that, will assess the activity status of the firm because it considers a firm active if it has an employment non-missing value. For instance, for the period for which such a firm appears with non-missing values of relevant variables such as employment, turnover, industry code, birth year, or local unit information.

A.2.0.0.2 Birth year The second assumption is in relation to the birth of a firm. The BSD includes birth year information for each firm. This can be outside (i.e., earlier than) the period covered by the business register. It can also be missing for some firms. In the latter case, DynEmp3 takes the first year in which the firm appears with non-zero employment in the register as the entry year, unless this year coincides with the first year of the business register (in which case the birth year is left missing). Thus, entering firms without birth year information will be only firms with some employment information; it cannot be missing values.

A.2.0.0.3 Industry classification Another critical assumption is the industry classification correspondence across years. The BSD contains information for industry classification based on UK Standard Industrial Classification (SIC) versions 1992, 2003, and 2007.

The SIC 2003 is essentially identical to SIC 1992 at the 3-digit level and above[112]. Thus, we assume that SIC 2003 is equivalent to SIC 1992 up to the 3-digit level. So, if a company has not had an available SIC 2003 classification for the years between 1998 and 2002, the SIC 2003 code for a company is to be assumed as the same SIC code it had between 1992 and 2002. The latest OECD DynEmp output, which included the UK in a

cross-country study, also made this assumption for years between 2001 and 2002 (OECD 2018).

Also, the International Standard Industrial Classification of All Economic Activities (ISIC) revision 4 has a one-to-one correspondence with the UK SIC 2007 up to the 3-digit level. Thus, we assume that the three-first digits of industry codes for companies' information between the years 2003 and 2007 are equivalent to the three-first digits according to an ISIC rev. 4 classification.[113]

DynEmp3 analysis excludes by default public administration sectors such as the public sector (Ind A38 = 84) and households and extraterritorial activities (Ind A38 = 97). The OECD DynEmp team recommends the exclusion for the national level and the manufacturing vs. services analysis sectors such as coke and refined petroleum products (19) 'manuf.', real estate activities (68) 'serv.' and construction (41). However, construction was not omitted in the OECD 2018's cross-country study on business dynamics and digitization. In Appendix B.7, additional UK-level figures are presented for some indicators excluding all these three sectors. Manufacturing includes (ind_{a38} codes 10 to 33) and services include (ind_{a38} codes 45 to 63 and 68 to 82). Services (also known in the program as Non-market Services) exclude ind_{a38}'s codes corresponding to finance because of measurement issues in this sector (ind_{a38} codes 64 to 66). No major differences were found.

The A7 macrosector classification (ind_{a7} =7: Non-Market Services), excludes from its non-market services macrosector, the ind_{a38}'s codes corresponding to finance because of measurement issues in this sector (64 to 66). This classification is not presented in this paper but serves as a reference.[114]

A.2.0.0.4 Preference for indicators: excluding self-employment firms? The OECD also recommends estimating national-level variables only for all firms, meaning with all employees in employment, when it is difficult to measure self-employment. In their 2018 cross-country study, they just showed employer firms' rates. In the present study, most of the statistics account for all firms and only selected indicators are estimated only for firms with at least one person engaged.

A.3 Comparability between U.K. and U.S. data for business dynamics analysis

The LBD feeds from the U.S. Business Register (B.R.), a comprehensive and secure database of all U.S. business establishments developed and maintained by the U.S. Census Bureau, with data from 1975 to the present.

In contrast to the U.K., the U.S. employment indicators report business dynamics excluding proprietors (employers in employment or self-employment) and partners of unincorporated businesses. Therefore, to make the various indicators comparable, the U.K. sample of firms is restricted to those with at least one person engaged; in other words, the U.K. sample for this comparative analysis excludes employers in employment or self-employment firms.

Figures presented here for the U.S. are estimated using the USBDS pre-estimated variables on yearly employment, the number of births, deaths, and active firms. I am

following the methodology of Decker et al. (2016b); De Loecker et al. (2018).

The leading U.S. business demographics indicators are primarily comparable with the ones presented for the U.K. for three reasons. First, the definitions of a firm and an establishment are similar. The USBDS data measures the net change in employment at the establishment level. An establishment is "a single physical location where business is conducted, or services or industrial operations are performed." This definition is comparable to the U.K.'s local or reporting unit definitions. In addition, according to the USBDS, a firm is a business organization consisting of one or more domestic establishments specified under common ownership or control. The firm and the establishment are the same for single-establishment firms. As explained earlier, the same is considered for the U.K. An enterprise is a unit that carries out activities at one or more locations through a local unit, which could be the firm itself.

Second, the definitions of Job Creation and Job Destruction in the USBDS are similar to those presented in the section 1.2.4. U.S. Job Creation is the sum of all employment gains from expanding establishments from year t–1 to year t, including establishment startups. U.S. Job destruction is the sum of all employment losses from contracting establishments from year t–1 to year t, including establishments shutting down. So, these definitions are equivalent to equations 3 and 4.

US $JCR_{c,t}$ indicates the Job Creation Rate for the U.S. and $USJDR_{c,t}$ displays the job destruction rate for the U.S.

Third, the U.S. defines births and deaths of firms in a similar fashion as in the present study. The USBDS defines a birth year as the year an establishment first reports positive employment in the LBD. This definition of birth corresponds to entering a firm with non-zero numbers of persons engaged at year $t, I_t^{e \text{ ent, per}}$.

That is why the U.S. statistics would be congruent with comparable statistics using the sample of firms labeled as "excluding employers in employment," where only firms with at least one person engaged besides the employer are considered. As explained earlier, the U.S. statistics could be comparable with the "employer birth" definition of the U.K. Business Demographics publication. Also, both databases consider similar definitions for the death of a firm or a firm's exit. The USBDS defines a firm death as an event when all the establishments associated with that firm cease all operations. So if a firm has 100 establishments linked and only 99 of those end operations, the firm is still active. This is similar to the BDS approach, which considers only a firm as active when it has positive turnover or employment and at least one administrative unit linked to it.

Finally, the USBDS considers that legal entities that cease to exist because of merger and acquisition activity do not qualify as firm exits in these data. This is comparable with the U.K. case because the U.K. BSD does not account for information on mergers and acquisitions in the dataset.

However, there is one difference. The U.S. definition of employment does not include proprietors, so their rates exclude employers in employment. DynEmp3 suggests using the value of total employment for a country. When that variable is unavailable and only the one indicating the number of persons engaged, the code automatically adds one to have a variable of total employment accounting for proprietors, too. Thus, the estimations of job reallocation for the U.K. also include employers in employment. Hence,

the comparison between the two countries needs to consider that job reallocation in the U.K. contains self-employment, and the U.S. does not. So, although they could not be directly comparable at the levels, they could be comparable in terms of trends because the patterns between rates with and without self-employment in the U.K. are similar.

A.4 Definitions of Entry and Exit Rates for the U.S.

Following the methodology by Javier Miranda from the U.S. Census Bureau, the firm entry rate refers to new firms as a percentage of all firms (US $ENR_{c,t}$) as defined in the follow equation:

$$USENR_{c,t} = \frac{I^{ent}_{c,t}}{I^{incb}_{c,t} + I^{ent}_{c,t}} * 100$$

Where I is the total number of firms. $I^{ent}_{c,t} = I^{age=0}_{c,t}$, where age = 0 indicates the group of startups firms that year. Birth year is the year an establishment first reports positive employment (equivalent to having at least one person engaged). $I^{incb}_{c,t} = I^{T}_{c,t} - I^{age=0}_{c,t}$, where $I^{T}_{c,t}$ is the total number of firms (of all ages) in that year t, and c is economy-wide for year t.

The U.S. entry rates in terms of employment $USENRE_{c,t}$ or also called employment-weighted annual firm entry rate, is new firm employment as a percent of all work:

$$USENRE_{c,t} = \frac{E^{ent}_{c,t}}{E^{incb}_{c,t} + E^{ent}_{c,t}} * 100$$

Where, E is total employment following the U.S. Census Bureau definition, $E^{ent}_{c,t} = E^{age=0}_{c,t}$ and $E^{incb}_{c,t} = E^{T}_{c,t} - E^{age=0}_{c,t}$, and c is economy wide for year t. The firm exit rate ($USEXR_{c,t}$) is exiting firms as a percent of all firms as defined per the equation:

$$US\,\mathrm{EXR}_{c,t} = \frac{I^{ext}_{c,t}}{I^{incb}_{c,t} + I^{ext}_{c,t}} * 100$$

And the U.S. exit rate in terms of employment or also called employment-weighted annual firm exit rate (US EXRE $E_{c,t}$) is exiting firms' employment as a percent of all work:

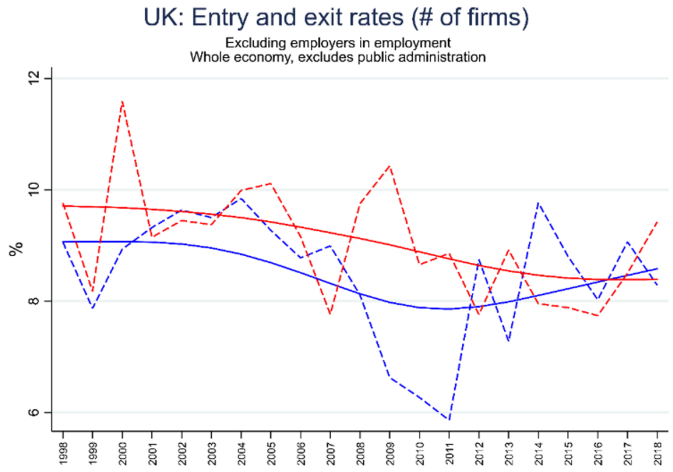
$$\text{US EXRE } E_{c,t} = \frac{E^{ext}_{c,t}}{E^{\text{incb}}_{c,t} + E^{ext}_{c,t}} * 100$$

Where $I^{ext}_{c,t} = I^{\text{deaths}}_{c,t}$, indicating the number of firm deaths as already estimated in the U.S. Census Bureau public database, and $E^{ext}_{c,t} = E^{\text{deaths}}_{c,t}$, or employment destroyed in t-1. It is essential to understand the differences between the UK and the US's entry (ENRE) and exit (EXRE) rates based on employment (right-hand side of Figure 16). DynEmp3 uses a particular approach to estimate EXRE. As mentioned in the entries and exits definitions section, this rate includes employment created by self-employment firms and includes in period t the employment of dead firms in t-1. In contrast, the U.S. figures use the Census definition of employment, which excludes self-employment firms. Thus, EXRE can have a lower denominator than ENRE and consequently appear higher than the latter for the UK. These methodological differences only permit inference over the trends between the two countries rather than precisely at the levels. Figure A.6

shows estimations for ENR and EXR for the UK that are comparable with US firm rate versions.

Specific methodological differences between the two countries' computations only permit inference over the trends between the two countries rather than at the levels. Figure A.6 shows estimations for ENR and EXR for the U.K. that are comparable with U.S. firm rate versions.

FIGURE A.6
Entry and Exit Rates excluding employers in employment



Notes: Both figures exclude the public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code = 97), coke and refined petroleum products (19), real estate activities (68), and construction (41).

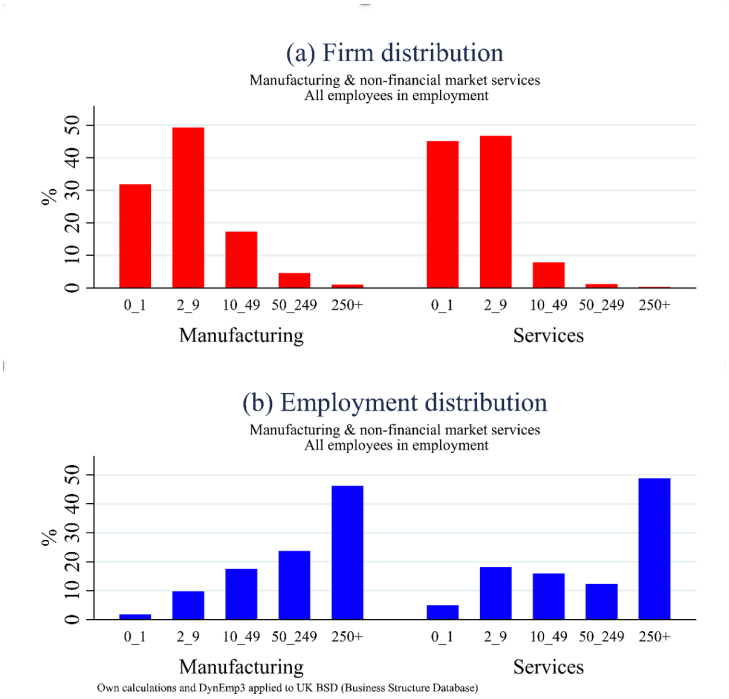
Appendix B DynEmp3

B.1 Fact BD: Uni-personal firms gains terrain in size distribution

Figure B.7 shows the 1998-2018 average size distribution for two macro sectors, manufacturing and services, considering all employees in employment. The upper graph (a) shows the average size composition of firms in each sector for the period, and the bottom graph (b) presents the average proportion of employees in each size class in each sector.

In both sectors, most of the firms are micro and medium-sized firms with between 2 and 49 employees; in services, however, self-employment firms $(0-1)$ account for about 45% of all firms on average. Smaller firms are more prevalent in the services sector than in the manufacturing industry; 95% of all firms (by number) in the services sector have between 1 and 9 employees in employment ($0-1$ and $2-9$ labels combined), whereas, in manufacturing, the proportion is lower, at 82%. Also, as in other industrialized countries, larger firms tend to employ most people in the economy. For both sectors, firms of 250+ employees hire around 50% of total employment. The services sector exhibits a slightly higher proportion than manufacturing.

FIGURE B.7
Size Distribution

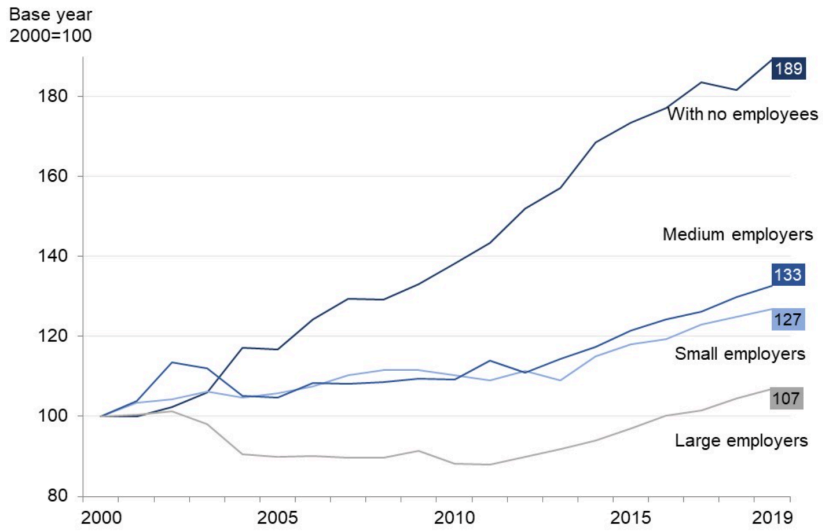


Notes: • This figure reports the size distribution in the UK for five size classes, separate for manufacturing and non-financial market services on average for the 1998-2018 period. Shares are calculated regarding (a) the number of units and (b) employment. • (a) and (b) exclude the public sector (ind_{a38} code =84) and households and extraterritorial activities (ind_{a38} code =97). • Manufacturing includes (ind_{a38} codes $10-33$) and services include (ind_{a38} codes 45-63, 68-82). The OECD DynEmp team recommends dropping from manufacturing, coke, refined petroleum products (19), and services real estate activities (68). Here, they were not omitted. There are no significant differences when these other sectors are excluded: Figure (a) shows the distribution of firms' size remains practically identical. Concerning figure (b) 250+ manufacturing group slightly decreases the average number of employees hired, increasing the gap relative to 250+ firms in the services sectors. Removing real estate from services does not alter much the hierarchy of the company sizes employing the most people. • Employee means one unit irrespective of the hours worked within the business

Furthermore, according to the UK's government business department, the UK private sector business population growth since 2000 has mainly been due to increasing numbers of non-employing businesses[115], equivalent to self-employment in this study, as is illustrated in Figure B.8. Thus, figure B.8 shows that since 2000, the number of non-employing companies has increased by 2.1 million (+89%), much more than the growth of other size classes, such as large, medium, and small firms. These trends and those portrayed in terms of the evolution of numbers of firms by size indicate that size distribution has transformed towards one composed chiefly of unipersonal firms. Moreover, considering B.7b, a few larger firms employ most people.

Accordingly, most companies in the U.K. economy are much smaller now than before the 2000s. This pattern could indicate a positive sign of dynamism since firms with only one person engaged have claimed territory against other size classes.

FIGURE B.8
Growth in the number of firms by size, 2000-2019



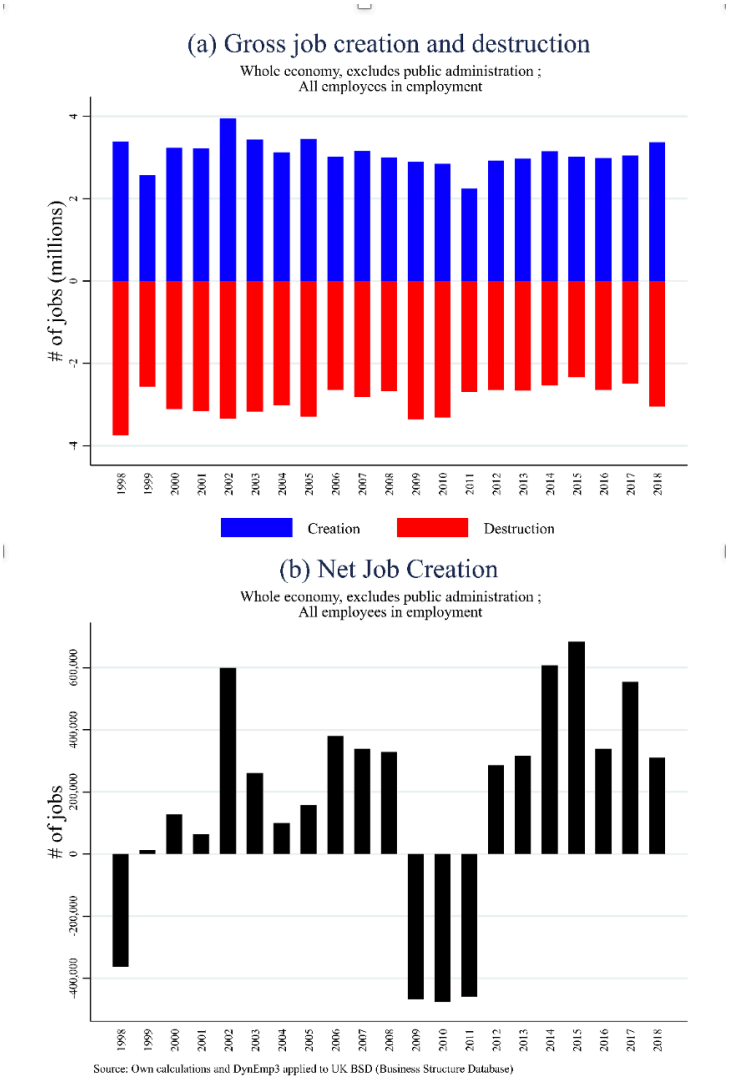
Source: Business population estimates for the U.K. and regions: 2019 statistical release (HTML). UK Department for Business, Energy & Industrial Strategy (2019)
https://www.gov.uk/government/statistics/business-population-estimates-2019/business-population-estimates-for-the-uk-and-regions-2019-statistical-release-html
Note: Index: base year = 2000

B.2 Fact BD: Job Creation and Destruction stable except during the financial crisis and Brexit vote

Figure B.9 presents yearly trends in gross job creation and destruction for the whole economy, considering all employees in employment. During the analysis period, net job creation was only negative right before the 2000s and during the financial crisis in the immediate years. On average, between 2009 and 2011, 467 thousand jobs were destroyed. Thus, assuming 261 working days per year, around 1790 jobs were lost every working day during those three years. Taking out the negative years of net job creation, on average, in the UK, every year, around 321 thousand new jobs have been created.

Net job creation exhibits similar trends when excluding employers in employment but seems different in pre-crisis years. As shown in Figure B.14 in the appendix, the "jobs creation power" during 2006-2008 seems to be explained mainly by self-employment.

FIGURE B.9
Job Creation and Destruction



Notes: (a) This figure reports the contribution of net job creation to aggregate employment change by year. It is defined as net job creation (i.e., The difference of total employment at time t and $t-1$ over 1998-2018 (b). This figure reports the number of jobs created and destroyed by all firms, including entrants, incumbents, and exiting firms, annually. Job creation = Number of jobs created by the entrant and incumbent firms in a given year. Job Destruction = Number of jobs destroyed by incumbent firms shrinking and firms exiting the market in a given year. Source: DynEmp3 outputs (level 3) and own calculations applied to UK Business Structure Database (BSD). (a) and (b) exclude public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code =97). This figure includes firms with 0 persons engaged, meaning only one employer can integrate firms in employment.

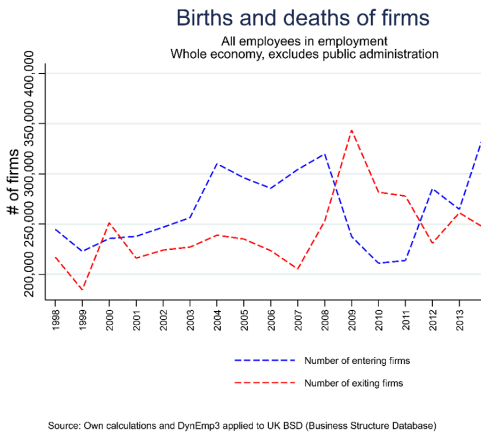
B.3 Fact BD: Births and Deaths of Firms Recover after financial crisis but Decline After Brexit vote

Figure B.10a depicts the number of births and deaths of firms in the U.K. between 1998 and 2018. After a hiatus during the financial crisis in 2009, the births of enterprises recovered the upward trend they had been experiencing years earlier. Meanwhile, the deaths of firms exhibit a more stable pattern for the period, increasing temporarily during the financial crisis but decreasing right after. Nonetheless, exits start rising again after 2016. When these numbers are estimated only for firms with at least one person engaged, the gap between the number of births and deaths is substantially reduced. This is shown when comparing Figure B.10a and Figure B.10b where the gap between births and deaths

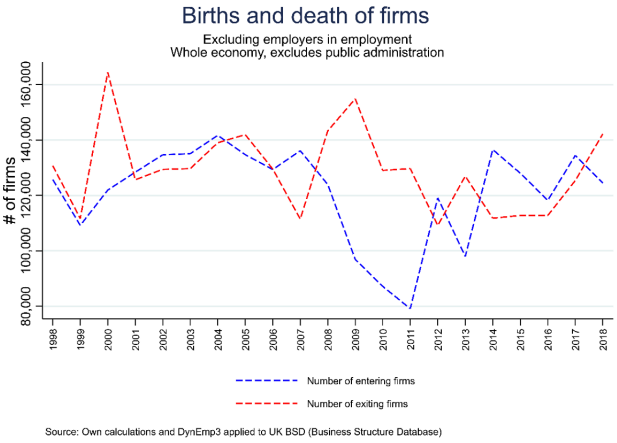
for the self-employment firms (or unipersonal, companies with no persons engaged) is narrower throughout the period, except for the financial crisis.

