DISCUSSION PAPER SERIES

DP17400

STARTUP TYPES AND MACROECONOMIC PERFORMANCE IN EUROPE

Ralph De Haas, Vincent Sterk and Neeltje van Horen

FINANCIAL ECONOMICS, INDUSTRIAL ORGANIZATION, INTERNATIONAL MACROECONOMICS AND FINANCE, MACROECONOMICS AND GROWTH AND PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

STARTUP TYPES AND MACROECONOMIC PERFORMANCE IN EUROPE

Ralph De Haas, Vincent Sterk and Neeltje van Horen

Discussion Paper DP17400
Published 21 June 2022
Submitted 15 May 2026

Centre for Economic Policy Research
187 boulevard Saint-Germain, 75007 Paris, France
2 Coldbath Square, London EC1R 5HL
Tel: +44 (0)20 7183 8801
www.cepr.org

STARTUP TYPES AND MACROECONOMIC PERFORMANCE IN EUROPE

Abstract

<p>We construct a new dataset of 1.3 million European startups. Cluster analysis identifies five startup types present across countries, industries, and cohorts. Initial differences across types persist, and each displays a characteristic life cycle in productivity, employment, and survival. We embed these startup types in a firm dynamics model with heterogeneous entry, investment, and financing, and quantify how policy can shift cohort composition. Differentiated corporate tax reforms favoring high-performance types yield substantial productivity gains. Uniform fiscal instruments differ sharply: capital subsidies raise output and productivity through composition effects, labor subsidies reduce them, and interest subsidies are roughly neutral.</p>

JEL Classification: D22, E24, E62, G32, G38, H25, L25, L26

Keywords: Startups, Firm dynamics, Capital structure, Corporate tax, Subsidies

Ralph De Haas - dehaasr@ebrd.com
EBRD, KU Leuven and CEPR

Vincent Sterk - v.sterk@ucl.ac.uk
Department of Economics, University College London and CEPR

Neeltje van Horen - n.vanhoren@uva.nl
University of Amsterdam and CEPR

Acknowledgements
We thank Filippo di Mauro for his enthusiastic support of this project, Victoria Marino for excellent research assistance, Marco Matani and Tommaso Bighelli for their expert help with running our code as part of constructing the 8th vintage of the CompNet dataset, and Tommaso Bighelli, Cagatay Bircan, John Haltiwanger, Sebnem Kalemli-Ozcan, Pete Klenow, Javier Miranda (discussant), Kjetil Storesletten and seminar participants at University College London, Institute for International Economic Studies (Stockholm), University of Regensburg, Danmarks Nationalbank, University of Bristol, University of Essex, Hong Kong University of Science and Technology, National University of Singapore, Hebrew University of Jerusalem, Toulouse School of Economics, Padova Macro Talks, Sao Paulo School of Economics, Nordic Summer Symposium in Macroeconomics, European Economic Association Annual Congress (Milan), and the CompNet ProdTalks for useful comments. The views expressed in this paper are ours and not necessarily those of the European Bank for Reconstruction and Development (EBRD).

Startup Types and Macroeconomic Performance in Europe*

Ralph De Haas[†], Vincent Sterk[‡], and Neeltje Van Horen[§]

May 15, 2026

Abstract

We construct a new dataset of 1.3 million European startups. Cluster analysis identifies five startup types present across countries, industries, and cohorts. Initial differences across types persist, and each displays a characteristic life cycle in productivity, employment, and survival. We embed these startup types in a firm dynamics model with heterogeneous entry, investment, and financing, and quantify how policy can shift cohort composition. Differentiated corporate tax reforms favoring high-performance types yield substantial productivity gains. Uniform fiscal instruments differ sharply: capital subsidies raise output and productivity through composition effects, labor subsidies reduce them, and interest subsidies are roughly neutral.

JEL classification: D22, E24, E62, G32, G38, H25, L25, L26

Keywords: Startups, firm dynamics, capital structure, entrepreneurship, corporate tax, subsidies

---

*We thank Filippo di Mauro for his enthusiastic support of this project, Victoria Marino for excellent research assistance, Marco Matani and Tommaso Bighelli for their expert help with running our code as part of constructing the 8$^{\text{th}}$ vintage of the CompNet dataset, and Tommaso Bighelli, Çağatay Bircan, John Haltiwanger, Şebnem Kalemli-Özcan, Pete Klenow, Javier Miranda (discussant), Kjetil Storesletten and seminar participants at University College London, Institute for International Economic Studies (Stockholm), University of Regensburg, Danmarks Nationalbank, University of Bristol, University of Essex, Hong Kong University of Science and Technology, National University of Singapore, Hebrew University of Jerusalem, Toulouse School of Economics, Padova Macro Talks, Sao Paulo School of Economics, Nordic Summer Symposium in Macroeconomics, European Economic Association Annual Congress (Milan), and the CompNet ProdTalks for useful comments. The views expressed in this paper are ours and not necessarily those of the European Bank for Reconstruction and Development (EBRD).

[†]EBRD, KU Leuven, and CEPR. Email: dehaasr@ebrd.com.

[‡]University College London and CEPR. Email: v.sterk@ucl.ac.uk.

[§]University of Amsterdam and CEPR. Email: n.vanhoren@uva.nl.

1 Introduction

In many advanced economies, politicians are increasingly concerned about lackluster macroeconomic performance, as reflected in anaemic productivity growth and, in some countries, low employment levels (OECD, 2015; Akcigit and Ates, 2021, 2023). Naturally, policymakers are looking for novel levers to structurally improve these macroeconomic outcomes. A long-standing literature has explored several directions that policy can take, including tax measures to stimulate R&D and innovation (Bloom, Griffith, and Van Reenen, 2002; Akcigit, Hanley, and Stantcheva, 2022) and structural reforms to reduce distortions in labor markets (Hopenhayn and Rogerson, 1993), financial markets (Buera and Shin, 2017), and product markets (Edmond, Midrigan, and Xu, 2023). Such policies aim to improve the productivity of existing firms or to strengthen the Darwinian selection process that weeds out inefficient ones (Syverson, 2011). This paper investigates an entirely different, and largely unexplored, policy lever: influencing the types of new firms that enter the economy.

The idea of improving the composition of new firm cohorts—as opposed to "fixing" established generations—appears attractive for two reasons. First, startups are key drivers of job creation and productivity growth (Foster, Haltiwanger, and Krizan, 2001; Haltiwanger, Jarmin, and Miranda, 2013; Adelino, Ma, and Robinson, 2017).[1] Recent evidence (discussed below) also suggests that ex ante heterogeneity among newly established firms helps predict performance later in life. It follows that structural policies that successfully shift the types of firms that enter the economy may generate substantial macroeconomic gains. Second, forward-looking policies to shift the composition of startup cohorts appear attractive given the high rates of firm entry and exit, typically around 10 percent annually. This means that most of the firms that will be active 20 years from now have yet to be founded, while many current firms will no longer exist by then.[2]

Policymakers are increasingly targeting specific types of startups. In March 2026, the European Commission proposed the EU Inc. framework, a harmonized pan-European corporate

[1] We define startups broadly, as all new firms that enter the economy, instead of using a narrow definition that focuses on a subset of innovative new firms that aim to scale up rapidly.

[2] For instance, the US Longitudinal Business Database (LBD) shows that, in 2023, 73 percent of all firms were 20 years old or younger. The start-up rate in that year was 9.1 percent. In the appendix, we show that European startup rates are comparable. Dent, Karahan, Pugsley, and Şahin (2016) use the LBD to show that about half of all sectoral employment reallocation in the US since the end of the 1980s reflects changes in start-up employment (that is, the entry margin) rather than employment shifts among incumbent firms.

legal form designed to make it easier for innovative startups to incorporate, scale, and attract capital across the Single Market.[3] Yet, in practice, government policy often lacks empirical guidance on how to effectively stimulate entrepreneurship. As a result, many such efforts have failed. These observations raise important questions: Can targeted policies structurally improve macroeconomic performance by altering the mix of startup types? And, if sizeable gains are possible, what types of startup should be encouraged and how to identify these types? Currently, there is no clear answer to these questions.

We tackle these issues using large-scale administrative data sets for ten European countries: Croatia, Denmark, Finland, France, Italy, Lithuania, the Netherlands, Slovenia, Spain, and Sweden. The data contain a rich set of variables, derived from the balance sheets and income statements of individual startups. We collected these data in close collaboration with the Competitiveness Research Network (CompNet), which uses a distributed micro-data approach to create regularly updated, micro-based, and internationally harmonized data on European firms. These data provide unique cross-country panel observations on more than 1.3 million European startups, achieving representative coverage of the full startup population while combining the breadth of administrative sources with the variable richness of balance-sheet data.[4]

A key challenge is that startup types are latent. We address this issue by using K-means clustering, an unsupervised machine learning algorithm that has recently gained popularity in the applied economics literature as a way of dealing with latent heterogeneity. Underlying our application of this method is the idea that a startup's type is revealed by a number of key choices entrepreneurs make when commencing operations. Exploiting the richness of our data, we classify startups based on five important choice variables in the initial year of operation: employment; the capital-to-labor ratio; total assets; the leverage ratio; and the cash-to-assets ratio. The practical advantage of this approach is that these variables are easily observed from tax information at the very beginning of a firm's life cycle. They can

[3]The EU Inc. proposal responds to recommendations in the Draghi Report on European competitiveness and echoes earlier national initiatives, such as the UK Future Fund (2020) and expanded R&D tax credits for pre-revenue startups under the US Inflation Reduction Act.

[4]This combination addresses well-known limitations of alternative data sources, including the under-representation of small firms in Compustat, the single-cohort design of the Kauffman Firm Survey, the limited variable coverage of the US Longitudinal Business Database, and the poor capture of firm entry and exit in Orbis (Bajgar, Berlingieri, Calligaris, Criscuolo, and Timmis, 2020; Kalemli-Özcan, Sørensen, Villegas-Sanchez, Volosovych, and Yeşiltaş, 2024).

therefore readily be used to differentiate startups and to facilitate targeted policies.

The clustering algorithm endogenously groups startups into five types. This classification captures most of the empirical variation in the clustering variables. The same five startup types emerge in each country: *Large*, *Capital-intensive*, *High-leverage*, *Cash-intensive*, and *Basic*. Using a "meta-clustering" analysis, we verify that each of these types has similar characteristics across countries. Monte Carlo simulations confirm that this cross-country similarity is not driven by mechanical features of the clustering variables' distributions. We also show that the distribution of startups across these five types is quite stable across countries, sectors, and cohorts. When we track these types over a decade, their life-cycle profiles confirm that initial cross-type differences are persistent. These findings therefore indicate that the clustering algorithm captures fundamentally different firm types.

Based on the clustering outcomes, we then document heterogeneity in performance. We find large and persistent differences in employment and productivity across the various types. This implies that a change in the composition of startups can potentially have large (positive or negative) macroeconomic effects. A number of salient patterns emerge. In particular, the performance of the *High-leverage* startup type tends to be consistently poor. Even when we compare startups within the same country and economic sector, the cluster of highly leveraged firms displays substantially lower labor productivity and total factor productivity (TFP). By contrast, startups that stand out because of their capital-intensive production technology, their high level of employment and assets, or their high cash-to-assets ratio, typically boast higher productivity levels.

To interpret these patterns and quantify their policy implications, we calibrate an equilibrium firm dynamics model in which types differ in their technology (including scalability) and in their initial equity. Firm entry, financing, and hiring decisions are jointly determined. By calibrating type-specific parameters to match the multidimensional life-cycle profiles in the data, the model provides a structural interpretation of what makes each startup type fundamentally different. The calibrated model rationalizes these performance differences as the joint outcome of technological heterogeneity (TFP, returns to scale, fixed costs) and heterogeneity in initial firm equity. Large firms stand out because of high scalability, high fixed costs, and high equity. Capital-intensive firms are relatively scalable in capital, and have moderately high fixed costs and initial equity. Cash-intensive firms are much less scalable, but have low fixed costs and high TFP. High-leverage firms have very low initial equity and low TFP, but also low fixed costs.

In this setting, the macroeconomic effects of policy changes are determined not only by the responses of incumbent firms but also, importantly, by changes in the composition of startup types. The composition effect is governed by three empirically estimable objects: the life-cycle profiles of each type, the type-specific entry elasticities (an important yet rarely measured input into firm dynamics models), and the immediate impact of the policy on firm profits. Each of these is recovered transparently from our data, so that the counterfactual results can be traced back to the underlying empirical inputs.

We conduct two sets of counterfactual policy experiments. First, we quantify the "policy space" opened by a budget-neutral differentiation of corporate income tax rates across startup types. Such a reform alters the incentives of different types to enter and thus shifts the composition of new cohorts. Searching the entire space of admissible policies, we find that even moderate differentiation can yield substantial macroeconomic gains: a budget-neutral reform that changes corporate tax rates by no more than ten percentage points at the type level raises aggregate labor productivity by up to 4.5 percent. The largest productivity gains come from raising the tax on *Basic* startups, which finances lower taxes on other, more productive types.[5]

Second, we use the model to compare three uniform fiscal instruments financed through corporate-tax adjustments: a capital subsidy, a labor subsidy, and an interest subsidy targeted at young firms. Although none of the three is differentiated by startup type, each affects the startup composition through differential exposure, with markedly different macroeconomic consequences. A capital subsidy raises aggregate output and labor productivity through favorable composition effects, whereas an equivalently sized labor subsidy reduces both. An interest subsidy is roughly neutral in the aggregate but tilts composition toward *High-leverage* startups, which perform relatively poorly. Taken together, these experiments show that the macroeconomic impact of fiscal policy on new firms depends in large part on how it interacts with startup heterogeneity: capital, labor, and interest subsidies generate sharply different composition effects that representative-firm frameworks cannot capture.

[5]Stimulating *Large* and *Capital-intensive* startups is also attractive from a welfare perspective, as these types tend to pay higher wages and maintain modest profit margins.

Related literature. We build on an emerging literature that documents the importance of ex ante heterogeneity for firms' performance over their life cycle. This heterogeneity manifests itself in vastly different growth expectations among new entrepreneurs (Hurst and Pugsley, 2011), the predictability of firm performance based on observable characteristics of entrepreneurs and businesses at the moment of startup (Schoar, 2010; Guzman and Stern, 2015; Belenzon, Chatterji, and Daley, 2017; Guzman and Li, 2023), strong cohort effects over the business cycle (Sedláček and Sterk, 2017), and systematic differences in financial structure at founding that reflect firm and founder characteristics (Robb and Robinson, 2014). Quantitatively, ex ante heterogeneity accounts for the majority of variation in firm-level employment, as shown by Sterk ⓡ Sedláček ⓡ Pugsley (2021), who estimate an employment process using micro data.

We contribute to this literature in four important ways. First, we are the first to analyze how policy can exploit ex ante firm heterogeneity to try to improve macroeconomic outcomes. In doing so, we shift the focus from the widely studied "startup deficit"—a decline in firm entry observed in several countries over the past decades—to the *composition* of new startup cohorts as a determinant of aggregate productivity and employment.[6] This provides a fresh perspective on the micro origins of productivity dispersion (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; Hsieh and Klenow, 2014; Midrigan and Xu, 2014; Moll, 2014; Gopinath, Kalemli-Özcan, Karabarbounis, and Villegas-Sanchez, 2017; Brandt, Kambourov, and Storesletten, 2026; De Loecker and Syverson, 2022).

Second, we treat ex ante heterogeneity as a multidimensional object. Most existing studies focus on just one dimension of heterogeneity, such as firm size.[7] We instead jointly characterize firms by five key choice variables at startup and consider several performance measures (labor productivity, TFP, profitability, and exit probability). Our data enable micro-to-macro mapping along these dimensions across ten countries, painting a rich characterization of the European startup landscape. Moreover, the clustering algorithm classifies firms into types based on observables in the first year of operation, making our approach

[6] On the startup deficit, see Decker, Haltiwanger, Jarmin, and Miranda (2017); Alon, Berger, Dent, and Pugsley (2018); Karahan, Pugsley, and Şahin (2024); Biondi, Inferrera, Mertens, and Miranda (2023).

[7] Recently, Bernard, Dhyne, Magerman, Manova, and Moxnes (2022), using Belgian data on firm-to-firm trade, have underlined the importance of multiple attributes to explain firm-level success and dispersion in the firm-size distribution. Other work assesses specific ex ante differences across startup founders, such as cognitive and non-cognitive personality traits (Levine and Rubinstein, 2017) and age (Azoulay, Jones, Kim, and Miranda, 2020).

straightforward for policymakers to implement. Other studies have not made firm types observable (Albert and Caggese (2021) and Sterk ⓡ Sedláček ⓡ Pugsley (2021)).

Third, we obtain empirical estimates for the elasticity of firm entry and show that these elasticities are heterogeneous across startup types. Such estimates are rare, even though they are a standard input in firm dynamics models, which often assume either fixed entry (that is, a zero elasticity) or a free entry condition (that is, an infinite entry elasticity).[8] Our estimates can thus be used to impose more empirical discipline on models in the tradition of Hopenhayn (1992) and models with firm entry more generally.

Fourth, we embed empirically identified, ex ante heterogeneous startup types in a firm dynamics model, allowing us to compare the macroeconomic effects of differentiated tax reforms as well as uniform fiscal instruments (capital, labor, and interest subsidies) through a common lens. Because the model captures compositional channels that representative-firm frameworks abstract from (Garicano, Lelarge, and Van Reenen, 2016; Akcigit, Hanley, and Stantcheva, 2022), it reveals that these channels are quantitatively first-order and can reverse the sign of macroeconomic effects predicted by the standard analysis.

2 Data and clustering

2.1 Startups in the CompNet database

We carry out a cross-country analysis of startups using confidential micro-data that we collected in close collaboration with the Competitiveness Research Network (CompNet). CompNet maintains a regularly updated, micro-based, and internationally harmonized competitiveness data set for 20 European countries. To preserve confidentiality at the firm level, and to improve cross-country comparisons, CompNet uses a distributed micro-data approach as developed by Bartelsman, Haltiwanger, and Scarpetta (2004). Data are annually updated by sending standardized code to national statistical agencies. These agencies then run the code on their confidential firm data and aggregate it to the sector-year level in a standardized fashion. The data are subsequently returned to CompNet with key statistical moments that describe the distribution of various firm characteristics at the sector-year level for each

[8]Other estimates can be found in Sedláček and Sterk (2017) and Gutiérrez, Jones, and Philippon (2021). Yet, they rely on full-blown structural DSGE models to estimate the entry elasticity, whereas our method is agnostic. Moreover, we estimate elasticities for different startup types.

country. Particular care is taken to ensure a high level of cross-country consistency to allow for international comparisons and the identification of idiosyncratic country effects.[9]

To collect harmonized cross-country data on European startups, we embedded additional code in the standard instructions sent to national authorities in preparation for the eighth CompNet vintage.[10] Our code extracted data on all firms that commence their operations in a particular country and year (that is, a startup cohort). We thus define the start year as the year in which a firm is first economically active. We also observe each firm's formal registration year and drop observations if the lag between firm registration and actual startup is more than four years and/or if registration occurs *after* the actual start year (this only happens in a handful of cases). We also exclude firms that enter with more than 50 employees. Given our definition of startups as firms that become economically active for the first time in the data, such large entrants are unlikely to reflect genuine de novo firm creation but instead capture cases such as mergers, spin-offs, or subsidiaries of existing firms. Once a firm enters our data set, we can track it for several years. This allows us to construct comparable data on how firms develop in terms of their employment generation, productivity, and survival—all areas on which cross-country evidence remains scant.

This approach results in a unique cross-country and cross-sector panel of 1,345,489 startups established during 2000–2019.[11] We aggregate the data at the country–industry–cohort level, using either a one-digit or two-digit NACE Rev.2 industry classification.[12] For each country–industry–type–cohort–age cell, we observe the mean of various firm characteristics (employment, capital intensity, cash ratio, leverage) as well as labor productivity and TFP.[13] The unit of observation throughout the empirical analysis is therefore a cell, not an individual firm. This aggregation is consistent with our empirical specifications. Both the lifecycle

[9]See Lopez-Garcia and di Mauro (2015) for more details on the CompNet project.

[10]We cannot use data for Germany, Poland, and the Slovak Republic as the national data sources exclude firms with fewer than 20 employees (10 in the case of Poland).

[11]We drop firms that are not observed every year between startup and exit. Appendix Table A1 presents the coverage per country.

[12]The one-digit industries are administration; construction; hospitality; ICT; manufacturing; professionals; trade; and transport. Our baseline uses the one-digit classification because (i) we are primarily interested in aggregate outcomes and (ii) the two-digit classification roughly halves the sample, as cells with fewer than ten underlying firm observations are dropped per CompNet confidentiality requirements. We report two-digit results as a robustness check throughout.

[13]Labor productivity is real value added per employee. TFP is based on a homogeneous production function estimated at the two-digit sectoral level via the control function of Ackerberg, Caves, and Frazer (2015), aggregated using employment shares.

regressions in Section 3 and the entry-elasticity regressions in Section 4.2 are specified at the cell level by construction: our framework links the *average* characteristics and *number* of startups within a type to macroeconomic aggregates, and our high-dimensional fixed effects absorb systematic cross-country, cross-industry and cross-cohort differences.[14]

Table A2 provides a detailed description of all variables. All monetary variables are denominated in thousands of euros and PPP-adjusted. Real variables are deflated with sectoral price indices. We also retrieve the data split by startup type. This classification of startups will be discussed in Section 2.2. We verify in Appendix A that our CompNet data on firm creation are comparable to those of Eurostat. Biondi et al. (2023) compare the total firm population in CompNet and Eurostat and conclude that the two data sets align well.

2.2 Identifying startup types

We use K-means cluster analysis (Everitt, Landau, Leese, and Stahl, 2011) as a data-driven approach to categorize firms by their startup strategy. Cluster analysis is a type of unsupervised machine learning that has recently gained traction in applied economics to study empirical settings with latent heterogeneity (e.g., Bonhomme, Lamadon, and Manresa (2022); Stantcheva (2021)). In our application, the heterogeneity is in firm startup types, and the idea of using the clustering algorithm is that choices made in the first year help to reveal the type of startup.

Let x_i be a vector of firm-level clustering variables, in practice converted into z-scores to avoid arbitrary scaling effects. The clustering algorithm allocates each individual firm i into one of $j = 1, .., k$ clusters, in order to solve $\min \sum_{j=1}^{k} \|x_i - \bar{x}_j\|^2$, where $\bar{x}_j$ is the cluster mean. The algorithm begins with k seed values as the initial group means. Each observation is then assigned to the group with the closest mean. Based on that initial categorization, new group means are determined and these iterative steps continue until no observations switch groups.

We experiment with different k's by calculating the statistic $\eta^2 \equiv 1 - \frac{WSS}{TSS}$, where $WSS \equiv \sum_{j=1}^{k} \sum_{i \in C_j} \|x_i - \bar{x}_j\|^2$ is the within-cluster sum of squares and $TSS \equiv \sum_{i=1}^{n} \|x_i - \bar{x}\|^2$ is the

[14] In the entry-elasticity regressions, the dependent variable (log firm value) and the independent variable (log number of entrants) are both cell-level objects, so identification occurs at precisely the level at which the data are aggregated.

total sum of squares, with $\bar{x}$ being the unconditional mean across all observations.[15] We let k vary between 1 and 10, as visualised in the scree plots in Appendix Figure A5. At $k = 5$, the η^2 statistic is around 0.6, which means five clusters capture around three-fifths of the total variation in the clustering variables. Beyond $k = 5$, the η^2 statistic still increases but levels off. The data therefore suggest that our startups are optimally clustered into five well-separated (non-overlapping) clusters, each representing a distinct startup strategy.

