Firm Cyclicality and Financial Frictions*

Alex Clymo† Filip Rozsypal‡

April 20, 2025

Abstract

How firms respond to shocks is informative about the frictions they face. Using administrative micro data we document heterogeneity in cyclicality with respect to size and age: for the youngest firms, small firms are more cyclical than large firms, but the reverse is true among older firms. Two mechanisms acting at different parts of the joint size-age firm can account for this heterogeneity: finance and heterogeneous returns to scale. Micro-level evidence supports the presence of these two mechanisms. We show that larger firms have higher returns to scale, compare the effect of leverage and returns to scale on cyclicality, and show that financial situation at entry affects firm survival. In a quantitative heterogeneous firm model, the existence of unconstrained high returns to scale firms fundamentally alters the propagation of financial frictions: Financial recessions are milder, but feature more pronounced missing generation effects.

Keywords: firm age, firm size, cyclicality, financial frictions, returns to scale
JEL codes: D22, E32, G32, L25

*The viewpoints and conclusions stated in this paper are the responsibility of the individual contributors, and do not necessarily reflect the views of Danmarks Nationalbank or Statistics Denmark. We thank Gian Luca Clementi, Michaela Elfsbacka Schmöller, Simon Gilchrist, Saleem Bahaj, Vasco Carvalho, Dino Palazzo, Tobias Renkin and Johannes Poeschl. as well as the participants of the Firm Heterogeneity 2021 workshop and 2nd Research ESCB Cluster conference for their helpful comments and feedback. Karl Bertelsen Robak provided outstanding research assistance.

†Department of Economics, University of Essex, and Paris School of Economics. Essex: Wivenhoe Park, Colchester CO4 3SQ, United Kingdom. Email: `a.clymo@essex.ac.uk`

‡Research Unit, Danmarks Nationalbank, Langelinie Allé 47, 2100 København Ø, Denmark. Email: `firo@nationalbanken.dk`

1 Introduction

We document systematic and significant differences in how firms of different ages and sizes react to the business cycle in administrative micro data. Why are certain firms more sensitive than others? Do size and age just act as a proxy for financial frictions which amplify responses to shocks, as suggested by Gertler and Gilchrist (1994)? In this paper we propose that the patterns in cyclicality can be explained by a combination of financial frictions and heterogeneous returns to scale, support this hypothesis by additional empirical evidence, and show that the interaction of returns to scale (henceforth, RTS) heterogeneity and financial frictions substantially affects the steady state and propagation of shocks using a quantitative heterogeneous firm model.

We use firm-level administrative and balance sheet data on the universe of Danish firms to measure cyclicality. We extend the methodology of Crouzet and Mehrotra (2020) by focusing on the differences across age as well as size, and measure cyclicality by using the co-movement of firm-level employment and sales with aggregate output. We find that among young firms cyclicality decreases with size, while it increases with size for older firms. We propose that two basic forces shape cyclicality across the age-size distribution. Young firms are more cyclical due to the presence of financial frictions, which bind more for the young. Among old firms, large firms are more cyclical if their size is driven by high returns to scale, which makes them more responsive to shocks.

To test the influence of returns to scale and finance on cyclicality we proceed in two steps. First, we directly estimate returns to scale (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Gandhi et al., 2020) and document a positive correlation with firm size: firms with 0-5 employees have RTS of 0.75 on average, while those with 120+ employees have essentially constant RTS, and this pattern is present even if we look only at currently young firms and measure size using future size. Second, using balance sheet data, we show that younger firms are more leveraged. Combining both estimated RTS and financial data in a joint regression we show that high leverage firms are more cyclical, as are high returns to scale firms. Furthermore, the effect of leverage on cyclicality is smaller among old firms, consistent with finance being more important for young firms. At the same time, RTS play a more important role for old than young firms.

Despite most output being created by large and old firms that are the least likely to be financially constrained, financial frictions can still be important, both because of general equilibrium effects and because they might affect which firms manage to outgrow their constraints. Exploiting the panel dimension of our data, we show that an important fraction of firms that grow to be large in old age were born small. If financial frictions slow down the growth of these firms while they are young, or stop unlucky entrants from reaching their potential, financial frictions could be very costly. We document wide heterogeneity in the financial situation of entrant firms in the data and show that finance is indeed important for firm survival: exit rates are persistently higher along the firm life-cycle the lower the initial net worth of a firm at entry. Since large firms have high returns to scale, this mechanism naturally suggests potential interactions between financial frictions and returns to scale over the life-cycle.

To better understand the interaction of heterogeneity in returns to scale and financial frictions, we build a quantitative model in the style of Khan and Thomas (2013), where the financial frictions take the form of a collateral constraint. The key ingredient of the model is to introduce permanent heterogeneity in returns to scale, where large firms are large due to high returns to scale and not necessarily due to high productivity, as is usually assumed. We carefully calibrate the model to our micro-data, including on the net worth of entrant firms and relationship between RTS and size. We validate the model by showing that it is able to replicate our results on the cyclicality by firm size and age. Motivated by this and our direct micro evidence, the remainder of the paper uses the model to investigate how heterogeneity in RTS alters the costs of financial frictions in this class of models. We highlight three key channels.

First, the composition of firms that are financially constrained changes. In the class of models following Khan and Thomas (2013), firms are born poor, and hence financially constrained, and grow out of these financial constraints as they age. Firms with higher returns to scale operate with lower profit margins and therefore need longer to outgrow their collateral constraints. In contrast, firms with low returns to scale reach their optimal steady state size much faster.[1] Outgrowing the financial constraints is thus the largest problem for the firms that would benefit the most from their relaxation.