FIGURE B.10
Entries and exits

(a) Employers in employment **included**



(b) Employers in employment **excluded**

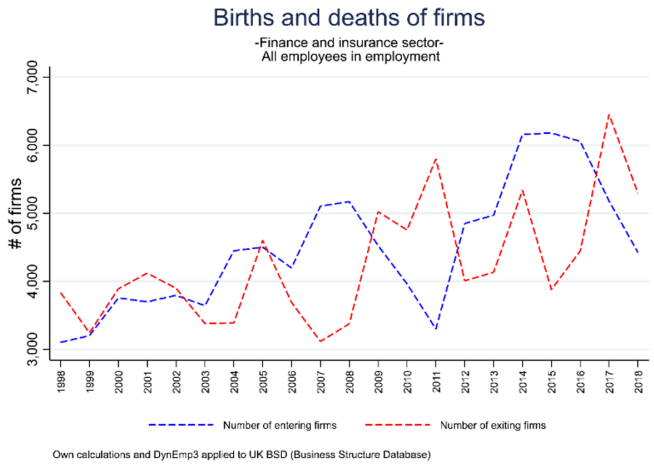


The rate versions of these indicators provide more information beyond the absolute values of current employment stocks.

B.4 Fact BD: The financial sector is more dynamic than the rest of the economy

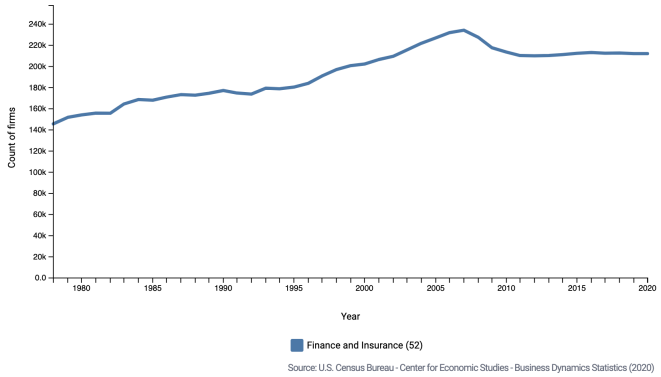
Figure B.11 presents births and deaths in the financial sector. Both curves show an upward trend for the period. Except for the decline and increase in entries and exits during the financial crisis, the industry exhibits positive dynamics. However, after the Brexit vote, such dynamism decreased again. This pattern is similar to the dynamism preserved for this sector in the U.S.; see Figure B.12.

FIGURE B.11
Entry and Exit excluding Employers in employment in the U.K.



Notes: Both figures exclude the public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code = 97), coke and refined petroleum products (19), real estate activities (68), and construction (41).

FIGURE B.12
Entry excluding Employers in employment (U.S.)



Notes: The figure includes the number of firm births in the U.S. as reported by the U.S. Census Bureau (2020).

B.5 Fact BD: Brexit's vote cuts up short business dynamics recovery

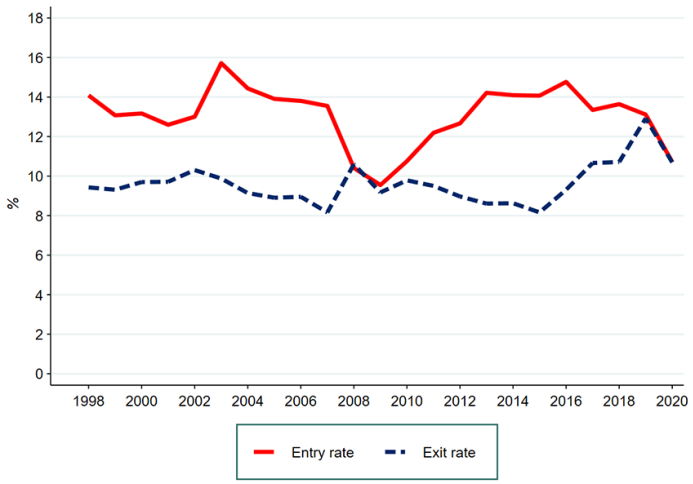
Like other countries presented in the current analysis, the U.K. experienced similar declines in dynamism throughout the period. However, the U.K. exhibited recovery trends after the financial crisis; a recovery apparently cut up short after 2016, the Brexit vote year. Entry rates seemed to recover and went down again, as did the share of net job creation of young and small firms; it was smaller in 2018 than in 2015, almost reaching the lower levels of 1998 (See Figure 7).

The assumptions used in this study to build the database imply a careful interpretation of results for the last three years, 2016-2018. Especially for those indicators that consider the deaths of firms. For Instance, DynEmp3 might inflate the number of deaths in 2018 because it considers a death of a firm in the year 2018 if the firm was present in 2017 but was not present in 2018; in contrast, the U.K. Business Demographics Report would only consider a death in 2018 if they wait for another year (2019) and confirm that it did not

appear active for a second consecutive year. Remember that the definition of death for that report is that if a firm in year t is not found involved in t+1 and t+2. But in the case of DynEmp, since we do not have that t+2 year of info to confirm the death, the death was still accounted for in 2018.

Notice, however, that subsequent studies replicating similar exercises with the Business Register report an increase in exits and a decrease in entries after 2016's Brexit vote. So, although exits are potentially higher in the present study for the final year of the sample, it is aligned with other findings (see Figure B.13 from the latest U.K. Competition and Markets Authority report on competition Competition and Markets Authority (2022)).

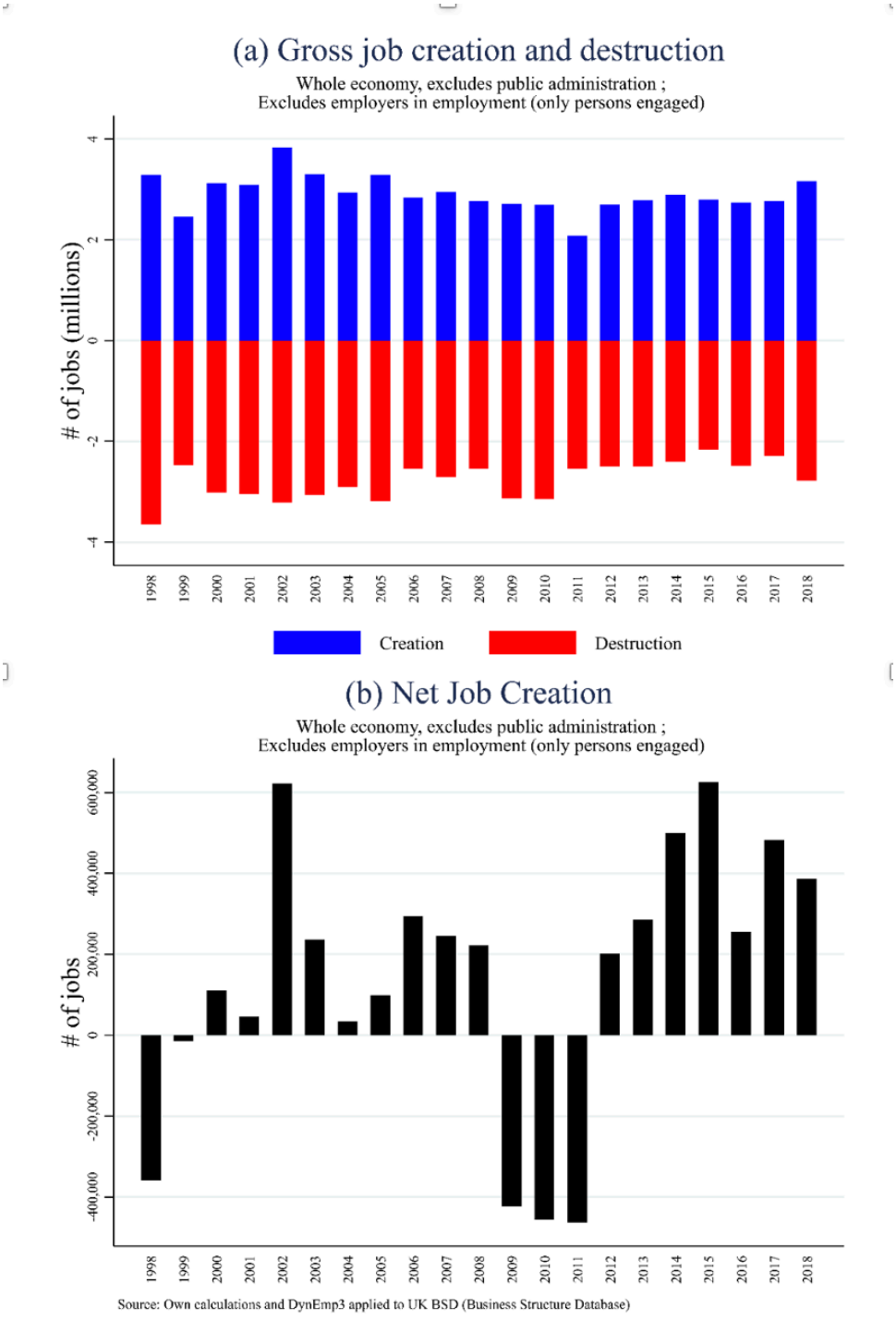
FIGURE B.13
Entry and Exit Rates according to the CMA



Source: CMA analysis of ONS BSD data Competition and Markets Authority (2022).

B.6 Job Creation and Destruction

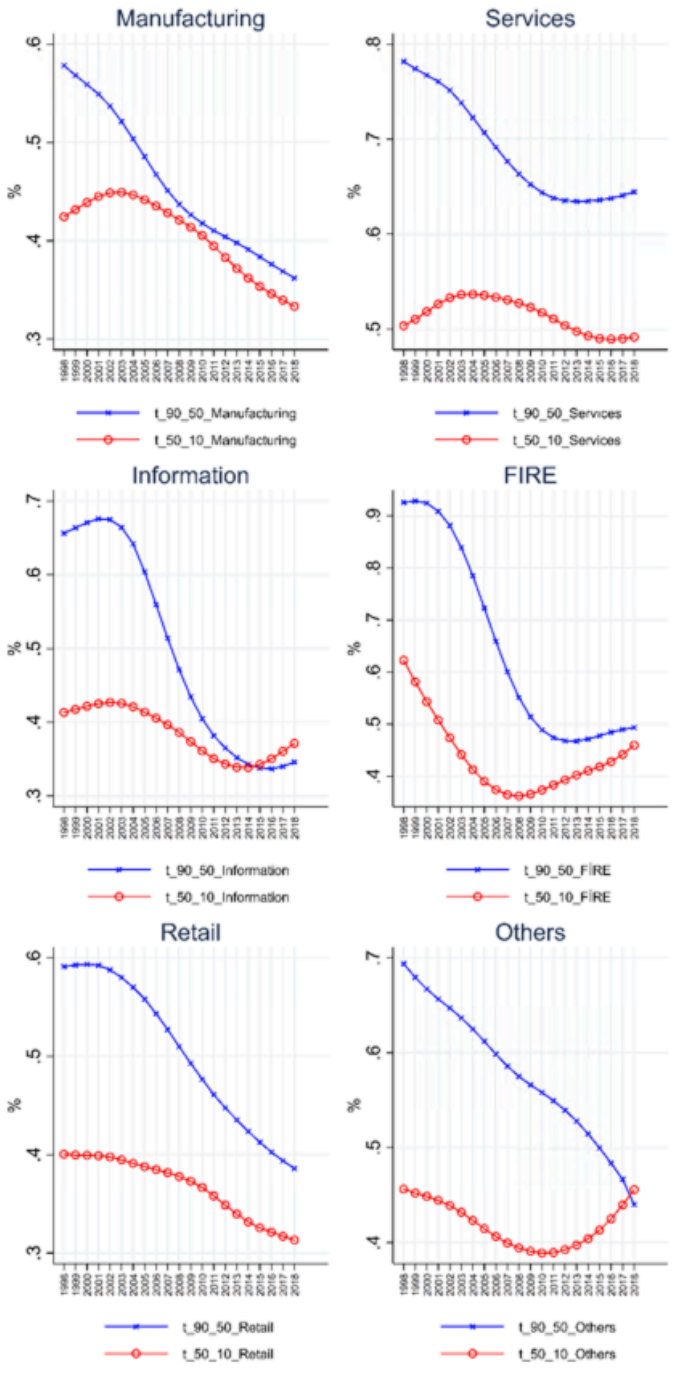
FIGURE B.14
Job Creation and Destruction



Notes: (a) and (b) exclude the public sector (ind_{a38} code = 84) and households, extraterritorial activities (ind_{a38} code =97). This figure only includes firms with at least one employee besides the owner. DynEmp3 estimates these variables based on their definition of employment (total number of persons engaged, meaning owner plus employees).

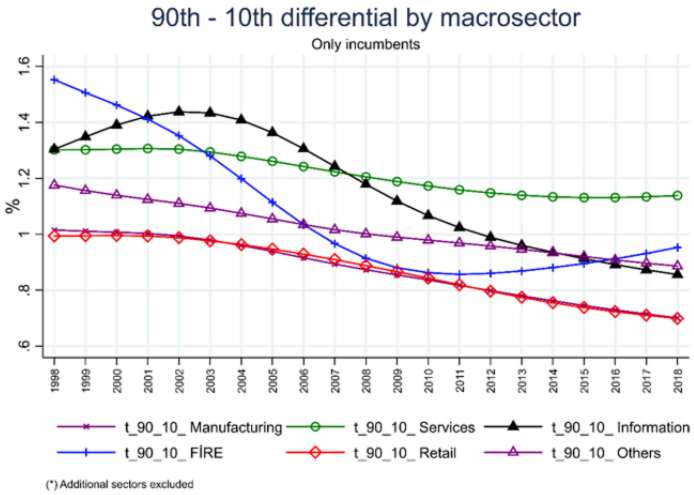
B.7 Excluded sectors

FIGURE B.15
Dispersion Differentials (excluded sectors)



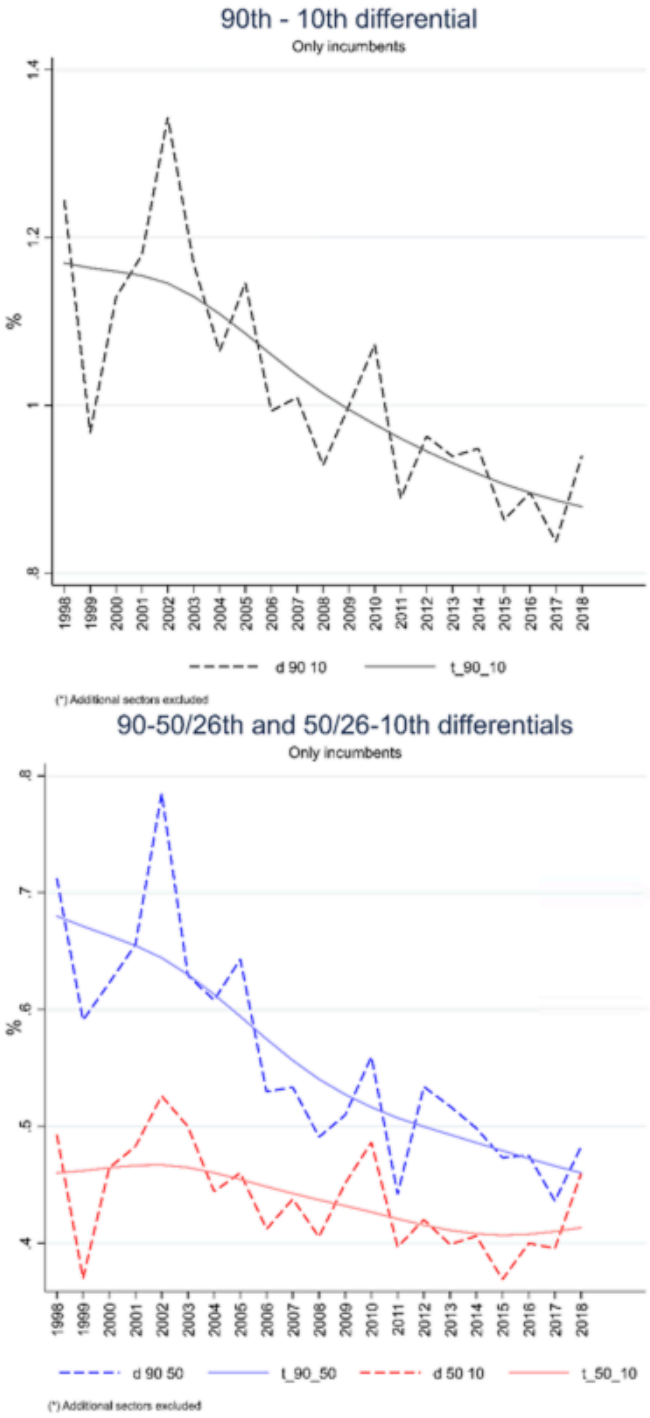
Notes: 1) Both figures exclude public sector (ind_a38 code = 84) and households, extraterritorial activities (ind_a38 code =97), coke and refined petroleum products (19), real estate activities (68) and construction (41). 2) Only include employer firms (employment >= 2).

FIGURE B.16
Dispersion Differentials by macrosector (excluded sectors)



Notes: 1) Both figures exclude public sector (ind_a38 code = 84) and households, extraterritorial activities (ind_a38 code =97), coke and refined petroleum products (19), real estate activities (68) and construction (41). 2) Only include employer firms (employment >= 2).

FIGURE B.17
Entry and exit (excluded sectors)



Notes: 1) Both figures exclude public sector (ind_a38 code = 84) and households, extraterritorial activities (ind_a38 code =97), coke and refined petroleum products (19), real estate activities (68) and construction (41). 2) Only include employer firms (employment >= 2).

Appendix C Regressions

C.1 Arellano Bond

For the AB strategy, we differentiate 19 on both sides to remove the fixed effects. To address the endogeneity between the first difference of the error and the first difference of the dependent variable, we apply a 2SLS-instrumental variables approach in a GMM framework.

Once we differentiate equation 19 the fixed effect μ_s disappears, but $\Delta y^s_{c,t-1} = y^s_{c,t-1} - y^s_{c,t-2}$, and all the first differences of other predetermined variables are correlated with the first difference error $\Delta\vartheta^s_{c,t}$; $y^s_{c,t-1}$ is a function of $\vartheta^s_{c,t-1}$, which is also in $\Delta\vartheta^s_{c,t} = \vartheta_{c,t} - \vartheta^s_{c,t-1}$.

Thus, we use the combined approach of using further lags of the differentiated dependent variables and further lags of the levels of the dependent variables as instruments, as per Anderson and Hsiao (1981) and revisited Arellano and Bond (1991) approaches. Thus for $\Delta y^s_{c,t-1}$, we use also lags such as $\Delta y^s_{c,t-2}$ or $\Delta y^s_{c,t-3}$. For the lagged levels of the dependent variable, we use $y^s_{c,t-2}, y^s_{c,t-3}$, which are expected to be uncorrelated with $\vartheta_{c,t}$ but also (hopefully) with $y^s_{c,t-1}$.

When considering endogenous or predetermined covariates, this also applies. For endogenous covariates, we use $\vec{x}^s_{c,t-2}, \vec{x}^s_{c,t-3}, \ldots$ as instruments, and for predetermined covariates, we can use as an additional instrument $\vec{x}^s_{c,t-1}$.

Thus, after removing the fixed effect, the lagged dependent variables' values (levels) constitute legitimate instruments for the first-differenced variable, provided that the residuals are free from second-order serial correlation. Arellano and Bond's (relatively limited) Monte Carlo simulations show that their AB (or *difference GMM*) method outperforms OLS and fixed effects (FE) estimators when the regression residuals are uncorrelated.

The estimators are constructed based on moment equations built from further lagged levels of $y^s_{c,t}$ and the first-differenced errors $\vartheta^s_{c,t}$. Since we have more instruments than parameters, we use the GMM framework.

Thus, we have the AB diff-GMM or first-differenced model from general equation 19:

$$\Delta y^s_{c,t} = \sum_{j=1}^{p_y} \theta_j \Delta y^s_{c,t-j} + \sum_{j=0}^{p_x} \Delta \vec{x}^s_{c,t-j} \vec{\beta}_j + \Delta\vartheta^s_{c,t} \tag{76}$$

For which we have the moment conditions of the form:

Lagged dependent variable:

$$E\left[y^s_{c,t-m} \Delta\vartheta_{c,t} c^s\right] = 0, \quad m = 2, 3, \ldots, t$$

Strictly exogenous covariates:

$$E\left[\vec{x}^s_{c,t-m} \Delta\vartheta^s_{c,t}\right] = \overrightarrow{0}, \quad t - m = 0, 1, \ldots, T$$

Predetermined covariates:

$$E\left[\vec{x}^{s}_{c,t-m}\Delta\vartheta^{s}_{c,t}\right] = \overrightarrow{0}, m = 1, 2, \ldots, t$$

Endogenous covariates:

$$E\left[\vec{x}^{s}_{c,t-m}\Delta\vartheta^{s}_{c,t}\right] = \overrightarrow{0}, m = 2, 3, \ldots, t \text{ with } t = m, \ldots, T.$$

Strict exogeneity of covariates requires that $\vartheta^{s}_{c,t}$ be unrelated to $\text{ShIntan}^{y,s}_{c,t}$, $\text{InvLabRatio}^{s}_{c,t}$ and $\text{wages}^{s}_{c,t}$ for any $m > t$, which rules out negative-economic shocks that simultaneously, for example, slow employment creation and impact negatively, in the future, investment in intangibles, wages or the ratio of investment to labor ratio. Because unforecastable errors today might affect future changes in the real wage and the capital stock, we suspect that the natural logarithm of the real product wage and the natural logarithm of the gross capital stock are predetermined instead of strictly exogenous. Thus, we treat investment, capital-to-labor ratio, and wages as predetermined and use their lagged levels as instruments.