We let the algorithm cluster startups using five key variables that entrepreneurs decide on when starting a business: the number of employees; real total assets; capital intensity (real fixed assets per employee); cash to total assets; and leverage (total debt to total assets). We thus cluster using variables that are decided at startup (but can be adjusted later on) rather than outcomes such as labor productivity. We choose these variables because prior literature identifies them as key startup decision parameters and because the microdata are complete across all sample countries.[16]

Note that real fixed assets are, by construction, the product of employment and capital intensity, introducing correlation among three of the five clustering inputs. However, all variables are converted into z-scores prior to clustering, which removes the multiplicative relationship that holds in levels. Moreover, the five variables capture conceptually distinct dimensions of startup strategy: labor-market scale, production technology, balance-sheet size, financing structure, and liquidity. Omitting total assets, for instance, would prevent the algorithm from distinguishing a firm that is capital-intensive due to a large balance sheet from one that is capital-intensive simply because it has very few employees. Table 1 confirms that the resulting clusters are not merely ordered by size: *Cash-intensive* and *High-leverage* startups are similar in employment and total assets yet differ sharply in financial structure. Finally, the clusters differ persistently along variables not used in the clustering (productivity, exit rates, wages, and profit margins, cf. Table 2) which would not occur if the classification were driven by a single redundant scale factor.

[15] This statistic has a similar interpretation to the R^2 reported in regression analysis.

[16] Earlier work has assessed the role of startup scale as measured by assets or employees (Albuquerque and Hopenhayn, 2004; Kerr and Nanda, 2010; Buera, Kaboski, and Shin, 2011) or set-up costs (Derrien, Mésonnier, and Vuillemey, 2025); liquidity and cash holdings (Bolton, Wang, and Yang, 2019); and leverage (Robb and Robinson, 2014; Derrien, Mésonnier, and Vuillemey, 2025). We also use capital intensity (fixed assets per employee) as a clustering variable because heterogeneity in production functions (and the resulting variation in the elasticity of substitution between capital and labor) is expressed in different choices about capital-to-labor ratios (Oberfield and Raval, 2021).

2.3 Comparing startup clusters across countries

We implement the cluster analysis using a separate micro data set for each country. There is therefore no a priori reason for the clustering outcomes to be similar across countries. Indeed, very different clusters may arise in different contexts. Moreover, even if the clusters would be similar, their shares might vary widely across countries. To assess the similarity of clusters, we run a second-stage "meta-clustering" analysis, which groups comparable clusters from different countries. This also provides an objective procedure to assign common names to similar clusters. Specifically, we repeat the clustering analysis while taking the cluster centers derived from each country's first-stage clustering as the observations.[17]

The four panels in Figure 1 summarize the outcome of this meta-clustering. Different meta-clusters are shown in different colours. The panels show five clusters arising in each country. For example, the first panel contains a red meta cluster of *Capital-intensive* startup clusters. Each of the individual red circles indicates a country-level cluster of startups that stand out nationally because of their high capital intensity. Such a distinct *Capital-intensive* cluster emerges in each country and the algorithm bunches them together in a single meta-cluster. Importantly, the variation between clusters within countries is much greater than the variation between countries in the same meta-cluster. In other words, the clusters in different countries are very similar. For the meta-clustering we obtain $\eta^2 = 0.96$. This indicates that variation between clusters explains the vast majority of the overall data variation. Only a very small contribution is left for cross-country variation within meta-clusters.[18]

One may worry there are mechanical reasons for the similar clusters in different countries, or that this similarity is a coincidence. To investigate this, we conduct a Monte Carlo experiment for the meta-clustering. We consider 1,000 random draws for the cluster variables, with means and standard deviations as observed in the data. However, in the experiment, these draws are i.i.d. so that no meta-clusters exist.[19] Each time we compute η^2. Appendix Figure A6 shows that these η^2 statistics are much lower in the experiment than the 0.96 observed

[17] That is, in the meta-clustering procedure, the units of observation are the z-scores of the first-stage cluster centers, averaged across years and industries. The z-scores are computed within each country to allow for measurement differences and institutional variation across countries.

[18] Another indication of the cross-country similarity of clusters is that, in each country, all clusters fall into different meta-clusters. This is not mechanically the case as the meta-clustering could have assigned multiple clusters in a country to the same meta-cluster.

[19] We assume log-normal distributions for these draws, as the cluster variables are non-negative.

in the real-world data. The Monte Carlo experiment hence supports our interpretation that the cluster outcomes in the data are indeed very similar across countries.

3 An anatomy of startup types

We now present three stylized facts that follow from our cluster analysis: five distinct startup types emerge across countries, industries, and cohorts (Section 3.1); initial differences between startup types are remarkably persistent over time (Section 3.2); and startups types differ strongly in terms of performance (Section 3.3). In Section 4, we develop a model of post-entry firm dynamics that accounts for the observed life-cycle profiles of these startup types in a coherent way.

3.1 Five startup types

Table 1 summarizes our clustering analysis. We label the five startup clusters that emerge *Basic* (49 percent of all startups); *Large* (4 percent); *Capital-intensive* (7 percent); *Cash-intensive* (26 percent); and *High-leverage* (14 percent). These labels reflect the key dimension along which a startup type differentiates itself. For example, *Large* startups employ, on average, 20 people when they begin operations, compared with an average of only three people for the other types. Likewise, *Cash-intensive* startups hold, on average, 54 percent of their assets as cash when they commence operations, whereas the average is just 12 percent for other startups. Startup types also consistently display their defining characteristic across countries. For example, as can be seen in the first panel of Appendix Figure A7, *Large* startups are consistently larger (in terms of assets) than other types.

Figure 2 (Panel A) shows that while the clustering algorithm yields the same startup types in each country, their local prevalence differs somewhat. For example, *Cash-intensive* startups are relatively common in Italy but less so in France.[20] Figure 2 (Panel B) views the startup composition through a one-digit sectoral lens. It shows that the five startup types also emerge within each main economic sector. The distribution over the five types

[20]This is consistent with evidence that Italian SMEs hold unusually large cash buffers due to financial constraints and fragmented local banking markets (Dottori and Micucci, 2018; Fasano and Deloof, 2021). The concentration of cash-intensive firms in the one country where structural cash holdings are well documented provides a useful consistency check and suggests the clustering captures genuine economic differences.

nevertheless differs somewhat between sectors. For example, *Large* startups are slightly overrepresented among manufacturing firms but underrepresented among information and communications technology (ICT) firms and enterprises that provide professional services. *Cash-intensive* firms are overrepresented in the ICT industry and professional services, while *High-leverage* startups are more common in the hospitality sector. This sectoral distribution also varies across countries (Appendix Figure A8). Figure A9 shows a similar distribution using two-digit sectors. Even at this more granular level, we observe the same five startup types within sectors. There is again some (intuitive) variation. For example, there are fewer *Capital-intensive* startups in "Administration, security and investigation" but more of them in "Manufacturing of basic pharmaceutical products".

To provide further evidence that our startup classification does not simply proxy for particular (types of) sectors, we classify all two-digit sectors into one of nine groups. We follow the Eurostat classification for the technology intensity (manufacturing) or knowledge intensity (services) of industries, as proxied by their R&D expenditure/value added. Appendix Figure A10 (Panel A) shows that the five startup types are consistently present across sectors with different technology and knowledge intensities. In Panel B, we follow Mian and Sufi (2014) and assign two-digit industries to one of four categories: construction, tradeable, non-tradeable, and other. We again find that the five startup types are present across these different economic sectors, with the *Capital-intensive* (*High-leverage*) type somewhat more common among tradeable (non-tradeable) firms.

Lastly, Figure 2 (Panel C) shows that the distribution of startups across the five types (aggregated over countries) is quite stable during the years 2000–2019. There is some decline in the share of *High-leverage* and *Basic* startups, while the share of *Cash-intensive* startups increases during the recovery after the global financial crisis.

An important question is what unites the firms within each cluster at a deeper economic level, given that they span diverse industries. The model calibration in Section 4 provides a structural answer: the five types reflect distinct combinations of production technology (scalability and TFP) and initial equity endowments, which jointly determine the balance-sheet composition at entry. Capital-intensive startups in hospitality and in pharmaceutical manufacturing, for instance, share the feature that their technology is highly scalable in capital but not in labor, generating high capital-to-labor ratios regardless of industry. Likewise, *Large* startups across basic metals and paper products share high overall scalability and high fixed costs that require large equity injections, features that our model interprets as a common technological and financial blueprint. It is this common blueprint, not the industry label, that determines their subsequent life-cycle trajectory and macroeconomic contribution.

That said, our classification is based on five balance-sheet variables at startup, and by construction does not capture all dimensions relevant for policy. In particular, our clustering variables do not directly measure technology intensity, innovation activity, or workforce composition. As a result, within-type heterogeneity along these dimensions may be substantial: a capital-intensive startup operating a highly automated production line differs from one relying on standard equipment, even if both share a similar balance sheet at birth. We view our approach as complementary to analyses that focus on these other dimensions, such as studies of high-growth "gazelles" (Decker, Haltiwanger, Jarmin, and Miranda, 2016) or innovative startups (Guzman and Stern, 2020). Our types capture the balance-sheet and production-factor structure at entry, not growth potential directly. Section 3.3 nonetheless shows that initial scale is a strong predictor of subsequent gazelle status. Richer data on innovation activity or founder characteristics could further refine the classification.

3.2 Early life cycles: Persistent differences

In Figure 3, we explore startups' development during the first years of their existence. We use a regression approach to estimate life cycle profiles in the cluster variables, controlling for various fixed effects. We find that, even as firms age, the differences between types often remain large and statistically significant.[21] First, while *Cash-intensive* startups, by definition, begin operations with substantially more cash (relative to total assets) than *Basic* and other types, they quickly reduce this cash intensity. Yet, even after 12 years, the cash intensity of this type remains about 10 percentage points higher than that of other startups. They also have persistently higher TFP. Second, while *Large* startups employ, on average, 18 more people than *Basic* startups when they commence operations, this difference widens to 24 employees over time. Third, we find clear evidence of convergence in leverage ratios across startup types. *High-leverage* startups are about 40 percentage points more leveraged

[21]We regress each cluster variable on dummies for four startup types (using the *Basic* type as benchmark), country $\times$ cohort fixed effects, and two-digit industry fixed effects. We use the full panel data set at the two-digit industry level. We run a separate regression for each age group (one-year-old firms, two-year-old firms, etc.) and plot the successive coefficients for the startup type dummies. Appendix Figure A11 shows that the life cycle patterns are very similar when aggregating data up to the one-digit industry level.

than *Basic* ones when they begin operations (even within the same country and industry). That excess leverage is reduced quickly so that the difference shrinks to just 5 percentage points after 12 years. Fourth, we also find (partial) convergence in terms of capital intensity. *Capital-intensive* startups begin production with a 47 percentage point higher capital-to-employee ratio. However, they quickly reduce this gap to 18 percentage points. Fifth, *Large* startups are not only bigger in terms of staff numbers but also in terms of total assets. Yet, while *Large* startups gradually expand their staff even further, they slightly shrink their balance sheet (relative to other types in the same country and industry). There is therefore some convergence in the average capital intensity of these large firms over time.

A natural question is whether this convergence reflects genuine economic adjustment or regression to the mean. Several features of our setting make the latter unlikely. First, our clustering is multidimensional: classification depends on a coherent combination of five variables, so measurement error in any single variable cannot generate the type-specific profiles we observe. Second, not all variables converge: employment for *Large* startups *diverges* over time, and TFP differences persist. Pure mean reversion would shrink gaps in all dimensions. Third, the observed patterns align with clear economic mechanisms rationalized by the structural model in Section 4. Finally, startup types differ persistently along variables *not* used in the clustering (productivity, exit rates, wages, and profit margins, see Table 2) which regression to the mean in the clustering variables cannot explain.

3.3 Performance of the five startup types

Table 2 explores the performance differences across startup types based on a set of panel regressions. Observations refer to cell averages for all new firms in a country, one-digit industry, startup year (cohort), age, and type. For example, an observation could refer to the average productivity of Spanish ICT firms of the *High-leverage* type, established in 2005, at age seven. We include dummies to indicate the startup types, again using *Basic* firms as the omitted group. We use high-dimensional fixed effects (FE) to flexibly control for unobservable drivers of firm performance that might correlate with startup type. Country-cohort FE absorb time-invariant characteristics common to startups established in a specific country and year; industry-cohort FE control for time-invariant traits common to startups in a specific industry and established in a given year; and country-industry FE absorb all time-invariant variation that characterizes startups in a specific industry and country. Finally, we add interactions between startups' age and their country, industry, and cohort. This accounts for startup traits specific to the country, sector, or year of establishment *and* that depend on firm age.[22]

The results in Panel A of Table 2 show that *Cash-intensive* and *Large* firms outperform *Basic* ones in terms of higher labor productivity (column 1), higher TFP (column 2), a lower likelihood of early exit (column 3) and higher wages (column 4). The same holds for *Capital-intensive* firms, which display a high labor productivity. *Large* startups stand out for their low exit rates but also their lower profit margins (column 5). *Cash-intensive* firms have particularly high TFP and profit margins.

The regressions also reveal poor performance of *High-leverage* firms, which consistently and strongly underperform in terms of labor productivity and TFP. They also operate with lower profit margins even though they pay lower wages compared to other startup types. In Panel B of Table 2, we replicate this analysis using only firms between five and nine years old. The results line up closely with those in Panel A, indicating that the performance divergence between startup types is not solely driven by transient differences in early life.

Lastly, in unreported cross-sectional linear-probability regressions, we examine which startups achieve "gazelle" status by their fifth year.[23] We classify gazelles as startups that demonstrate above-median average TFP growth during their first five years while simultaneously reaching the top quintile of the overall TFP distribution by age five. Consistent with our earlier discussion on the importance of initial firm size, our findings indicate that only *Large* startups demonstrate an above-average probability of attaining gazelle status.

[22]We provide robust but unclustered standard errors as we use data on the full population of startups across sectors and countries rather than a random sample of these (Abadie, Athey, Imbens, and Wooldridge, 2023). All results are robust to clustering at the sector-country-cohort level to account for possible correlation of model errors over time at that level.

[23]These regressions include country-cohort FE, industry-cohort FE, and country-industry FE and can only be run for Italy and Spain as they require extracting additional information from CompNet.

4 Startup composition: A theoretical framework

4.1 Model environment

In this section, we present a firm dynamics model calibrated to match the empirical patterns documented above. The model serves two purposes. First, by comparing type-specific parameter values, it provides a structural interpretation of what makes the five types fundamentally different (Section 4.3). Second, it enables the policy counterfactuals in Section 5. There, we evaluate both budget-neutral differentiation of corporate tax rates across types and three uniform fiscal instruments (capital, labor, and interest subsidies) whose incidence varies across types and therefore shifts cohort composition with markedly different macroeconomic consequences.

The model period is one year and we denote firm age by a. We distinguish between three age categories: entrants, young firms, and old firms. The age at which firms turn old is denoted a^*. Each entrant belongs to one of a finite number of types. Startups decide on an initial capital investment that cannot be reversed in subsequent years, in the spirit of the literature on lumpy investment (Doms and Dunne, 1998; Cooper and Haltiwanger, 2006). Entrants also face a financial friction, which shapes how they set up the financial structure of the firm. Old firms are not financially constrained.[24] Firms maximize the present value of profits, with a discount factor $\beta \in (0, 1)$.

Technology and real rigidities. Firms use capital, k, and labor, l. Their production function is given by $y = A_a k^\alpha l^\gamma$, where A_a denotes TFP and varies by age and across firm types. The production function elasticities $\alpha > 0$ and $\gamma > 0$ also vary across types, but not with age. Capital depreciates at a rate $\delta \in (0, 1)$ which is common across firms. Production requires a fixed cost $c \geq 0$, which may differ across types. The wage w is common across firms. Entrants decide on an initial level of capital but cannot further adjust the capital stock until they turn old at age a^*, after which they can adjust freely. We abstract from productivity shocks.

[24] This assumption eases computations but is by no means crucial (especially since our focus is on young firms). Also, Figure 4 shows that after a decade, financial leverage is low among all types.

Assets and financial constraints. Firms can invest in short-term liquid assets (cash), denoted m, which earn an interest rate $r = \frac{1}{\beta} - 1$. However, entrants cannot borrow in this type of asset: they face a constraint $m_0 \geq 0$. This constraint does not apply to old firms, which are therefore financially unconstrained. Entrants are born with an initial amount of equity m_{-1}, which may vary across types.

To finance the initial capital investment, entrants can take out a long-term loan $b \geq 0$ with a duration a^*. The loan is to be repaid in equal installments plus a per-period interest rate $r_b(b)$. We assume that $r_b(b)$ is weakly increasing in the loan amount b, that $r_b(b) \geq \widetilde{r}$, and that $r_b(0) = \widetilde{r}$. Here, $\widetilde{r} = \frac{1}{\sum_{a=0}^{a^*-1} \Lambda^{a+1}} - \frac{1}{a^*}$ is the frictionless rate for a loan with this repayment schedule, consistent with the subjective discount factor and the interest rate on cash. In other words, whenever the firm borrows, it pays an external premium that increases in loan size. This setup captures in a simplified way the external finance premium arising in models of financial frictions, such as models of costly state verification (e.g., Townsend (1979)).

Old firms ($a \geq a^*$). We now spell out the old firms' decision problem. Old firms are free to adjust the capital stock as they see fit and they face no financial constraints. For simplicity, we assume that once a firm reaches old age, TFP remains constant and firms exit at an exogenous rate ρ. Because old firms face no constraints, their decision problem is equivalent to one in which they statically rent capital and labor from households. No arbitrage implies that the rental rate of capital is given by $r + \delta$.[25] The profits of an old firm are given by $\pi = Ak^{\alpha}l^{\gamma} - (r + \delta)\, k - wl - c$. Old firms statically choose k and l to maximize profits. The firm value (excluding assets) is given by $V^* = \frac{\pi}{1-\beta(1-\rho)}$. Old firms are indifferent between holding assets (cash and/or capital) or paying dividends.

Young firms ($a < a^*$). Young firms cannot adjust their long-term loan and capital stock, but they must fulfill loan obligations and pay for the upkeep of the capital stock. Young firms do not pay any dividends, i.e. they retain all profits in the form of cash.[26] For simplicity, we assume that firms do not exit during this phase. The young firms' decision problem can

[25] Equivalently, we could assume that the firms own the capital.

[26] If the firm is currently constrained, or may be constrained in the future, it is optimal not to pay dividends. It may be the case that a young firm is fully unconstrained. It is then indifferent between paying dividends and accumulating cash. For those cases, we assume that the firm does not pay dividends.

be expressed as:

$$
\begin{aligned}
V_a(m_{a-1}, k, b) = \max_{l_a} \beta V_{a+1}(m_a, k, b) \\
s.t.\ \ m_a = (1+r)m_{a-1} + \pi_a - \left(\frac{1}{a^*} + r_b(b)\right) b, \\
\pi_a = A_a k^\alpha l_a^\gamma - \delta k - w l_a - c, \\
l_a \geq 0.
\end{aligned}
$$

The value of a firm turning old, including assets, is given by $V_{a^*}(m_{a-1}, k, 0) = V^* + k + (1 + r)m_{a^*-1}$. Note that at that point, the firm has fully repaid its loan ($b = 0$).

Entrants ($a = 0$). Just after entry, a firm chooses the amount of capital, long-term debt, and cash to maximize firm value subject to a budget constraint and financial constraints:

$$
\begin{aligned}
V(m_{-1}) = \max_{k,b,m_0} V_0(m_0, k, b) \\
s.t.\ \ k + m_0 = m_{-1} + b, \\
b, m_0 \geq 0,
\end{aligned}
$$

where the initial amount of cash, m_{-1}, is exogenous.

Startup types and entry decisions. We index firm types by $j = 1, 2, .., J$. Types differ in terms of technology ($\{A_a\}_{a=0}^{a=a^*}, \alpha, \gamma, c$) and in terms of their initial equity (m_{-1}). From now on, we denote by $V_j \equiv V(m_{-1}; \{A_a\}_{a=0}^{a=a^*}, \alpha, \gamma, c)$ the entry value of a firm of type j.

For any type j, there are a certain number of *potential* entrants in any given period. Potential entrants know their type and, conditional on type, the expected firm value V_j. However, they do not know any idiosyncratic realization that may distinguish them from other entrants of the same type. The only source of within-type heterogeneity that matters for the entry decision is the entry cost $\theta_{i,j}$, which varies across potential entrants. Both the number of potential entrants and the distribution of the entry costs may vary across types.

Optimal decision-making implies that a firm of type j starts whenever $V_j \geq \theta_{i,j}$. That is, for each type there exists a cutoff value of $\theta_{i,j}$ such that only potential entrants with an entry cost below the cutoff start a firm.[27] We can now express the actual number of entrants of type j as $n_j = g(V_j, \Gamma_j)$ where g is an increasing function in V_j, and where Γ_j denotes the number of potential entrants of type j and their distribution over entry costs. Thus, the number of entrants is a function of the expected firm value and shocks to the number of potential entrants.

Households and equilibrium. Finally, we add a household sector to the model, allowing us to solve for the general equilibrium. We will focus on steady-state (i.e. long-run) equilibria. The household has a subjective discount factor β, so that in the steady state $r = \frac{1}{\beta} - 1$ is given as above. We assume for tractability that households have Greenwood–Hercowitz–Huffman preferences over leisure and consumption. Optimal labor supply then takes the form $l = \kappa_0 w^{\kappa_1}$, where κ_1 is the labor supply elasticity, which can be calibrated externally, and κ_0 is a scaling parameter. Labor market clearing requires: $l = \sum_j n_j \left(\sum_{a=0}^{a^*-1} l_{a,j} + \frac{l_{*,j}}{1-\rho} \right)$ while the asset and capital market clear trivially.

4.2 Calibration

4.2.1 Post-entry parameters

Recall that startup types differ in terms of technology ($\{A_a\}_{a=0}^{a=a^*}, \alpha, \gamma, c$) and initial equity ($m_{-1}$). Our goal is to interpret the five types by finding a parametrization that fits the empirical profile of each type. We can then compare the parameter values for different types to understand how their fundamentals differ. The quantitative solution method for the model is discussed in Internet Appendix B.

For each type, we feed into the model the empirically observed TFP profiles by age, plotted in the top left panel of Figure 4. We externally set $\Lambda = 0.96$, corresponding to an annual real interest rate of four percent, and $\delta = 0.1$, a depreciation rate of ten percent. We further assume that $r_b(b) = \widetilde{r} + \zeta b$, where we calibrate ζ to target a 10 basis point spread based on average debt across all categories in the data. We set the wage, and scale the levels of TFP, such that the model matches (by construction) the capital-to-labor ratio and

[27] *Within* a startup type, all entrants have the same expectations prior to drawing the demand and production function. This implies there are no entry selection effects within types (only across types). While this modeling assumption is standard in the literature following Hopenhayn (1992), we verify it empirically in Section 4.4.

the employment level of old *Basic* firms.[28] Next, we calibrate for each type the parameters $(m_{-1}, \alpha, \gamma, c)$, targeting the respective empirical profiles, presented in Figure 4. That is, we target the profiles for employment, capital intensity, leverage, the cash ratio, and profits. The latter we include because profits enter directly into the entry condition, which is at the center of the policy counterfactuals.

4.2.2 Entry elasticity

To study the fundamentals driving the post-entry outcomes for each type, we only need to calibrate the parameters described above. However, in order to conduct policy counterfactuals, we also need to set pre-entry parameters that determine how the number and type composition of entrants reacts to a change in firm values. To this end, a first-order approximation (in logs) of this function gives the following expression for the percentage change in the number of entrants of type j:

$$d \ln n_j = \varepsilon_j \cdot d \ln V_j + \gamma_j \tag{1}$$

Here, $\varepsilon_j > 0$, is the elasticity of the number of entrants of type j with respect to the ex ante expected firm value, which is given by the mass of entrants at the entry cutoff, relative to the mass of entrants below the cutoff. Moreover, γ_j denotes shocks to the number and distribution of potential entrants of type j (that is, shocks to Γ_j). Thus, the number of entrants of a certain type may increase either because of an increase in the expected firm value, or because of a shock to the number of potential entrants and/or their distribution over entry costs.[29]

To estimate entry elasticities, we run a regression based on the entry condition, Equation (1). A challenge, however, is that in the data we observe the ex post average realization, denoted $\ln \tilde{V}_j$, while it is the ex-ante expected value $\ln V_j$ that pins down the entry. The difference between the two arises due to common shocks that occur *after* entry. If we were to estimate Equation (1) based on ex post realizations, the residual would contain $\ln \tilde{V}_j - \ln V_j$,

[28]For this initial calibration step, we assume $\alpha = 0.3$ and $\gamma = 0.4$, close to the values found in the next step. Also, the parameter value for the interest rate function is $\zeta = 2.2741e - 06$.