The stacked moment conditions look as follows:

$$E\left[\vec{q}_{s}(\delta)\right] = E\left[\vec{Z}^{D\prime}_{s}\Delta\vec{\vartheta}_{s}\right] = 0 \tag{77}$$

Where $\vec{\delta} = [\vec{\theta}, \vec{\beta}]\Delta\vec{\vartheta}_{S} = (\Delta\vartheta_{s2}\Delta\vartheta_{S3}\ldots\Delta\vartheta_{ST})$ and $\vec{Z}^{D}_{s} = \left(\vec{Z}^{D}_{ys}, \vec{Z}^{D}_{xs}\right)$, where is the GMM-type instruments matrices with $\vec{Z}^{D}_{yS}[T \times (T-2)]$ and $\vec{Z}^{D}_{xs}[T \times (T-2)]$.

When the ϑ_{st} is serially uncorrelated and homoscedastic, the optimal weighting matrix is independent of $\vec{\delta}$ such that we can use the one-step instead of the two-step-estimator. However, when there is a risk of heteroskedasticity, one can compute panel-robust or cluster-robust standard errors with STATA, and the one-step GMM estimator remains consistent. Yet, it is no longer efficient [116].

To exemplify the total number of instruments that could emerge in this model, let's assume that there is only one lag of the dependent variable, k=3 regressors that are considered exogenous, T=21 available years in total, and p = T-2 = 21-2 = 19 periods from which to form moment conditions.

Suppose we have k=3 regressors (without including year dummies) that are strictly exogenous. In that case, each contributes with one instrument; in addition, there are 21 – 2 instruments available in periods p = 3, 4, ..., 21 in this way: in period 3, y^{c}_{i1} is a valid instrument for y^{c}_{i3}; in period 4, y^{c}_{i1} and y^{c}_{i2} valid instrument for $y_i 4^{c}$, and so on.

Thus, with unlimited GMM-type instruments, there are at least 193 = 3 + 19*(19+1)/2 moment conditions for the sys-GMM AB model. The dimension of the GMM-type instrument matrix grows as the number of periods and endogenous regressors expands. A large number of instruments, very close to the number of total observations, may imply over-identification problems, so we do apply some collapse (Instead of the "GMM-type" instruments, use "standard" instruments) and curtailing (only using a limited number of lags as instruments) techniques to limit instruments used.

C.2 Blundell Bond

We also test the Blundell-Bond method or *system GMM* estimation. This system estimates a two-equation system of the regression in levels and first differences. According to studies on best empirical strategies for addressing dynamic panels, BB is one of the most accurate methodologies.

In AB, the lagged levels may provide little information about the first differenced variable, particularly if they are serially correlated (Arellano and Bover (1995), Blundell and Bond (1998)). The instruments $y_{s,t-2}$,$y_{s,t-3}$,... are weakly correlated with the first difference lagged dependent variable $\Delta y_{s,t-1}$ when θ_j -> 1. In particular, when T is small, the diff-GMM estimator could be substantially biased. We do not think that this can be a problem in the setting of the data used here because we have fewer panels than periods (N=16 < T=21), so we do not have a particular small T; however, when conducting unit root tests accounting for panel data, we find that some of the Y's series are non-stationary[117]. BB assumes the absence of second-order serial correlation. Although the latter might be a problem, we do not observe second-order serial correlation when testing the models. Hence, we expect not to impact the performance of these BB estimators.

BB, in addition to the first differencing used by AB, uses the lagged first differences as instruments in the non-transformed (levels) equation, equation 19. For this system-GMM, we add to the moment conditions specified in the AB section for the first-differenced model in 76, the following moment conditions for the level model in 19:

Lagged dependent variable:

$$E\left[\Delta y_{s,t-1}^{c}\left(\mu_s+\vartheta_{st}\right)\right]=0, \quad t=2,3,\ldots,T$$

Strictly exogenous covariates:

$$E\left[\Delta \vec{x}_{s,t}\left(\mu_s+\vartheta_{st}\right)\right]=\overrightarrow{0}, \quad t=1,2,\ldots,T$$

Endogenous covariates:

$$E\left[\Delta \vec{x}_{s,t-1}\left(\mu_s+\vartheta_{st}\right)\right]=\overrightarrow{0}, t=2,3,\ldots,T$$

The stacked moment conditions are:

$$E\left[\vec{q}_s(\delta)\right]=E\left[\begin{pmatrix} \vec{Z}_s^{\mathrm{D}'} & \Delta\vec{\vartheta}_s \\ \vec{Z}_S^{L'} & \vec{\varepsilon}_S \end{pmatrix}\right]=0$$

Where $\overrightarrow{\varepsilon_s}=\left(\mu_s+\vartheta_{s1},\mu_s+\vartheta_{s2},\ldots,\mu_s+\vartheta_{sT}\right)'$ and $\vec{Z}_s^L=\left(\vec{Z}_{yS}^L,\vec{Z}_{xS}^L\right)$ where is the GMM systemtype instruments matrices with $\vec{Z}_{ys}^L[T\times(T-1)]$ and $\frac{L}{xs}[T\times(T-1)]$.

C.3 Long difference estimations

In part because the AB and BB instruments are invalidated by second-order autocorrelation (Baltagi (2008)), Hahn et al. (2007) derived a *long difference* instrumental

variable estimation technique. Assuming balanced panels, Hahn et al. show that combining multi-period differencing with longer lagged instrument choices can produce less biased estimates than the AB or BB approaches. For this reason, we also test this econometric technique.

The LD estimator improves on the first differences. To explicitly test the long-term effects of intangibles investment intensity, we mimic long differences estimators; these models analyse only two observations, separated by a long period, and measure the effect of changes in the independent variable (share of intangibles) on changes in the dependent variable (any business dynamics variable) over many years. To maximize the number of cases included, we compare sectors employment growth rate dispersion differential in 1998 (share of intangibles and other independent variables as well) with sectors' values in the same variable in 2016 (a 16 year gap). Models include sector and year-fixed effects and estimate standard errors clustered by sector.

$$y_{it} = \gamma y_{it-1} + x_{it}\beta + \mu_i + \vartheta_{it}, \quad t = 2, \ldots, T$$

LD yields to

$$y_{iT} - y_{i2} = \gamma\left(y_{iT-1} - y_{i1}\right) + \left(x_{iT} - x_{i2}\right)\beta + \vartheta_{iT} - \vartheta_{i2}$$

Where y_{i1} is available as an instrumental variable for $y_{iT-1} - y_{i1}$ And, $y_{i1}, y_{iT-1} - \hat{\gamma} * y_{iT-2}, \ldots, y_{i2} - \hat{\gamma} * y_{i1}$ permits to estimate with a 2 SLS instrumental variables approach and use them as additional instruments in an iterative process using the preliminary consistent estimator $\hat{\gamma}$ to obtain residuals as instruments, based on

$$\begin{aligned}
&E\left[y_{i1} * \left(y_{iT} - y_{i2} - \gamma\left(y_{iT-1} - y_{i1}\right)\right)\right] = 0 \\
&E\left[\left(y_{iT} - \gamma y_{iT-1}\right) * \left(y_{iT} - y_{i2} - \gamma\left(y_{iT-1} - y_{i1}\right)\right)\right] = 0 \\
&\vdots \left[\left(y_{i2} - \gamma y_{i1}\right) * \left(y_{iT} - y_{i2} - \gamma\left(y_{iT-1} - y_{i1}\right)\right)\right] = 0 \\
&DD^c_{iT} - DD^c_{i1} = \beta\left(DD^c_{iT-1} - DD^c_{i0}\right) + \left(\boldsymbol{X}^c_{iT} - \boldsymbol{X}^c_{i0}\right)' + \left(\varepsilon_{iT} - \varepsilon_{i1}\right)
\end{aligned}$$

So after initial IV estimation, we obtain the estimated residuals from the first iteration, re-estimate the LD model using the y_{i1} and the residual from the first iteration, and so on; typically, only three iterations are enough. Hahn (2007) shows superior performance when the coefficient gamma tends to 1.

Thus, we consider this model if we assume that we do not expect yearly or even changes every decade in share in intangible investments to have their full effect on employment dispersion differentials or other economic variables related to business dynamics; we will estimate in the long differences; that is a panel including only two dates $t_0$and t_1(in practice, 1998 and 2016/2018). The long-difference regressions make interpretation easier because they directly measure the effect of change in shares of intangibles between two dates on the difference in business dynamic variables between the exact dates. Since we have only two dates, it is equivalent algebraically to estimate:

$$\Delta y_i = \pi \Delta x_i + \Delta\mu + \Delta Z_i'\beta + \Delta\varepsilon_i$$

Where $\Delta y_i \equiv y_{it_1} - y_{it_0}$, and the rest of deltas are defined similarly. Long-difference specifications with two observations per sector, one for the initial date and one for the final date. We use DD0 as an instrument for the lagged difference for LD. For LD4, we use a yit18 as the initial instrument.

The LD4 Four Period Long Differencing replicates the Huang and Ritter (2009) implementation of Hahn et al.'s (2007) estimator. They consider several differencing windows, but we implement their 4-period variation in our simulations because it excludes the fewest sectors.

C.4 Dispersion differentials

TABLE C.2

Dispersion Differences: Long Differences Regression 90th-50th

	(1) D 90th-50th([t] - [t-17])	(2) D 90th-50th([t] - [t-16])	(3) D 90th-50th([t] - [t-15])	(4) LD4 = D 90th-50th([t] - [t-5]))
L.D 90th-50th([t] - [t-17])	0.795*			
	(0.451)			
L.D 90th-50th([t] - [t-16])		0.457*		
		(0.254)		
L.D 90th-50th([t] - [t-15])			0.129	
			(0.319)	
L.D 90th-50th([t] - [t-4])				0.124
				(0.113)
Share of intangibles ([t] - [t-17])	-1.128			
	(0.836)			
Share of intangibles ([t] - [t-16])		-2.763***		
		(0.617)		
Share of intangibles ([t] - [t-15])			-0.958	
			(1.213)	
Share of intangibles ([t] - [t-4])				-0.250
				(0.210)
ln(labor compensation) ([t] - [t-17])	-0.177			
	(0.250)			
ln(labor compensation) ([t] - [t-16])		-0.295**		
		(0.142)		
ln(labor compensation) ([t] - [t-15])			0.0512	
			(0.239)	
ln(labor compensation) ([t] - [t-4])				-0.287**
				(0.115)
Ratio total investment to labor ([t] - [t-17])	-0.515***			
	(0.173)			
Ratio total investment to labor ([t] - [t-16])		-0.589***		
		(0.156)		
Ratio total investment to labor ([t] - [t-15])			-0.481**	
			(0.225)	
Ratio total investment to labor ([t] - [t-4])				-0.241***
				(0.0869)
Constant	0.206	0.0593	-0.161	0.00508
	(0.229)	(0.100)	(0.166)	(0.0245)
Observations	16	16	16	224
R-squared (adj.)	0.332	0.733	-0.00425	0.0302
Degrees of freedom	4	4	4	4
Chi-2 statistic	18.92	61.60	6.325	12.93

Standard errors in parentheses

L. indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods. Dependent variable is: D 90th-50th

* p<0.10, ** p<0.05, *** p<0.010

Note: Dispersion differential expressed in percentages.

TABLE C.3
Dispersion Differences: Long Differences Regression 50th-10th

	(1) D 50th-10th([t] - [t-17])	(2) D 50th-10th([t] - [t-16])	(3) D 50th-10th([t] - [t-15])	(4) LD4 = D 50th-10th([t] - [t-5]))
L.D 50th-10th([t] - [t-17])	-0.215			
	(0.225)			
L.D 50th-10th([t] - [t-16])		0.766		
		(0.565)		
L.D 50th-10th([t] - [t-15])			0.133	
			(0.175)	
L.D 50th-10th([t] - [t-4])				-0.0268
				(0.142)
Share of intangibles ([t] - [t-17])	-1.187***			
	(0.446)			
Share of intangibles ([t] - [t-16])		0.428		
		(0.689)		
Share of intangibles ([t] - [t-15])			-1.171***	
			(0.405)	
Share of intangibles ([t] - [t-4])				0.102
				(0.139)
ln(labor compensation) ([t] - [t-17])	-0.0528			
	(0.131)			
ln(labor compensation) ([t] - [t-16])		-0.112		
		(0.166)		
ln(labor compensation) ([t] - [t-15])			0.0466	
			(0.0973)	
ln(labor compensation) ([t] - [t-4])				-0.190**
				(0.0762)
Ratio total investment to labor ([t] - [t-17])	-0.0430			
	(0.0967)			
Ratio total investment to labor ([t] - [t-16])		-0.0491		
		(0.141)		
Ratio total investment to labor ([t] - [t-15])			0.107	
			(0.0903)	
Ratio total investment to labor ([t] - [t-4])				-0.101*
				(0.0612)
Constant	0.0976	-0.0140	-0.0343	0.0179
	(0.0976)	(0.116)	(0.0702)	(0.0157)
Observations	16	16	16	224
R-squared (adj.)	-0.0453	-0.249	0.174	0.0202
Degrees of freedom	4	4	4	4
Chi-2 statistic	7.689	2.499	12.03	9.808

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods. Dependent variable is: D 50th-10th
* p<0.10, ** p<0.05, *** p<0.010

Note: Dispersion differential expressed in percentages.

C.5 Job Reallocation Rate

TABLE C.4
Job Reallocation Rates

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Job Reallocation Rate			0.580***	0.355***	0.371***	0.757***
			(0.0581)	(0.0622)	(0.103)	(0.0905)
L2.Job Reallocation Rate			0.234***	0.0682		
			(0.0556)	(0.0579)		
Share of intangibles	0.0802***	-0.114*	0.0196	-0.0357	-0.0248	0.0913**
	(0.0238)	(0.0637)	(0.0610)	(0.0688)	(0.0782)	(0.0377)
L.Share of intangibles			0.0800	0.0537		
			(0.0780)	(0.0734)		
L2.Share of intangibles			-0.0903	-0.121*	-0.107*	-0.0683
			(0.0608)	(0.0620)	(0.0541)	(0.0403)
Ratio total investment to labor	-0.0455***	-0.0248	-0.00548	-0.00777	0.00841	0.0387
	(0.0161)	(0.0213)	(0.00945)	(0.0199)	(0.0126)	(0.0259)
L.Ratio total investment to labor					-0.0312	-0.0546
					(0.0204)	(0.0334)
ln(Labor compensation)	-0.0387***	-0.0629**	-0.00639	-0.0460*	-0.0459	-0.00859
	(0.00634)	(0.0284)	(0.00394)	(0.0266)	(0.0340)	(0.00685)
Constant	0.616***	0.924***	0.136***	0.646**		0.0920
	(0.0701)	(0.315)	(0.0465)	(0.300)		(0.0824)
Year FEs		X	X	X	X	X
Observations	288	288	288	288	288	288
R-squared (adj.)	0.116	0.398	0.712	0.494		
Degrees of freedom	3	35	24	39	0	23
AR1 (p-value)					0.00301	0.00236
AR2 (p-value)					0.936	0.502
Sargan (p-value)					0.0806	0.492
Hansen (p-value)					1	1
Chi-2 statistic						
F statistic	13.53	11.23	30.51	13.31	.	219.7

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Job Reallocation Rate
* p<0.10, ** p<0.05, *** p<0.010

Note: JRR is expressed in percentages

C.6 Entry and Exit Rates

TABLE C.5
Entry rates (Number of companies)

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Entry Rate (N)			0.956***	0.811***	0.732***	0.914***
			(0.0685)	(0.0712)	(0.0291)	(0.0250)
L2.Entry Rate (N)			-0.0480	-0.118*		
			(0.0700)	(0.0698)		
Share of intangibles	-0.00126	0.0466	0.00356	0.00772	0.00392	0.0226
	(0.0143)	(0.0389)	(0.0278)	(0.0325)	(0.0171)	(0.0165)
L.Share of intangibles			0.0224	0.0141		
			(0.0354)	(0.0345)		
L2.Share of intangibles			-0.0305	-0.0486*	-0.0455	-0.0293
			(0.0276)	(0.0292)	(0.0332)	(0.0173)
Ratio total investment to labor	0.00834	0.0697***	-0.00261	0.0118	0.0147***	0.00668**
	(0.00966)	(0.0131)	(0.00429)	(0.00993)	(0.00411)	(0.00257)
L.Ratio total investment to labor					-0.00555	-0.00948
					(0.00920)	(0.00614)
ln(Labor compensation)	-0.0102***	0.0652***	-0.00154	0.00908	0.00323	-0.000797
	(0.00381)	(0.0174)	(0.00176)	(0.0129)	(0.00702)	(0.00131)
Constant	0.199***	-0.659***	0.0389*	-0.0596		0.0315**
	(0.0421)	(0.193)	(0.0200)	(0.144)		(0.0134)
Year FEs		X	X	X	X	X
Observations	288	288	288	288	288	288
R-squared (adj.)	0.0903	0.218	0.830	0.610		
Degrees of freedom	3	35	24	39	0	23
AR1 (p-value)					0.0298	0.0323
AR2 (p-value)					0.0509	0.0411
Sargan (p-value)					0.0000571	0.000000212
Hansen (p-value)					1	1
Chi-2 statistic						
F statistic	10.49	5.761	59.31	20.37	.	1299.3

Standard errors in parentheses
L indicates a lagged variable 1 period, L2 indicates lagged variable 2 periods
Dependent variable is: Entry Rate (N)
* p<0.10, ** p<0.05, *** p<0.010

TABLE C.6
Entry rates: Number of employees

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Entry Rate (Empl.)			0.344***	0.159***	0.160	0.642***
			(0.0490)	(0.0577)	(0.199)	(0.200)
L2.Entry Rate (Empl.)			0.514***	0.380***		
			(0.0469)	(0.0531)		
Share of intangibles	0.0138**	-0.0163	0.0167	0.00621	0.00151	0.0214
	(0.00615)	(0.0182)	(0.0163)	(0.0188)	(0.0224)	(0.0163)
L.Share of intangibles			0.000912	0.000400		
			(0.0208)	(0.0201)		
L2.Share of intangibles			-0.0181	-0.0310*	-0.0376*	-0.0204
			(0.0162)	(0.0171)	(0.0196)	(0.0147)
Ratio total investment to labor	-0.0226***	-0.00630	-0.00378	-0.00443	0.000181	-0.0161**
	(0.00415)	(0.00609)	(0.00266)	(0.00541)	(0.00256)	(0.00648)
L.Ratio total investment to labor					-0.00649	0.00412
					(0.00644)	(0.00629)
ln(Labor compensation)	-0.00954***	0.0145*	-0.00163	0.0140*	0.0132	-0.00408
	(0.00164)	(0.00812)	(0.00109)	(0.00728)	(0.0121)	(0.00327)
Constant	0.133***	-0.108	0.0263**	-0.117		0.0611
	(0.0181)	(0.0900)	(0.0124)	(0.0816)		(0.0432)
Year FEs		X	X	X	X	X
Observations	288	288	288	288	288	288
R-squared (adj.)	0.129	0.143	0.696	0.330		
Degrees of freedom	3	35	24	39	0	23
AR1 (p-value)					0.0441	0.0862
AR2 (p-value)					0.146	0.191
Sargan (p-value)					0.0617	0.578
Hansen (p-value)					1	1
Chi-2 statistic						
F statistic	15.16	4.149	28.40	7.527	.	58.32

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Entry Rate (Empl.)
* p<0.10, ** p<0.05, *** p<0.010

Note: Robust standard errors are in parentheses for models 3 to 6.

TABLE C.7
Exit rates: Number of companies

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Exit Rate (N)			0.711***	0.600***	0.608***	0.835***
			(0.0604)	(0.0634)	(0.0950)	(0.0740)
L2.Exit Rate (N)			0.139**	0.00922		
			(0.0596)	(0.0634)		
Share of intangibles	-0.00386	-0.0133	-0.0172	-0.0350	-0.0226	0.00551
	(0.0100)	(0.0294)	(0.0238)	(0.0279)	(0.0231)	(0.0297)
L.Share of intangibles			0.0168	0.0190		
			(0.0304)	(0.0300)		
L2.Share of intangibles			-0.00328	-0.00517	0.00608	-0.00821
			(0.0237)	(0.0253)	(0.0189)	(0.0257)
Ratio total investment to labor	-0.00935	-0.0147	-0.00689*	-0.0162**	-0.0175***	-0.0240***
	(0.00676)	(0.00985)	(0.00367)	(0.00802)	(0.00466)	(0.00657)
L.Ratio total investment to labor					0.00865	0.0165**
					(0.00668)	(0.00650)
ln(Labor compensation)	-0.00259	-0.00655	-0.00116	-0.000336	-0.00103	-0.00153
	(0.00267)	(0.0131)	(0.00150)	(0.0108)	(0.0155)	(0.00152)
Constant	0.126***	0.167	0.0618***	0.0568		0.0215
	(0.0295)	(0.145)	(0.0166)	(0.121)		(0.0181)
Year FEs		X	X	X	X	X
Observations	288	288	288	288	288	288
R-squared (adj.)	-0.00373	0.212	0.718	0.483		
Degrees of freedom	3	35	24	39	0	23
AR1 (p-value)					0.0220	0.0229
AR2 (p-value)					0.426	0.330
Sargan (p-value)					0.0567	0.0669
Hansen (p-value)					1	1
Chi-2 statistic						
F statistic	0.644	5.608	31.49	12.78	.	1043.5

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Exit Rate (N)
* p<0.10, ** p<0.05, *** p<0.010

TABLE C.8
Exit rates: Number of employees

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Exit Rate (Empl.)			0.392***	0.170***	0.142*	0.542***
			(0.0573)	(0.0630)	(0.0731)	(0.145)
L2.Exit Rate (Empl.)			0.273***	0.0843		
			(0.0522)	(0.0567)		
Share of intangibles	0.0201**	-0.0399	-0.0249	-0.0410	-0.0313	0.0203
	(0.00867)	(0.0281)	(0.0285)	(0.0323)	(0.0312)	(0.0204)
L.Share of intangibles			0.0533	0.0382		
			(0.0363)	(0.0346)		
L2.Share of intangibles			-0.0251	-0.0351	-0.0143	-0.0125
			(0.0284)	(0.0293)	(0.0205)	(0.0215)
Ratio total investment to labor	-0.00785	-0.00541	-0.000878	-0.00275	-0.00586	-0.00289
	(0.00585)	(0.00941)	(0.00439)	(0.00932)	(0.0159)	(0.0194)
L.Ratio total investment to labor					0.00133	-0.00181
					(0.00535)	(0.0151)
ln(Labor compensation)	-0.00594**	-0.0169	-0.000842	-0.0134	-0.0106	-0.00116
	(0.00231)	(0.0125)	(0.00179)	(0.0125)	(0.0149)	(0.00315)
Constant	0.103***	0.236*	0.0306	0.187		0.0101
	(0.0255)	(0.139)	(0.0200)	(0.141)		(0.0343)
Year FEs		X	X	X	X	X
Observations	288	288	288	288	288	288
R-squared (adj.)	0.0241	0.259	0.475	0.283		
Degrees of freedom	3	35	24	39	0	23
AR1 (p-value)					0.0193	0.00651
AR2 (p-value)					0.0575	0.0169
Sargan (p-value)					0.0222	0.272
Hansen (p-value)					1	1
Chi-2 statistic						
F statistic	3.361	6.768	11.82	6.345	.	20.83

Standard errors in parentheses
L indicates a lagged variable 1 period, L2|indicates lagged variable 2 periods
Dependent variable is: Exit Rate (Empl.)
* p<0.10, ** p<0.05, *** p<0.010

Note: Robust standard errors are in parentheses for models 3 to 6.

C.7 Shares of employment

TABLE C.9
Young and small firms

(a) Units

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Share of young*SME firms	0.993***	0.911***	1.022***	0.640***	0.757***	1.154***
	(0.0112)	(0.0261)	(0.0329)	(0.208)	(0.199)	(0.107)
L2.Share of young*SME firms			-0.127***	-0.201***	-0.218***	-0.224***
			(0.0249)	(0.0592)	(0.0522)	(0.0546)
Share of intangibles	-0.497	-2.811	-3.385	3.678	2.340	
	(0.511)	(1.833)	(2.127)	(4.512)	(3.799)	
L.Share of intangibles			2.389	2.355	1.331	3.332
			(2.366)	(2.658)	(2.966)	(2.567)
L2.Share of intangibles			-1.713	-3.193	-0.371	-3.005
			(1.895)	(3.144)	(3.502)	(2.525)
Ratio total investment to labor	-0.0433	0.0997	-0.418	3.155**	1.189	-2.262
	(0.274)	(0.627)	(0.834)	(1.394)	(0.914)	(1.934)
L.Ratio total investment to labor			0.444	1.247	1.092	2.716
			(0.772)	(1.838)	(1.539)	(2.310)
ln(Labor compensation)						
Constant	0.319	3.027***	3.428***			1.853
	(0.493)	(1.164)	(1.269)			(1.292)
Year FEs						
Observations	288	288	288	288	288	288
R-squared (adj.)	0.966	0.870	0.881			
Degrees of freedom	3	35	39	0	0	23
AR1 (p-value)				0.0203	0.00795	0.0307
AR2 (p-value)				0.652	0.443	0.227
Sargan (p-value)				1.13e-24	3.49e-13	0.106
Hansen (p-value)				1	1	1
Chi-2 statistic				.	.	
F statistic	2707.4	98.13	90.57			175.3

Standard errors in parentheses
L. indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Share of young*SME firms
* p<0.10, ** p<0.05, *** p<0.010

(b) Employment

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Share of empl. by young*SME	0.994***	0.944***	1.085***	0.973***	1.033***	1.173***
	(0.00794)	(0.0240)	(0.0414)	(0.121)	(0.103)	(0.0650)
L2.Share of empl. by young*SME			-0.164***	-0.186***	-0.196***	-0.183**
			(0.0388)	(0.0502)	(0.0505)	(0.0670)
Share of intangibles	-0.148	-1.033	-0.903	1.343	0.977	
	(0.285)	(1.085)	(1.268)	(2.215)	(1.520)	
L.Share of intangibles			1.671	1.326	1.370	1.385
			(1.411)	(1.731)	(1.947)	(2.124)
L2.Share of intangibles			-2.665**	-1.285	-1.164	-1.137
			(1.132)	(1.630)	(1.934)	(2.247)
Ratio total investment to labor	0.138	0.304	0.00680	0.415	0.0558	-0.521
	(0.163)	(0.366)	(0.493)	(0.482)	(0.270)	(0.488)
L.Ratio total investment to labor			0.220	-0.0238	0.527	0.850
			(0.460)	(0.471)	(0.517)	(0.713)
ln(Labor compensation)						
Constant	0.174	1.400**	2.107***			0.427
	(0.231)	(0.671)	(0.747)			(0.314)
Year FEs						
Observations	288	288	288	288	288	288
R-squared (adj.)	0.985	0.869	0.879			
Degrees of freedom	3	35	39	0	0	23
AR1 (p-value)				0.00552	0.00502	0.0126
AR2 (p-value)				0.341	0.490	0.721
Sargan (p-value)				0.00136	0.142	0.912
Hansen (p-value)				1	1	1
Chi-2 statistic				.	.	
F statistic	6196.3	97.06	88.17			16.19

Standard errors in parentheses
L indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Share of empl. by young*SME
* p<0.10, ** p<0.05, *** p<0.010

Note: Robust standard errors are in parentheses for models 3 to 6.

TABLE C.10
Mature and large

(a) Units

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Share of mature*large firms	0.925***	0.926***	1.214***	1.162***	1.186***	1.338***
	(0.00854)	(0.0131)	(0.0533)	(0.0780)	(0.0797)	(0.0348)
L2.Share of mature*large firms			-0.284***	-0.216***	-0.233***	-0.379***
			(0.0520)	(0.0723)	(0.0757)	(0.0319)
Share of intangibles	0.0648	0.150	0.395*	0.420	0.381	
	(0.0486)	(0.202)	(0.230)	(0.534)	(0.456)	
L.Share of intangibles			-0.646**	-0.634	-0.696	-0.347
			(0.256)	(0.550)	(0.540)	(0.244)
L2.Share of intangibles			0.133	0.252*	0.182	0.362
			(0.209)	(0.144)	(0.116)	(0.244)
Ratio total investment to labor	0.0133	0.0402	0.0378	0.00273	0.0591	-0.0137
	(0.0287)	(0.0754)	(0.0921)	(0.134)	(0.0783)	(0.0621)
L.Ratio total investment to labor			-0.0107	-0.00188	-0.0712	0.0179
			(0.0841)	(0.145)	(0.0641)	(0.0400)
ln(Labor compensation)						
Constant	-0.00718	-0.0693	0.0561			0.0299**
	(0.0343)	(0.118)	(0.126)			(0.0133)
Year FEs						
Observations	288	288	288	288	288	288
R-squared (adj.)	0.982	0.961	0.966			
Degrees of freedom	3	35	39	0	0	23
AR1 (p-value)				0.139	0.143	0.119
AR2 (p-value)				0.179	0.188	0.166
Sargan (p-value)				2.45e-08	0.00129	0.0874
Hansen (p-value)				1	1	1
Chi-2 statistic				.	.	
F statistic	5243.1	354.4	337.7			14331.2

Standard errors in parentheses
L. indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Share of mature*large firms
* p<0.10, ** p<0.05, *** p<0.010

(b) Employment

	(1) OLS	(2) FE	(3) OLS Dynamic	(4) FE Dynamic	(5) Arellano Bond (endo.)	(6) Arellano Bond (exog.)
L.Share of empl. by mature*large	0.996***	0.780***	0.884***	0.727***	0.737***	1.279***
	(0.00537)	(0.0352)	(0.0524)	(0.156)	(0.124)	(0.184)
L2.Share of empl. by mature*large			-0.112***	-0.151*	-0.161*	-0.290
			(0.0400)	(0.0874)	(0.0901)	(0.189)
Share of intangibles	-0.884	-2.973	-4.168	-8.459***	-7.913***	
	(0.582)	(2.544)	(3.044)	(3.242)	(2.777)	
L.Share of intangibles			-0.199	-1.740	-0.706	-4.222
			(3.380)	(3.309)	(3.295)	(2.943)
L2.Share of intangibles			3.157	1.102	-0.00695	3.160
			(2.706)	(3.262)	(3.438)	(2.657)
Ratio total investment to labor	-0.493	0.267	-0.877	-2.504**	-0.437	0.583
	(0.334)	(0.848)	(1.181)	(1.277)	(1.354)	(0.790)
L.Ratio total investment to labor			1.330	1.839	1.456	-1.004
			(1.112)	(1.543)	(1.610)	(0.828)
ln(Labor compensation)						
Constant	0.792*	11.79***	11.27***			1.746*
	(0.476)	(2.412)	(2.552)			(0.890)
Year FEs						
Observations	288	288	288	288	288	288
R-squared (adj.)	0.993	0.675	0.683			
Degrees of freedom	3	35	39	0	0	23
AR1 (p-value)				0.0110	0.00444	0.0105
AR2 (p-value)				0.0431	0.0242	0.0509
Sargan (p-value)				0.415	0.00188	0.505
Hansen (p-value)				1	1	1
Chi-2 statistic				.	.	
F statistic	14437.7	31.56	27.41			18.93

Standard errors in parentheses
L. indicates a lagged variable 1 period, L2. indicates lagged variable 2 periods
Dependent variable is: Share of empl. by mature*large
* p<0.10, ** p<0.05, *** p<0.010

Note: Robust standard errors are in parentheses for models 3 to 6.

Appendix D Intangible Investments Data Matching with Business Indicators By Sector

TABLE D.11
Sector Classification

STAN aggregation based on ISIC rev. 4 classification (compatible with SIC 2003 and SIC 2007 at the 3-digit level)			
SIC Rev. 2007 compatible with ISIC rev. 4 (2-digits)	**IND A38**	**A7 Macro sector**	**Description**
01-03	1	Agriculture, forestry, and fishing	AGRICULTURE, FORESTRY AND FISHING [A]
05-09	5	Mining and quarrying	Mining and quarrying [B]
10-12	10	Manufacturing	Food products, beverages, and tobacco [C
13-15	13	Manufacturing	Textiles, wearing apparel, leather and
16-18	16	Manufacturing	Wood and paper products, and printing
19	19	Manufacturing	Coke and refined petroleum products [CD
20	20	Manufacturing	Chemicals and chemical products [CE]
21	21	Manufacturing	Basic pharmaceutical products and pharm
22-23	22	Manufacturing	Rubber and plastics products, and other
24-25	24	Manufacturing	Basic metals and fabricated metals prod.
26	26	Manufacturing	Computer, electronic and optical products
27	27	Manufacturing	Electrical equipment [CJ]
28	28	Manufacturing	Machinery and equipment n.e.c. [CK]
29-30	29	Manufacturing	Transport equipment [CL]
31-33	31	Manufacturing	Furniture; other manufacturing; repair
35	35	Electricity, gas, water, and waste	Electricity, gas, steam, and air condition
36-39	36	Electricity, gas, water, and waste	Water supply; sewerage, waste management
41-43	41	Construction	CONSTRUCTION [F]
45-47	45	Non-financial market services	Wholesale and retail trade, repair of m
49-53	49	Non-financial market services	Transport and storage
55-56	55	Non-financial market services	Accommodation and food service activity
58-60	58	Non-financial market services	Publishing
61	61	Non-financial market services	Telecommunications [JB]
62-63	62	Non-financial market services	IT and other information services [JC]
64-66	64	EXCLUDED	FINANCIAL AND INSURANCE ACTIVITIES [K]
68	68	Non-financial market services	REAL ESTATE ACTIVITIES
69-71	69	Non-financial market services	Legal and accounting services' activities of head offices, management consultancy activities. Architecture and engineering activities, technical testing, and analysis
72	72	Non-financial market services	Scientific research and development [MB
73-75	73	Non-financial market services	Advertising and market research; other
77-82	77	Non-financial market services	Administrative and support service activities
84	84	EXCLUDED	Public administration and defence
85	85	Non-market services	Education [P]
86-88	86	Non-market services	Human health and social work services
90-93	87	Non-market services	Arts, entertainment, and recreation [R]
94-96	90	Non-market services	Other services activities
97-99	94	EXCLUDED	Households and extraterritorial activities

TABLE D.12
Intangibles Dataset Sector Classification

Industry (NACE Rev. 2)		Institutional sector			Industry groups					
		TOT	MKT	NonMKT	All	Goods	Trade OSvc	BizSvc	Public	NFBiz
Core Section	Description / Codes	1	1.1+1.2+1.4	1.3+1.5	0	0.1	0.2	0.3	0.4	0.5
A	*Agriculture, forestry and fishing*	*X*	*X*	*X*	*X*					
B	Mining and quarrying	y	y	y	y	Y				Y
C	Manufacturing	y	y	y	y	Y				Y
D	Electricity, gas, steam and air conditioning supply	y	y	y	y				Y	
E	Water supply; sewerage, waste management and remediation activities	y	y	y	y				Y	
F	Construction	y	y	y	y	Y				Y
G	Wholesale and retail trade; repair of motor vehicles and motorcycles	y	y	y	y		Y			Y
H	Transportation and storage	y	y	y	y		Y			Y
I	Accommodation and food service activities	y	y	y	y		Y			Y
J	Information and communication	y	y	y	y			Y		Y
K	Financial and insurance activities	y	y	y	y			Y		Y
L	*Real estate activities*	*X*	*X*	*X*	*X*					
M	Professional, scientific and technical activities	y	y	y	y			Y		Y
N	Administrative and support service activities	y	y	y	y			Y		Y
O	Public administration and defence; compulsory social security	y	y	y	y				Y	
P	Education	y	y	y	y				Y	
Q	Human health and social work activities	y	y	y	y				Y	
R	Arts, entertainment and recreation	y	y	y	y				Y	
S	Other service activities	y	y	y	y		Y			Y
T	*Activities of households*	*X*	*X*	*X*	*X*					
U	*Activities of extraterritorial organisations and bodies*	*X*	*X*	*X*	*X*					

Note: Notice that INTAN-INVEST does include Ag, but excludes L, O,P,Q and T,U TOT = MKT + NonMKT
market (MKT) = (1.1) non-financial corporation (market producer) + (1.2) financial corporation (market producer) + (1.4) households as entrepreneurs (market or private producer for own final use) non-market (NonMKT) = (1.3) general government (public non-market producer) + (1.5) Non-profit institutions serving households (private non-market producer)

Appendix E Additional definitions for U.S. estimations of business dynamics indicators

- US: A firm is a business organization consisting of one or more domestic establishments specified under common ownership or control. The firm and the establishment are the same for single-establishment firms.
- US: Business Register (BR) – A comprehensive database of all U.S. business establishments developed and maintained by the U.S. Census Bureau, with data beginning in 1975 and continuing to the present. This is restricted data and is the source data for the Longitudinal Business Database (LBD).
- US: Business Dynamics Statistics (BDS) – A research dataset based on the LBD that describes United States business dynamics across various measures. Disclosure analysis is performed before release to the public to protect the confidentiality of LBD data.
- US: Firm Exit – Firm exit identifies events where all the establishments associated with a particular firm and the firm itself cease all operations. Note firm legal entities that cease to exist because of merger and acquisition activity are not classified as firm exits in these data.
- US: Employment – Paid employment consists of full and part-time employees, including salaried officers and executives of corporations, who were on the payroll

in the pay period including March 12. Included are employees on sick leave, holidays, and vacations; proprietors and partners of unincorporated businesses are not included.

Appendix F Dictionary of keywords for tech and intangible variables

TABLE F.13
Technology bigrams

	keywords
threedprinting	**3d printing , 3d printed , 3d printer , 3d printers , 3d print , additive manufacturing , inkjet , bioprinting , binder jetting , rapid prototyping , 3dprinting , 3dprinted , 3dprinter , 3dprinters , 3dprint , additivemanufacturing , inkjet , bioprinting , binderjetting , rapidprototyping , impresiones 3d , impreso 3d , impresora 3d , impresoras 3d , impresion 3d , fabricación aditiva , chorro de tinta , bioimpresión , inyección de aglomerante , prototipos rápidos , impresiones3d , impreso3d , impresora3d , impresoras3d , impresion3d , fabricaciónaditiva , chorrodetinta , bioimpresion , inyecciónde aglomerante , prototiposrápidos , fabricacion aditiva , fabricacionaditiva , prototipado rapido , prototipadorapido**
ai_bigrams	**artificial intelligence , machine learning , neural networks , deep learning , predictive analytics , language processing , artificialintelligence , machinelearning , neuralnetworks , deeplearning , predictiveanalytics , languageprocessing , inteligencia artificial , aprendizaje automático , Redes neuronales , aprendizaje profundo , análisis predictivo , procesamiento del lenguaje , inteligenciaartificial , aprendizajeautomático , Redesneuronales , aprendizajeprofundo , análisispredictivo , procesamientodellenguaje , aprendizaje automatico , aprendizajeautomatico**
bluetooth	**bluetooth**

TABLE F.13
Technology bigrams *(continued)*

	keywords
cloud	**cloud computing , cloud technology , cloud resources , cloud storage , cloud applications , cloud services , public cloud , cloud business , private cloud , cloud service , hybrid cloud , cloud platform , cloud infrastructure , cloud solutions , cloud providers , cloud offering , cloud revenue , cloud offerings , cloud solution , cloud based , service cloud , cloud customers , cloud data , enterprise cloud , cloud environment , mobile computing , cloud security , cloud environments , cloud deployments , cloud application , cloud platform , cloudcomputing , cloudtechnology , cloudresources , cloudstorage , cloudapplications , cloudservices , publiccloud , cloudbusiness , privatecloud , cloudservice , hybridcloud , cloudplatform , cloudinfrastructure , cloudsolutions , cloudproviders , cloudoffering , cloudrevenue , cloudofferings , cloudsolution , cloudbased , servicecloud , cloudcustomers , clouddata , enterprisecloud , cloudenvironment , mobilecomputing , cloudsecurity , cloudenvironments , clouddeployments , cloudapplication , cloudplatform , computación en la nube , tecnología en la nube , recursos en la nube , almacenamiento en la nube , aplicaciones en la nube , servicios en la nube , nube pública , negocio en la nube , nube privada , servicio de almacenamiento en la nube , nube híbrida , plataforma en la nube , infraestructura en la nube , soluciones en la nube , Proveedores de nubes , ofrenda de nubes , Ingresos de la nube , Ofertas en la nube , solución en la nube , basado en la nube , servicio de la nube , clientes de la nube , datos en la nube , nube empresarial , ambiente de la nube , informática móvil , seguridad en la nube , ambientes en la nube , implementaciones en la nube , aplicación en la nube , cloudplatform , computaciónenlanube , tecnologíaenlanube , recursosenlanube , almacenamientoenlanube , aplicacionesenlanube , serviciosenlanube , nubepública , negocioenlanube , nubeprivada , serviciodealmacenamientoenlanube , nubehíbrida , plataformaenlanube , infraestructuraenlanube , solucionesenlanube , Proveedoresdenubes , ofrendadenubes , Ingresosdelanube , Ofertasenlanube , soluciónenlanube , basadoenlanube , serviciodelanube , clientesdelanube , datosenlanube , nubeempresarial , ambientedelanube , informáticamóvil , seguridadenlanube , ambientesenlanube , implementacionesenlanube , aplicaciónenlanube , computacion en la nube , tecnologia en la nube , nube publica , nube hibrida , solucion en la nube , informatica movil , aplicacion en la nube , computacionenlanube , tecnologiaenlanube , nubepublica , nubehibrida , solucionenlanube , informaticamovil , aplicacionenlanube**
disk_drive	**disk drive , disc drive , hard drive , hard disk , hard disc , diskdrive , discdrive , harddrive , harddisk , harddisc , unidad de disco , disco duro , discos duros , disco externo , unidaddedisco , discoduro , discosduros , discoexterno**

TABLE F.13
Technology bigrams *(continued)*

	keywords
driverless	**autonomous car , autonomous cars , selfdriving car , selfdriving cars , selfdriving truck , selfdriving trucks , autonomous trucks , autonomous vehicles , selfdriving vehicles , autonomous driving , autonomous vehicle , automated driving , driverless car , driverless cars , driverless vehicle , driverless vehicles , autonomouscar , autonomouscars , selfdrivingcar , selfdrivingcars , selfdrivingtruck , selfdrivingtrucks , autonomoustrucks , autonomousvehicles , selfdrivingvehicles , autonomousdriving , autonomousvehicle , automateddriving , driverlesscar , driverlesscars , driverlessvehicle , driverlessvehicles , coche autónomo , autos autónomos , carro autonomo , automoviles autonomos , camion autonomo , camiones autonomos , camiones autónomos , vehículos autónomos , conduccion autonoma , conducción autónoma , vehículo autónomo , conducción automatizada , coche sin conductor , autos sin conductor , vehículo sin conductor , Vehículos sin conductor , cocheautónomo , autosautónomos , carroautonomo , automovilesautonomos , camionautonomo , camionesautonomos , camionesautónomos , vehículosautónomos , conduccionautonoma , conducciónautónoma , vehículoautónomo , conducciónautomatizada , cochesinconductor , autossinconductor , vehículosinconductor , coche autonomo , cocheautonomo , vehiculosautonomos , vehiculos autonomos , vehiculo autonomo , vehiculoautonomo , conduccion automatizada , conduccionautomatizada , vehiculos sin conductor , vehiculosinconductor**
electric_cars	**electric car , electric vehicle , electric cars , electric vehicles , electric hybrid , hybrid car , hybrid bus , electric bus , hybrid vehicle , hybrid electric , hybrid vehicles , electric drive , plugin hybrid , hybrid cars , electric buses , vehicle charging , electrical vehicles , electrical vehicle , electriccar , electricvehicle , electriccars , electricvehicles , electrichybrid , hybridcar , hybridbus , electricbus , hybridvehicle , hybridelectric , hybridvehicles , electricdrive , pluginhybrid , hybridcars , electricbuses , vehiclecharging , electricalvehicles , electricalvehicle , coche eléctrico , vehículo eléctrico , coches eléctricos , vehículos eléctricos , híbrido eléctrico , Auto Hibrido , autobús híbrido , autobús eléctrico , vehículo híbrido , eléctrico híbrido , vehículos híbridos , transmisión eléctrica , complemento híbrido , carros híbridos , autobuses eléctricos , carga de vehículos , vehículos eléctricos , vehículo eléctrico , coche eléctrico , vehículo eléctrico , coches eléctricos , vehículos eléctricos , electricibrid , Auto Hibrido , vehículo híbrido , hibrideléctrico , híbridos , electricidad , recarga del vehiculo electrico , carros híbridos , puestos de electricidad , vehículo eléctrico , carro electrico , carros electricos , carro eléctrico , carros eléctricos , recarga del coche electrico , recarga del carro electrico , recarga del automovil electrico , cocheeletrico , vehículoeléctrico , cocheseléctricos , vehículoseléctricos , híbridoeléctrico , AutoHibrido , autobúshíbrido , autobúseléctrico , vehículohíbrido , eléctricohíbrido , vehículoshíbridos , transmisióneléctrica , complementohíbrido , carroshíbridos , autobuseseléctricos , cargadevehículos , vehículoseléctricos , vehículoeléctrico , cocheeléctrico , vehículoeléctrico , cocheseléctricos , vehículoseléctricos , electricibrid , AutoHibrido , buseléctrico , buselectrico , vehículohíbrido , hibrideléctrico , híbridos , electricidad , recargadelvehiculoelectrico , carroshíbridos , puestosdeelectricidad , vehículoeléctrico , carroelectrico , carroselectricos , carroeléctrico , carroseléctricos , recargadelcocheelectrico , recargadelcarroelectrico , recargadelautomovilelectrico**