[29]An alternative setup to arrive at the same result is to assume an unlimited number of entrants and to assume that the entry cost is homogeneous across entrants, but an increasing function of the number of entrants (within a type). In that case, the elasticity ε_j is determined by the shape of the entry-cost function. See, for example, Gutiérrez, Jones, and Philippon (2021).

giving rise to correlation between the residual and the right-hand side variable, $\ln V_j$. This would introduce a bias. A second issue is that, under a null hypothesis of free entry (a baseline assumption in the theoretical literature), the entry elasticity is infinite, making the regression specification ill-defined. To circumvent these problems, we rearrange Equation (1) as:

$$d \ln V_j = \frac{1}{\varepsilon_j} d \ln n_j + \xi_j + u_j \tag{2}$$

where $\xi_j \equiv \frac{\gamma_j}{\varepsilon_j}$ is a fixed effect due to entry shocks, for which we control in the regression. Moreover, $u_j \equiv d \ln \tilde{V}_j - d \ln V_j$, which is orthogonal to n_j since ex post shocks are unknown at entry, and thus cannot affect the number of entrants, n_j.

We use observations at the country(c)-industry(i)-type(j)-cohort(t) level, with corresponding indices between brackets. We control for shocks to the number and distribution of entrants at the country-industry-type level and at the country-industry-year level by using interactive fixed effects at these levels. The former is important as the number of potential entrants may naturally vary by country, industry, and type. The latter captures that, in a specific industry and country, the number of potential entrants may fluctuate over time. We run these regressions both at the one-digit and at the two-digit industry level.[30] This results in the following regression specification, estimated via ordinary least squares (OLS):

$$\ln V_{c,i,j,t} = \beta_{0,j} + \beta_{1,j} \ln n_{c,i,j,t} + \xi_{c,i,j} + \xi_{c,i,t} + u_{c,i,j,t} \tag{3}$$

We note that $\frac{1}{\beta_{1,j}}$ is the type-specific entry elasticity and that, under free entry (that is, an infinite elasticity), it holds that $\beta_{1,j} = 0$. We estimate this equation separately for each startup type j.

A final question is how to measure $\ln V_j$ in the data. We do so using each startup type's life-cycle profile for profits and exit rates, and the number of entrants for each country-industry-type-cohort in our data set.[31] We assume a discount factor $\Lambda = 0.96$, corresponding to an annual discount rate of about 4 percent. We only use cohorts that we observe for at least seven years (that is, those established before 2011). For age twelve and onward, we

[30]Implicit in our specification is the assumption that, within a country-industry cell, shocks over time to the number of *potential* startups are common across types.

[31]We reconstruct profits by multiplying profits (Earnings Before Interest and Tax, EBIT) per unit of revenue with revenues, at the country-industry-startup type-cohort level.

assume that profits and the year-on-year exit rate remain fixed.[32]

Figure 5 plots the estimates for the inverse entry elasticity, β_1. Without conditioning on startup type, the overall estimate, based on the data set at the one-digit sector level, is 0.51, implying an entry elasticity of around 2. This estimate is significantly different from zero at the 1 percent level, thus rejecting the free-entry assumption. When we use our data at the two-digit sector level, the estimate is 0.85, implying an entry elasticity of about 1.17. These estimates are in the ballpark of those from estimated structural DSGE models.[33]

The figure shows the estimates of the inverse elasticity by firm type, illustrating substantial heterogeneity. Regardless of whether we use the data at the one-digit industry level (red dots) or at the two-digit level (blue), *Basic* startups show the highest entry elasticity (about 4 at the one-digit level and 1.7 at the two-digit level). *Large* startups instead have an entry elasticity of between 1.3 (one-digit level) and 0.67 (two-digit level). As we discuss in Section 5.2, this cross-type heterogeneity in entry elasticities is an important reason why even an undifferentiated, uniform policy change has a different impact on different startup types.

Basic startups react quite flexibly to changing economic conditions as their creation is mostly uncomplicated. Other types may require more specific expertise, which fewer potential entrepreneurs possess. For example, founding a *Large* startup likely requires a more complex business plan than starting a smaller and simpler firm.

4.3 Interpreting the five startup types

We now use the calibrated model to interpret the different startup types.[34] We present the model fit in Figure 4 and parameter values in Table 3. In the table, we divide initial assets and the fixed cost by the wage per worker for ease of interpretation. Overall, the model accounts reasonably well for the empirical profiles, considering that for each startup type we

[32] That is, we compute firm value at age zero as: $V_0 = \sum_{a=0}^{11} \Lambda^a s_{0,a}\pi_a + s_{0,12}\pi_{12} \sum_{a=12}^{\infty} \Lambda^a s_{11,12}^a = \sum_{a=0}^{11} \Lambda^a s_{0,a}\pi_a + \Lambda^{12} s_{0,12} \frac{\pi_{12}}{1-\Lambda s_{11,12}}$ where $s_{k,l}$ is the survival rate between age k and $l \geq k$. We verified that changing the age cutoff at which profits are held constant hardly affects the results. This reflects that the component of the value beyond age twelve is relatively unimportant for the startup value, due to discounting and high exit rates in young age groups.

[33] Gutiérrez, Jones, and Philippon (2021) and Sedláček and Sterk (2017) find estimates of about 1.5 and 5.5, respectively.

[34] For this exercise, the estimated entry elasticities are irrelevant.

target 50 moments with just four parameters.[35] As for the parameter values, we observe large differences in startup types both in terms of their physical structure (production function, fixed cost, TFP) and their startup equity. This observation, combined with the satisfactory model fit, suggests that heterogeneity in technologies and initial equity are major drivers of differences between startup types. We now consider in more detail how the fundamentals of the five types differ.

Basic startups. *Basic* firms have moderate TFP and modest returns to scale ($\alpha+\gamma = 0.7$). As they age, their TFP and employment levels increase, and their capital intensity (capital-to-labor ratio) declines. *Basic* types in the model have no leverage and modest cash holdings, in line with the data. The presence of many *Basic* startups, with an undifferentiated strategy and mediocre performance, fits with recent work on the productivity gap between, on the one hand, a small group of well-performing firms that operate at the technological frontier and, on the other hand, a long tail of less-productive laggards (Hsieh and Klenow, 2009; Andrews, Criscuolo, and Gal, 2016). Our clustering analysis indicates that such an uneven productivity distribution already emerges when new firm cohorts are born.

Large startups. *Large* firms have the highest overall scalability of all types ($\alpha + \gamma = 0.75$). Strictly speaking, we model scalability as a technological feature, but it could also be interpreted as scalability of customer demand (Hottman, Redding, and Weinstein, 2016). The TFP levels of *Large* startups are also relatively high, suggesting they combine physical and human capital in a relatively productive way. These fundamentals make entrants of this type poised to grow even bigger. Indeed, they have more than five times as many employees as other startups, and their *level* of capital is highest among all types (their capital intensity is the second highest, after the *Capital-intensive* startups). In the data, we observe that *Large* firms are more prevalent in sectors like the manufacturing of basic materials (metals, paper). *Large* startups also remain persistently larger than the other types, consistent with empirical findings in Sterk ⓡ Sedláček ⓡ Pugsley (2021), who use US data to document the importance of ex ante heterogeneity for firm-level employment.

[35] One dimension along which the model fits the data less well is that it predicts an increasing cash ratio profile for *Capital-intensive* and *High-leverage* firms, whereas in the data this profile is flat. Note further that the TFP profiles are the same in the data and the model, by construction.

The scalable technology of *Large* firms comes with fixed costs almost four times as high as those of *Basic* startups. The high initial fixed investment and high fixed cost imply a strong financing need. Indeed, according to the model, the initial equity of *Large* entrants is about four times as high as that of *Basic* firms. This substantial initial equity allows *Large* firms to operate without high leverage and with some cash holdings, in line with the data.

Capital-intensive startups. Compared to *Basic* startups, the scalability of *Capital-intensive* firms is much higher for capital (α) but lower for labor (γ). As a result, capital intensity (the capital-to-labor ratio) is much higher. As they become more productive and hire more workers over time, the capital intensity of these firms remains elevated compared to other types.[36]

The fixed cost of *Capital-intensive* firms is more than twice that of *Basic* firms, consistent with their prevalence in manufacturing, transport, hospitality, and rental & leasing. These high capital requirements are reflected in initial equity roughly twice that of *Basic* firms. This equity finances the initial capital stock, leaving *Capital-intensive* firms with relatively little cash, especially early in life (again in line with the data).

Cash-intensive startups. These startups are less scalable, especially with respect to capital ($\alpha = 0.24$). On the other hand, they have the highest TFP of all types, while their fixed cost is moderate. These may be businesses producing "niche" goods or services, being small but profitable. While *Cash-intensive* firms are present across all sectors, they are more common in ICT, professional services, and construction.

According to the model, the initial equity of these firms is similar to that of *Basic* ones but larger than that of *High-leverage* startups. Their low scalability with respect to capital means that *Cash-intensive* startups initially invest relatively little. This allows them to retain a significant amount of initial equity as cash on their balance sheet and to avoid high leverage. Indeed, these firms initially hold almost half of their assets as cash, about five times as much as other startup types.

[36] In the data, the capital intensity of these startups is particularly high in the first year, which the model does not fully capture. This may reflect adjustment costs in hiring workers, which we do not include in the model.

High-leverage startups. In terms of scalability, high-leverage startups are similar to *Basic* ones. However, they have significantly lower TFP, especially when young. In the data, high-leverage firms are prevalent in hospitality, retail trade, and professional services.

High-leverage firms begin operations by taking on substantial debt: their average leverage ratio is 116 percent. According to the model, high-leverage firms have very little initial equity, so they need to borrow substantially to finance their first investments. As they age, they pay off initial debt and their leverage ratio almost converges to that of other types after a decade. This deleveraging aligns with empirical evidence in Dinlersoz, Kalemli-Özcan, Hyatt, and Penciakova (2019) and Kochen (2025), while the gradual nature of this convergence is in line with Lemmon, Roberts, and Zender (2008), who document that leverage heterogeneity is quite persistent.

4.4 Within-type selection at entry

A key assumption of the model is that entry selection operates across types, not within them (Section 4.1). We now test this assumption empirically. We run age-partitioned regressions in which observations are at the country–two-digit sector–type–cohort level (for example, Spanish beverage manufacturers of the *Cash-intensive* type established in 2017). For each age, we regress a performance metric (labor productivity, TFP, an exit dummy, or employment) or a cluster variable (size, capital intensity, cash intensity, or leverage) on the log number of firms in the cohort, to estimate the relationship between cohort size and later startup performance. We saturate the specification with country-sector-type, country-sector-cohort, country-type-cohort, and sector-type-cohort fixed effects, and cluster standard errors at the country-sector-cohort level. If positive within-type selection were material, larger cohorts would systematically underperform later in life. Figures A12–A13 show that the estimated elasticities are economically small, statistically insignificant at most horizons, and show no systematic trend, at both the two- and one-digit sector definitions. Within-type entry selection is thus empirically negligible, validating the modeling assumption in Section 4.[37]

[37] Hombert, Schoar, Sraer, and Thesmar (2020) show that a 2003 French reform providing downside insurance to unemployed entrepreneurs increased firm creation without reducing average firm quality, consistent with limited selection effects at entry.

5 Counterfactual policy experiments

We now perform counterfactual policy experiments to study how changes in taxation can influence specific macroeconomic outcomes by shifting the composition of new startup cohorts. We do this using the model framework developed in the previous section.

5.1 A budget-neutral tax policy differentiated by startup type

We start by introducing a corporate income tax. The after-tax profits of a firm are now given by $\pi_{a,j} = (1-\tau_{\pi,j})\left(A_a k_{a-1}^{\alpha} l_a^{\gamma} - k_a + (1-\delta)k_{a-1} - wl_a - c\right)$, where $\tau_{\pi,j}$ is the corporate tax rate which may vary across types. Our goal is to examine the aggregate effects of a differentiated corporate income tax rate. To impose discipline, we restrict ourselves to policies that are revenue-neutral.[38] This implies that, if some startup types are to be taxed less, other types need to be taxed more. We further impose equilibrium, i.e., the labor market clearing condition holds. Lastly, we limit ourselves to policies of no more than a 10 percentage points change in the corporate tax rate for each type. Aside from these restrictions, we search the entire space of policy changes differentiated by type, and evaluate the macroeconomic implications for any admissible policy.

Our policy experiments are based on a numerical Monte Carlo procedure. We consider a large number of uniform random draws for the tax policies. For any draw, denoted $\tilde{\tau}_{\pi,j}$, we set the tax rates as $\tilde{\tau}_{\pi,j} + \bar{\tau}$, where $\bar{\tau}$ is a uniform adjustment factor used to ensure budget neutrality. We then jointly solve for firm decisions, $\bar{\tau}$ and the wage w such that the policy is budget neutral and the labor market clears.

Based on the estimated entry elasticities, we evaluate the change in the number of firms within each type. This allows us to construct counterfactuals for the number of startups by age and type, i.e. $n_{a,j}$. Similarly, we solve for the counterfactual choices in each of these bins. We then aggregate to compute counterfactual aggregate outcomes.

After evaluating the macro outcomes for the entire space of admissible policies, we group policies into bins according to their intensity, as measured by the maximum absolute change in the tax rate across types. The idea is that larger policy changes may bring about greater macroeconomic effects, but may also face stronger political resistance. It is therefore useful

[38] I.e., we impose that $\sum_a \sum_j n_{a,j} \frac{\tau_{\pi,j}}{1-\tau_{\pi,j}} \pi_{a,j} = 0$, where $n_{a,j}$ is the number of firms of age a and type j.

to study the policy space conditional on a certain intensity of the policy. For each bin, we compute policies that bring about the largest positive and the largest negative change in a macro variable (aggregate productivity or employment). We label this the policy frontier and present the tax-rate changes that generate the policy upper bound. We perform this analysis twice: once using entry elasticities derived from one-digit sector data (Figure 6) and once using elasticities based on two-digit sector data (Figure A14). The latter approach is more precise as it uses more granular variation in entry elasticities, but comes at the cost of using less information (as the two-digit data contain more empty cells) and hence being less representative of the overall startup population.

Are large aggregate gains possible? Figure 6 summarizes our policy experiment. Panel A presents the policy space for aggregate labor productivity (using entry elasticities based on a one-digit sector classification).[39] The horizontal axis measures the intensity of the potential policy change (that is, the maximum change in the tax rate across all firms). Moving from left to right, warmer colors indicate stronger corporate tax-rate differentiation. The solid (dashed) line plots the policy upper bound (lower bound). This is the largest possible aggregate labor productivity increase (decrease) given a certain policy intensity.

The figure shows that substantial macroeconomic gains are possible. For the maximum policy intensity of 0.1 (that is, when the tax rate does not change by more than 10 percentage points for any firm) labor productivity can be increased by about 4.5 percent. At the same time, substantial macroeconomic losses are also possible for policies that shift the composition away from high-productivity startup types; the maximum productivity decline is more than four percent.

Panel B of Figure 6 again plots the policy space, this time for aggregate employment. Small employment gains are possible: 0.2 percent for a policy intensity of 0.1. At the same time, much larger employment losses—of around 1.6 percent—are possible. This asymmetry reflects that the high-productivity types the optimal policy encourages are relatively capital-intensive, so they boost output per worker more than headcount, while taxing *Basic* startups (the type with the highest entry elasticity and the largest population share) can sharply reduce the total number of firms.

[39]We focus on labor productivity rather than TFP, since the former also captures the macro effects of differences in capital investment between different types. That said, the exercise can also be performed for TFP (or any observable). Figure 8 considers TFP alongside several other outcomes.

Which startup types to encourage? The two panels on the left of Figure 7 plot the corporate tax policies associated with the policy upper bounds. Likewise, the two panels on the right show the associated firm shares. Which startup types should policy encourage to increase productivity or employment? The upper left panel shows that, to achieve the productivity upper bound, taxes should be lowered for *Capital-intensive*, *Cash-intensive*, *High-leverage* and *Large* types, financed by a tax increase for *Basic* types. The upper right panel shows that, as a result, the share of *Capital-intensive* startups increases, as do the shares of *Cash-intensive*, *High-leverage* and *Large* startups. Accordingly, the share of *Basic* startups falls. Since labor productivity is relatively low among *Basic* firms, the shift in startup composition increases aggregate labor productivity.

The lower left panel of Figure 7 shows the tax rates associated with the employment frontier. To boost aggregate employment, taxes should be lowered for the *Basic* type, and increased for all other types. This leads to an increased share of these types (lower right panel). Intuitively, the entry elasticity among *Basic* startups is relatively high. A tax cut for this type therefore has a relatively large impact on the overall number of entrants, and therefore employment.

Policy trade-offs. The previous analysis suggests a trade-off between maximizing productivity and employment. To gauge this trade-off more directly, Figure 8 provides scatter plots that depict pairs of changes in aggregate or average macroeconomic outcomes resulting from each potential policy.[40] The top panel shows that there is generally some trade-off between increasing aggregate employment and labor productivity: the two outcomes are negatively correlated across possible policy changes. Yet, there exist specific reforms that simultaneously boost (or reduce) aggregate labor productivity and employment, e.g. by increasing the shares of *Large* and *Cash-intensive* startups as much as possible (given a certain policy intensity). Interestingly, there is also a fairly strong negative trade-off between aggregate employment and TFP (middle panel). The trade-off between aggregate labor productivity and TFP appears more limited (bottom panel).

[40]We again analyze policies with a maximum intensity of 0.1 while using entry elasticities based on the one-digit sector classification.

5.2 Subsidies on labor, capital and interest

So far, we have only considered corporate income taxes, differentiated by startup type. We now consider three uniform fiscal instruments: a flow subsidy on capital per unit of installed capital, a labor subsidy that reduces the effective wage paid by firms, and an interest subsidy that lowers the borrowing rate for young firms (those below age ten). The first two apply to all firms; the third is restricted by age. While uniform across types, each instrument's incidence varies across types and therefore affects startup composition and macroeconomic outcomes differently. In Internet Appendix B, we introduce these instruments into the model.

5.2.1 Capital subsidy

First, we consider a flow subsidy on capital, denoted $\tau_k < 0$, per unit of installed capital. To finance the subsidy, we assume that corporate income taxes are adjusted uniformly, so that the overall intervention is budget neutral. The policy is scaled such that financing it requires a five-percentage-point increase in the corporate income tax rate, yielding a capital subsidy of 0.29 percent. As before, we also impose that wages adjust to clear the labor market.

Table 4 presents the effects of the capital subsidy of 0.29 percent. Panel B shows the effects on startup composition. As expected, the subsidy increases the share of capital-intensive startups, as this type benefits the most from the subsidy. The subsidy also stimulates capital accumulation. Panel C shows the aggregate effects of the policy change. The total number of firms declines, which is due to the increase in the corporate income tax required to finance the subsidy, driven to a large extent by *Basic* firms, which have a relatively high entry elasticity. Despite this decline, the effect on output remains positive, as does the effect on labor productivity.

Panel C of Table 4 also reports the macroeconomic effects that result purely from changes in the number and composition of startups. Both effects are quantitatively significant. Moreover, whereas the change in the number of firms reduces aggregate output and employment, the composition effect boosts these macro outcomes.

5.2.2 Labor subsidy

Next, we consider a subsidy on labor, denoted $\tau_l < 0$. The effective wage paid by the firm is $w(1-\tau_l)$, while households continue to receive w. As before, we impose labor market clearing and budget neutrality, and scale the subsidy such that it requires a five–percentage-point increase in the corporate income tax rate.

Table 4 shows that this subsidy policy has sizable negative effects on both output and labor productivity. These effects are driven by a decline in the number of firms and by negative composition effects, both of which are quantitatively large. In particular, the policy significantly reduces the share of *Capital-intensive* startups.

Although the exercise considers a labor *subsidy*, the results suggest that labor *taxes* may be less detrimental to the macroeconomy than often assumed. In particular, our counterfactual results indicate that if higher labor taxation is used to lower corporate income taxes, it may generate positive effects through changes in the number and composition of startups.

5.2.3 Interest subsidy

Finally, we consider an interest subsidy, denoted $\tau_b < 0$, for young firms (firms up to age ten). We consider a subsidy of 0.29 percent, the same magnitude as in the capital subsidy experiment. The policy requires a small increase in the corporate income tax rate. As shown in Panel C of Table 4, the macroeconomic effects of the policy are small but positive. The number of startups increases, which raises output. However, this increase is partly offset by a negative composition effect, as the policy increases the share of high-leverage startups, which exhibit relatively poor performance in both the data and the model.

5.3 Tax differentiation for startups in practice

Our counterfactual exercises in Section 5.1 show how tax differentiation across startup types can in principle yield relatively large macroeconomic gains. While such differentiation should in principle be possible, as it distinguishes startup types on the basis of a few observable traits, the practical implementation or political feasibility may be challenging.[41] Yet, it may not be necessary to differentiate policies explicitly by startup type. As the subsidy experiments in Section 5.2 demonstrate, unconditional fiscal instruments shift startup composition through their heterogeneous incidence across types. For example, the capital subsidy disproportionately benefits *Capital-intensive* startups, boosting their cohort share and raising

[41] For example, startups may alter their initial choices to mimic types that are more favorably taxed. In practice, this can be costly as the observed choice differences across startups are large.

aggregate productivity — even though the policy does not distinguish between types. More generally, any unconditional policy proposal can be evaluated through our framework for its composition effects.

Several countries have implemented policies that, either implicitly or explicitly, distinguish between startup types (or between startups and more mature firms). Belgium, France, Hungary, Latvia, Lithuania, Luxembourg, the Netherlands, Poland, Portugal, Slovakia, Spain, and the US all apply reduced corporate tax rates to firms below a certain size threshold (Bergner, Brautigam, Evers, and Spengel, 2017). A few countries differentiate tax rates among startups specifically. For example, the rate that applies to Indian manufacturing startups depends on their annual turnover (Kalra, 2019). Other countries engage in more implicit tax differentiation among young and small businesses. Measures such as tax credits, breaks, and exemptions can lower the effective tax rate for such businesses substantially.

Our results indicate that tax policies that penalize large firms can adversely affect the startup composition and reduce macroeconomic performance. Indeed, we show that *Large* startups not only create many new jobs but also pass their relatively high productivity levels on to workers and consumers. Similarly, our results suggest that size-contingent regulations—such as less onerous entry regulation or labor legislation for small firms—may adversely affect the composition of new firm cohorts, as they effectively discourage the entry of larger startups.

In some countries, tax benefits only apply to startups with a particular funding structure. These include special credit guarantees and loans for startups. Other measures help startups to attract equity funding by offering tax relief to investors who buy new shares in startups. Our results suggest that schemes to help startups access debt or equity can have strong impacts, through the startup composition channel, on the performance of new generations of firms. In particular, we find that *High-leverage* startups tend to perform very poorly compared to other types. Policies that facilitate startups' access to credit may therefore have very different impacts on firm performance than measures to help startups issue equity.[42]

This discussion illustrates how many countries already operate corporate tax or other policies that in some way differentiate, sometimes implicitly, between startup types. The

[42]This concern complements evidence that credit misallocation toward weak European firms has adverse macroeconomic consequences, including for productivity and inflation dynamics (Acharya, Crosignani, Eisert, and Eufinger, 2024).

resulting variation in effective tax rates is often substantial and reflective of policy goals such as job creation or productivity growth. Many of these real-world examples of corporate tax differentiation can be easily translated to, or nested in, our more general policy exercise.