TABLE F.13
Technology bigrams *(continued)*

	keywords
fracking	**horizontal drilling , coalbed methane , unconventional gas , hydraulic fracturing , tight sandstones , shale rock , shale gas , shale oil , unconventional shale , oil shale , gas sands , shale plays , gas shale , hydraulic fracturing , fracturing activity , fracking , gps global positioning , horizontaldrilling , coalbedmethane , unconventionalgas , hydraulicfracturing , tightsandstones , shalerock , shalegas , shaleoil , unconventionalshale , oilshale , gassands , shaleplays , gasshale , hydraulicfracturing , fracturingactivity , fracking , perforación horizontal , carbón metano , gas no convencional , fracturamiento hidráulico , areniscas apretadas , roca de esquisto , gas de esquisto , aceite de esquisto bituminoso , esquisto no convencional , lutita de petróleo , arena de gas , Playes de lutita , lutita de gas , fracturamiento hidráulico , actividad de fractura , fracking , Posicionamiento global del GPS , perforaciónhorizontal , carbónmetano , gasnoconvencional , fracturamientohidráulico , areniscasapretadas , rocadeesquisto , gasdeesquisto , aceitedeesquistobituminoso , esquistonoconvencional , lutitadepetróleo , arenadegas , Playesdelutita , lutitadegas , fracturamientohidráulico , actividaddefractura , fracking , PosicionamientoglobaldelGPS , carbonmetano , fracturamientohidraulico**
lcds	**flat panel , lcd , panel display , panel tv , flatpanel , paneldisplay , paneltv , pantalla plana de lcd , lcd , visualización del panel , TV de panel , visualizacion del panel , pantalla lcd , television lcd , tv lcd , lcd tv , pantallaplanadelcd , lcd , visualizacióndelpanel , TVdepanel , visualizaciondelpanel , pantallalcd , televisionlcd , tvlcd**
machine_vision	**machine vision , machinevision , visión de máquina , vision de maquina , visióndemáquina , visiondemaquina**
monoclonal_-antibody	**monoclonal antibody , monoclonal antibodies , monoclonalantibody , monoclonalantibodies , anticuerpo monoclonal , anticuerpos monoclonicos , monoclonalanticuerpo , anticuerpos monoclonicos , anticuerpomonoclonal , anticuerposmonoclonicos , monoclonalanticuerpo , anticuerposmonoclonicos**
oled	**oled**
optical_fiber	**optical fiber , optical fibre , fiber optic , fibre optic , fiber optics , fibre optics , optical transport , optical networks , fiber access , fiber channel , fiber cable , opticalfiber , opticalfibre , fiberoptic , fibreoptic , fiberoptics , fibreoptics , opticaltransport , opticalnetworks , fiberaccess , fiberchannel , fibercable , fibra óptica , fibra optica , fibras opticas , optica de fibra , acceso a fibra óptica , transporte óptico , redes ópticas , acceso a fibra optica , canal de fibra , cable de fibra , fibraóptica , fibraoptica , fibrasopticas , opticadefibra , accesoafibraóptica**
rfid	**rfid**
smartcard	**smart card , smart cards , smartcard , smartcards , tarjeta electrónica , tarjetas inteligentes , tarjeta electrónica , tarjetas inteligentes , tarjeta electronica , tarjetaelectrónica , tarjetasinteligentes , tarjetaelectrónica , tarjetasinteligentes , tarjetaelectronica**
smartphone	**smart phone , smart phones , samsung galaxy , android phone , android phones , smartphones tablets , smartphone , smartphones , iphone , smartphone , smartphones , samsunggalaxy , androidphone , androidphones , smartphonestablets , smartphone , smartphones , teléfono inteligente , teléfonos inteligentes , Samsung galaxy , teléfono Android , teléfonos Android , tabletas de teléfonos inteligentes , teléfono inteligente , teléfonos inteligentes , iPhone , teléfono inteligente , teléfonos inteligentes , Samsung galaxia , telefono Android , telefonos Android , telefono inteligente , teléfono inteligente , telefonos inteligentes , tabletas inteligentes , teléfonointeligente , teléfonosinteligentes , Samsunggalaxy , teléfonoAndroid , teléfonosAndroid , tabletasdeteléfonosinteligentes , teléfonointeligente , teléfonosinteligentes , iPhone , teléfonointeligente , teléfonosinteligentes , Samsunggalaxia , telefonoAndroid , telefonosAndroid , telefonointeligente , Huwaei**

TABLE F.13
Technology bigrams *(continued)*

	keywords
solar	**solar power , solar energy , solar business , solar cell , solar projects , solar market , solar capital , solar industry , solar cells , solar panels , solar project , solar panel , solar pv , solar module , solar modules , solar side , solar technology , solar farm , solar products , solar assets , solar systems , solar industries , solar generation , solar pwr , solar markets , solar wafer , solar plant , solar grade , solar capacity , solar facility , solar farms , solar companies , solar plants , solar installations , solar equipment , solar product , solar technologies , solar thermal , solar customers , solar segment , solar applications , solar installation , solar materials , solar facilities , solar sales , solarpower , solarenergy , solarbusiness , solarcell , solarprojects , solarmarket , solarcapital , solarindustry , solarcells , solarpanels , solarproject , solarpanel , solarpv , solarmodule , solarmodules , solarside , solartechnology , solarfarm , solarproducts , solarassets , solarsystems , solarindustries , solargeneration , solarpwr , solarmarkets , solarwafer , solarplant , solargrade , solarcapacity , solarfacility , solarfarms , solarcompanies , solarplants , solarinstallations , solarequipment , solarproduct , solartechnologies , solarthermal , solarcustomers , solarsegment , solarapplications , solarinstallation , solarmaterials , solarfacilities , poder solar , energía solar , industria de energia solar , célula solar , proyectos solares , mercado solar , capital solar , industria solar , células solares , paneles solares , proyecto solar , panel solar , solar fotovolta , modulo solar , módulos solares , lado solar , tecnología solar , granja solar , productos solares , activos solares , sistemas solares , industrias solares , generación solar , solar PWR , mercados solares , oblea solar , planta solar , grado solar , capacidad solar , instalación solar , granjas solares , compañías solares , plantas solares , instalaciones solares , equipo solar , producto solar , tecnologías solares , solar térmica , clientes solares , segmento solar , aplicaciones solares , instalación solar , materiales solares , instalaciones solares , ventas solares , podersolar , energíasolar , industriadeenergiasolar , célulasolar , proyectossolares , mercadosolar , capitalsolar , industriasolar , célulassolares , panelessolares , proyectosolar , panelsolar , solarfotovolta , modulosolar , módulossolares , ladosolar , tecnologíasolar , granjasolar , productossolares , activossolares , sistemassolares , industriassolares , generaciónsolar , solarPWR , mercadossolares , obleasolar , plantasolar , gradosolar , capacidadsolar , instalaciónsolar , granjassolares , compañíassolares , plantassolares , instalacionessolares , equiposolar , productosolar , tecnologíassolares , solartérmica , clientessolares , segmentosolar , aplicacionessolares , instalaciónsolar , materialessolares , instalacionessolares , ventassolares , generacionsolar , instalacionsolar , empresassolares , compañiassolares , empresasolar , solartermica , instalacionsolar , modulosolares , tecnologiasolar , tecnologiassolares , energiasolar , celulasolar , celulassolares**
touch_screen	**touch screen , touch screens , touchscreen , touchscreens , pantalla táctil , pantallas táctiles , pantalla tactil , pantallas tactiles , pantallatáctil , pantallastáctiles , pantallatactil , pantallastactiles**
virtual_reality	**virtual reality , augmented reality , mixed reality , virtualreality , augmentedreality , mixedreality , realidad virtual , realidad aumentada , realidad mixta , realidad virtual , realidad aumentada , Realidad mixta , realidadvirtual , realidadaumentada , realidadmixta , realidadvirtual , realidadaumentada , Realidadmixta**
gps	**global positioning , gps , gps global positioning , gpsglobalpositioning , globalpositioning , posicionamiento global , GPS , Posicionamiento global del GPS , posicionamientoglobal , GPS , PosicionamientoglobaldelGPS**

TABLE F.14
Intangible Bigrams

	keywords
training	**training , entrenamiento laboral , entranimientolaboral , capacitacion , formacionlaboral , formacion laboral , formación laboral , capacitación , entramiento de los empleados , entrenamientodelosempleados , entrenamientodelempleado , entrenamiento del empleado , capacitacion laboral , capacitaciónlaboral , capacitación laboral , capacitacionlaboral , entrenamiento**
software	**software , cloud computing , cloud technology , cloud resources , cloud storage , cloud applications , cloud services , public cloud , cloud business , private cloud , cloud service , hybrid cloud , cloud platform , cloud infrastructure , cloud solutions , cloud providers , cloud offering , cloud revenue , cloud offerings , cloud solution , cloud based , service cloud , cloud customers , cloud data , enterprise cloud , cloud environment , mobile computing , cloud security , cloud environments , cloud deployments , cloud application , cloud platform , cloudcomputing , cloudtechnology , cloudresources , cloudstorage , cloudapplications , cloudservices , publiccloud , cloudbusiness , privatecloud , cloudservice , hybridcloud , cloudplatform , cloudinfrastructure , cloudsolutions , cloudproviders , cloudoffering , cloudrevenue , cloudofferings , cloudsolution , cloudbased , servicecloud , cloudcustomers , clouddata , enterprisecloud , cloudenvironment , mobilecomputing , cloudsecurity , cloudenvironments , clouddeployments , cloudapplication , cloudplatform , computación en la nube , tecnología en la nube , recursos en la nube , almacenamiento en la nube , aplicaciones en la nube , servicios en la nube , nube pública , negocio en la nube , nube privada , servicio de almacenamiento en la nube , nube híbrida , plataforma en la nube , infraestructura en la nube , soluciones en la nube , Proveedores de nubes , ofrenda de nubes , Ingresos de la nube , Ofertas en la nube , solución en la nube , basado en la nube , servicio de la nube , clientes de la nube , datos en la nube , nube empresarial , ambiente de la nube , informática móvil , seguridad en la nube , ambientes en la nube , implementaciones en la nube , aplicación en la nube , cloudplatform , computaciónenlanube , tecnologíaenlanube , recursosenlanube , almacenamientoenlanube , aplicacionesenlanube , serviciosenlanube , nubepública , negocioenlanube , nubeprivada , serviciodealmacenamientoenlanube , nubehíbrida , plataformaenlanube , infraestructuraenlanube , solucionesenlanube , Proveedoresdenubes , ofrendadenubes , Ingresosdelanube , Ofertasenlanube , soluciónenlanube , basadoenlanube , serviciodelanube , clientesdelanube , datosenlanube , nubeempresarial , ambientedelanube , informáticamóvil , seguridadenlanube , ambientesenlanube , implementacionesenlanube , aplicaciónenlanube , computacion en la nube , tecnologia en la nube , nube publica , nube hibrida , solucion en la nube , informatica movil , aplicacion en la nube , computacionenlanube , tecnologiaenlanube , nubepublica , nubehibrida , solucionenlanube , informaticamovil , aplicacionenlanube**
reputation	**product launch , launching , trade fair , new in , introducing , iso certification , lanzamiento de producto , lanzamiento de servicio , feria comercial , introduciendo el producto , introduciendo el servicio , certificacion iso , lanzamientos de productos , lanzamientos de servicios**
randd	**research and development , r&d , prototype , clinical trial , patent , copyright , trademark , investigación y desarrollo , investigacion y desarrollo , I+D , prototipo , prototipado , ensayo clinico , patente , derechos de autor , derecho de autor , marca comercial , ensayo clínico , ensayos clínicos , ensayos clinicos , researchanddevelopment , patentes , marca registrada , clinicaltrial , marcas registradas , marcaregistrada , marcasregistradas , investigacionydesarrollo , investigacionydesarrollo , ensayoclinico , derechosdeautor , derechodeautor , marcacomercial , ensayoclínico , ensayosclínicos , ensayosclinicos**
design	**industrial design , Diseño Industrial , DiseñoIndustrial , industrialdesign , disenoindustrial**

TABLE F.14
Intangible Bigrams *(continued)*

	keywords
organizational	**key performance indicators , supply chain management , knowledge management , organizational strategy , organizational efficiency , lean production , business process , indicadores de desempeño , indicador de desempeño , cadena de abastecimiento , cadenas de abastecimiento , administración del conocimiento , estrategia organizacional , eficiencia organizacional , producción liviana , proceso de negocio , keyperformanceindicators , supplychainmanagement , knowledgemanagement , organizationalstrategy , organizationalefficiency , leanproduction , businessprocess , indicadoresdedesempeño , indicadordedesempeño , cadenadeabastecimiento , cadenasdeabastecimiento , administracióndelconocimiento , estrategiaorganizacional , eficienciaorganizacional , producciónliviana , procesodenegocio , produccionliviana , administraciondelconocimiento , indicadoresdedesempeno , indicadordedesempeno , kpi**

Appendix G Twox Latin America

TABLE G.15
Twitnovative Tweets in 2022: Latin America

Sector	Handle	Tweet
Information	claroperu	hola recuerda que nuestras promos se renuevan constantemente mantente atenta que te sorprenderemos
Retail Trade	idea_antioquia	rt por primera vez llega el pavimento a villa german y buenos aires mas de habitantes del sector esperaron esta
Public	gobernacionsai	la secretaria de cultura invita a la comunidad a participar de una serie de actos conmemorativos y culturales en celebracion del dia oficial la lengua kriol los esperamos
Education	sanjosecolegio	mencion especial del excmo ayuntamiento de estepona al colegio san jose en los actos conmemorativos del dia de andalucia por la labor realizada desde que comenzo la pandemia #diadeandalucia #mencio...
Support Services	corferiasbogota	rt contigo #sonamosdigital y #lohacemosreal si quieres asistir a en y disfrutar
Manufacturing	empresaspolar	llega al mercado el nuevo toddy pudin
Utilities	energiapereira	actualiza tus datos con energia de pereira y asi podremos avisarte via mensaje de texto todas las novedades en el servicio de energia ingresa a y diligencia el formulario #actualizatusdatos #energ...
Arts/Entertainment	millosfcoficial	vallas invictas en partidos jugados

TABLE G.16
Correlation between Twox Share, Customer Engagement and Bigrams - Latin America -

	(1) Twox	(2) Twox	(3) Twox	(4) Twox	(5) Twox	(6) Twox	(7) Twox
ln(1 + Number of words)	-0.030***	-0.030***	-0.037***	-0.059***	-0.061***	-0.060***	-0.061***
	(0.005)	(0.005)	(0.006)	(0.010)	(0.010)	(0.009)	(0.009)
ln(1 + Number of mentions)	0.001	0.001	0.000	-0.002	-0.003		
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)		
ln(1 + Rate of Response perc.)	-0.007	-0.001	0.003	0.009	0.010		
	(0.004)	(0.007)	(0.007)	(0.023)	(0.024)		
ln(1 + Rate of Response (weighted))		-0.720	-0.706	-0.277	-0.414		
		(0.563)	(0.582)	(2.225)	(2.232)		
training			0.000***	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
software			-0.000	0.001	0.000	0.003	0.003
			(0.000)	(0.003)	(0.004)	(0.004)	(0.004)
research			-0.000	-0.001	-0.001	-0.001	-0.001
			(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
design			-0.015	0.078	0.080	0.031	0.030
			(0.015)	(0.113)	(0.111)	(0.087)	(0.087)
reputation			0.001	-0.021	-0.020	-0.014	-0.014
			(0.002)	(0.027)	(0.027)	(0.025)	(0.025)
organizational			0.001	-0.006	-0.005	-0.006	-0.006
			(0.002)	(0.011)	(0.011)	(0.010)	(0.010)
ln(Number of Employees)				-0.000	-0.011	-0.000	-0.005
				(0.004)	(0.012)	(0.004)	(0.012)
ln(Number of Employees) sq.					0.001		0.001
					(0.001)		(0.001)
Bigrams	None	None	All	All	All	All	All
Full Sample	Yes	Yes	Yes	Yes	Yes	Mentioned	Mentioned
Country FEs X Sectors FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	824	824	824	259	259	265	265
R2 within	0.172	0.174	0.201	0.343	0.345	0.340	0.340
F-test TechBigrams	No	No	1.767	2.713	2.830	No	No

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: Twox Share .999

Note: The table presents regressions at the firm level, including industry-fixed effects. The number of employees is at the yearly level per firm.

Rate of Response Perc.: RR_{jht} defined in tweets addressed within a month.

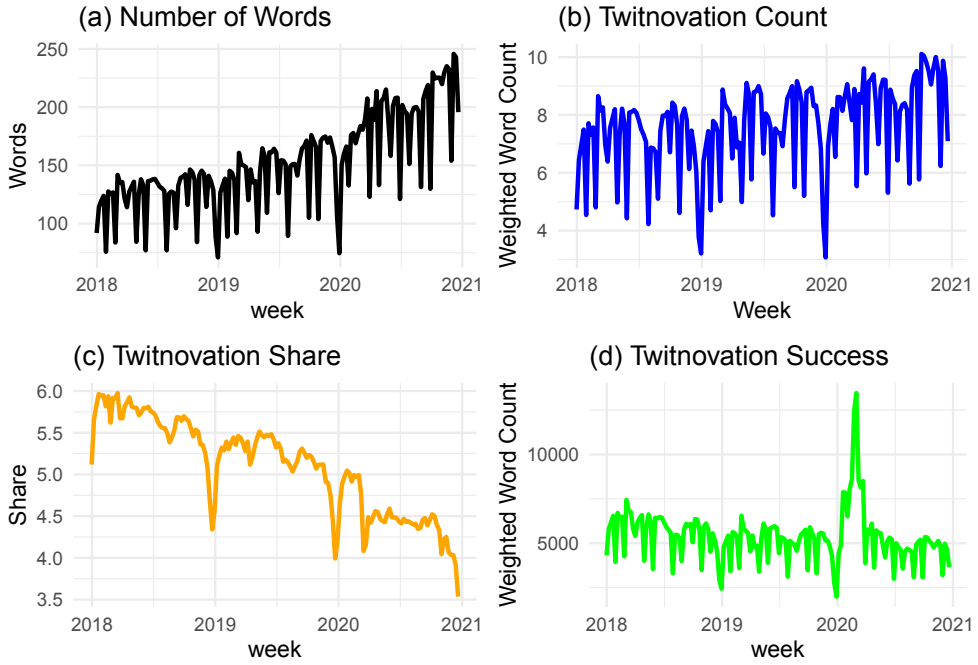
Rate of response (weighted): $RR_{jt}^{\text{weighted}}$ where time γ is defined in hours to reply.

Each cell referring to an intangible (e.g., training) represents the correlation of each firm-week value of the respective Twox variable (Twox Share) and the appearance in that firm's history (between 2017 and 2021) of the respective intangible-bigram one or more times (e.g., *training* is a dummy variable = 1 if the firm mentioned any word from the dictionary associated to training in the year to which that week belongs to.).

The number of employees is at the yearly level per firm.

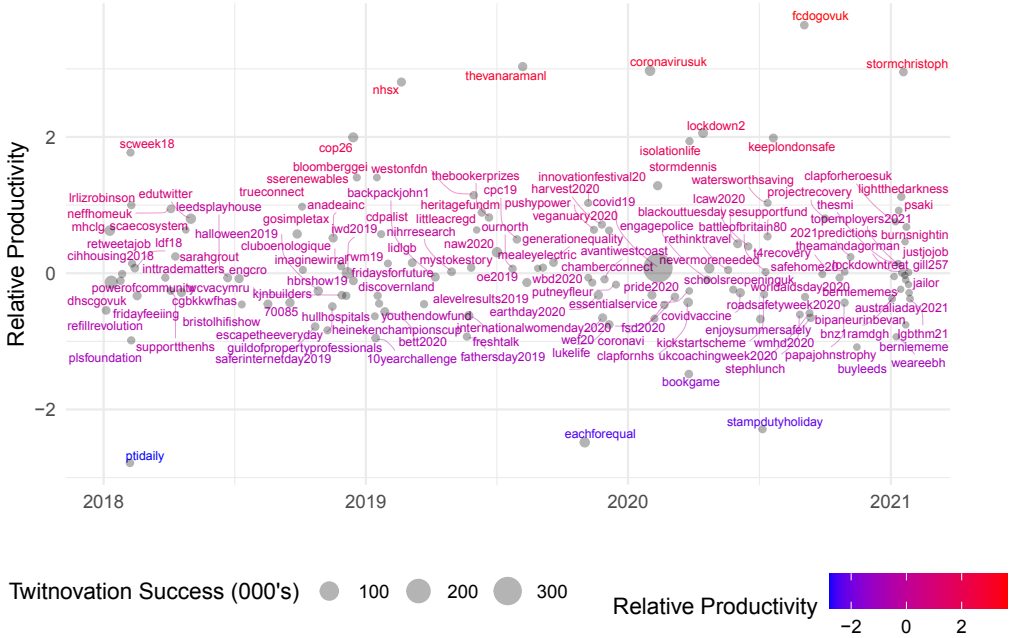
Appendix H Twox UK

FIGURE H.18
Trend Over Time for the Average Firm



Note:
This Figure plots Twitnovation variables for the average firm in the sample from 2018 to 2021. Panel A starts by plotting the average number of words firms post weekly. We see that this is increasing over time from about 100 words in early 2018 to almost 250 by the end of 2020. This illustrates that firms have intensified their Twitter activity over time. Panel b plots the Twitnovation counts of the average firm in each week over time. Since our measure TwoxCount sums over all words posted in a given week, it will be more prominent when firms post more tweets or longer tweets. The values fluctuate between 4 and 10, much lower than the word counts shown on Panel A. The upward trend also appears less intense than on Panel a implies that, as firms are tweeting more, the novelty is decreasing. This is confirmed on Panel C which plots the Twitnovation Share of the average firm.
On Panel c, the values can be interpreted as the percentage of words tweeted that are twit- innovative. We see that values fluctuate between 3% and 6% and are downward trending. We can think of this as an extreme example of decreasing returns to idea generation or innovation. In the early days, it is easy to generate a Twitnovation. In fact, on the first day that Twitter was launched, every new tweeted word is a Twitnovation. As time goes by, such low-hanging fruits 13 are exploited, and Twitnovation is harder to come by.
Panel d plots the average firm's Twitnovation success. We saw a prominent peak when the COVID pandemic emerged in the UK, around March 2020. This indicates the appearance of new words that diffuse and get adopted by many other users. On Panel c, the period around March 2020 coincides with an abrupt decrease in the Twitnovation Share. This is most likely because, despite some new words appearing at the beginning of the pandemic, many users very quickly adopted those new words wearing off their novelty faster than can be observed on the graph. Interestingly, this period also seems to coincide with an increase in Twitter activity on Panel a. Finally, we note seasonal effects on Panel a, b, and c. The end of calendar years leads to a relative drop in the number of words tweeted (Panel a) and a drop in the average twitnovativeness of these words (Panel c). Naturally, this leads to relatively lower Twitnovation counts (Panel b). Possibly, this is because an increasing amount of the corporate tweet budget is allocated to the traditional "Merry xmas" and "a happy new year."

FIGURE H.19
Twitnovative words according to Twinovation Success
and correlation with Relative Productivity:
United Kingdom



Note: The Figure reports a scatter plot of within transformations - within (2 digit) sector×country cells - for labor productivity (revenue over employment) vs Twitnovation Success, i.e. the difference between a particular value and the sector× country cell average.

Appendix I Labor Productivity, Twox and Patents

TABLE I.17
Regressions of Patent Counts
- Latin America -

	(1) ln(patents)	(2) ln(patents)	(3) ln(patents)	(4) ln(patents)	(5) ln(patents)
ln(1 + Number of words)	0.028*	0.031**	0.029**	0.013	0.018*
	(0.014)	(0.013)	(0.013)	(0.008)	(0.009)
Mean Twox .9 Success 1 hour	1.155	0.097			-0.143
	(0.844)	(0.653)			(0.671)
Mean Twox .9 Success 30 days	-0.232**				
	(0.093)				
Mean Twox .9	-0.284				
	(0.264)				
training			0.000		-0.000
			(0.000)		(0.000)
software			-0.000		-0.001
			(0.000)		(0.001)
research			0.001		0.002
			(0.001)		(0.001)
design			-0.078**		-0.069**
			(0.034)		(0.032)
reputation			-0.021		-0.036**
			(0.015)		(0.017)
organizational			0.015*		0.006
			(0.009)		(0.009)
rfid				-0.006	-0.006
				(0.006)	(0.006)
ai_long				0.000	0.000
				(0.001)	(0.001)
ai_short				-0.001	-0.001
				(0.001)	(0.001)
lcds				-0.029	-0.020
				(0.020)	(0.020)
threedprinting				0.014	0.015*
				(0.010)	(0.008)
solar				0.001	0.001
				(0.001)	(0.001)
oled				0.000	0.000
				(0.000)	(0.000)
virtual_reality				0.011	0.017
				(0.009)	(0.011)
gps				0.006	0.005
				(0.012)	(0.012)
driverless				-0.015**	-0.017**
				(0.007)	(0.008)
smartphone				-0.002	-0.003
				(0.002)	(0.002)
fracking				0.028	0.028
				(0.026)	(0.025)
disk_drive				-0.015	-0.026
				(0.017)	(0.019)
cloud				-0.002	0.003
				(0.003)	(0.004)
electric_cars				-0.002	-0.005
				(0.002)	(0.004)
smartcard				-0.014	0.001
				(0.014)	(0.019)
bluetooth				0.034	0.030
				(0.032)	(0.031)
optical_fiber				0.014	0.017
				(0.011)	(0.012)
touch_screen				0.040	0.033
				(0.025)	(0.025)
Country FEs X Sectors FEs	X	X	X	X	X
Observations	1,186	1,186	1,186	1,186	1,186
R2 Within	0.011	0.008	0.013	0.072	0.081

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(1 + Number of Patents Ever Filed)

Appendix J Aboutnovation

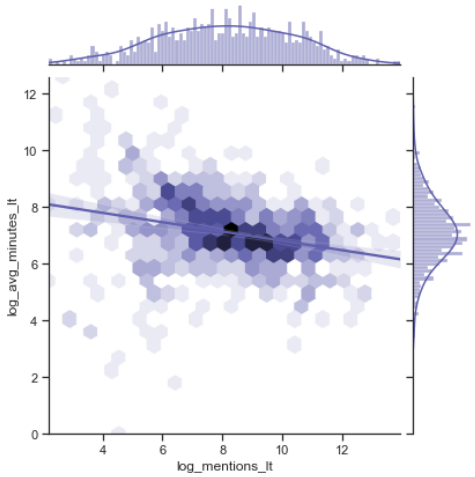
J.1 Descriptives

TABLE J.18
User Mentions
- Latin America -

	Total	Addressed	Rate of response (%)
Mentions	29,277,701	766,111	2.6
Customers	5,721,208	345,492	6.0
Firms	1,205	824	68.4

Note: Tweets without a valid date are ignored. Counts are at the handle level and for mentions between 2017 and 2021.

FIGURE J.20
Response Time and Number of Mentions: Latin America



Note: The x-axis shows the total number of mentions (in log) received by firms between 2017 and 2021. The y-axis plots the average number of minutes (in log) it took for the firm to respond, including any time between 2017 and 2021.

J.2 Types of mentions - Examples

TABLE J.19
Mentions by type and subtypes - Latin America - English Examples

Company	Tweet (Mention)	Reference Type	Retweet Type
mckinsey	RT @McKinsey: Leading Off: Odysseus had one. Oprah is one. The universal lessons of why mentoring matters for leaders and those they lead https://t.co/c55OVseJgb	Independent	Direct
aimgroupint	RT @AIMGroupInt: #AIMazing that our mascot AIMy is featured in your highlights video! Great projects with great people at @Shocklogic's! Our wish is to have an even more exciting 2019! #eventprofs #digitalmarketing https://t.co/6LQLFLuGzH	Quoted	Direct
xiaomi	RT @xiaomi: Do you love our planet? Send us you best pics of Mother Earth! #Xiaomi #MotherEarthDay https://t.co/KPV2QV4jX2	Replied to	Direct
xiaomi	RT @Xiaomi: Here's an unboxing review of #MiNote10 from our super Mi Fan Richmond! Richmond is one of the lucky winners of our #MiSteryBOX…	Retweeted	Direct
accenture	RT @agoeskoes: Unlocking The Power of Digital. #by @antgrasso #infographic #infosec #FutureofWork #business #management #leadership https://t.co/pv3q777xnl @Accenture #Innovation #4IR #Industry40 #DigitalTransformation https://t.co/sV3XLiKQIh	Independent	Indirect
mckinsey	RT @chidambara09: RT @MasterofIoT: #ArtificialIntelligence - the next global frontier #AI #ML #DL #BigData #fintech #Insurtech #Marketing #Analytics #Robotics #Healthtech #MarTech #Robots #DeepLearning #MachineLearning via @McKinsey cc @DeepLearn007 @Kai… https://t.co/JlAIAKyOrV	Quoted	Indirect
xiaomi	RT @xtremeskinsuk: With @ThreeUK announcing that @xiaomi are coming to the UK, we've decided to giveaway an Xiaomi Mi 6 wrapped in some goodness. Simply RT+F to ENTER. Ends 11/05/18 23:59. Further info https://t.co/HtECkekyYV (UK entrants ONLY) https://t.co/S41cJTWdsb	Replied to	Indirect
maccosmetics	RT @adorechristian: this touched my entire heart today!! he was asking me for a q tip and I handed him my brushes. He said with him having…	Retweeted	Indirect
hiltonhotels	I'm at Starbucks - @hiltonhotels in San Juan, NA https://t.co/TVi2qKbctj	Independent	Rest
sgs_sa	Regards to @WindfallGeotek and @SGS_SA! MMI geochemistry as first-pass investigation for CARDS targets? Learn more, https://t.co/tQcv2PUEEe	Quoted	Rest
hiltonhotels	@Kpedeaux @HiltonQuebec @HiltonHotels It didn't feel that cold this afternoon but it does now that the sun has gone down. I swear New Orleans damp cold is worse.	Replied to	Rest
xiaomi	@rainbownstoran's account is temporarily unavailable because it violates the Twitter Media Policy. Learn more.	Retweeted	Rest

Note:
N. Indirect RT: Number of Indirect Retweets of the company
N. Direct RT: Number of Direct Retweets of a company's Tweet
Rest: All tweets that are not re-posting someone else's tweet.
Some retweets fall in the category of rest because of the inclusion of withheld tweets in our collection of tweets. According to countries' policies, Twitter can withhold access to the tweet, preventing us from retrieving the original text. For instance, mention of **sports_england** was withheld, and the only text retrieved was *This Tweet from @VGAGymnastics has been withheld in: United Kingdom.*, so we do know **sports_england** was mentioned but do not know how.
See https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country. "Many countries, including the United States, have laws that may apply to Tweets and/or Twitter account content. In our continuing effort to make our services available to people everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to withhold access to certain content in a particular country from time to time." Twitter, 2023.

TABLE J.20
Mentions by type and subtypes - Latin America - Spanish Examples

Company	Tweet (Mention)	Reference Type	Retweet Type
antevenio	RT @antevenio: ¿Sabes qué es exactamente el píxel de Facebook Ads y cómo funciona? https://t.co/wXdLxxyBR8 https://t.co/igpcC6GMdo	Independent	Direct
epsonlatin	RT @EpsonLatin No te pierdas la innnovación que Epson trae para ti, #Refacciones #Computo #Impresoras https://t.co/yUSCYpMrQS	Quoted	Direct
elcomercio	RT @elcomercio @monterocnn Todo lo que dijo #DonaldTrump su primera conferencia (FOTOS) https://t.co/l8SjaFCQXM https://t.co/GuhFIWOdqt	Replied to	Direct
ideamcolombia	RT @IDEAMColombia: Por medio de una actividad de discusión y compartir conocimientos, se busca actualizar y definir los roles en los partic…	Retweeted	Direct
globant	RT @EndeavorArg: Hoy escuchamos hablar de #InteligenciaArtificial en todos lados pero, ¿sabemos realmente qué es? El co-fundador y CEO de @Globant, Martin @Migoya, nos deja 10 puntos para entender esta tecnología >> https://t.co/5jqqxQ1BFA https://t.co/Ja1abrOVau	Independent	Indirect
electrohuila	RT @COCIER_: @Electrohuila Acompañamos a @Electrohuila con la #EficienciaEnergética https://t.co/4Dd8yw3yQm	Quoted	Indirect
alstom	RT @omarfayad Me reuní con Henri Poupart-Lafarge CEO de @Alstom , empresa que construirá los vagones del @TrenMayaMX en Ciudad Sahagún, región en donde este proyecto generará más de 11 mil 500 empleos,en coordinación con el @GobiernoMX. https://t.co/R831oZxZPi	Replied to	Indirect
gobvalle	RT @PoliciaCali: Así cuidamos a los ciudadanos de #CaliCo. Nuestras balizas advierten que estamos presentes en cada rincón de la ciudad act…	Retweeted	Indirect
havanna_arg	El @Havanna_arg de Caballito (Pasaje El Maestro y Rivadavia) es pequeño pero súper acogedor, y las personas que trabajan allí súper amigables. #Merienda	Independent	Rest
sodimacperu	Buena campaña, realmente genera conciencia @SodimacPeru @bomberosPE https://t.co/4Vo2JTfN7B	Quoted	Rest
electricaribesa	@ElectricaribeSA @laguajirahoy @LAGUAJIRAGOB @BETTYPERIODISTA @LaGuajira_ @laguajiracol_ @SGuerreroB @DiarioLaGuajira @DiarioDelNorte @kennethrivadene A que hora se restablece el servicio	Replied to	Rest
tecnoglasssa	This Tweet from @angiecervantes8 has been withheld in: United Kingdom.	Retweeted	Rest

Note:
N. Indirect RT: Number of Indirect Retweets of the company
N. Direct RT: Number of Direct Retweets of a company's Tweet
Rest: All tweets that are not re-posting someone else's tweet.
Some retweets fall in the category of rest because of the inclusion of withheld tweets in our collection of tweets. According to countries' policies, Twitter can withhold access to the tweet, preventing us from retrieving the original text. For instance, mention of **sports_england** was withheld, and the only text retrieved was *This Tweet from @VGAGymnastics has been withheld in: United Kingdom.*, so we do know **sports_england** was mentioned but do not know how.
See https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country. "Many countries, including the United States, have laws that may apply to Tweets and/or Twitter account content. In our continuing effort to make our services available to people everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to withhold access to certain content in a particular country from time to time." Twitter, 2023.

TABLE J.21
Mentions by type and subtypes - United Kingdom - English Examples

Company	Tweet (Mention)	Reference Type	Retweet Type
artisuk_	RT @ArtisUK_: Happy Easter! Want to #win a Squashy Mug? RT+F for your chance! #freebiefriday #competition https://t.co/O2QTuuDKwc	Independent	Direct
saintgobainuk	RT @SaintGobainUK. Find out how Saint-Gobain are embarking on a #MissiontoEarth to bring #sustainability and #innovation into #construction; creating more comfortable living and working environments https://t.co/UvjApYWxnl	Quoted	Direct
foxsportsasia	RT @FOXSportsAsia: More countries have issues! Now Thailand and Cambodia complain about the logistical trouble in Philippines. …	Retweeted	Direct
korn_ferry	RT @FieldNation ""The reasons [for the growth of the gig economy] include cost savings, access to high-calibre talent and ease of managing gig economy professionals" @Korn_Ferry #GigEconomy" https://t.co/tCN8nnwzyE	Independent	Indirect
commscope	RT @petergjones: This was fun, BUT - if you are making physical infrastructure choices, read this! @cs_morrisonfell from @CommScope. Thanks for being an @NBASETAlliance member. https://t.co/fCGQH7Uo4h	Quoted	Indirect
toptrumps	RT @littleboo_21: [AD-GIFTED] Time for #WinItWednesday! * Follow @littleboo_21 and @toptrumps & RT to win a Friends Top Trumps Quiz https:…	Retweeted	Indirect
sheffieldhosp	Cardiologists and #radonc come together to introduce the UK #cardiacSABR experience in the treatment of ventricular tachycardia. A truly multidisciplinary effort across several centres @HattonMQF @shepherde1711 @NcccOncology @SheffieldHosp @SouthTees #UKSABR19 https://t.co/HraC73oBo3	Independent	Rest
cambjunction	Our friends @TrinityBristol & @CambJunction have a £3k commission available for D/deaf or Disabled artist(s) to create brand-new performance / exhibition for an online audience. Deadline: 14 August. Find out more here: https://t.co/DQtLQHHx52 https://t.co/iRRx6BxA22	Quoted	Rest
sohohouse	@manyokawo's account is temporarily unavailable because it violates the Twitter Media Policy. Learn more.	Retweeted	Rest

Note:
N. Indirect RT: Number of Indirect Retweets of the company
N. Direct RT: Number of Direct Retweets of a company's Tweet
Rest: All tweets that are not re-posting someone else's tweet.
Some retweets fall in the category of rest because of the inclusion of withheld tweets in our collection of tweets. According to countries' policies, Twitter can withhold access to the tweet, preventing us from retrieving the original text. For instance, mention of **sports_england** was withheld, and the only text retrieved was *This Tweet from @VGAGymnastics has been withheld in: United Kingdom.*, so we do know **sports_england** was mentioned but do not know how.
See https://help.twitter.com/en/rules-and-policies/tweet-withheld-by-country. "Many countries, including the United States, have laws that may apply to Tweets and/or Twitter account content. In our continuing effort to make our services available to people everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to withhold access to certain content in a particular country from time to time." Twitter, 2023.

TABLE J.22
Sample mentions - Latin America

Mention text	Company	Reference Type
RT @copetranvirtual: ¿Qué les parece nuestros nuevos Doble piso Más? https://t.co/KbtTY245FM	copetranvirtual	retweeted
RT @guerrerorozco: Una vez finalizada la votación para la elección del nuevo Comité Directivo del Fondo Nacional del Turismo-Fontur, este…	iata	retweeted
@rialjorge @Aerolineas_AR No puede ser que los azafatas/os ganen más que los maestros. Con este nuevo gobierno están todos esperanzados. Desastreeeeee	aerolineas_ar	replied_to
@CRIC_Cauca @Matador000 @SuperPopayan @NotivisionCauca @GobCauca @ASOINCACAUCA @elnuevoliberal Gracias, como los necesitamos	gobcauca	replied_to
@GarcesRojasJuan presidente de la @VCtieneasamblea felicita al diputado y nuevo presidente de la Mesa Directiva de la Comisión Accidental de Paz y Post Conflicto, @dannielhoyos. Cuenta con todo el respaldo de la corporación para ser un departamento pionero de PAZ. @GobValle https://t.co/xD12VrbvAN	gobvalle	quoted
Estás cifras de nuevos contagios, 2.193 casos nuevos, prendieron las alarmas nuevamente al punto que podrían volver a decretar cuarentena total. @Charly_zapata1 @AlcaldiaDeCali @GobValle @JorgeIvanOspina @ClaraLuzRoldan https://t.co/1jqjOpaipP	gobvalle	quoted

Note: This table presents examples of tweets mentioning companies and the word “new”.

TABLE J.23
Sample mentions - UK

Mention text	Company	Reference Type
RT @FOXSportsAsia: Finally some GOOD NEWS coming in from #SEAGames2019! @CeresNegrosFC owner has helped the organisers arrange 17 new b…	foxsportsasia	retweeted
RT @nutanix: It's earnings time! In Q1 FY20 we continued our transition to subscription revenue, saw momentum with our new products, closed…	nutanix	retweeted
Our friends @TrinityBristol & @CambJunction have a £3k commission available for D/deaf or Disabled artist(s) to create brand-new performance / exhibition for an online audience. Deadline: 14 August. Find out more here: https://t.co/DQtLQHHx52 https://t.co/iRRx6BxA22	cambjunction	quoted
#NPC #npcindia #NPCwebinars #npccertificate #NPC_INDIA_GOV #KYP @Meghmani5 @HPCL @IFFCO_PR @BASF @BayerPharma @IndianOil_Delhi @AjantaPharmaLtd @TorrentPharmaIn @Cadila_Pharma @drreddys @2Hazira @AmerChemSociety @EastmanChemCo @IFDCnews @FertilizerNews @Fertilizer_Inst https://t.co/Wsshtqska0	basf	quoted
66% of #automotive #manufacturers believe their go-to-market cycles will get faster in the next five years. @Jabil's new report explores why: https://t.co/nh3Rng0TOp #auto #autonomousvehicle #plm #digitalplm #DigitalTransformation #study https://t.co/4s7ZfAFDUg	jabil	Independent
TEAM NEWS: Jordan Rossiter returns to the starting XI for the first time since August as #RangersFC face @HibsOfficial in the final match of the season. Full team news to follow on https://t.co/gOwDjobecr https://t.co/n8UsPyX3fg	hibsofficial	Independent

Note: This table presents examples of tweets mentioning companies and the word "new".

J.3 Language

TABLE J.24
Mentions by language - Latin America -

Language	N. Tweets	N. Firms
es	17,632,032	1,152
en	7,886,185	924
others	2,650,202	1,058
pt	1,326,095	764

Note: The table presents the top three "language" attributes for the mentions and others. It reports the Number of mentions (N. Tweets) and number of different companies by language (N. Firms). Languages refer to pt: Portugues, es: Spanish, en: English, others: Other languages. It includes all mentions, not all for which sentiment was estimated. It includes only mentions between 2017 and 2021.

TABLE J.25
Mentions by language - UK -

Language	N. Tweets	N. Firms
en	30,648,026	6,289
others	2,107,381	4,479
und	677,809	3,748
qme	552,648	4,080

Note: The table presents the top three "language" attributes for the mentions and others. It reports the Number of mentions (N. Tweets) and number of different companies by language (N. Firms). Languages refer to en: English, qme: for tweets with media links only, und: if no language could be detected. , others: Other languages. It includes all mentions, not all for which sentiment was estimated. It includes only mentions between 2017 and 2021.