6 Conclusions

This paper combines novel and comprehensive data on startups in multiple European countries with a firm dynamics model to quantify the macroeconomic gains from policies that shift the composition of new startup cohorts. Using unsupervised machine learning, we find that similar clusters of startups (identified by the choices they make at entry) emerge in each of these countries. Moreover, the subsequent performance of these types—in terms of employment, productivity, and survival—differs markedly. There are therefore potential macroeconomic gains (or losses) to be had from policies that alter the composition of new startup cohorts.

Two sets of policy experiments based on the calibrated model quantify these impacts. First, a budget-neutral differentiation of corporate income tax rates (bounded at ten percentage points per type) can raise aggregate labor productivity by up to 4.5 percent. The *Basic* type has by far the highest entry elasticity, so taxing it more heavily produces the largest compositional shift toward the higher-productivity types. The same logic implies that uniform interventions are dampened by selection across types, making appropriately targeted reforms substantially more cost-effective. Targeting the highest-productivity types (in particular *Large* and *Cash-intensive*) is also attractive on broader welfare grounds, since these firms tend to pass productivity gains on to workers through higher wages.

Second, even uniform fiscal instruments differ sharply in their macroeconomic incidence once compositional channels are accounted for. A capital subsidy raises output and labor productivity through favorable composition effects, while an equivalently sized labor subsidy reduces both. An interest subsidy aimed at young firms is roughly neutral in the aggregate but tilts composition toward *High-leverage* startups, which perform poorly in both data and model. The choice of instrument therefore matters as much as the scale of intervention, a finding that the standard representative-firm framework cannot recover.

These results speak directly to current policy debates on how to revive business dynamism. In the United States, concerns about secular declines in startup entry and the disappearance of high-growth young firms have prompted a combination of industrial policy and deregulatory responses. Laws such as the CHIPS and Science Act and the Inflation Reduction Act provide subsidies and tax credits that reduce fixed costs and ease financing constraints for entrants in capital-intensive sectors. In parallel, deregulatory proposals seek to lower administrative burdens that may disproportionately deter new firm formation. In Europe, similar ambitions are reflected in the EU Inc. initiative announced by the European Commission, which aims to harmonize the legal and fiscal environment for startups across member states. Our framework provides a tool to assess the macroeconomic effects of such reforms ex ante, with particular emphasis on their impact on the composition of entering firms. More broadly, our approach connects to the renewed interest in industrial policy (Juhász, Lane, and Rodrik, 2024): by explicitly modeling how policies shape the composition of startup cohorts, it enables a quantitative evaluation of both the benefits and potential distortions associated with targeted interventions.

Increasingly, statistical agencies make rich micro data on firms publicly available, so our methodology can readily be applied to a wider set of countries. A natural question is to what extent differences in startup composition account for cross-country differences in macroeconomic performance. On the methodological side, it would be interesting to explore whether more granular startup classifications can further sharpen the policy implications. A related extension is to design optimal combinations of fiscal instruments (rather than evaluating them one at a time) while accounting for how the optimal mix depends on the composition of the existing startup population. We leave these issues for future research.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2023. "When Should You Adjust Standard Errors for Clustering?." *Quarterly Journal of Economics* 138 (1): 1–35.

Acharya, Viral V., Matteo Crosignani, Tim Eisert, and Christian Eufinger. 2024. "Zombie Credit and (Dis-)Inflation: Evidence from Europe." *Journal of Finance* 79 (3): 1883–1929.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–2451.

Adelino, Manuel, Song Ma, and David Robinson. 2017. "Firm Age, Investment Opportunities, and Job Creation." *Journal of Finance* 72 (3): 999–1038.

Akcigit, Ufuk, and Sina T. Ates. 2021. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." *American Economic Journal: Macroeconomics* 13 (1): 257–98.

Akcigit, Ufuk, and Sina T. Ates. 2023. "What Happened to US Business Dynamism?." *Journal of Political Economy* 131 (8): 2059–2124.

Akcigit, Ufuk, Douglas Hanley, and Stefanie Stantcheva. 2022. "Optimal Taxation and R&D Policies." *Econometrica* 90 (2): 645–684.

Albert, Christoph, and Andrea Caggese. 2021. "Cyclical Fluctuations, Financial Shocks, and the Entry of Fast-growing Entrepreneurial Startups." *Review of Financial Studies* 34 (5): 2508–2548.

Albuquerque, Rui, and Hugo A. Hopenhayn. 2004. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies* 71 (2): 285–315.

Alon, Titan, David Berger, Robert Dent, and Benjamin Pugsley. 2018. "Older and Slower: The Startup Deficit's Lasting Effects on Aggregate Productivity Growth." *Journal of Monetary Economics* 93: 68–85.

Andrews, Dan, Chiara Criscuolo, and Peter Gal. 2016. "The Best versus the Rest. The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy." Productivity Working Papers No. 5, Organisation for Economic Co-operation and Development, Paris.

Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda. 2020. "Age and High-Growth Entrepreneurship." *American Economic Review: Insights* 2 (1): 65–82.

Bajgar, Matej, Giuseppe Berlingieri, Sara Calligaris, Chiara Criscuolo, and Jonathan Timmis. 2020. "Coverage and Representativeness of Orbis Data." OECD Science, Technology and Industry Working Papers No. 6, Organisation for Economic Co-operation and Development, Paris.

Bartelsman, Eric, John C. Haltiwanger, and Stefano Scarpetta. 2004. "Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries." *World Bank Policy Research Working Paper No. 3464*. World Bank, Washington, D.C.

Belenzon, Sharon, Aaron Chatterji, and Brendan Daley. 2017. "Eponymous Entrepreneurs."

American Economic Review 107 (6): 1638–1655.

Bergner, Soren M., Rainer Brautigam, Maria T. Evers, and Christoph Spengel. 2017. "The Use of SME Tax Incentives in the European Union." *ZEW Discussion Paper 17-006.* ZEW—Leibniz Centre for European Economic Research, Mannheim.

Bernard, Andrew, Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes. 2022. "The Origins of Firm Heterogeneity: A Production Network Approach." *Journal of Political Economy* 130 (7): 1765—1804.

Biondi, Filippo, Sergio Inferrera, Matthias Mertens, and Javier Miranda. 2023. "Declining Business Dynamism in Europe: The Role of Shocks, Market Power, and Technology." *IWH Discussion Paper No. 19-2023.* Halle Institute for Economic Research, Halle (Saale).

Bloom, Nick, Rachel Griffith, and John Van Reenen. 2002. "Do R&D Tax Credits Work? Evidence from a Panel of Countries 1979–1997." *Journal of Public Economics* 85 (1): 1–31.

Bolton, Patrick, Neng Wang, and Jinqiang Yang. 2019. "Investment under Uncertainty with Financial Constraints." *Journal of Economic Theory* 184: 104912.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2022. "Discretizing Unobserved Heterogeneity." *Econometrica* 90 (2): 625–643.

Brandt, Loren, Gueorgui Kambourov, and Kjetil Storesletten. 2026. "Barriers to Entry and Regional Economic Growth in China." *The Review of Economic Studies* 93 (1): 286–326.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review* 101 (5): 1964–2002.

Buera, Francisco J, and Yongseok Shin. 2017. "Productivity Growth and Capital Flows: The Dynamics of Reforms." *American Economic Journal: Macroeconomics* 9 (3): 147–85.

Cooper, Russell W., and John C. Haltiwanger. 2006. "On the Nature of Capital Adjustment Costs." *Review of Economic Studies* 73 (3): 611–633.

De Loecker, Jan, and Chad Syverson. 2022. "An Industrial Organization Perspective on Productivity." *Handbook of Industrial Organization (Vol. IV).*

Decker, Ryan A., John C. Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2017. "Declining Dynamism, Allocative Efficiency, and the Productivity Slowdown." *American Economic Review* 107 (5): 322–26.

Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2016. "Where

Has All the Skewness Gone? The Decline in High-Growth (Young) Firms in the U.S..” *European Economic Review* 86 (C): 4–23.

Dent, Robert C, Fatih Karahan, Benjamin Pugsley, and Ayşegül Şahin. 2016. “The Role of Startups in Structural Transformation.” *American Economic Review Papers and Proceedings* 106 (5): 219–23.

Derrien, François, Jean-Stéphane Mésonnier, and Guillaume Vuillemey. 2025. “Setup Costs and the Financing of Young Firms.” *Management Science*.

Dinlersoz, Emin, Sebnem Kalemli-Özcan, Henry Hyatt, and Veronika Penciakova. 2019. “Leverage over the Firm Life Cycle, Firm Growth, and Aggregate Fluctuations.” Working Paper 2019-18, Federal Reserve Bank of Atlanta.

Doms, Mark, and Timothy Dunne. 1998. “Capital Adjustment Patterns in Manufacturing Plants.” *Review of Economic Dynamics* 1 (2): 409–429.

Dottori, Davide, and Giacinto Micucci. 2018. “Corporate Liquidity in Italy and Its Increase in the Long Recession.” *Economia Politica* 35 (3): 981–1014.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. “How Costly are Markups?.” *Journal of Political Economy* 131 (7): 000–000.

Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl. 2011. *Cluster Analysis.*: Wiley.

Fasano, Francesco, and Marc Deloof. 2021. “Local Financial Development and Cash Holdings in Italian SMEs.” *International Small Business Journal* 39 (8): 781–799.

Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. “Aggregate Productivity Growth: Lessons from Microeconomic Evidence.” In *New Developments in Productivity Analysis*, edited by Edward Dean, Michael Harper and Charles Hulten, 303–372: University of Chicago Press.

Garicano, Luis, Claire Lelarge, and John Van Reenen. 2016. “Firm Size Distortions and the Productivity Distribution: Evidence from France.” *American Economic Review* 106 (11): 3439–79.

Gopinath, Gita, Şebnem Kalemli-Özcan, Loukas Karabarbounis, and Carolina Villegas-Sanchez. 2017. “Capital Allocation and Productivity in South Europe.” *Quarterly Journal of Economics* 132 (4): 1915–1967.

Gutiérrez, Germán, Callum Jones, and Thomas Philippon. 2021. “Entry Costs and Aggregate

Dynamics." *Journal of Monetary Economics* 124: S77–S91.

Guzman, Jorge, and Aishen Li. 2023. "Measuring Founding Strategy." *Management Science* 69 (1): 101–118.

Guzman, Jorge, and Scott Stern. 2015. "Where is Silicon Valley?." *Science* 347 (6222): 606–609.

Guzman, Jorge, and Scott Stern. 2020. "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." *American Economic Journal: Economic Policy* 12 (4): 212–243.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda. 2013. "Who Creates Jobs? Small versus Large versus Young." *Review of Economics and Statistics* 95 (2): 347–361.

Hombert, Johan, Antoinette Schoar, David Sraer, and David Thesmar. 2020. "Can Unemployment Insurance Spur Entrepreneurial Activity? Evidence from France." *Journal of Finance* 75 (3): 1247–1285.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (5): 1127–1150.

Hopenhayn, Hugo A., and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101 (5): 915–938.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *Quarterly Journal of Economics* 131 (3): 1291–1364.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124 (4): 1403–1448.

Hsieh, Chiang-Tai, and Peter J. Klenow. 2014. "The Life Cycle of Plants in India and Mexico." *Quarterly Journal of Economics* 129: 1035–1084.

Hurst, Erik, and Benjamin W. Pugsley. 2011. "What Do Small Businesses Do?." *Brookings Papers on Economic Activity* 43 (2): 73–142.

Juhász, Réka, Nathan J Lane, and Dani Rodrik. 2024. "The New Economics of Industrial Policy." *Annual Review of Economics* 16: 213–242.

Kalemli-Özcan, Şebnem, Bent E Sørensen, Carolina Villegas-Sanchez, Vadym Volosovych, and Sevcan Yeşiltaş. 2024. "How to Construct Nationally Representative Firm-Level Data from the Orbis Global Database: New Facts on SMEs and Aggregate Implications for Industry Concentration." *American Economic Journal: Macroeconomics* 16 (2): 353–374.

Kalra, Aarti. 2019. "Understanding the Corporate Tax Structure for Domestic Startup Companies." available at https://www.startupindia.gov.in/ 1764, Government of India.

Karahan, Fatih, Benjamin Pugsley, and Ayşegül Şahin. 2024. "Demographic Origins of the Startup Deficit." *American Economic Review* 114 (7): 1986–2023.

Kerr, William R., and Ramana Nanda. 2010. "Banking Deregulations, Financing Constraints, and Firm Entry Size." *Journal of the European Economic Association* 8 (2-3): 582–593.

Kochen, Federico. 2025. "Finance over the Life Cycle of Firms." *CEMFI Working Paper No. 2518*. CEMFI, Madrid.

Lemmon, Michael L., Michael R. Roberts, and Jaime F. Zender. 2008. "Back to the Beginning: Persistence and the Cross-section of Corporate Capital Structure." *Journal of Finance* 63 (4): 1575–1608.

Levine, Ross, and Yona Rubinstein. 2017. "Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More?." *Quarterly Journal of Economics* 132 (2): 963–1018.

Lopez-Garcia, Paloma, and Filippo di Mauro. 2015. "Assessing European Competitiveness: The New CompNet Microbased Database." Working Paper 1764, European Central Bank.

Mian, Atif, and Amir Sufi. 2014. "What Explains the 2007–2009 Drop in Employment?." *Econometrica* 82 (6): 2197–2223.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review* 104 (2): 422–58.

Moll, Benjamin. 2014. "Productivity Losses from Financial Frictions: Can Self-financing undo Capital Misallocation?." *American Economic Review* 104 (10): 3186–3221.

Oberfield, Ezra, and Devesh Raval. 2021. "Micro Data and Macro Technology." *Econometrica* 89 (2): 703–732.

OECD. 2015. "The Future of Productivity." *Organisation for Economic Cooperation and Development*. Paris.

Restuccia, Diego, and Richard Rogerson. 2013. "Misallocation and Productivity." *Review of Economic Dynamics* 16 (1): 1–10.

Robb, Alicia M., and David T. Robinson. 2014. "The Capital Structure Decisions of New Firms." *Review of Financial Studies* 27 (1): 153–179.

Schoar, Antoinette. 2010. "The Divide between Subsistence and Transformational Entrepreneurship." *Innovation Policy and the Economy* 10 (1): 57–81.

Sedláček, Petr, and Vincent Sterk. 2017. "The Growth Potential of Startups over the Business Cycle." *American Economic Review* 107 (10): 3182–3210.

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?." *Quarterly Journal of Economics* 136 (4): 2309–2369.

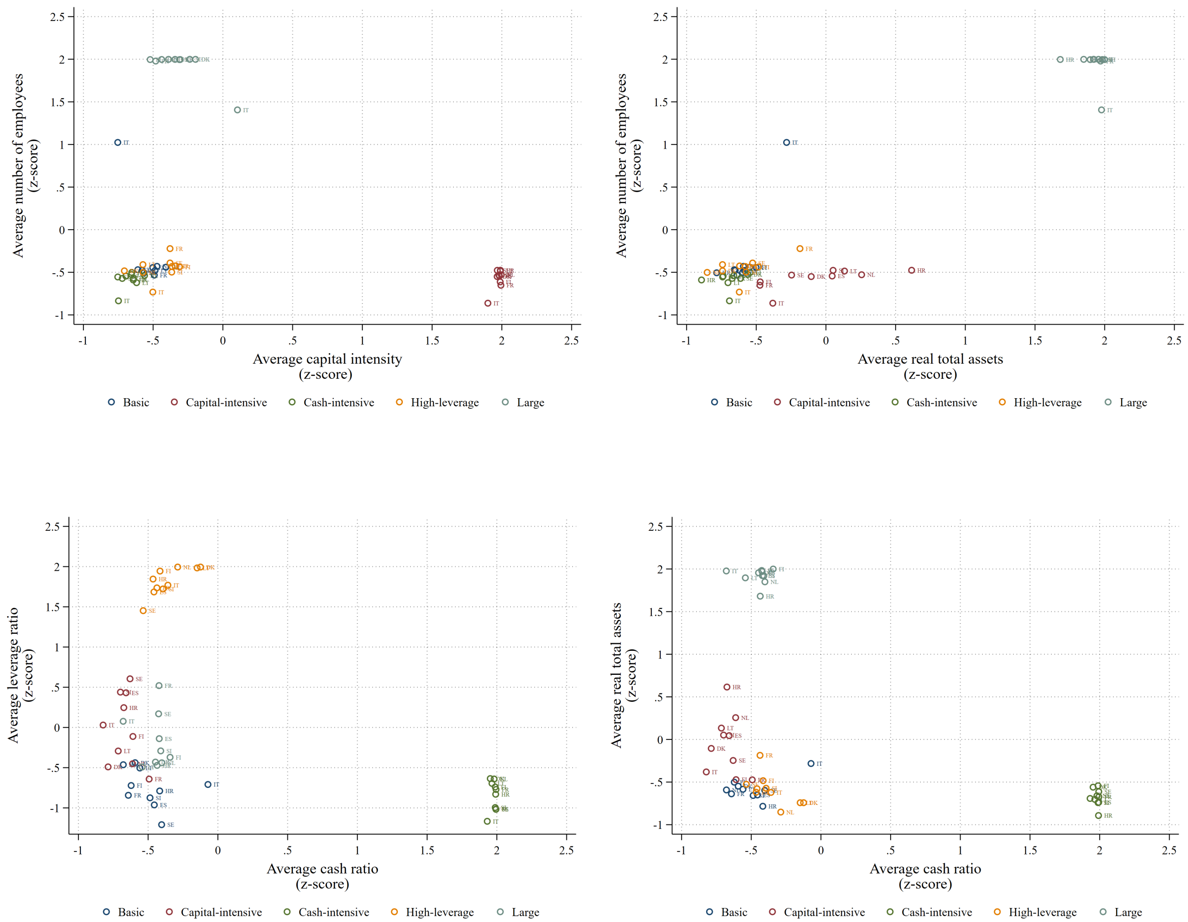
Sterk, Vincent ⓡ Sedláček, Petr ⓡ Benjamin W. Pugsley. 2021. "The Nature of Firm Growth." *American Economic Review* 111 (2): 547–579.

Syverson, Chad. 2011. "What Determines Productivity?." *Journal of Economic Literature* 49 (2): 326–365.

Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory* 21 (2): 265–293.

Tables and Figures

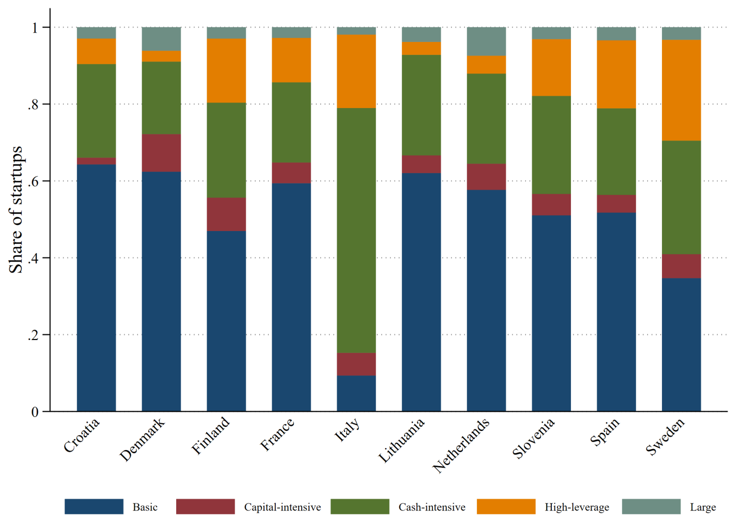
Figure 1. Meta-clustering of startup types



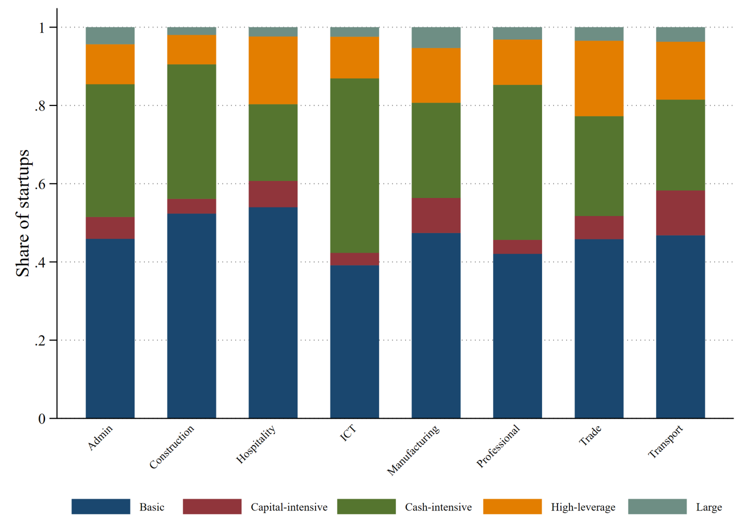
Notes: The four panels in this figure summarize the meta-clustering procedure. Different meta-clusters are denoted by different colors. The meta-clustering groups comparable clusters from different countries, taking the first-stage cluster centers derived from each country's clustering procedure as the observations. More specifically, in the meta-clustering procedure, the units of observation are z-scores of the first-stage cluster centers, averaged across years and industries.