J.4 Sentiment of Mentions

TABLE J.26
Sentiment of mentions in English - Latin America (Vader)

	N Tweets	Pct.
NEGATIVE	491,735	13.0
NEUTRAL	1,044,470	27.7
POSITIVE	2,241,101	59.3
TOTAL	3,777,306	100.0

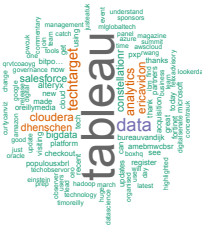
TABLE J.27
Sentiment of mentions in English - UK (Vader)

	N Tweets	Pct.
NEGATIVE	1,737,294	14.6
NEUTRAL	3,574,268	30.1
POSITIVE	6,568,471	55.3
TOTAL	11,880,033	100.0

J.5 N-grams Aboutnovation

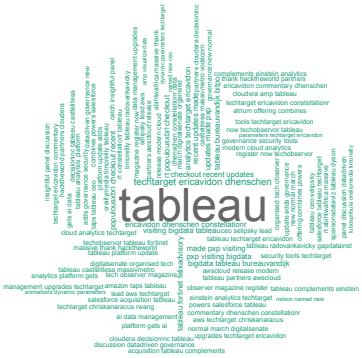
FIGURE J.21
Examples of Aboutnovation Words

(a) 1-gram



(b) 1-gram

(c) 3-gram



(d) 3-gram

Note: The wordclouds show words correspondent to their frequency in the whole history of mentions of the company from 2017 to 2022

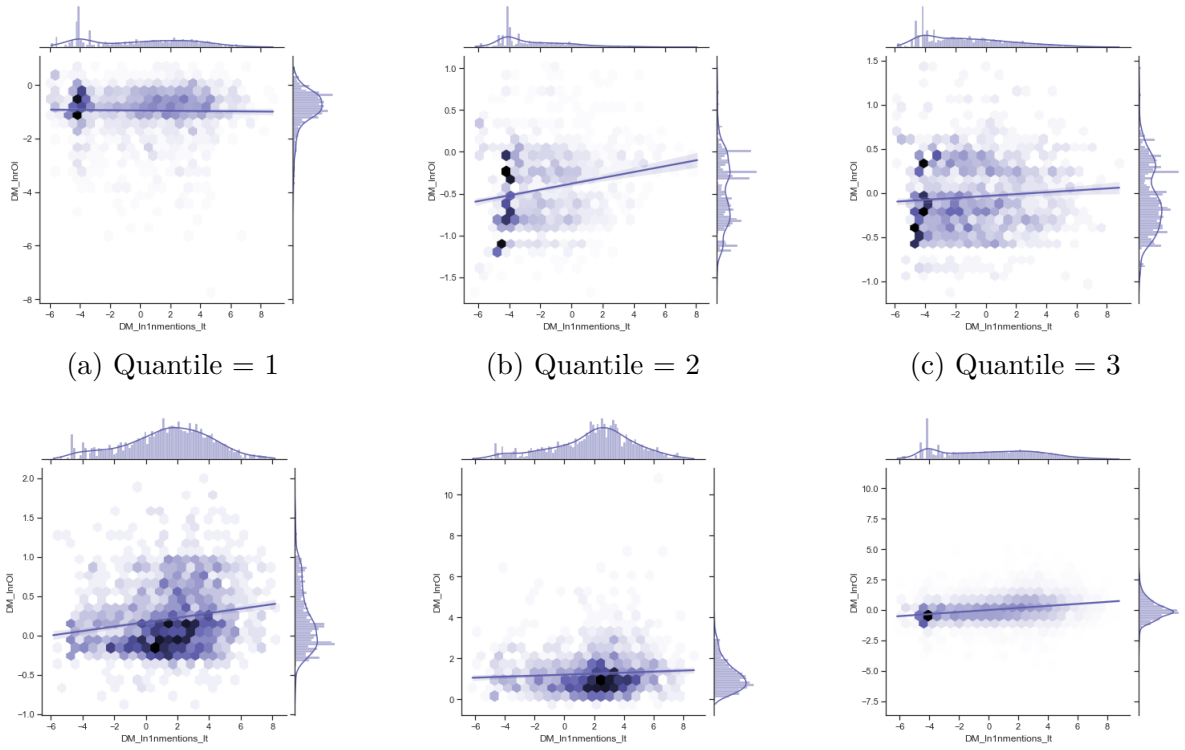
J.6 Labor Productivity and Mentions

TABLE J.28
Regressions of Labor Productivity on Customer Engagement
- United Kingdom -
- Likes and Retweets -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)	(6) ln(Rev/Empl)
ln(1+Number of Mentions)	0.084***	0.067***	-0.013			
	(0.004)	(0.018)	(0.027)			
ln(1+Number of Words C.)		0.004	0.033**	0.025***	0.058***	0.040***
		(0.014)	(0.013)	(0.009)	(0.004)	(0.010)
ln(1+avg. likes)		0.122***			0.104***	-0.051
		(0.026)			(0.024)	(0.059)
ln(1+avg. rt.)		0.013			0.029**	0.109**
		(0.014)			(0.013)	(0.048)
ln(1+sum likes)			0.051***	0.057***		0.106***
			(0.014)	(0.013)		(0.033)
ln(1+sum rt.)			0.003	0.001		-0.073**
			(0.011)	(0.009)		(0.037)
ln(Number of Employees)				0.004	-0.005	0.000
				(0.023)	(0.023)	(0.023)
ln(Number of Employees)^2				-0.006**	-0.004*	-0.005**
				(0.002)	(0.002)	(0.002)
Full Sample	X	X	X	X	X	X
Sectors FEs	X	X	X	X	X	X
Observations	7748	7748	7745	7745	7748	7745
R2 within	.0743	.0775	.0763	.0811	.0803	.0824
R2	0.259	0.262	0.261	0.265	0.264	0.266

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)
Note: Words C.: Words resulting from the count after cleaning the text of the mention (tweet)
Words: Words resulting from the count of the raw text of the mention (tweet).
ln(1 + x): ln(1+ number of tweets with number of type x)
Likes: Number of likes a Tweet (mention) has for the firm i
Rt.: Number of retweets a Tweet (mention) has for the firm i
Avg.: Average of the type of tweet by the firm Sum.: Sum of type for the whole firm

FIGURE J.22
Regressions of Labor Productivity Quantiles on Customer Engagement
United Kingdom



(a) Quantile = 1 (b) Quantile = 2 (c) Quantile = 3

(d) Quantile = 4 (e) Quantile = 5 (f) All

Note: The Figure reports scatter plots of within transformations - within (2 digits) sector×country cells - for labor productivity (revenue over employment) vs natural logarithm of one plus the number of mentions; i.e., the difference between a particular value and the sector× cell average. Panels (a) to (e) report on Twox Success (within 1 hour). Panel (b) reports Twox to count. Panel (c) reports Twox Success for the sample for which we have customer engagement data. Panel (d) reports the (log) of the mentions counts. Panels a to c present the results of a robust regression to control for outliers where observations with large residuals tend to be down-weighted or dropped, We graph the data points with the weight information as the size of circles.

TABLE J.29
Regressions of Labor Productivity on Customer Engagement II
- United Kingdom -

	(1) ln(Rev/Empl)	(2) ln(Rev/Empl)	(3) ln(Rev/Empl)	(4) ln(Rev/Empl)	(5) ln(Rev/Empl)
ln(1+Number of Words C.)	0.026**	-0.015	-0.060***	0.019*	-0.044*
	(0.013)	(0.010)	(0.020)	(0.011)	(0.022)
ln(1+Number of Mentions)	0.057***				
	(0.017)				
ln(Number of Employees)	0.006	0.013	0.012	0.013	0.012
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
ln(Number of Employees)^2	-0.005**	-0.006***	-0.006***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln(1+direct ret.)		-0.136***	-0.123***		
		(0.015)	(0.015)		
ln(1+indirect ret.)		-0.023***	-0.014		
		(0.008)	(0.009)		
ln(1+rest of types)		0.266***	0.290***		
		(0.021)	(0.023)		
Mentioned (1: Yes)			0.187***		0.243***
			(0.068)		(0.071)
ln(1+independent)				0.190***	0.221***
				(0.020)	(0.023)
ln(1+quoted)				0.010	0.013
				(0.017)	(0.017)
ln(1+retweeted)				-0.118***	-0.087***
				(0.017)	(0.019)
Full Sample	X	X	X	X	X
Sectors FEs	X	X	X	X	X
Observations	7748	7748	7748	7748	7748
R2 within	.0789	.0985	.0992	.0932	.0942
R2	0.263	0.279	0.279	0.274	0.275

Linear regression. Standard errors in parentheses
clustered at the user level.
Outcome variable: ln(LP = Revenues/Number of Employees)

Note:

Words C.: Words resulting from the count after cleaning the text of the mention (tweet)
Words: Words resulting from the count of the raw text of the mention (tweet).
ln(1 + x): ln(1+ number of mentions of type x)
Independent: N. of mentions (tweets) of the independent type
Quoted: N. of mentions (tweets) of the quoted type
Retweeted: N. of mentions (tweets) of the retweeted type
Direct ret: N. of retweets of a company's tweet'
Indirect ret: N. of retweets of a user's tweet mentioning a company
Rest of types = N. of mentions - Direct ret. - Indirect ret.

Appendix K Summary statistics: Twitnovation and Aboutnovation

TABLE K.30
Summary Statistics
- Latin America -

	Sum	Mean	SD	Min	Max	N
ln(LP = Revenues/Number of Employees)	4,681	11.5021	2.0799	-0	20	407
ln(1 + Number of words)	8,433	4.3222	1.4778	1	10	1,951
ln(1 + Number of mentions)	10,194	7.5176	2.9398	0	15	1,356
ln(sentiment-EN-Qu)	232	0.2458	0.1888	-1	1	942
ln(sentiment-ES-Qu)	244	0.2173	0.0955	0	1	1,122
ln(sentiment-EN-Re)	181	0.3522	0.2335	-0	1	513
ln(sentiment-ES-Re)	161	0.1933	0.1527	0	1	832
ln(1+likes-Qu)	8,591	7.5692	3.2309	0	16	1,135
ln(1+likes-Re)	8,591	7.5692	3.2309	0	16	1,135
ln(1+retweets-Re)	4,694	4.1356	2.9474	0	12	1,135
ln(1+retweets-Qu)	11,397	10.0410	4.3684	0	22	1,135
Rate of Response perc.	5,724	4.3432	6.4993	0	55	1,318
Rate of Response (Weighted)	9	0.0068	0.0102	0	0	1,318
ln(1 + Rate of Response perc.)	1,602	1.2159	0.9214	0	4	1,318
ln(1 + Rate of Response (weighted))	9	0.0067	0.0100	0	0	1,318
Mean Twox .9 Success 30 days	104	0.0532	0.1041	0	2	1,951
Mean Twox .9 Success 1 hour	8	0.0042	0.0144	0	0	1,951
rfid	142	0.0728	1.4966	0	46	1,951
ai_long	13,238	6.7852	28.5958	0	573	1,951
ai_short	5,259	2.6955	15.7035	0	318	1,951
lcds	81	0.0415	0.4459	0	8	1,951
threedprinting	762	0.3906	2.7218	0	54	1,951
solar	3,648	1.8698	22.9317	0	535	1,951
oled	6,141	3.1476	26.8243	0	873	1,951
virtual_reality	987	0.5059	2.5560	0	47	1,951
gps	550	0.2819	2.3117	0	73	1,951
driverless	215	0.1102	1.7162	0	51	1,951
smartphone	3,691	1.8919	14.2795	0	299	1,951
fracking	440	0.2255	2.8959	0	83	1,951
disk_drive	158	0.0810	1.1704	0	28	1,951
cloud	1,162	0.5956	8.2870	0	202	1,951
electric_cars	2,338	1.1984	7.3651	0	128	1,951
smartcard	40	0.0205	0.2186	0	3	1,951
bluetooth	187	0.0958	0.7514	0	14	1,951
optical_fiber	392	0.2009	1.6628	0	25	1,951
touch_screen	145	0.0743	0.8748	0	21	1,951
monoclonal_antibody	0	0.0000	0.0000	0	0	1,951
machine_vision	1	0.0005	0.0226	0	1	1,951
training	22,119	11.3373	42.1949	0	491	1,951
software	5,471	2.8042	20.7697	0	427	1,951
research	4,514	2.3137	14.7594	0	443	1,951
design	9	0.0046	0.0750	0	2	1,951
reputation	378	0.1937	1.1847	0	35	1,951
organizational	429	0.2199	1.0848	0	15	1,951

Note: Sentiment = Sentiment Score using Vader (Python) for EN (English) tweets or Sentiment Analysis for ES (Spanish) tweets.
Qu: Mentions of user
Re: Responses by Company
likes: Number of likes
retweets: Number of retweets of all mentions

TABLE K.31
Summary Statistics
- United Kingdom -

	Sum	Mean	SD	Min	Max	N
ln(LP = Revenues/Number of Employees)	96,905	12.5071	1.1141	5	24	7,748
ln(1+Number of Words C.)	50,954	6.5764	4.2022	0	15	7,748
ln(1+Number of Words)	52,236	6.7418	4.2144	0	15	7,748
Mentioned (1: Yes)	6,206	0.8010	0.3993	0	1	7,748
ln(1+Number of Mentions)	34,119	4.4035	3.2927	0	12	7,748
ln(1+direct ret.)	29,023	3.7459	3.1629	0	12	7,748
ln(1+indirect ret.)	6,062	0.7824	2.0022	0	12	7,748
ln(1+rest of types)	30,020	3.8745	3.0338	0	12	7,748
ln(1+pos. GBH)	22,219	3.5802	2.5204	0	12	6,206
ln(1+neg. GBH)	10,754	1.7329	2.0422	0	11	6,206
ln(1+neu. GBH)	33,045	5.3247	2.7415	0	12	6,206
ln(Number of Employees)	27,421	3.5391	1.7582	0	12	7,748
ln(Number of Employees)^2	120993	15.6161	14.7660	0	141	7,748

Note: GBH: Go et al. (2009) methodology to classify English tweets sentiment.
Words C.: Words resulting from the count after cleaning the text of the mention (tweet)
Words: Words resulting from the count of the raw text of the mention (tweet).
ln(1 + x): ln(1+ number of mentions of type x)
Independent: N. of mentions (tweets) of the independent type
Quoted: N. of mentions (tweets) of the quoted type
Retweeted: N. of mentions (tweets) of the retweeted type
Direct ret: N. of retweets of a company's tweet'
Indirect ret: N. of retweets of a user's tweet mentioning a company
Rest of types = N. of mentions - Direct ret. - Indirect ret.
retweets: Number of retweets of all mentions

Appendix L Assumptions of (De Loecker et al., 2020) for estimation of elasticities using COMPUSTAT dataset

1. DLW: Keep only observation for one industry (some firms are in several industries)
2. DLW: Drop firms without industry information
3. DLW: Take into account obs with industry code obs for which only d-1 digits in the d-category!!!
4. DLW: TRIM : no negative values of sales, cogs and xsga expenses

```
[breaklines=true]
            drop if sale_D<0
            drop if cogs_D<0
            drop if xsga<0
```

5. DLW: trim on sales-cogs ratio as mu_0 is simply $0.85 * sales/cogs$

```
            gen s_g = sale/cogs
            keep if s_g>0
            gen trim=0
            keep if year>1949
```

6. DLW: trim the sample on the bottom and top 1% of the sample of the sales to cogs ratio

```
* main results for 1% trim (below)
* robustness for appendix: change to
p(x) p(y) with x=2-5 and y=95-98
* robustness 2% and 5%
forvalues t=1/5 {
    bysort year: egen s_g_p_`t'  = pctile(s_g), p(`t')
                }
forvalues s=95/99 {
    bysort year: egen s_g_p_`s'  = pctile(s_g), p(`s')
                 }

    keep if s_g> s_g_p_1 & s_g< s_g_p_99
```

7. DLW: trim on costshares: on the bottom and top 1% of the sample, if missing or equal to 0

```
    xsga: Operating expense: Selling, General, and
    Administrative Expense
    cogs: Cost of goods sold
    kexp: capital expenses

    gen costshare1 = cogs_D/(cogs_D+kexp)
    label var costshare1 "cogs_D/(cogs_D+kexp)
    gen costshare2 = cogs_D/(cogs_D+xsga_D+kexp)
    label var costshare2 "cogs_D/(cogs_D+xsga_D+kexp)

    forvalues s=1/2 {
        bysort year: egen cs`s'_p1=pctile(costshare`s'), p(1)
        bysort year: egen cs`s'_p99=pctile(costshare`s'), p(99)
        drop if costshare`s'==0 | costshare`s'==.
        drop if costshare`s' > cs`s'_p99
        drop if costshare`s' < cs`s'_p1
        }
```

8. DLW: Replace thetas (estimated elasticities) for the rolling window models for a constant estimate for years before 1970

```
    bysort  year:   egen TOTSALES  = sum(sale_D)
    gen share_firm_agg  = sale_D/TOTSALES

    bysort ind4d year: egen TOTSALES_IND_4 = sum(sale_D)
    gen share_IND4 = TOTSALES_IND_4/TOTSALES
    gen share_ind_4  = sale_D/TOTSALES_IND_4
```

```
gen _b_lcogs    = .
egen nrind2 = group(ind2d)
forvalue  s = 1/23 {
reg y lcogs i i2 i3 k k2 k3 ki share_firm share_ind_4 ///
if year<1970 & nrind2==`s'
replace _b_lcogs = _b[lcogs] if year<1970 & nrind2==`s'
}
collapse (firstnm)  _b_lcogs, by(ind2d)
```

Appendix M Additional assumptions for estimation of elasticities for markups

1. In addition to assumption L5 we eliminate observations with cogs shares greater than one where it is found greater than sales to eliminate other extreme values for the new model. This step is only applied for the estimations

```
drop if cogshare=. if cogs_D>sale_D & cogs_D!=. & sale_D!=.
```

Appendix N Estimations of markups

TABLE N.32
Comparing Variables for Prodest Estimation: De Loecker et al.

	Mean	Median	SD	Min	Max	N	Mean	Median	SD	Min	Max	N
ln(Sales Deflated)	11.97	11.97	2.33	-0.11	19.95	241,619						
ln(COGS Deflated)	11.43	11.46	2.43	-0.11	19.83	241,619						
ln(Capital Deflated (K))	11.07	11.02	2.61	-0.11	19.94	240,508						
COGS Share in Sales (Adj.)	0.62	0.65	0.19	0.04	1.00	241,619						
Left Hand Side							0.19	0.09	0.77	-16.38	19.63	214,923
$\ln(Kt-1)$- $\ln(Kt)$							0.10	0.05	0.40	-10.08	8.24	217,331
$(\ln(Kt-1)-\ln(Kt))\,t-1$							0.10	0.06	0.37	-8.44	8.24	196,585
Left Hand Side $(t-1)$							0.20	0.09	0.72	-16.38	19.63	193,161

Note: Summary statistics for U.S. firms from 1955 to 2016. Data are obtained from Compustat (balance sheet and income statement variables). Nominal values are deflated using US GDP and are expressed in thousands of 2015 dollars. Sales are measured as firm-deflated revenue. The dataset contains 21,393 unique firms across 23 NACE Rev. 2 sectors at the two-digit level. All variables are winsorized at the 1% tail of the sales-to-cogs ratio.

Appendix O Forlani Method: Cobb-Douglas

To do so, note that the inverted loglinear version (the simplest case) of the demand function can be expressed as:

$$p_{it} = \frac{1}{\mu_{it}} d_{it} - \frac{1}{\eta_{it}} q_{it} \tag{O.78}$$

where $\mu_{it} = \frac{1}{1-\frac{1}{\eta_{it}}}$ and η_{it} has an inverse relationship with markups, μ_{it}. To do this we express the demand equation 51 in logarithmic terms:

$$q_{it} = (1 - \eta_{it})\, d_{it} - \eta_{it} p_{it} \tag{O.79}$$

The demand function can be expressed in this way, considering first, the maximization of profits problem where marginal revenue is equal to marginal cost, so markups can be expressed as a function of the elasticity of the reverse demand that firms face (the more inelastic, the greater the markup). Second, by replacing in the log-demand equation O.79, such expression for markups. Thus we can obtain a log-linear demand function expressed in terms of its markups, quantity of output, and demand shocks, as in equation O.78. A simpler log-linear case from demand equation 51 is one trivial that can generalize to various cases: $p_{it} = \frac{1}{\mu_{it}} q_{it}^{-\frac{1}{\eta_{it}}}$ where $\mu_{it} = \frac{1}{1-\frac{1}{\eta_{it}}}$

Now, by substituting p_{it} according to the revenue expression from equation 44, taking into account the expression for markup $\mu_{it} = \frac{1}{1-\frac{1}{\eta_{it}}}$, we obtain an expression of revenue in logarithmic terms that can have a first-order linear approximation around the profit maximization solution; it uses the inverse relationship between markup and elasticity of demand. Thus, we can also express the log-revenue of the firm as. This also implies that using the actual quantity of product q_{it} and revenue sold by a firm r_{it}, the profit-maximizing markup μ_{it} measured with production function parameters, and the cost minimization condition as explained above, one can recover the firm-specific demand measure of d_{it} as:

$$d_{it} \sim \mu_{it} r_{it} - q_{it} \tag{O.80}$$

Thus,

$$r_{it} \sim \frac{1}{\mu_{it}} (q_{it} + d_{it}) \tag{O.81}$$

By introducing the log-linear production function from equation 43 into this log-linear revenue equation O.81, a firm's revenue can be expressed as a function of its inputs of production and a decomposed shock (omega) $\omega_{it} = (a_{it} + d_{it})$. As mentioned above, d_{it} represents the demand shock and a_{it} is a technological shifter:

$$r_{it} = \frac{\alpha_M}{\mu_{it}} (m_{it} - k_{it}) + \frac{\alpha_L}{\mu_{it}} (l_{it} - k_{it}) + \frac{\gamma}{\mu_{it}} k + \frac{1}{\mu_{it}} (a_{it} + d_{it}) \tag{O.82}$$

Rearranging the terms of the above equation, the expression can be split between variables that can be observed in the data on the left-hand side (LHS) and the portion that needs to be estimated to obtain the elasticity of the flexible inputs (β's) in equation 45, on the righthand side (RHS). The latter is in the function of the quasifixed input (capital) and the stochastic error omega.

We cannot estimate equation O.82 if we allow markups to vary across observations. However, from equation 42, we know that $s_{Mit} = \frac{\alpha_M}{\mu_{it}}$ so by dividing this equation through s_{Mit} we get the expression to estimate econometrically:

$$LHS_{it} \equiv \frac{r_{it} - s_{Mit} (m_{it} - k_{it})}{s_{Mit}} = \frac{\gamma}{\alpha_M} k_{it} + \frac{\alpha_L}{\alpha_M} (l_{it} - k_{it}) + \frac{1}{\alpha_M} \omega_{it} \tag{O.83}$$

Thus, we can observe all variables from the LHS as we have data on the share of expenditures in materials s_{Mit}. This equation applies when we assume labor is not a variable input and, like capital, is predetermined before period t. De Loecker's estimations consider COGS, which includes materials and labor costs, as the variable input. To compare with De Loecker's estimations, we also assume that labor is a variable input and it is not predetermined as capital. We can assume that equation 42 also holds for labor. Thus, $\alpha_L = s_{Lit}\mu_{it}$, where μ_{it} and we know that $\frac{\alpha_L}{\alpha_M} = \frac{s_{Lit}}{s_{Mit}}$. By replacing this ratio in equation O.83, and reorganizing to the left we get:

Appendix P Forlani method: Translog

Thus, the loglinear version of the production function in the context of Model 3 is:

$$\begin{aligned} q_{it} = a_{it} + \sum_{X \in M,L,K} \left[\alpha_x ln X_{it} + \frac{1}{2}\alpha_{xx} ln(X_{it})^2 \right] \\ + \alpha_{MK} ln M_{it} ln K_{it} + \alpha_{ML} ln M_{it} ln L_{it} + \alpha_{LK} ln L_{it} ln K_{it}. \end{aligned} \tag{P.1}$$

As with equation O.82 in model 2, revenues can be expressed as the equation below.

$$\Delta r_{it} \approx \frac{\gamma}{\bar{\mu}_{it}} + \bar{s}_{Mit}\left(\Delta m_{it} - \Delta k_{it}\right) + \bar{s}_{Lit}\left(\Delta l_{it} - \Delta k_{it}\right) + \frac{1}{\bar{\mu}_{it}}\Delta\omega_{it} \tag{P.2}$$

Where, $\Delta r_{it} = \ln R_{it} - \ln R_{it-1}$, $\Delta m_{it} = \ln m_{it} - \ln m_{it-1}$, $\Delta l_{it} = \ln L_{it} - \ln L_{it-1}$ and $\Delta k_{it} = \ln K_{it} - \ln K_{it-1}$. The share of expenditures of an input X is the average between the present and lagged value of such share $\bar{s}_{Xit} = \frac{s_{Xit} + s_{Xit-1}}{2}$; X represents labor L or cost of materials M. In this case, ω combines TFPQ and demand shifters.

In the translog case, equation 41 in Model 1, which emerges from the cost minimization, implies that:

$$\hat{\alpha_{M_{it}}} = \alpha_M + \alpha_{KM} k_{it} + \alpha_{LM} l_{it} + \alpha_{MM} m_{it} \tag{P.3}$$

Appendix Q Descriptives of markups

TABLE Q.33
Comparing Raw Variables: De Loecker et al.