Figure 2. Distribution of startup types by country, industry, and cohort

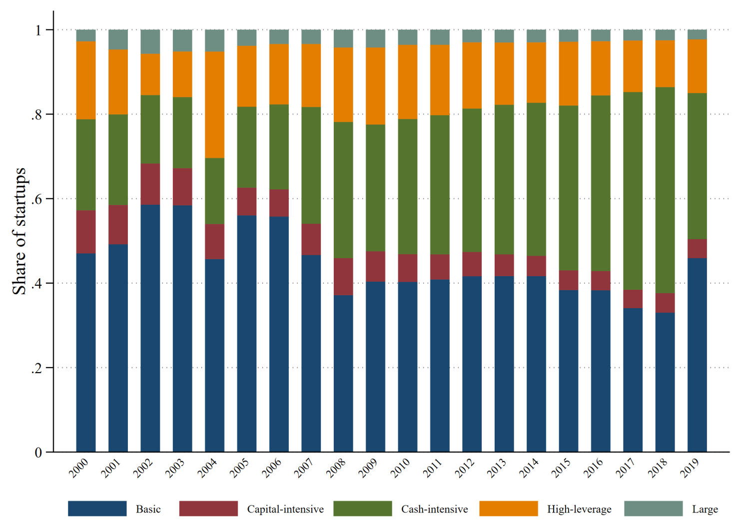
Panel A: Distribution of startup types by country



Panel B: Distribution of startup types by one-digit industry

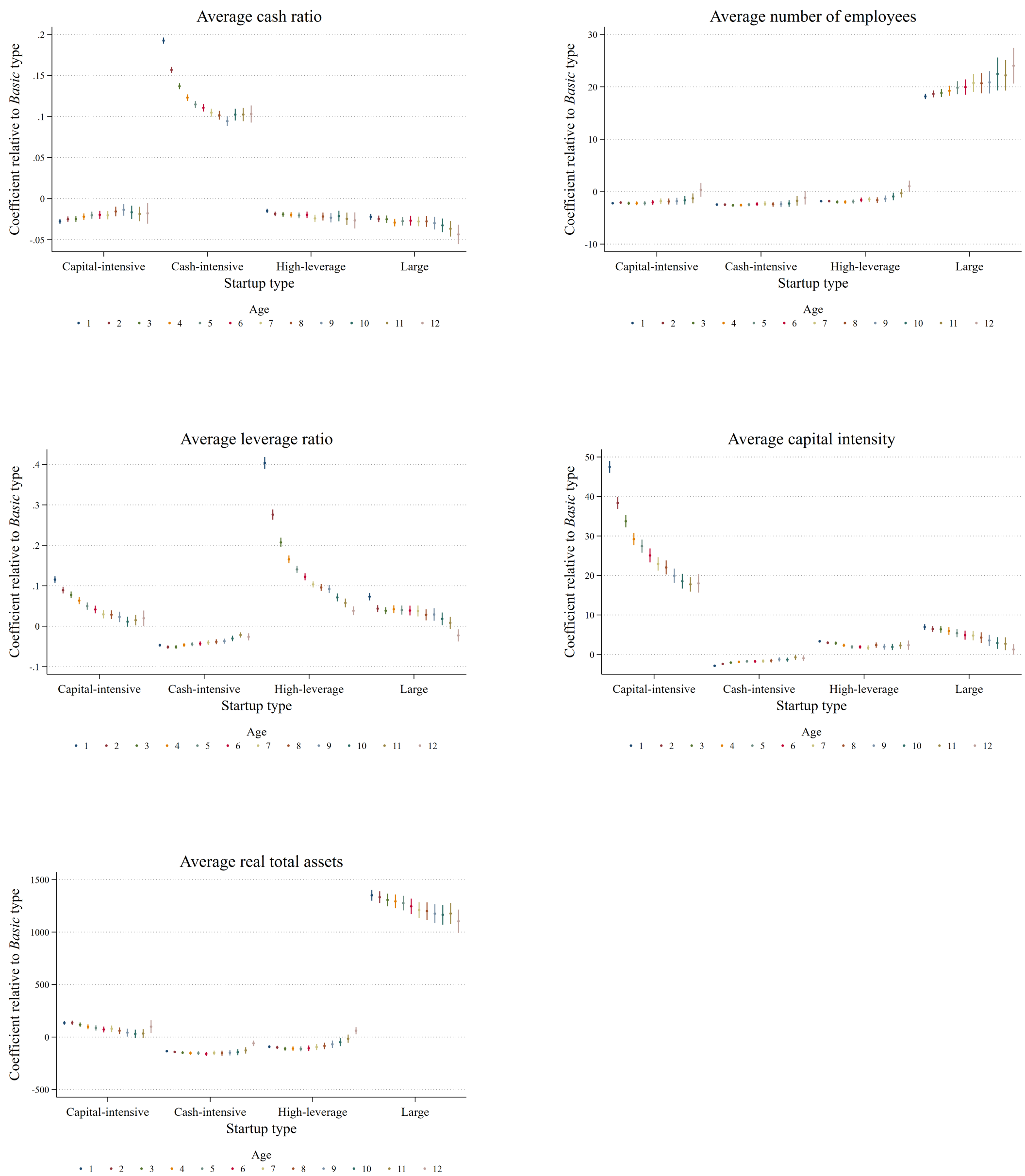


Panel C: Distribution of startup types by cohort



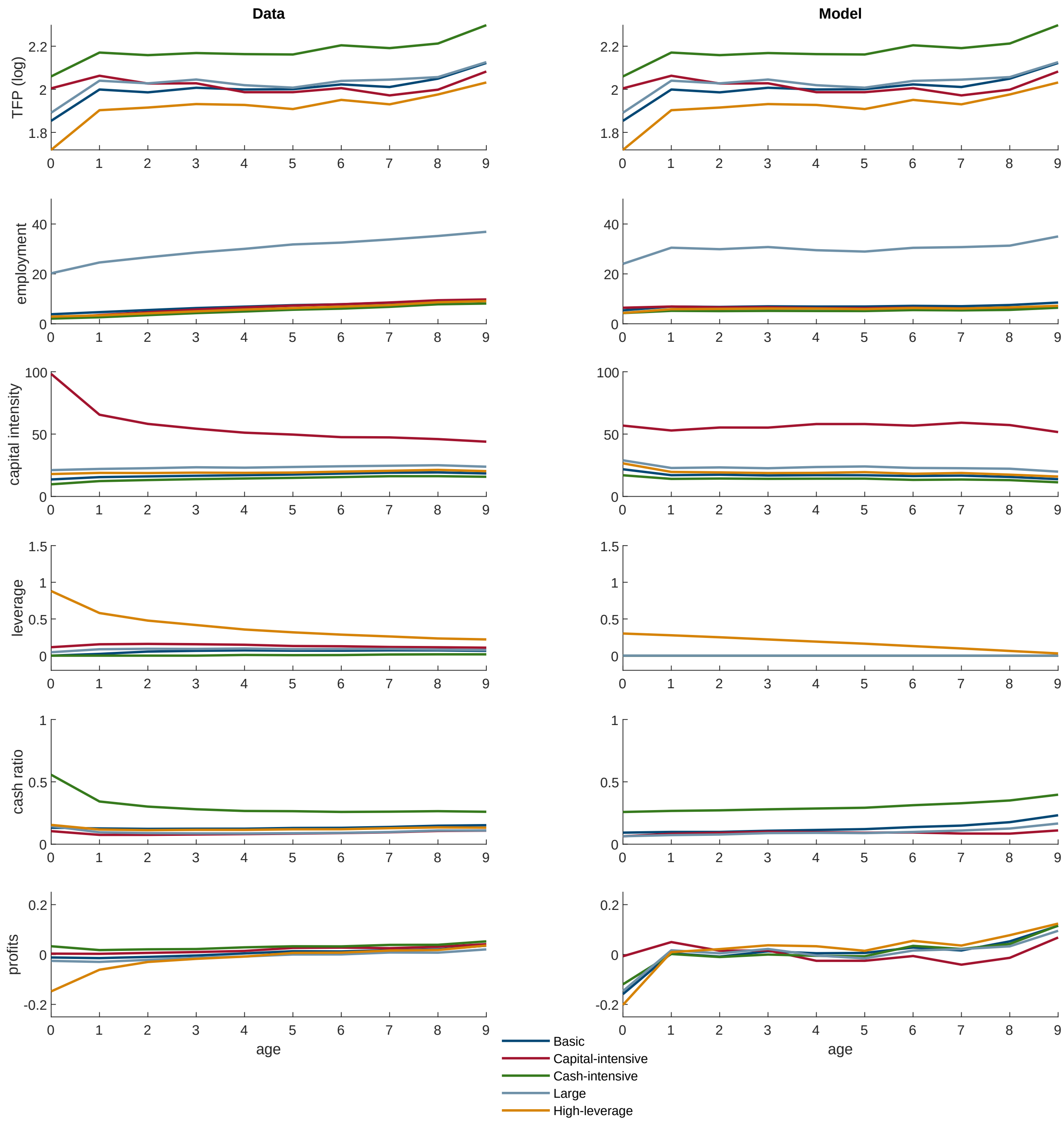
Notes: This figure illustrates the distribution of the startup population across the five startup types for individual countries (Panel A), one-digit NACE Rev.2 industries (Panel B), and cohorts (Panel C). The startup population comprises firms in their year of entry (age 0) in all cohorts available for each country.

Figure 3. The life cycle of startup types



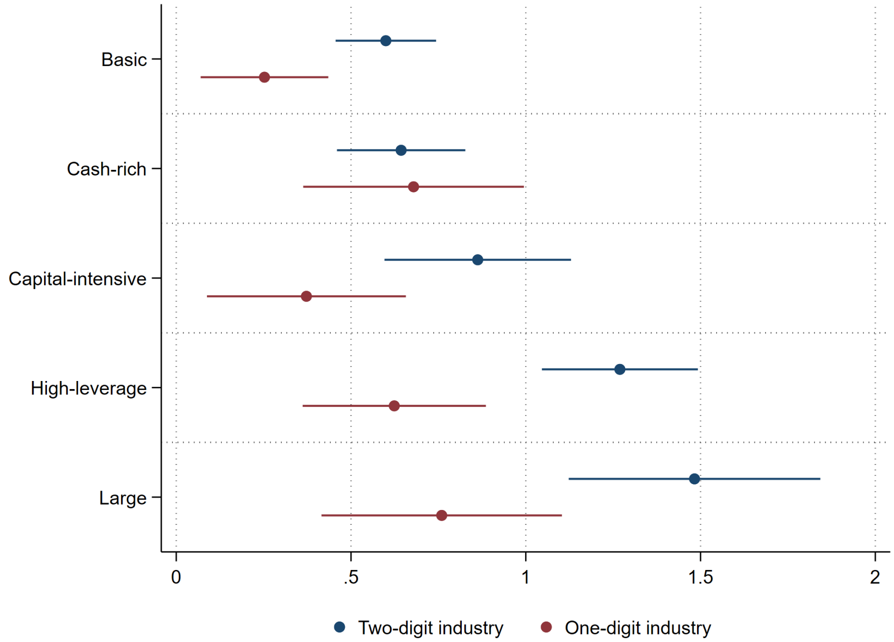
Notes: The panels in this figure summarize how the startup types develop during the first 12 years of their life in terms of the five clustering variables. Each panel corresponds to one clustering variable and plots the coefficients from 12 separate regressions where the dependent variable is this clustering variable. Each regression is then run for each age (age is 1, 2, ...,12 years). For example, the first panel summarizes regressions in which the *Average cash to total assets* ratio is regressed on dummy variables for the startup types (the *Basic* type is omitted) as well as *Country* × *Cohort* and *Industry* fixed effects. The sample is the full panel data set at the two-digit industry level. Standard errors are clustered at the *Industry* × *Country* × *Cohort* level, and whiskers indicate 95 percent confidence intervals.

Figure 4. Startup profiles: Data versus model



Notes: See the main text and appendix for a description of the model and calibration procedure. The charts on the left show unweighted averages across countries and cohorts. Starting from data aggregated to the country-startup type-cohort-age level, we regress outcomes on *Type x Age FE* and *Country x Age FE*. For each outcome, we plot the total effect for each type at each age.

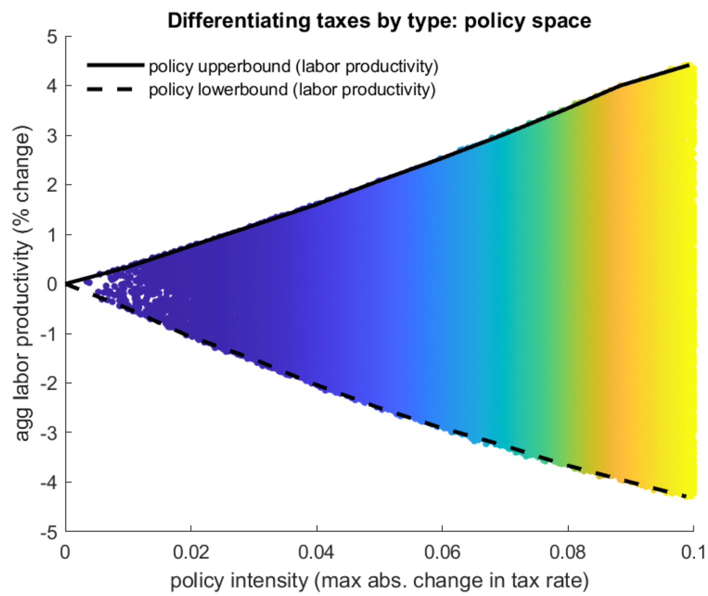
Figure 5. (Inverse) entry elasticities by startup type



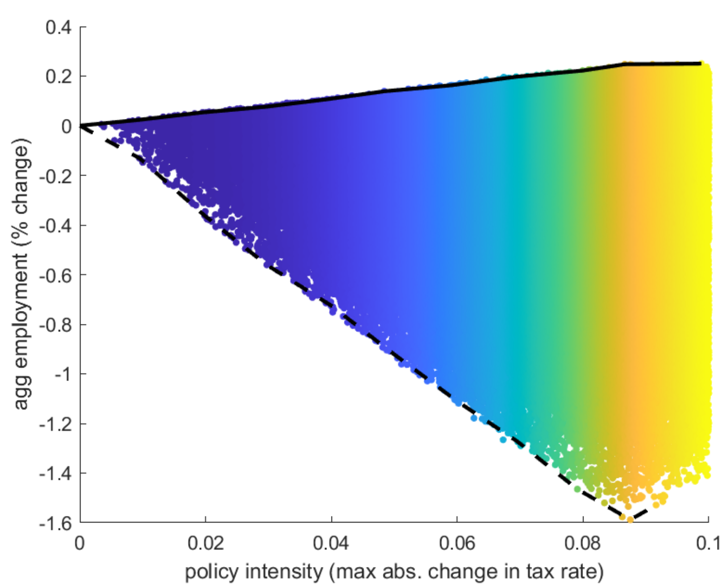
Notes: This figure shows coefficients from two OLS regressions where the dependent variable is the log net present value of firm profits. The estimated coefficients represent the inverse entry elasticity, β_1, for each startup type. *Country* × *Startup type* × *Industry* and *Country* × *Industry* × *Cohort* fixed effects are included throughout. Blue and red markers depict regressions at the two-digit and one-digit NACE Rev.2 industry level, respectively. The data set only includes cohorts observed for more than seven years (those founded in or before 2011). For age 12 and onward, profits and the year-on-year exit rate are assumed fixed. Standard errors are clustered at the *Industry* × *Country* × *Cohort* level, and whiskers indicate 95 percent confidence intervals.

Figure 6. Policy experiment: Differentiation of corporate income tax

Panel A: Labor productivity

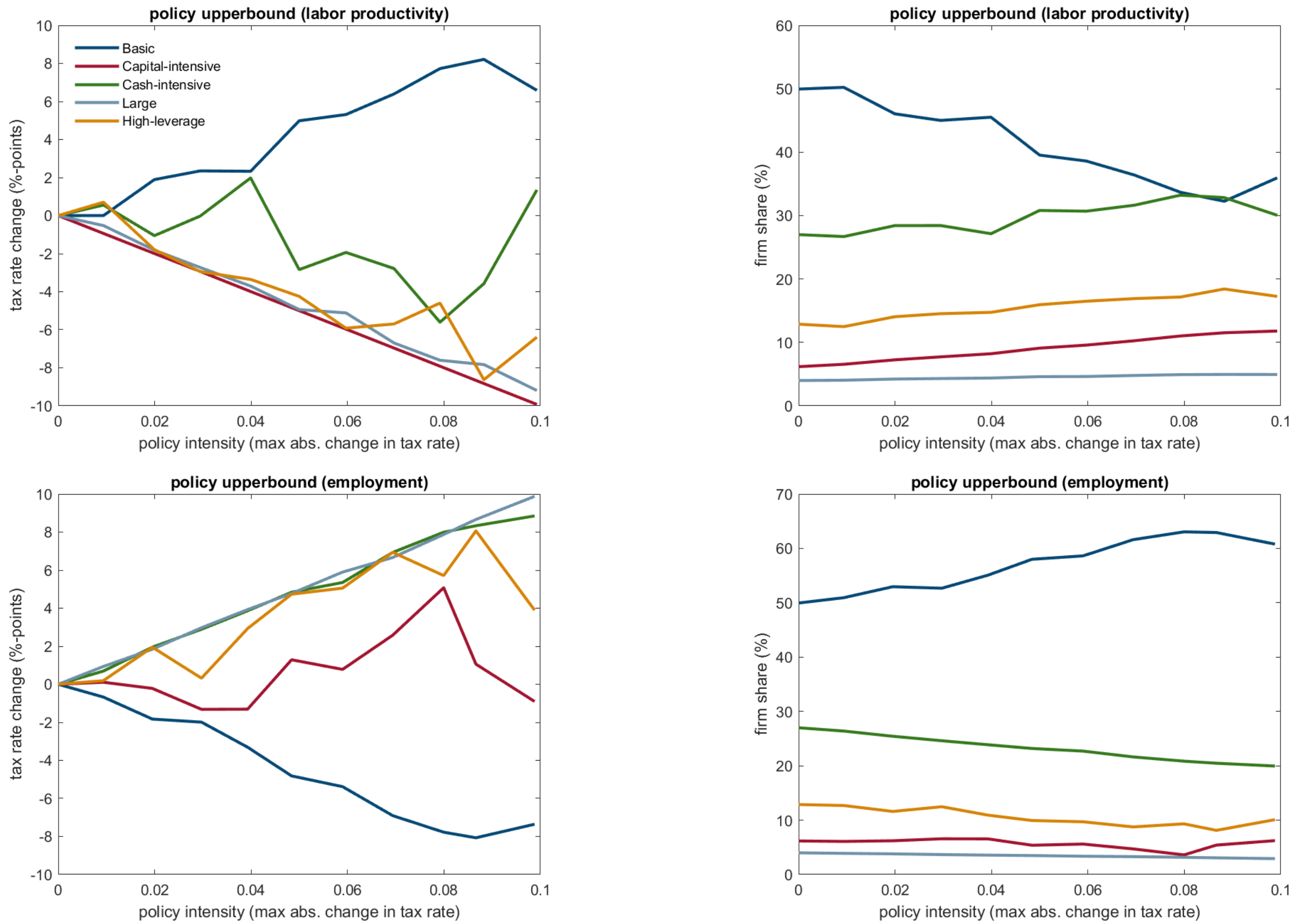


Panel B: Employment



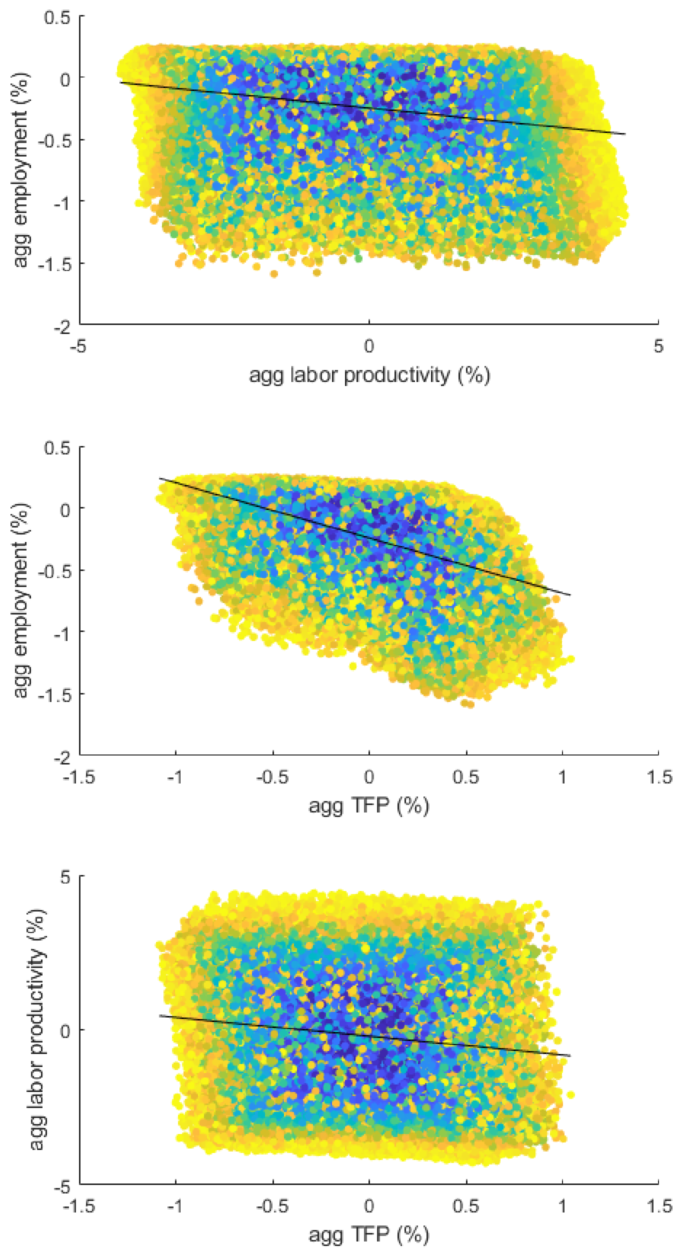
Notes: This figure summarizes the policy experiment. Panel A (B) shows the policy space for aggregate labor productivity (aggregate employment). The horizontal axis measures the intensity of the potential policy change as the maximum absolute change in the tax rate. Warmer colors indicate stronger corporate tax rate differentiation. The solid line plots the "policy upper bound": the largest possible aggregate labor productivity increase given a certain policy intensity. The dashed line plots the corresponding "policy lower bound". Both panels use entry elasticities based on one-digit industry data.

Figure 7. Policy experiment: Upper bounds of macro impacts due to tax differentiation



Notes: This figure shows the specific tax policies and resulting startup shares associated with the two policy upper bounds from Figure 6. The two panels on the left (right) plot the corporate tax rate policies (startup shares) associated with the two policy upper bounds. The upper left (right) panel shows the tax rates (startup shares) associated with the labor productivity upper bound. The lower left (right) panel shows the tax rates (startup shares) associated with the employment upper bound.

Figure 8. Policy experiment: Macroeconomic trade-offs



Notes: These scatter plots each depict pairs of changes in aggregate or average macroeconomic outcomes resulting from each potential policy (up to a policy intensity of 0.1). Wages per employee are in thousands of euros. Linear regression lines are shown in black. Warmer colors indicate higher policy intensity, as in Figure 6.

Table 1. Characteristics of startup types at time of entry

	(1)	(2)	(3)	(4)	(5)
Start-up type	No. of employees	Capital intensity	Real total assets	Cash ratio	Leverage ratio
Basic	4	8.56	166.59	0.12	0.23
Capital-intensive	2	93.18	405.12	0.09	0.41
Cash-intensive	2	4.67	92.46	0.54	0.18
High-leverage	3	12.90	122.97	0.14	1.18
Large	20	16.05	1488.30	0.13	0.34

Notes: This table presents the cross-country means of the cluster variables for each of the five startup types in the year of establishment. Means are unweighted and based on all available cohorts and countries.

Table 2. Startup types and firm outcomes

Panel A: All firms, aged 0-9

	(1)	(2)	(3)	(4)	(5)
	Aggregate labor productivity	Aggregate TFP	Average exit probability	Average wage per employee	Average profit margin
Capital-intensive	0.308	0.047	-0.061	2.359	0.012
	(0.004)	(0.004)	(0.003)	(0.091)	(0.001)
Cash-intensive	0.030	0.050	-0.009	1.104	0.022
	(0.003)	(0.002)	(0.002)	(0.056)	(0.001)
High-leverage	-0.044	-0.038	-0.000	-1.494	-0.030
	(0.003)	(0.003)	(0.002)	(0.062)	(0.001)
Large	0.165	0.045	-0.125	3.391	-0.018
	(0.004)	(0.003)	(0.003)	(0.092)	(0.001)
Constant	3.362	2.216	0.717	27.904	0.042
	(0.002)	(0.001)	(0.001)	(0.032)	(0.000)
R-squared	0.904	0.978	0.639	0.910	0.604
N	28,991	20,786	31,240	30,259	29,992

Panel B: All firms, aged 5-9

	(1)	(2)	(3)	(4)	(5)
	Aggregate labor productivity	Aggregate TFP	Average exit probability	Average wage per employee	Average profit margin
Capital-intensive	0.190	0.011	-0.062	0.832	0.013
	(0.007)	(0.006)	(0.003)	(0.114)	(0.001)
Cash-intensive	0.018	0.031	-0.024	1.174	0.013
	(0.005)	(0.004)	(0.003)	(0.088)	(0.001)
High-leverage	-0.025	-0.022	0.001	-1.328	-0.008
	(0.005)	(0.005)	(0.002)	(0.097)	(0.001)
Large	0.151	0.048	-0.140	2.669	-0.013
	(0.007)	(0.006)	(0.003)	(0.144)	(0.001)
Constant	3.473	2.251	0.817	30.289	0.049
	(0.003)	(0.003)	(0.001)	(0.048)	(0.001)
R-squared	0.904	0.981	0.618	0.929	0.648
N	10,160	6,993	11,167	10,664	10,583
Country × Cohort FE	✓	✓	✓	✓	✓
Industry × Cohort FE	✓	✓	✓	✓	✓
Country × Industry FE	✓	✓	✓	✓	✓
Age × Country FE	✓	✓	✓	✓	✓
Age × Industry FE	✓	✓	✓	✓	✓
Age × Cohort FE	✓	✓	✓	✓	✓

Notes: This table shows OLS regressions where the dependent variable is indicated in the column heading. Observations are at the country × sector × startup type × cohort × age level. The dependent variables in the first two columns are aggregate outcomes (i.e. employment-weighted averages) by startup type. The dependent variables in the last three columns are simple averages by startup type. Industries are defined at the one-digit NACE Rev.2 level. Regressions are based on the full panel of firms aged 0-9 in Panel A and firms aged 5-9 in Panel B. Robust standard errors are in parentheses. Wages are denominated in thousands of euros.