	Mean	Median	SD	Min	Max	N
ln(Sales Deflated)	11.87	11.91	2.41	-0.11	19.95	248,358
ln(COGS Deflated)	11.36	11.40	2.46	-0.11	19.83	248,358
ln(Capital Deflated (K))	11.02	10.98	2.63	-0.11	19.94	247,082
COGS Share in Sales (Raw)	0.68	0.66	0.83	0.04	80.59	248,358

Note:
Summary statistics for U.S. firms from 1955 to 2016. Data are obtained from Compustat (balance sheet and income statement variables). Nominal values are deflated using US GDP and are expressed in thousands of 2015 dollars. Sales are measured as firm-deflated revenue. The dataset contains 21,393 unique firms across 23 NACE Rev. 2 sectors at the two-digit level. All variables are winsorized at the 1% tail of the sales-to-cogs ratio.

TABLE Q.34
Comparing Raw Variables: De Loecker et al.

	p10th	p50th	p75th	p90th	p95th	p99th	N
cogs_share	0.34	0.66	0.77	0.86	0.91	1.75	248,358

Note:
This table shows the percentiles of the original cogs shares of the distribution.

TABLE Q.35
Cogshares percentiles

	mean	N	min	max
1	0.24	24,836	0.04	0.34
2	0.40	24,836	0.34	0.46
3	0.50	24,836	0.46	0.54
4	0.58	24,836	0.54	0.61
5	0.64	24,835	0.61	0.66
6	0.69	24,836	0.66	0.71
7	0.73	24,836	0.71	0.75
8	0.77	24,836	0.75	0.80
9	0.83	24,836	0.80	0.86
10	1.39	24,835	0.86	80.59
Total	0.68	248,358	0.04	80.59

Note: This table shows the percentiles of the original cogs shares of the distribution.

TABLE Q.36
Descriptives: Global sample

	mean	**sd**	min	max	**sd:between**	**sd:within**
(1) μ	1.56	157.24	0.41	260,800.27	32.39	149.83
(2) μ	1.31	124.50	0.28	206,524.88	25.72	118.63
(3) μ	1.10	1.78	-1,703.23	77.20	2.78	0.97
(1) μ_R	1.56	158.12	0.40	262,264.59	32.57	150.67
(2) μ_R	1.32	116.06	0.20	192,463.25	24.08	110.58
(3) μ_R	1.11	1.12	-833.82	67.53	1.24	0.75

(a) Full sample

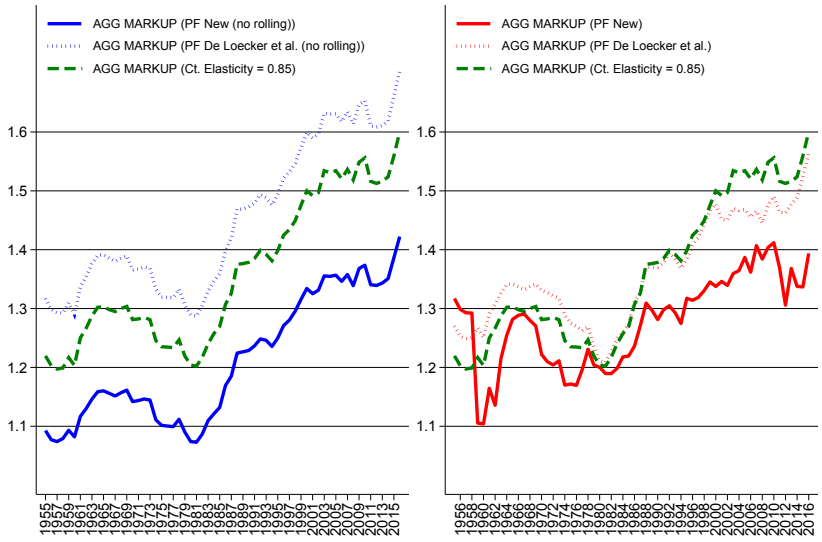
	mean	**sd**	min	max	**sd:between**	**sd:within**
(1) μ	1.91	34.98	0.63	13,386.44	24.02	28.39
(2) μ	1.20	2.07	0.41	624.69	2.50	1.14
(3) μ	1.12	3.20	-1,447.09	17.65	2.95	2.26

(b) J-Tag dissagregation

Note: μ_R: Rolling Markups.
(1) De Loecker et al. style - Model 1
(2) Model 2 - New - Cobb Douglas
(3) Model 3 - New - Translog

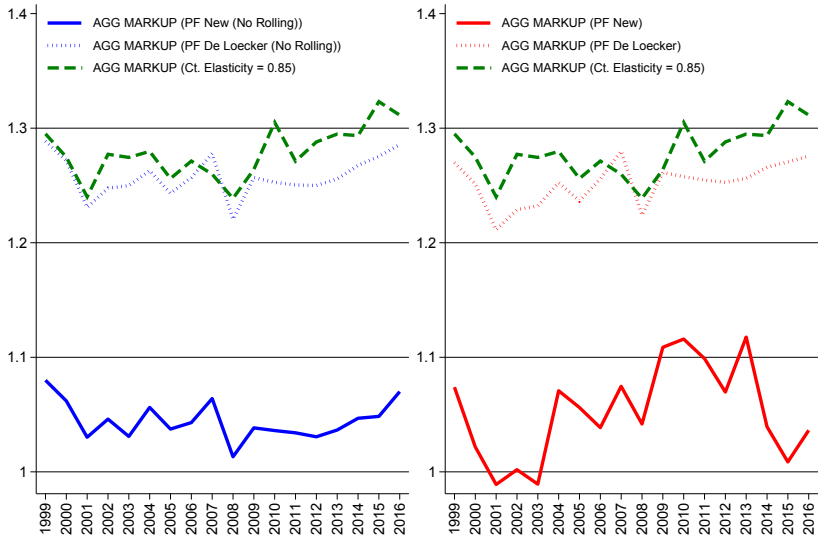
Appendix R Version for US (Markups): All Rolling Vs. Non-Rolling Both Models as in Figure 30b

FIGURE R.23
Average U.S. Markup (weighted by sales). De Loecker et al. vs. New - (1955 - 2016)



Note: The figure presents aggregate markups (weighted by sales) for both models' rolling and no rolling estimations. **Left panel** represents all no-rolling estimations or cases in which constant elasticity only varies by sector or is constant for the whole economy. **Right panel** presents the rolling estimations, contrasting models 1 and 2 plus adding the dashed green line with constant elasticity as reference.

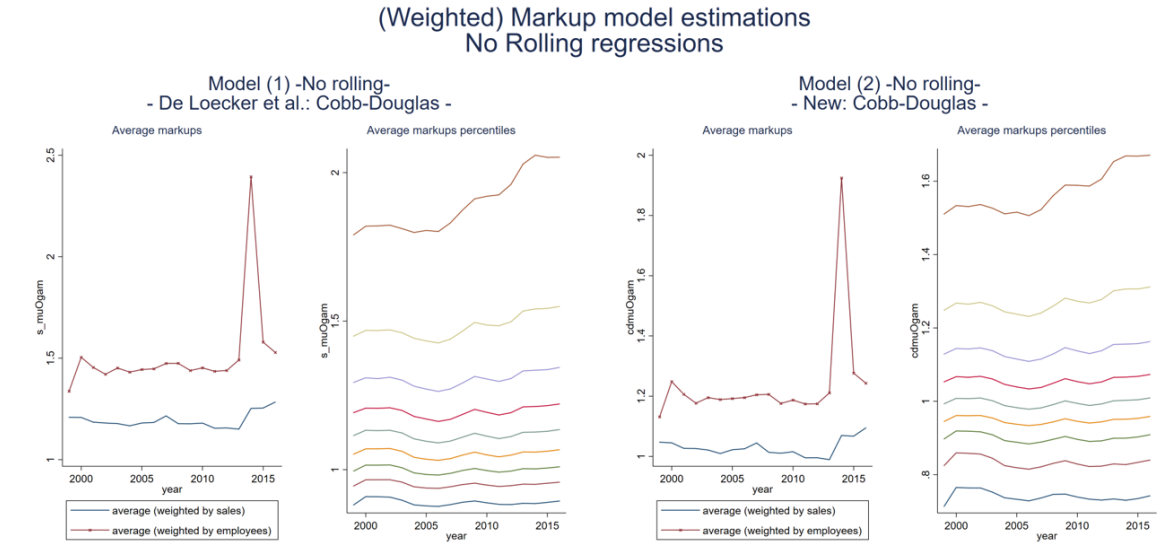
FIGURE R.24
Average European Markup (weighted by sales). De Loecker et al. vs. New (1999 - 2016)



Note: The figure presents aggregate markups (weighted by sales) for both models' rolling and no rolling estimations. **Left panel** represents all no-rolling estimations or cases in which constant elasticity only varies by sector or it is constant for the whole economy. **Right panel** presents the rolling estimations, contrasting models 1 and 2 plus adding the dashed green line with constant elasticity as reference.

Appendix S Weighted version for 131 countries (Markups)

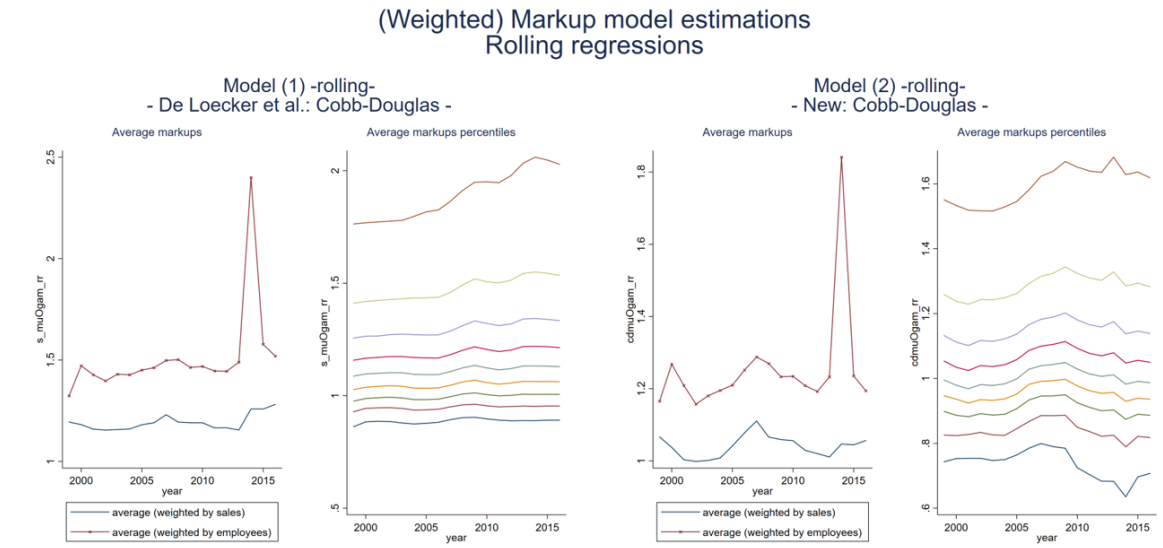
FIGURE S.25
De Loecker Style vs. Model 2



Notes:
(1) Markups obtained by industry sector: Nace Rev. 2.(1 to 20 sectors). Weighted by sales.
(2) s_muOgam: De Loecker style, cd_muOgam: simple cobb douglas transformed model
(3) Each line represents a percentile (lower 10, higher 90).
(4) Countries included: 131

ORBIS data

FIGURE S.26
De Loecker Style vs. Model 2, technologies change

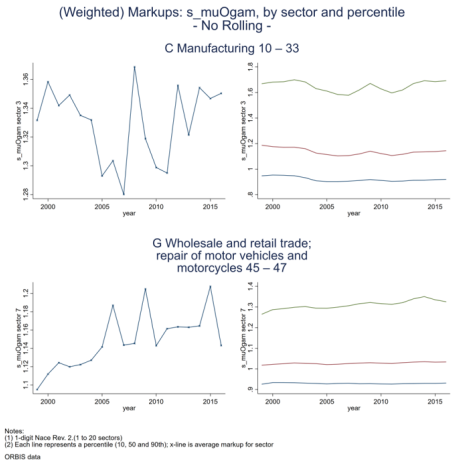


Notes:
(1) Moving windows of five years. Markups obtained by industry sector: Nace Rev. 2.(1 to 20 sectors). Weighted by sales.
(2) s_muOgam: De Loecker style, cd_muOgam: simple cobb douglas transformed model
(3) Each line represents a percentile (lower 10, higher 90).
(4) Countries included: 131

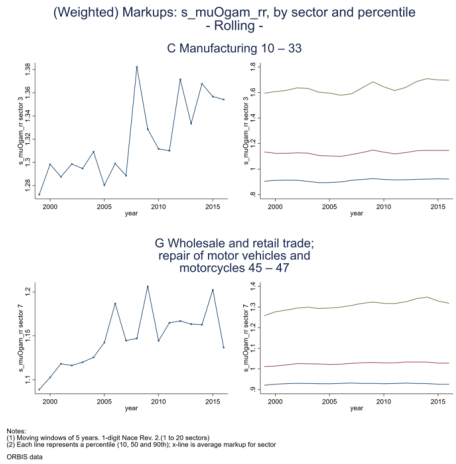
ORBIS data

Appendix T Weighted version for 131 countries (Markups) - By Industry

FIGURE T.27
De Loecker Style (Model 1) - Manufacturing and Wholesale

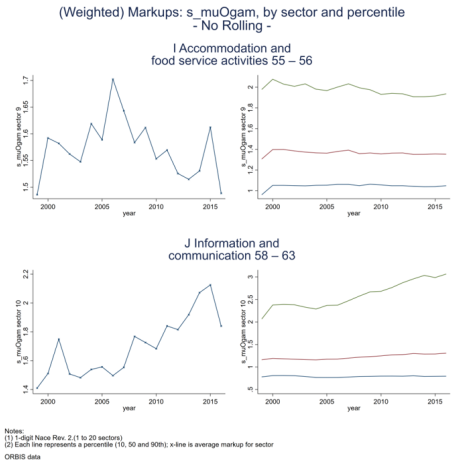


(a) No Rolling

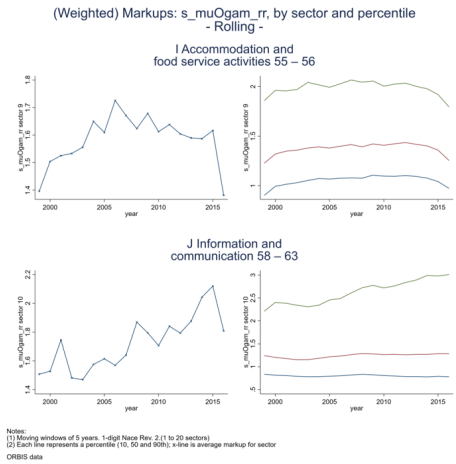


(b) Rolling

FIGURE T.28
De Loecker Style (Model 1): Accommodation and Information

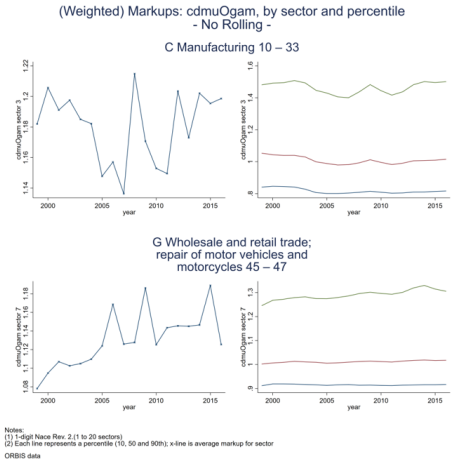


(a) No Rolling

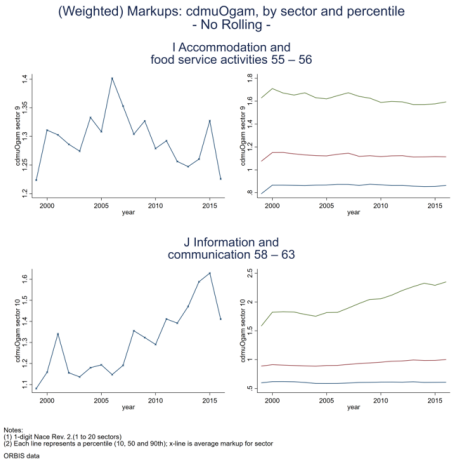


(b) Rolling

FIGURE T.29
Model 2: Manufacturing and Wholesale retail

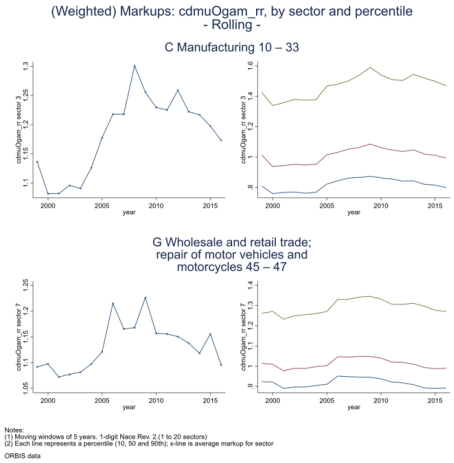


(a) No Rolling

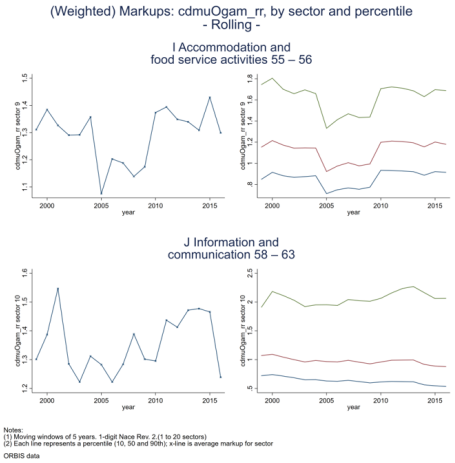


(b) Rolling

FIGURE T.30
Model 2: Accommodation and Information



(a) No Rolling



(b) Rolling

Appendix U Weighted version for 131 countries (Markups) - By Tech Sector

FIGURE U.31
Tech sector by sub-sector: Model (1)

Notes:
(1) No rolling windows. For each tech sector: if techsecx==sss
(2) CD: Cobb Douglas. Trans: Trans log
(3) sss is equal to a main tech sector, according to J tag OECD taxonomy for tech industries
(4) Each line represents a percentile (lower 10, higher 90).
(5) Countries included:

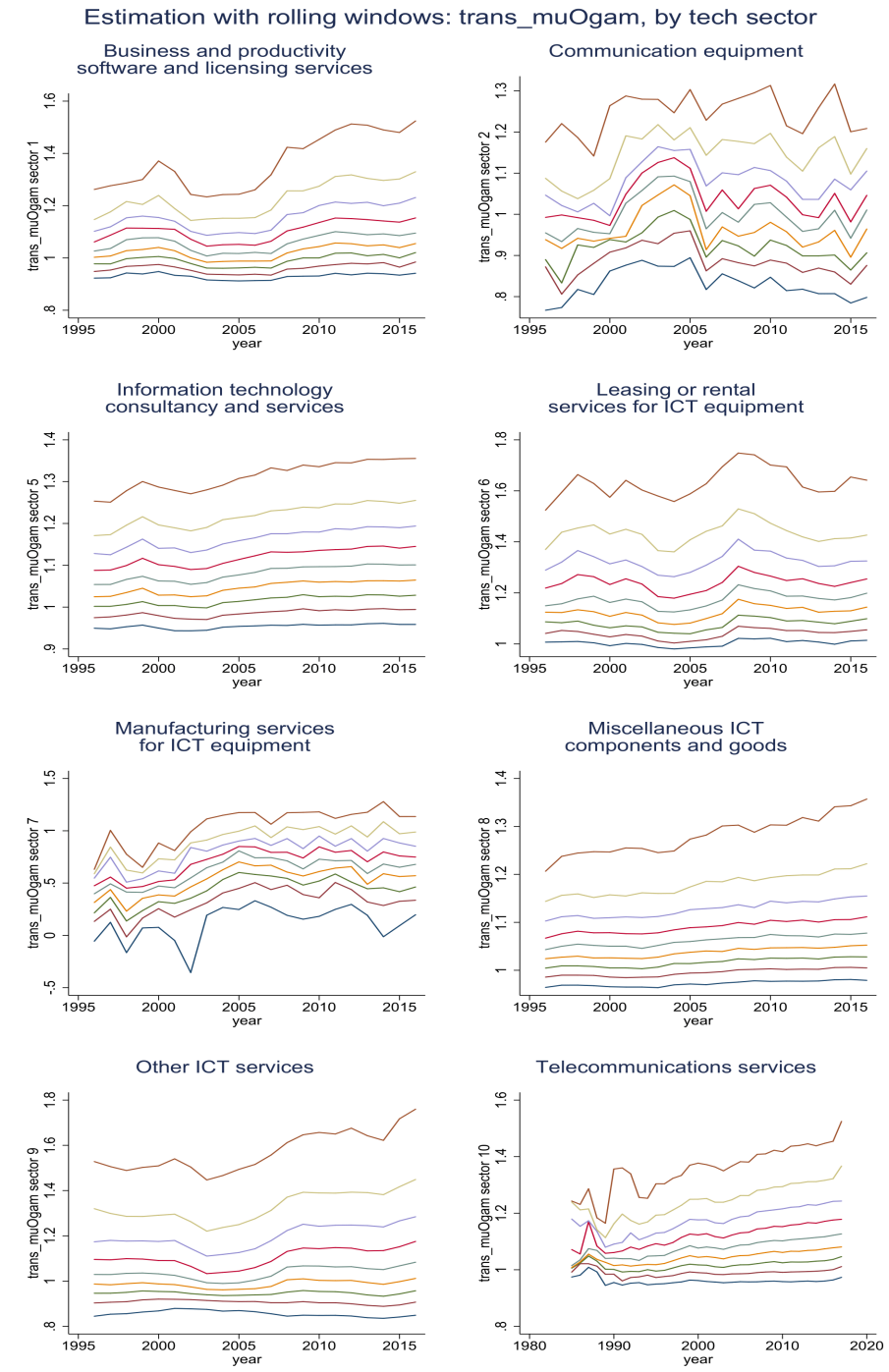
ORBIS data

FIGURE U.32
Tech sector by sub-sector: Model (2)

Notes:
(1) No rolling windows. For each tech sector: if techsecx==sss
(2) CD: Cobb Douglas. Trans: Trans log
(3) sss is equal to a main tech sector, according to J tag OECD taxonomy for tech industries
(4) Each line represents a percentile (lower 10, higher 90).
(5) Countries included:

ORBIS data

FIGURE U.33
Tech sector by sub-sector: Model (3)



Notes:
(1) No rolling windows. For each tech sector: if techsecx==sss
(2) CD: Cobb Douglas. Trans: Trans log
(3) sss is equal to a main tech sector, according to J tag OECD taxonomy for tech industries
(4) Each line represents a percentile (lower 10, higher 90).
(5) Countries included:

ORBIS data