Table 3. Type-specific parameters in the post-entry model

	α	γ	$\alpha + \gamma$	m_{-1}/w	c/w
	Capital elasticity	Labor elasticity	Scalability	Initial equity	Fixed cost
Basic	0.29	0.41	0.70	35.5	6.8
Capital-intensive	0.44	0.18	0.62	100.8	19.4
Cash-intensive	0.24	0.41	0.65	26.6	5.5
High-leverage	0.31	0.38	0.70	21.5	31.2
Large	0.37	0.38	0.75	201.1	6.2

Notes: See the main text and appendix for a description of the model and calibration procedure. The initial equity and fixed costs are scaled by the annual wage per worker for ease of interpretation.

Table 4. Macroeconomic effects of capital, labor, and interest subsidies

	Baseline	Capital subsidy	Labor subsidy	Interest subsidy
Panel A. Policy rates (%)				
τ_k	0.00	−0.29	0.00	0.00
τ_l	0.00	0.00	−1.68	0.00
τ_b	0.00	0.00	0.00	−0.29
τ_π	0.00	5.00	5.00	0.02
Panel B. Startup shares (percentage-point change vs. baseline)				
Basic		−0.57	0.54	−0.33
Capital-intensive		0.62	−1.51	−0.02
Cash-intensive		−0.10	1.11	−0.06
Large		0.09	0.03	0.02
High-leverage		−0.06	−0.17	0.42

Panel C. Aggregate effects (% change vs. baseline)

	Capital subsidy			Labor subsidy			Interest subsidy		
	Total	Comp.	# firms	Total	Comp.	# firms	Total	Comp.	# firms
Output (Y)	0.67	0.82	−1.29	−3.10	−1.25	−3.13	0.03	−0.06	0.07
Capital (K)	2.65	1.33	−1.29	−0.29	−2.35	−3.13	0.04	−0.04	0.07
Employment (L)	0.05	0.25	−1.29	−0.01	0.10	−3.13	0.01	−0.06	0.07
Labor prod. (Y/L)	0.63	0.57	0.00	−3.09	−1.35	0.00	0.02	0.00	0.00
# firms	−1.29	0.00	−1.29	−3.13	0.00	−3.13	0.07	0.00	0.07

Notes: This table reports the effects of three alternative fiscal interventions financed through a uniform adjustment in the corporate income tax rate τ_π, with wages adjusting to clear the labor market. The capital subsidy τ_k is a flow subsidy per unit of installed capital; the labor subsidy τ_l reduces the effective wage paid by the firm; and the interest subsidy τ_b reduces the interest rate on long-term debt and is restricted to firms below age ten. Each intervention is scaled so that the financing requirement corresponds to a five-percentage-point increase in τ_π, except for the interest subsidy, whose total cost is small because of the age restriction. Panel A reports the policy rates implied by each intervention, with all rates expressed in percent. Panel B reports the resulting shifts in the share of each startup type, in percentage points relative to baseline. Panel C reports aggregate effects in percent relative to baseline: "Total" is the overall effect; "Comp." is the contribution from changes in startup composition holding the number of firms fixed; and "# firms" is the contribution from changes in the total number of firms holding composition fixed. Because of non-linearities, the decomposition is non-additive. The baseline is the calibrated economy with zero subsidies and a zero corporate income tax rate.

Internet Appendix for

Startup Types and Macroeconomic Performance in Europe

Ralph De Haas
EBRD, KU Leuven & CEPR

Vincent Sterk
UCL & CEPR

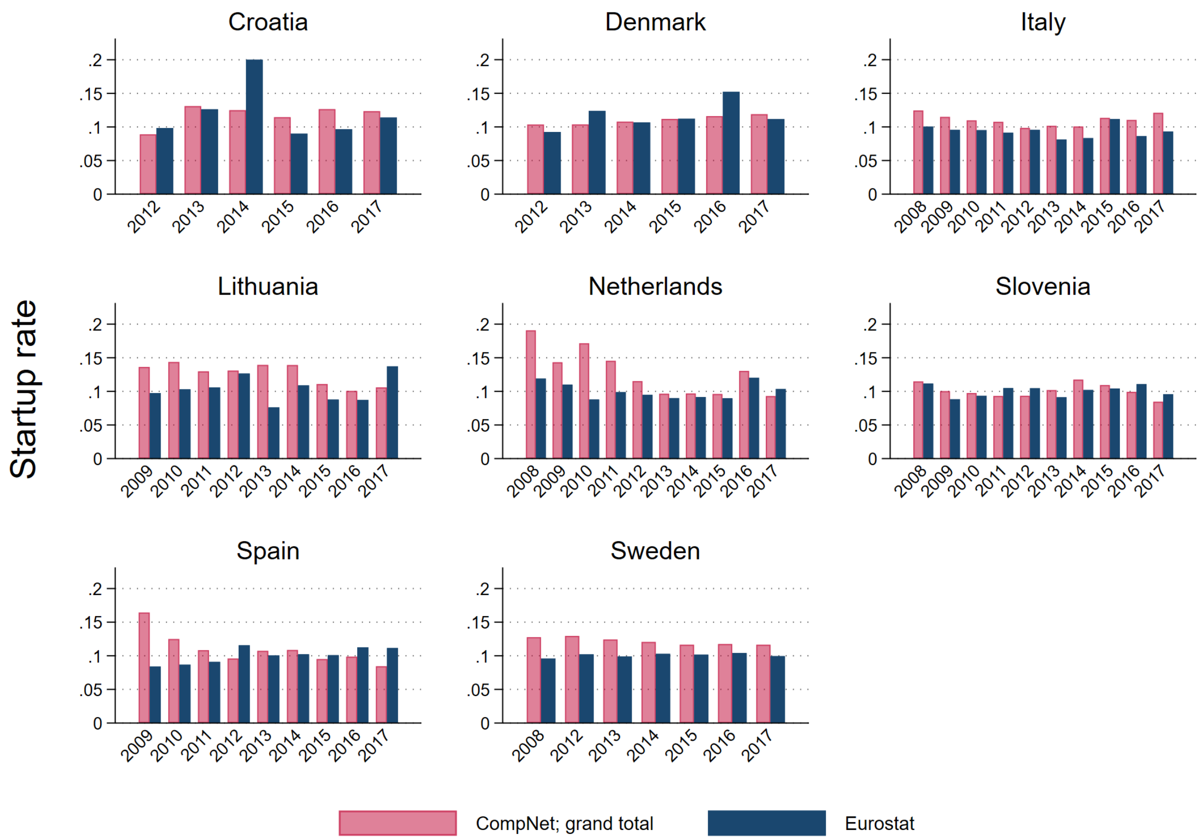
Neeltje Van Horen
University of Amsterdam & CEPR

Internet Appendix A: Additional Figures and Tables

Comparison to Eurostat. In Figures A1–A4, we validate our data by comparing our startup population to Eurostat's Business Demography Statistics (BDS) on startups (while excluding sole proprietorships for consistency). Eurostat is the statistical office of the European Union (EU) and collects the BDS from the national statistical institutes of EU member states. Figure A1 shows that startup rates (the number of startups in a year as a fraction of the total firm population) are very comparable between our CompNet-based data set and the aggregate data published by Eurostat. The same holds, by and large, when we compare average employment growth during the first five years after startup (Figure A2). In most countries, trend growth in both data sets is very similar. In a few cases—such as the Netherlands and Sweden—there are gaps in the average *level* of reported employees. In those cases, a comparison with other countries suggests the Eurostat data are anomalous, rather than the CompNet data. Next, Figure A3 compares exit rates (firm death) over time. The five-year cumulative exit rate is comparable at 45 percent (56 percent) in the CompNet (Eurostat) data set. Moreover, the trends as firms age are very similar in both cases, although in Sweden, CompNet tends to overreport firm exit.

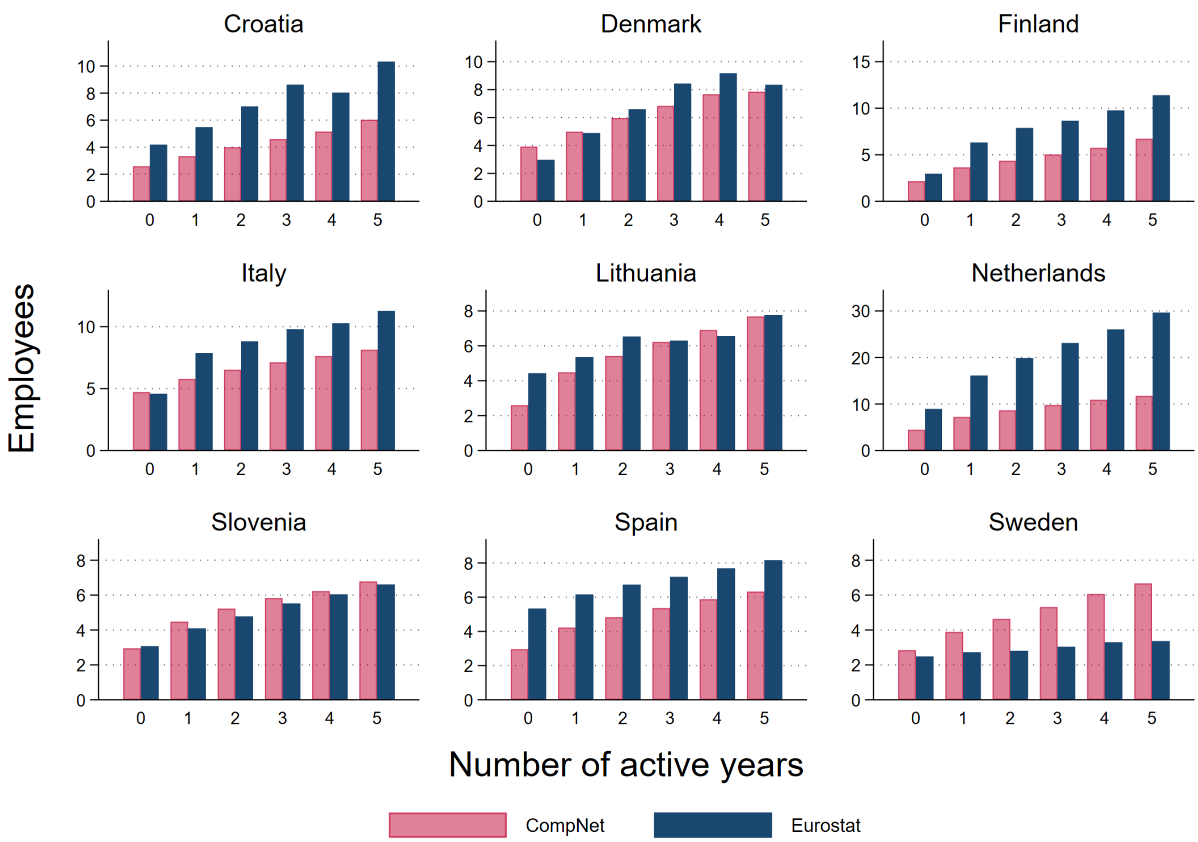
Finally, Figure A4 compares the sectoral composition of startups. We focus on the number of startups in each one-digit industry as a share of total startups across all industries, using pooled data over 2008–2018. While there is cross-country heterogeneity in the sectoral composition of startups, this national composition is very similar when using CompNet or Eurostat data. Overall, we conclude that the firm population that underpins the statistical moments we derive from CompNet is representative of the firm population in our sample countries.

Figure A1. Startup rates by cohort and country: CompNet versus Eurostat



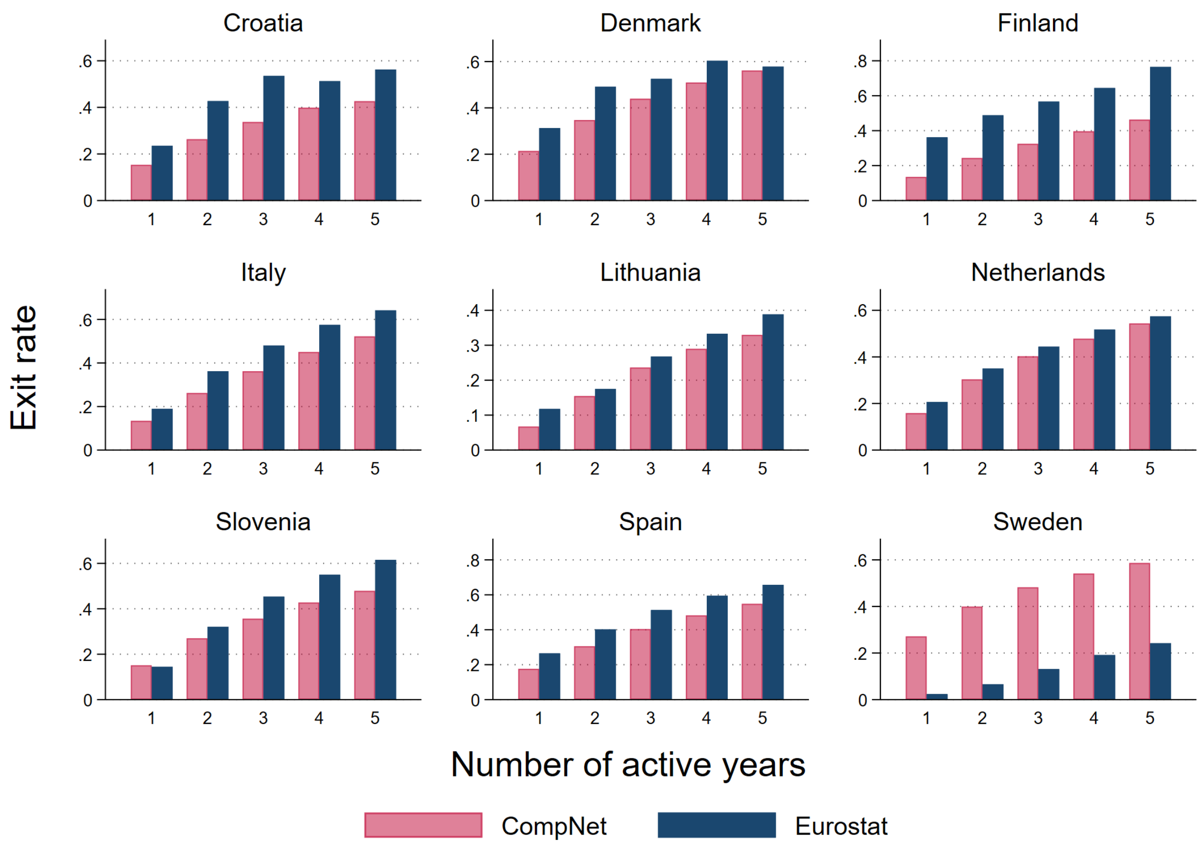
Notes: This figure compares the total number of startups in the CompNet database (pink bars) with establishment of firms in Eurostat (blue bars). Cohorts reported are subject to data availability. Finland is omitted in this chart because we cannot access the total number of startups in the CompNet database for that country. France does not appear as Eurostat does not report firm entry prior to 2008.

Figure A2. Employment by firm age: CompNet versus Eurostat



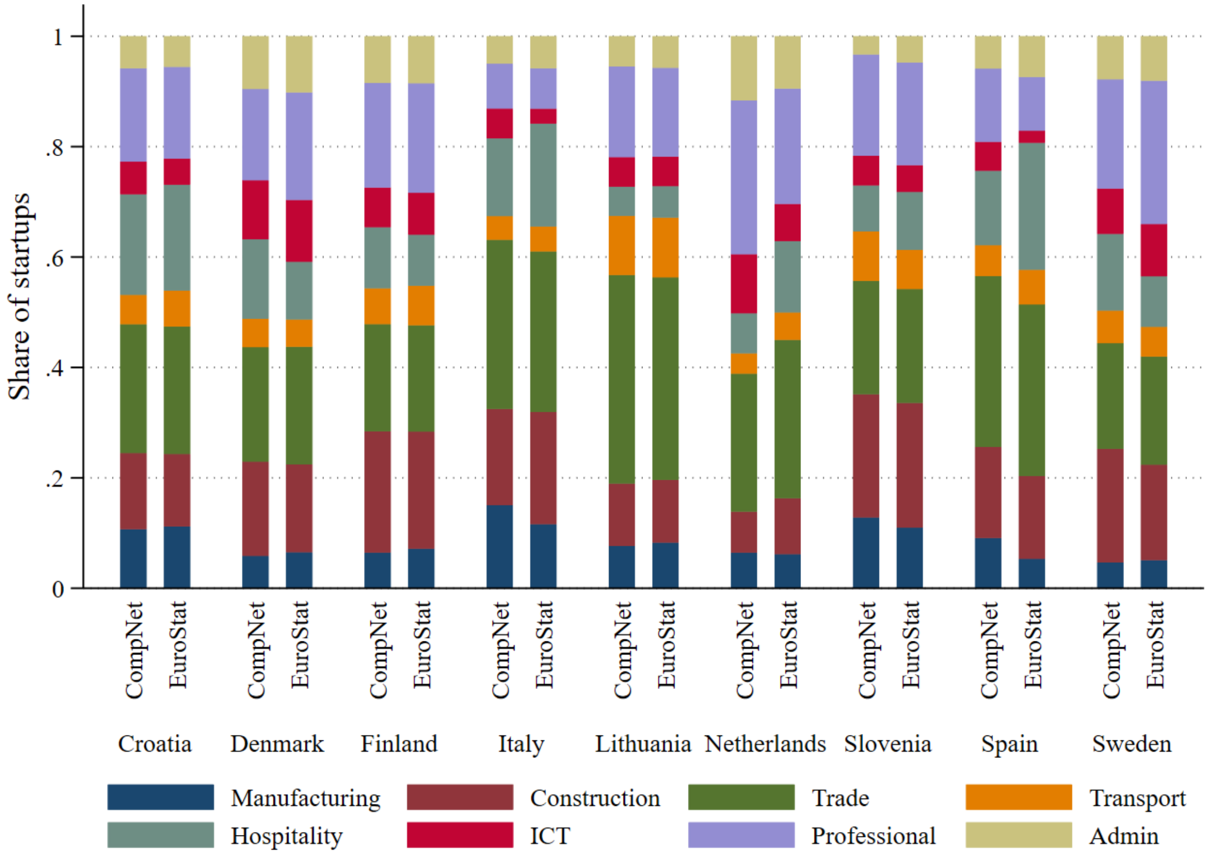
Notes: This figure compares growth in number of persons employed by startups as reported by CompNet (pink bars) and Eurostat (blue bars). CompNet and Eurostat data are matched based on the startup cohort and age group. For comparison purposes, we adjust the Eurostat data such that sole proprietorship firms are removed and we adjust for the average number of persons employed by sole proprietorship firms. The x-axis depicts *Startup age*, which is the number of years a startup has been active. *Employees* on the y-axis is averaged over cohorts. France does not appear as Eurostat does not report firm entry prior to 2008.

Figure A3. Cumulative exit rates by firm age: CompNet versus Eurostat



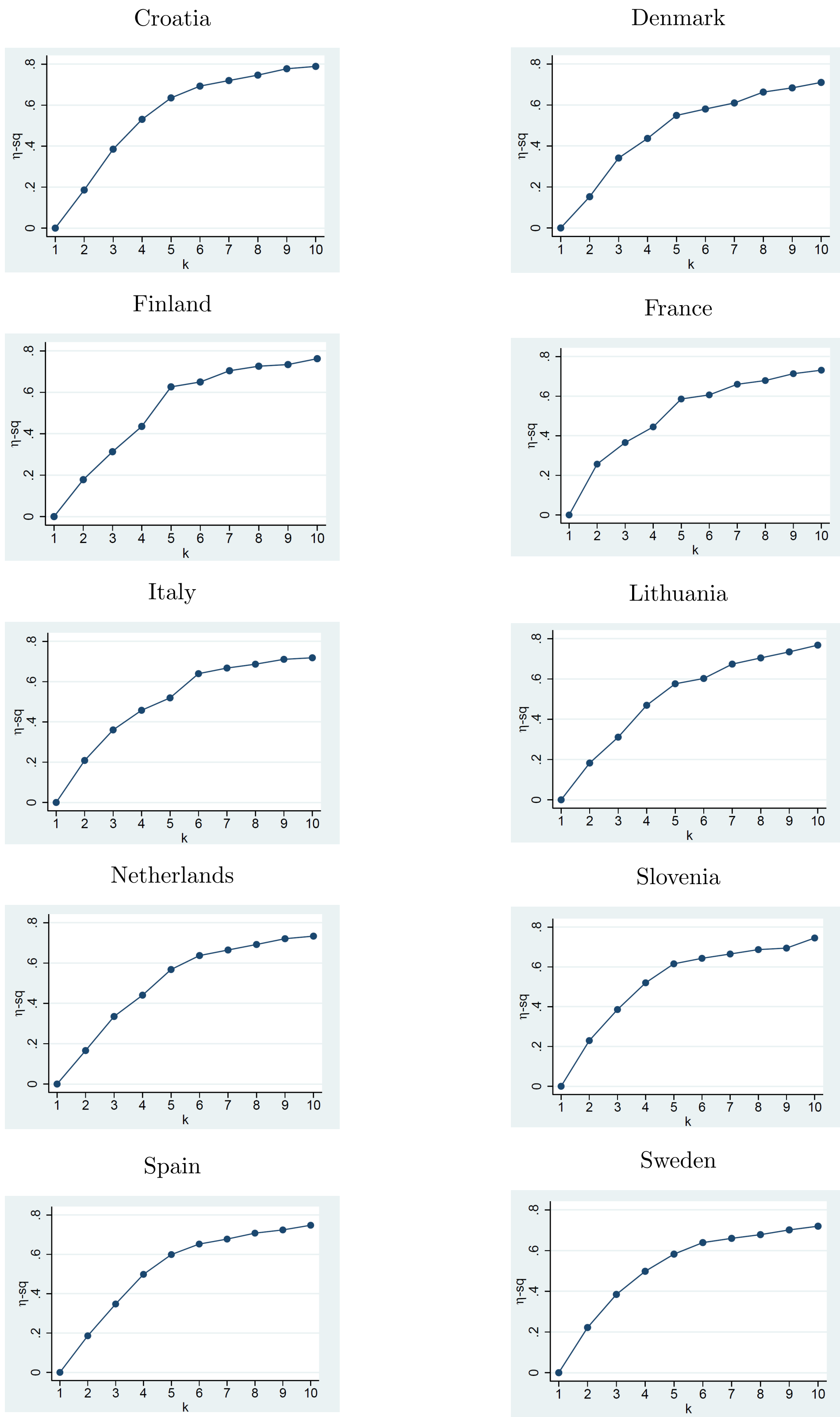
Notes: This figure compares the cumulative exit rate of startups in CompNet (pink bars) and Eurostat (blue bars). CompNet and Eurostat data are matched based on the startup cohort and age group. The x-axis depicts *Startup age*, which is the number of years a startup has been active. *Exit rate* is the average cumulative exit rate over all cohorts for each startup age group. France does not appear as Eurostat does not report firm entry prior to 2008.

Figure A4. Startup shares by industry: CompNet versus Eurostat



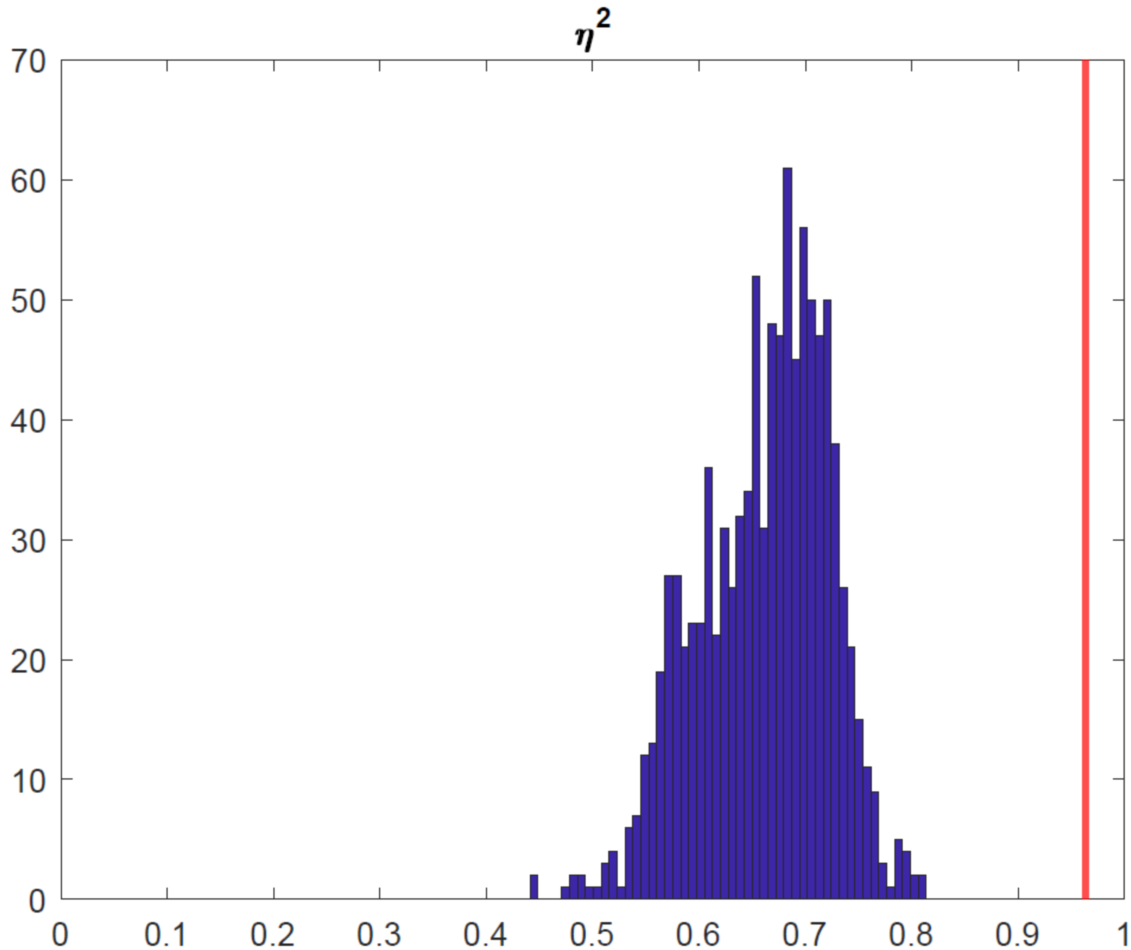
Notes: This figure compares the number of startups in each industry as a share of total startups across the industries. The data is pooled over the years 2008–2018; country × year × industry cells only enter the calculations if the cell is non-missing in both CompNet and Eurostat.

Figure A5. Scree plots used to choose the number of clusters



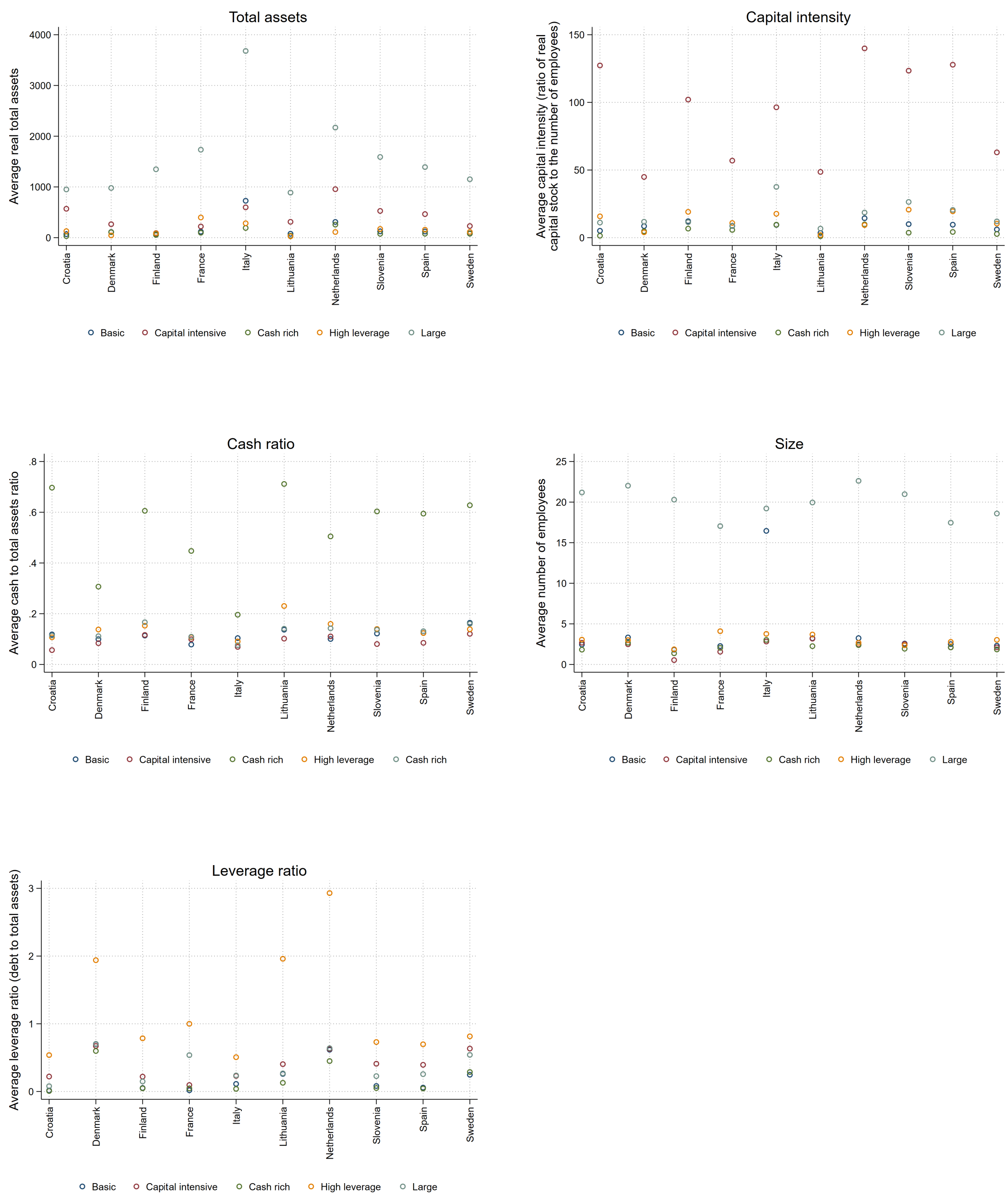
Notes: This figure shows scree plots resulting from the k-means cluster algorithm at the firm-level, for each country in the sample. On the x-axis, k indicates the number of clusters. The η^2 coefficient on the y-axis measures the proportional reduction of the within sum of squares for each cluster solution k compared with the total sum of squares.

Figure A6. Monte Carlo validation of the meta-clustering procedure



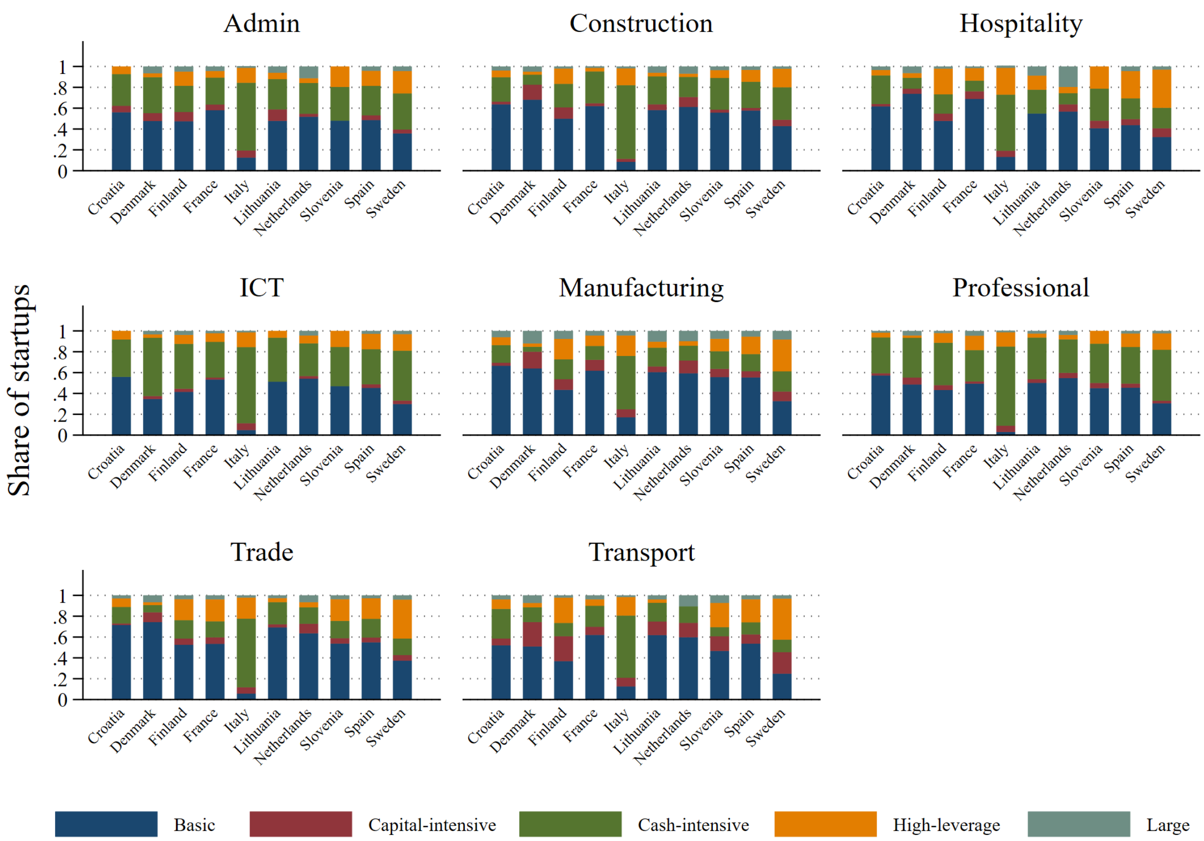
Notes: This histogram summarizes a Monte Carlo experiment consisting of 1,000 random draws for the cluster variables, with means and standard deviations as observed in the data. These draws are i.i.d. so that no clusters exist in the experimental data. Each time η^2 is computed (blue bars). The vertical red line indicates the true η^2 statistic based on the actual data.

Figure A7. Startup characteristics at the time of firm entry, by type and country



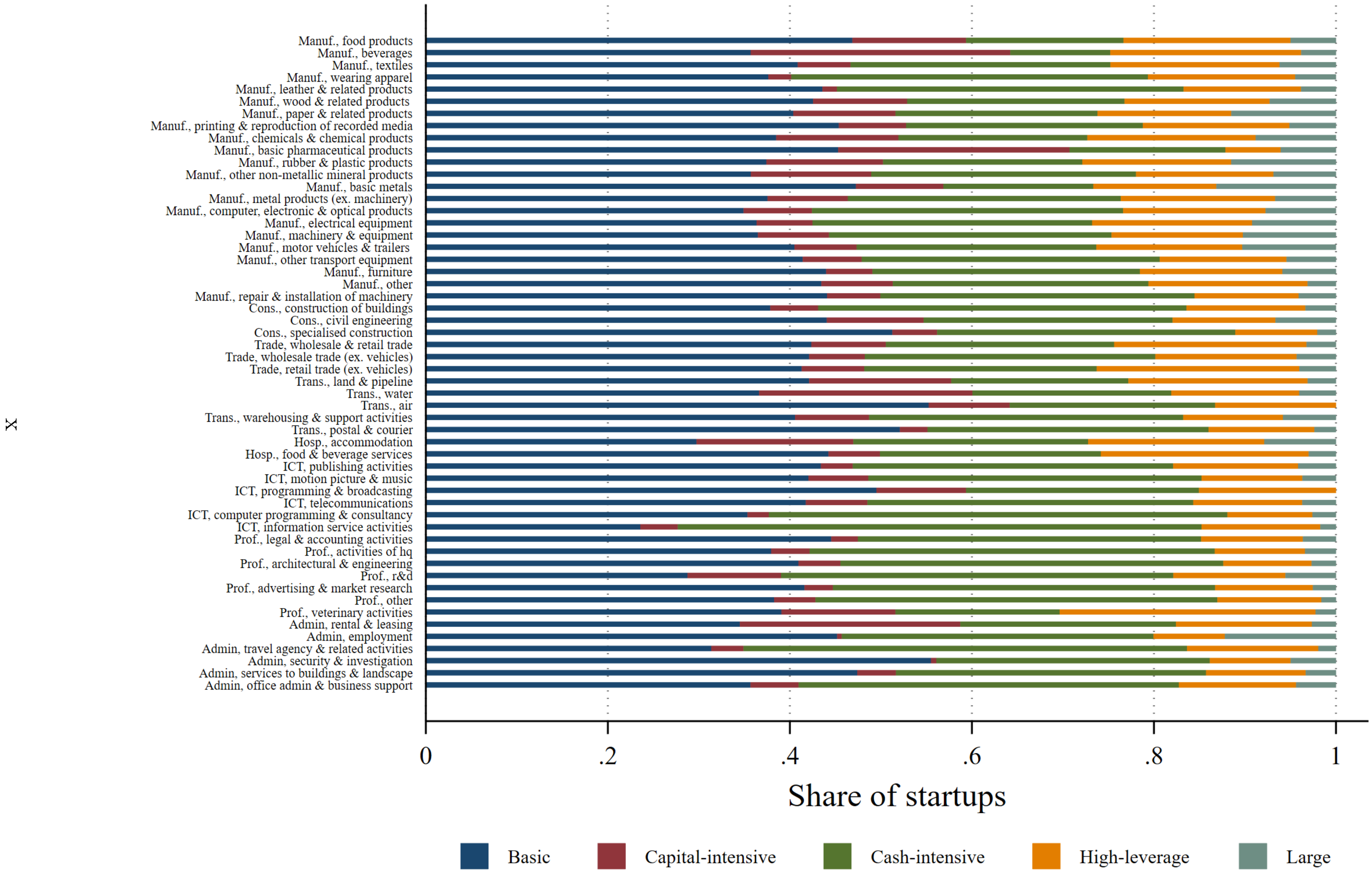
Notes: This figure presents the country-level mean of the five cluster variables for the five startup types in the year of firm entry.

Figure A8. Distribution of startup types by industry and by country



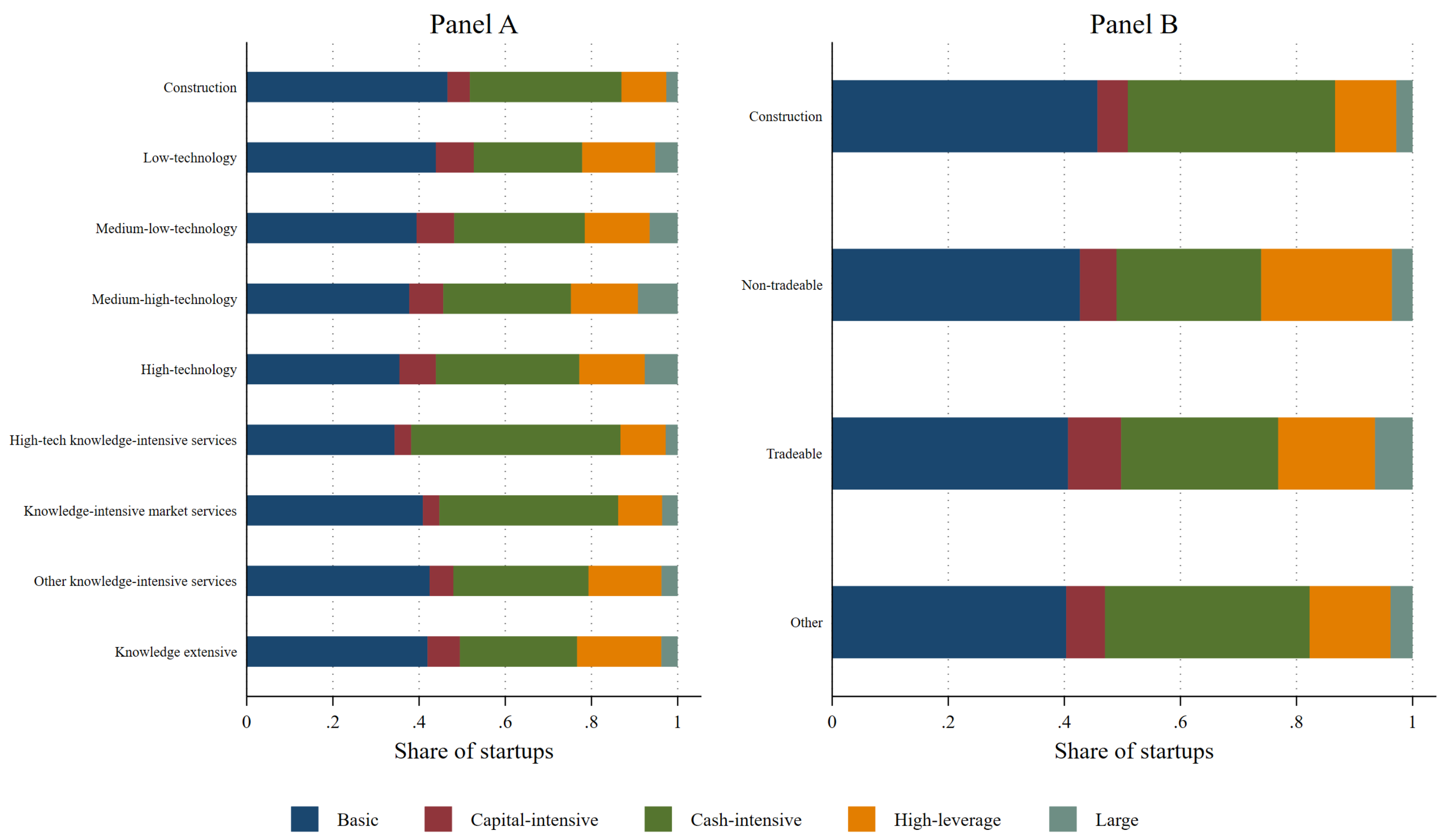
Notes: This figure presents the distribution of the startup population for individual one-digit NACE Rev.2 industries and countries across the five startup types. The startup population comprises startups that are just born (age 0) and shares are averaged over all cohorts available for each country.

Figure A9. Distribution of startup types by two-digit industry



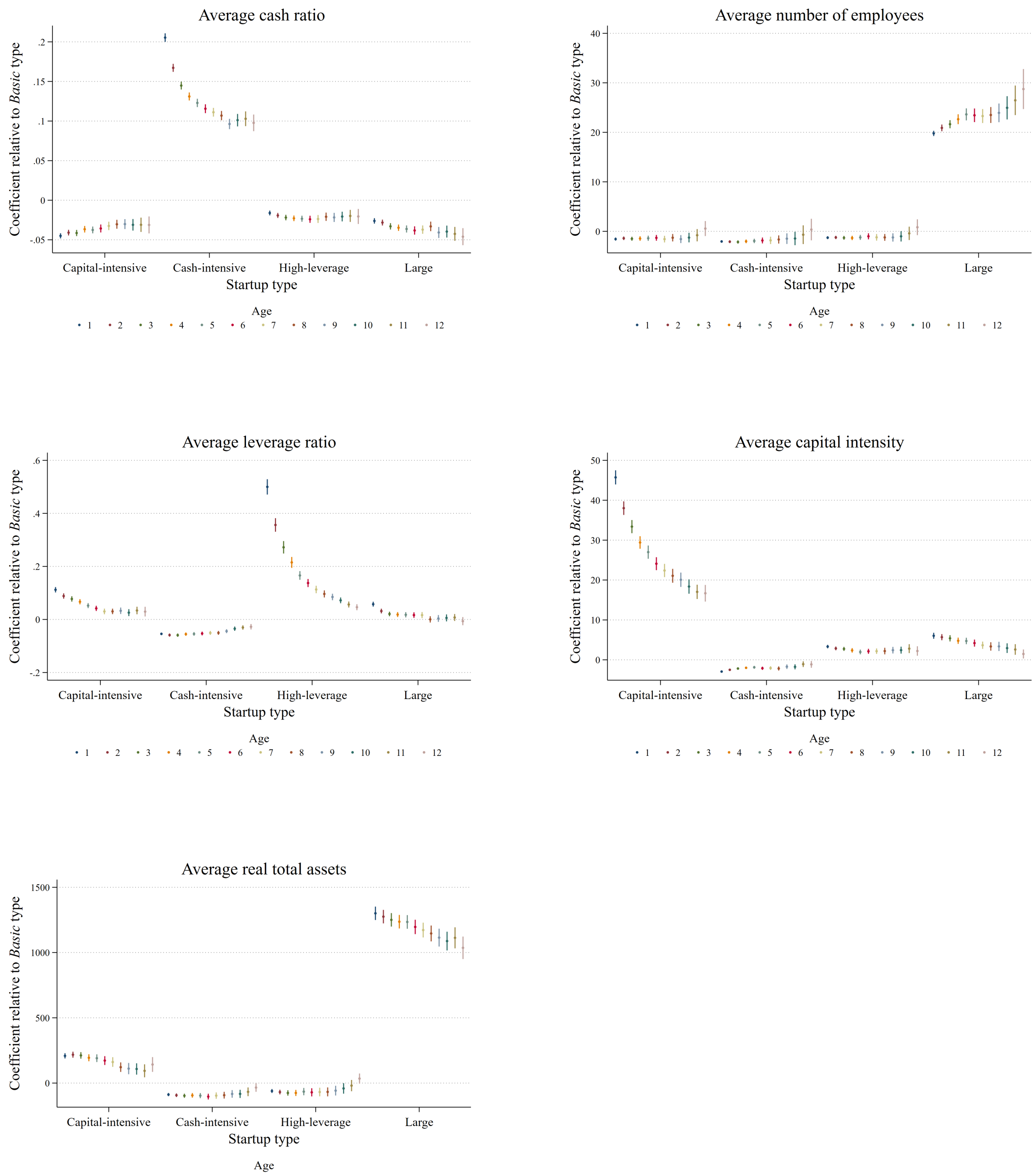
Notes: This figure illustrates the distribution of the startup population for two-digit NACE Rev.2 industries across the five startup types. The startup population comprises startups that are just born (age 0) in all cohorts available for each country.

Figure A10. Distribution of startup types by alternative industry classification



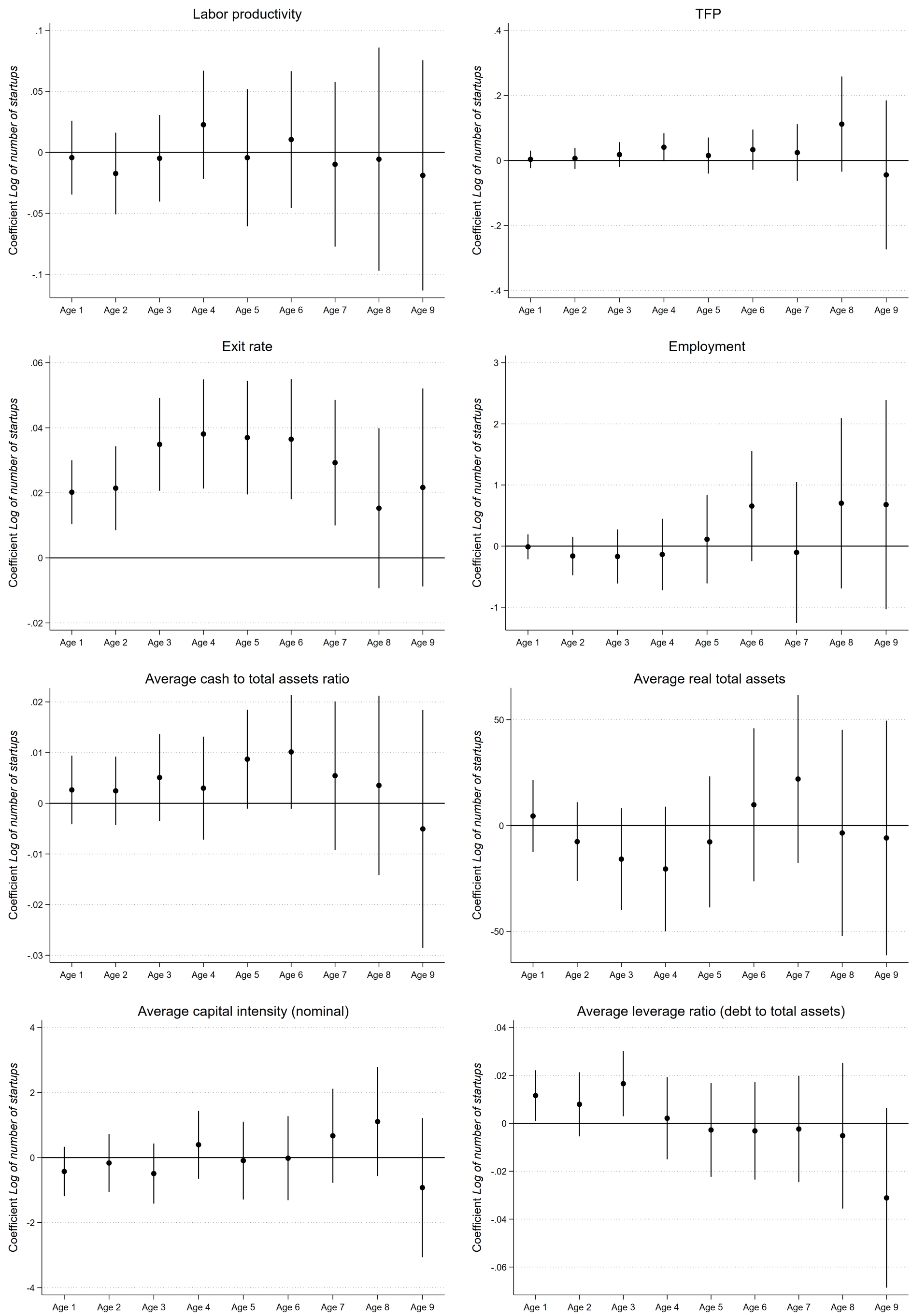
Notes: This figure illustrates the startup distribution for two different industry classifications. Panel A draws on Eurostat's classification of high-tech manufacturing industries and knowledge-intensive services. Using these classifications, we allocate two-digit industries into one of nine categories. Our data also includes 15 two-digit industries in construction, trade, transport, hospitality and administration that cannot be categorised this way. For the construction industries, we create a separate category while the remaining industries are grouped in a "knowledge extensive" category. Panel B draws on Mian and Sufi (2014) to assign two-digit industries to construction, tradable, non-tradable, and other. This is straightforward for 33 two-digit industries as all underlying four-digit industries have the same Mian and Sufi (2014) classification. For 22 other industries the underlying four-digit industries indicate at least two Mian and Sufi classifications. We then assign the categorisation that is attached to the majority of four-digit industries within the more aggregated two-digit industry. The population comprises startups that are just born (age 0) in all cohorts available for each country.

Figure A11. The life cycle of startup types (one-digit industry level)



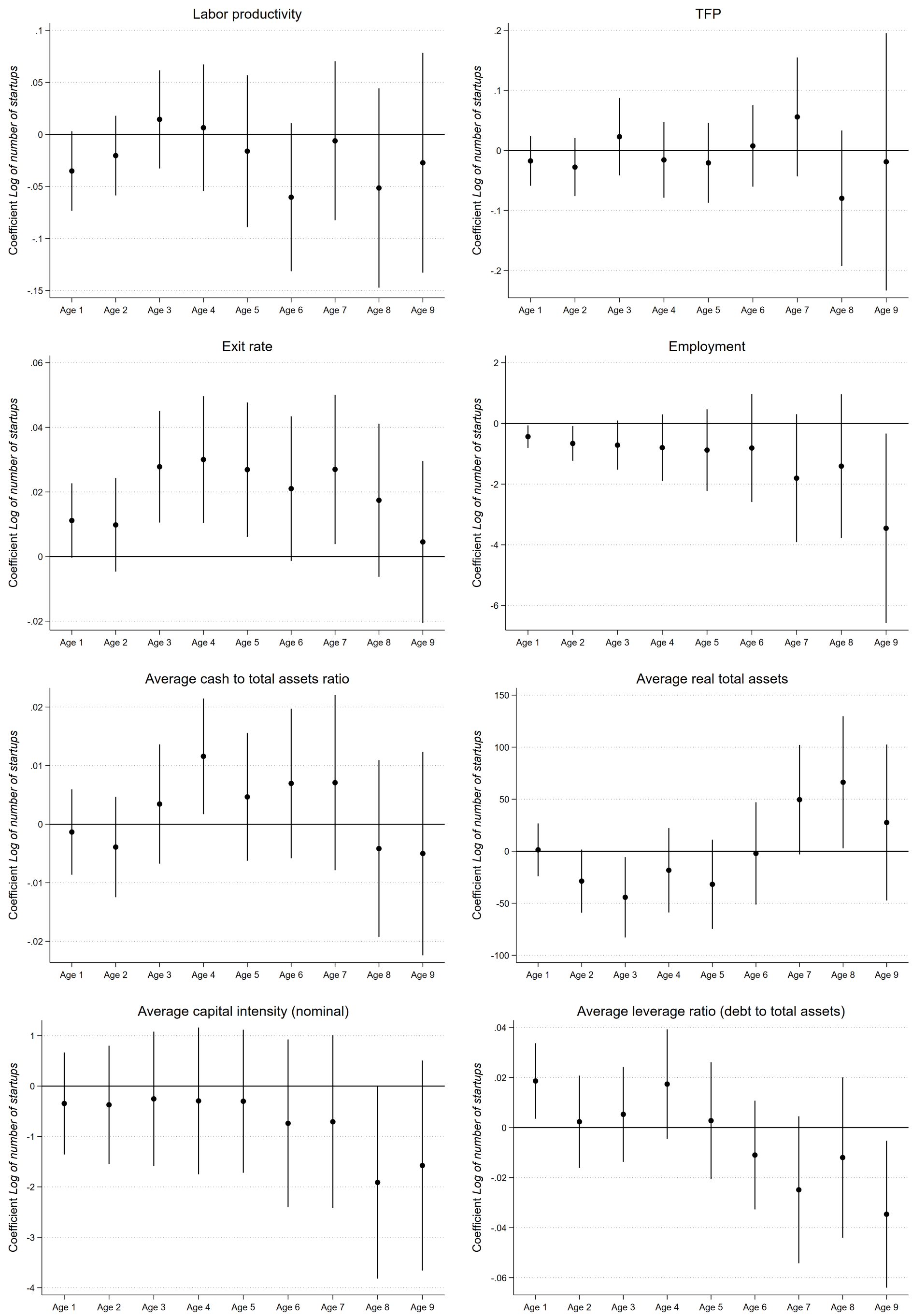
Notes: The panels in this figure summarize how the startup types develop during the first 12 years of their life in terms of the five clustering variables. Each panel corresponds to one clustering variable and plots the coefficients from 12 separate regressions where the dependent variable is this variable. Each regression is then run for an age group (age is 1, 2, ...12 years). For example, the first panel summarizes regressions in which the *Average cash to total assets* ratio is regressed on dummy variables for the startup types (the *Basic* type is omitted) as well as *Country* × *Cohort* and *Industry* fixed effects. The sample is the full panel data set at the one-digit industry level. Standard errors are clustered by the *Industry* × *Country* × *Cohort* and whiskers indicate 95 percent confidence intervals.

Figure A12. Entry selection: Cohort size and startup performance over time (2-digit sector)



Notes: This figure summarizes OLS regressions where the dependent variable is indicated in the figure titles. The independent variable of interest is the log number of startups in the cohort. *Exit rate* is the cumulative exit rate defined as 1 minus the ratio of number of startups that survived until t divided by the number of startups in *t-0*. Each dot corresponds to a partitioned regression for all startups of a particular age, as indicated on the x-axis, with interactive fixed effects for *Industry* × *Startup type* × *Country*; *Industry* × *Country* × *Cohort*; *Startup type* × *Country* × *Cohort*; and *Industry* × *Startup type* × *Cohort*. The sample is the full panel data set partitioned by age and at the two-digit industry level. Standard errors are clustered by the *Industry* × *Country* × *Cohort* and whiskers indicate 95 percent confidence intervals.

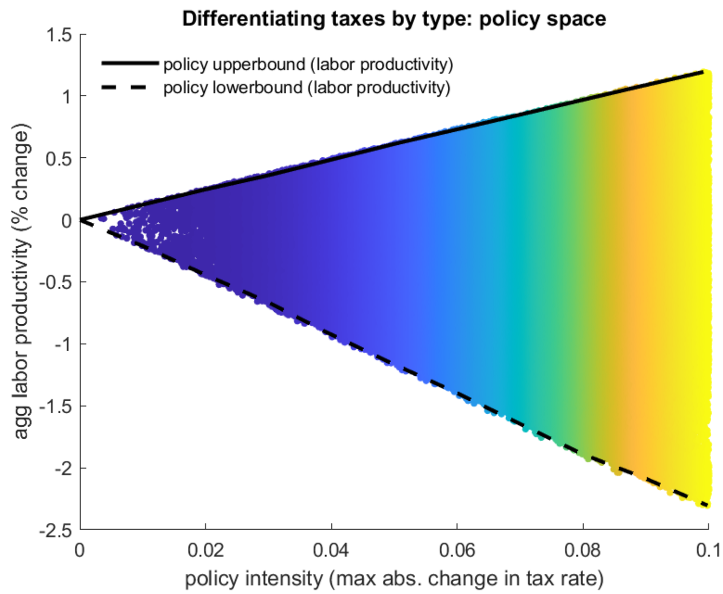
Figure A13. Entry selection: Cohort size and startup performance over time (1-digit sector)



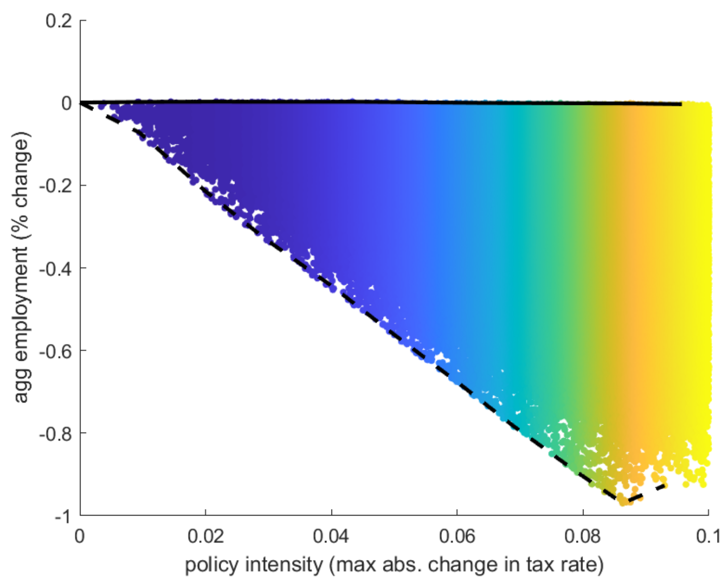
Notes: This figure summarizes OLS regressions where the dependent variable is indicated in the figure titles. The independent variable of interest is the log number of startups in the cohort. *Exit rate* is the cumulative exit rate defined as 1 minus the ratio of number of startups that survived until *t* divided by the number of startups in *t-0*. Each dot corresponds to a partitioned regression for all startups of a particular age, as indicated on the x-axis, with interactive fixed effects for *Industry* × *Startup type* × *Country*; *Industry* × *Country* × *Cohort*; *Startup type* × *Country* × *Cohort*; and *Industry* × *Startup type* × *Cohort*. The sample is the full panel data set partitioned by age and at the one-digit industry level. Standard errors are clustered by the *Industry* × *Country* × *Cohort* and whiskers indicate 95 percent confidence intervals.

Figure A14. Differentiation of corporate income tax: policy space (two-digit elasticities)

Panel A: Labor productivity



Panel B: Employment



Notes: This figure summarizes the policy experiment. Panel A shows the policy space for aggregate labor productivity. The horizontal axis measures the intensity of the potential policy change as the maximum absolute change in the tax rate. Warmer colors indicate stronger corporate tax rate differentiation. The solid line plots the "policy upper bound": the largest possible aggregate labor productivity increase given a certain policy intensity. Panel B plots the policy space for aggregate employment. Both panels use entry elasticities based on two-digit industry data.

Table A1. Sample composition by country

Country	Sample period	Number of startups
Croatia	2003-2019	64,760
Denmark	2002-2018	114,195
Finland	2000-2019	126,554
France	2005-2007	210,033
Italy	2007-2018	322,893
Lithuania	2001-2017	47,322
Netherlands	2008-2018	63,729
Slovenia	2006-2019	23,435
Spain	2009-2018	236,192
Sweden	2004-2019	136,376
Total		*1,345,489*

Notes: This table shows the sample composition of the full panel of startups just born.

Table A2. Variable definitions

Variable	Definition
Capital intensity	Average ratio of real capital stock to the number of employees
Cash ratio	Average cash to total assets ratio
Employment	Average number of employees
Exit probability	1 minus the ratio of number of firms that survived until t divided by the number of startups at t=0
Labor productivity	Logarithm of employment-weighted average labor productivity, where firm-level labor productivity is defined as real value-added divided by number of employees
Leverage ratio	Average ratio of debt to total assets
Profit margin	Average ratio of operating profit (Earnings Before Interest and Tax (EBIT)) to revenue
Real total assets	Average real total assets (thousands of euros)
TFP	Employment-weighted average total factor productivity based on a GMM estimation following Ackerberg, Caves and Frazer (2015) and assuming a Cobb–Douglas production function
Wage per employee	Average wage per employee

Notes: All monetary variables are PPP-adjusted.

Internet Appendix B: Model details

To rationalise the existence of the different startup types, as well as their performance differences, we now propose a model of heterogeneous firms. Specifically, we allow for heterogeneity in terms of (i) Total Factor Productivity, (ii) scalability with respect to capital and/or labor, and (iii) entry costs, fixed costs and initial assets. We introduce consider policy variables:

- A corporate income tax (possibly differentiated by types): τ_π
- A labor tax: τ_l
- A capital tax: τ_k
- A loan tax: τ_b

All of these may be negative (subsidies).

1. Environment.

The model period is one year and we denote firm age by a. We distinguish between three age categories: (i) potential entrants ($a = -1$), (ii) young firms ($0 \leq a < a^*$), (iii) old firms ($a \geq a^*$). Entrants and young firms are subject to real and financial rigidities, whereas old firms are unrestricted. Entrepreneurs have a discount factor $\beta \in (0, 1)$ and they can save outside of the firm in one-year bonds with a rate $r = \frac{1}{\beta} - 1$.

Technology. Let the production function of the firm be given by $y = A_a k^\alpha l^\gamma$, where TFP A_a varies positively with age (a) and is heterogeneous across firms. Moreover, α and γ are heterogeneous across firms (but constant over time). Capital depreciates at a rate $\delta \in (0, 1)$ per year. Production also requires the payment of an age-dependent fixed cost c_a. Concretely, we assume that the fixed cost may be higher in the initial year, reflecting firm establishment costs. Subsequently, the fixed cost remains constant. TFP evolves as:

$$A_a = z + z_a$$

where z is a firm-specific but time-invariant component, and z_a is an age-dependent component which is constant across firms of the same age (and known to all).

Entrants. Starting a firm requires the payment of an entry cost θ, which is heterogeneous across entrepreneurs. Any firm starts with zero capital, and some initial level of assets m_{-1}. A potential entrant observes its individual characteristics $\{z, \alpha, \gamma, \theta\}$ and decides whether to enter or not. If the firm enters, the firm must decide its capital level k which cannot be adjusted until becoming old (the firm pays an upkeep δk per period to compensate for depreciation). Upon entry, the firm can also take out a long-term loan with duration a^*, which is to be repaid in equal installments and comes with an interest rate $r_b(b) \geq \widetilde{r}$. We assume that $r_b(b)$ is weakly increasing in the loan amount b, and that $r_b(b) = \widetilde{r}$ when $b \leq 0$ (i.e. when the firm has positive savings). Here, $\widetilde{r} = \frac{1}{\sum_{a=0}^{a^*-1} \beta^{a+1}} - \frac{1}{a^*}$ is the "frictionless rate" for this kind of loan that is consistent with the subjective discount factor and the interest rate on savings outside the firm. In other words, whenever the firm borrows it pays a premium, which is increasing in the size of the loan and which captures financial frictions. No further loans can be taken out and the firm cannot default on a loan or pre-pay it. Concretely we assume that $r_b(b) = \widetilde{r} + (\zeta + \tau_b)\, b$ for $b \geq 0$, where $\zeta + \tau_b > 0$.

Young firms. Having made the initial decisions, the firm enters at age $a = 0$. As mentioned above, until age a^* the firm cannot re-optimise its capital and loan decisions. However, the firm can freely adjust its employment, l, in every period.

Old firms. At age $a = a^*$ a firm becomes full-grown. The firm's TFP level remains constant from then onwards. Moreover, the firm becomes financially unconstrained (e.g. it can raise equity) and can freely adjust its capital stock.

2. Decision problems.

Let us now formalize the firm's decision problems in the various phases. We will solve the model backwards and start with the full-grown firms.

Old firms ($a \geq a^*$). The old firm's decision problem is equivalent to maximizing economic profits, which can be expressed as $\pi = (1 - \tau_\pi)\left(Ak^\alpha l^\gamma - (r + \delta + \tau_k)\, k - w(1 + \tau_l)l - c_{a \geq 0}\right)$.

$$r + \delta + \tau_k = \alpha \frac{y}{k},$$
$$w\,(1 + \tau_l) = \gamma \frac{y}{l}.$$

We will denote the profits of a full-grown firm by $\pi^*(z, \alpha, \gamma)$. Old firms exit exogenously at a rate ρ.

Young firms ($a = 0, 1, .., a^* - 1$). Now consider the firms in the "scale-up" phase, which lasts for a^* years. The firm does not exit during this phase. Recall that in this stage the firm can only adjust labor inputs, denoted l_a. It does so to statically maximise profits, given by:

$$\pi_a = (1-\tau_\pi)\left(A_a k^\alpha l_a^\gamma - w(1+\tau_l)l_a - (\delta+\tau_k)\,k - c_a\right) - \left(\frac{1}{a^*} + r_b(b)\right) b,$$

where $\frac{b}{a^*}$ captures repayment of the loan. The profit-maximizing choice of labor is again given by:

$$l_a = \left(\frac{w\,(1+\tau_l)}{\gamma A_a} k^{-\alpha}\right)^{\frac{1}{\gamma-1}}.$$

Taken together, the above two equations define a profit function for young firms, which we denote $\pi_a(A_a, k, b, w)$. The firm's short-term asset positions evolve as:

$$m_a = (1+r)m_{a-1} + \pi_a$$

where $a = 1, 2, .., a^*$ and initial assets satisfying $m_0 = m_{-1} + b - k + \pi_0$.

We define financial variables as follows. Let $\mathbb{I}_m$ be an indicator function such that $\mathbb{I}_m = 1$ if $m_a > 0$ (the firm has positive cash) and $\mathbb{I}_m = 0$ otherwise (the firm is in short-term debt). We measure cash as $cash_a = \mathbb{I}_m m_a$ and short-term debt as $debt_a^{ST} = -(1-\mathbb{I}_m)m_a$. Moreover, long-term debt is given by $debt_a^{LT} = b_a(1-\frac{a}{a^*})$. Accordingly, we measure the leverage ratio as $\frac{debt_a^{ST}+debt_a^{LT}}{k+cash_a}$ and the cash ratio as $\frac{cash_a}{k+cash_a}$.

Entrants. Immediately after entry, a firm must take its decisions on the capital stock and the loan, which cannot be re-optimised until age a^*. The value of an entrant firm is given by:

$$V(z,\alpha,\gamma,m_{-1}) = \max_{b,k} \sum_{a=0}^{a^*-1} \beta^a \pi_a(A_a,k,b) + \beta^{a^*}\left(\frac{\pi^*(z,\alpha,\gamma)}{1-\beta\,(1-\rho)} + k\right) + m_{-1} + b - k$$

subject to the following financing constraint:

$$k + c_0 \leq m_{-1} + b.$$

The constraint states that the firm must be able to finance the initial capital stock plus the initial fixed cost in production. The financing constraint is slack when $b \leq 0$.[43]

We can solve for the optimal choices from the first-order conditions, conditional upon the

[43] In this case, the firm is at the margin indifferent between holding funds within the firm or outside the firm, given that the interest rate is the same, and there is no implication for the firm's revenues when the firm is financially unconstrained.

firm being in debt or not. A firm enters if and only if

$$V(z, \alpha, \gamma, k, m_{-1}) > m_{-1} + \theta.$$

We can now solve the problem as follows:

1. Solve for k from the following first-order condition, which holds conditional on the firm not being financially constrained ($b \leq 0$):

$$1 - \beta^{a^*} = \sum_{a=0}^{a^*-1} \beta^a \left((1 - \tau_\pi) \left(\alpha A_a k^{\alpha-1} l_a^\gamma - \delta - \tau_k \right) \right)$$

where $l_a = \left(\frac{w(1+\tau_l)}{\gamma A_a} k^{-\alpha} \right)^{\frac{1}{\gamma - 1}}$.

2. If for this solution $b = k + c_0 - m_{-1} > 0$, then instead solve for k from the following first-order condition, which holds conditional on the firm being in debt ($b > 0$).

$$\begin{aligned} - \beta^{a^*} = \sum_{a=0}^{a^*-1} \beta^a \Big(&(1 - \tau_\pi) \left(\alpha A_a k^{\alpha-1} l_a^\gamma - \delta - \tau_k \right) \\ &- \left(\frac{1}{a^*} + r_b(k + c_0 - m_{-1}) \right) - \frac{\partial r_b(k + c_0 - m_{-1})}{\partial b} \cdot (k + c_0 - m_{-1}) \Big), \end{aligned}$$

where again $l_a = \left(\frac{w(1+\tau_l)}{\gamma A_a} k^{-\alpha} \right)^{\frac{1}{\gamma - 1}}$.

3. Compute $b = k + c_0 - m_{-1}$ and compute profits at any age.

4. Compute the value of an entrant. The firm enters if and only if $V(z, \alpha, \gamma, m_{-1}) > m_{-1} + \theta$.

3. Equilibrium.

We now add a household sector to the model, allowing us to solve for the general equilibrium. The household has a subjective discount factor β, so that in the steady state $r = \frac{1}{\beta} - 1$ is given as above. We assume for tractability that households have GHH preferences over leisure and consumption. Optimal labor supply then takes the form:

$$l = \kappa_0 w^{\kappa_1}$$

where κ_1 is the labor supply elasticity, which can be calibrated externally, and κ_0 is a scaling parameter. The labor market clears iff:

$$l = \sum_j \left(\sum_{a=0}^{a^*-1} l_{a,j} + l_{*,j} \right)$$

while the asset and capital market clear trivially. We consider budget neutral policy, imposing the following budget constraint:

$$\sum_j \tau_{l,j} \left(\sum_{a=0}^{a^*-1} l_{a,j} + l_{*,j} \right) + \sum_j \tau_{k,j} \left(\sum_{a=0}^{a^*-1} k_{a,j} + k_{*,j} \right) + \sum_j \tau_{\pi,j} \left(\sum_{a=0}^{a^*-1} \pi_{a,j} + \pi_{*,j} \right) + a^* \sum_j \tau_{b,j} \mathbb{I}_b b_j = 0,$$

where $\mathbb{I}_b$ is an indicator which equals one if and only if $b > 0$. In the policy exercises we jointly solve for w and one of the policy instruments to make the labor market clear and the government budget constraint hold. Asset markets clear trivially.