Second, the presence of even a small number of old, financially unconstrained firms with high RTS can greatly dampen the aggregate output cost of financial frictions, while

[1]Specifically, these statements are conditional on a given starting size of firms, relative to their optimal size, and are not due to the fact that high RTS firms are larger. Consider two firms with the same optimal size, one with high RTS and low productivity and one with low RTS but high productivity. Conditional on the same starting size the firm with higher RTS grows more slowly than a firm with lower RTS.

also making the constrained firms, including entrants, worse off. Tightened borrowing constraints, either in steady state or as a temporary financial shock, lead to slower growth of financially constrained firms and reduced firm entry, putting downward pressure on wages. In the model with heterogeneous RTS, the high RTS unconstrained firms respond strongly to the fall in wages, and "crowd in", decreasing the fall in aggregate output. This effect is present but severely muted in standard calibrations where all firms have the same RTS, whereas here the largest firms are not only responsible for most of the aggregate output, but are also the most responsive among the unconstrained firms. A model with heterogeneous RTS can be dominated by the behavior of a handful of high RTS firms and, in some respects, behave more like a model with constant RTS than a model where all firms have a common decreasing RTS.[2]

Third, heterogeneous RTS amplifies "missing generation" effects following a financial crisis. Firm entry declines more and more persistently than in the standard model, meaning that heterogeneous RTS amplifies the firm-level responses of both young (constrained) and old (unconstrained) firms. This is due to a general equilibrium feedback from old to young firms: the closer to constant RTS the old large unconstrained firms are, the smaller the factor price declines they need to induce them to absorb the resources shed by constrained firms. While this dampens the output contraction following a financial shock, it also worsens the situation of new entrants who suffer most from the shock and do not get the benefit of the larger factor price fall induced in the standard model. Putting these channels together, the larger reallocation from young to old firms in our model leads to a more persistent aggregate productivity decline following a financial recession.

Related literature. We connect to three broad strands of literature. Firstly, to the firm dynamics literature studying firm age and size over the business cycle. Different papers find different, sometimes conflicting, results, which is partly driven by different samples of firms available in a given dataset. Gertler and Gilchrist (1994) investigate the cyclicality of small versus large firms and find that small firms are more sensitive to periods of credit market tightening than large firms. Khan and Thomas (2013), building on insights of Bernanke and Gertler (1989); Kiyotaki and Moore (1997); Bernanke et al. (1999) and Jer-

[2]Our model also has small unconstrained firms with lower RTS than the standard calibration, who are less responsive. The high responsiveness of large, high RTS firms dominates, and heterogeneity in RTS amplifies the crowding in effect. Since this channel operates through general equilibrium it complements the finding of Winberry (2021) that the responsiveness of factor prices to shocks can change the predictions of heterogeneous firm models.

mann and Quadrini (2012), show that small firms contracted more than large firms during the financial crisis, and Gavazza et al. (2018) show that the vacancy yield was more cyclical at small than large firms during this same period. On the other hand, Moscarini and Postel-Vinay (2012) find that larger firms (in terms of number of employees) are more cyclical, when aggregate conditions are measured using the (HP-filtered) level of the unemployment rate. Similarly, Mian and Sufi (2014) show that larger establishments contracted more in areas with larger declines in house prices. More recently, it has been shown that firm age is a more important predictor of both the average level and cyclicality of firm growth than firm size (see Fort et al. (2013); Haltiwanger et al. (2013), for evidence from the US). Fort et al. (2013) also study cyclicality by joint age-size bin. They find that young firms are more cyclical than old firms, and that this difference is much more important than the differential between small and large firms. They use state-level house price data to argue that financial frictions may drive this result,[3] while we directly use firm-level financial data. Sedláček and Sterk (2017) highlight the role of high growth "gazelle" firms over the cycle, and we argue that one cost of financial frictions is to interfere with the growth of such firms.

Secondly, we connect to the empirical literature investigating how finance affects firm cyclicality. At the aggregate level, the recoveries from financial crises tend to be particularly slow (Reinhart and Rogoff, 2014; Sufi and Taylor, 2022), and so understanding the firm-level causes of this persistence is particularly important. Due to data limitations, much of the knowledge about cyclicality and firm finance is based on large publicly traded firms. Sharpe (1994) uses Compustat data to document that high-leverage firms are more cyclical than low-leverage firms. Giroud and Mueller (2017) combine Compustat data with establishment-level employment data to show that the decline in house prices during the Great Recession affected higher leverage firms more strongly. Conversely, Ottonello and Winberry (2020) use Compustat data and find that firms with low default risk, including those with low debt burdens, are the most responsive to monetary shocks. Cloyne et al. (2023) use Compustat data to show that younger, non-dividend paying firms exhibit the largest changes in investment following monetary policy shocks. Publicly traded firms are

[3] Another strand of literature examines the cyclicality of firm financing, both in terms of empirics and also model building. For example, see Jermann and Quadrini (2012) (investigate the cyclicality of debt and equity issuance), Covas and Haan (2011) (the cyclicality of financing is different across firms of different sizes, with the procyclicality of equity issuance decreasing monotonically with firm size), Crouzet (2017) (the choice of bank and bond financing), Begenau and Salomao (2018) (firm size and debt/equity cyclicality), Jensen et al. (2017) (size and cyclicality of financing and probability of default), Nikolov et al. (2018) (size and source of financial constraints), Poeschl (2023) (size and cyclicality of debt maturity), Drechsel (2023) (earnings-based borrowing constraints), or Casiraghi et al. (2021) (entrepreneur-banker relationship building).

only a small subset[4] and as such are not representative of the whole firm population, and we complement this literature by using financial data even for young, small, unlisted firms.

More recently, new, often administrative, datasets allow going further than publicly listed firms to achieve wider firm coverage.[5] An early example is Crouzet and Mehrotra (2020), who find that only the largest firms (99th percentile and above measured by assets) are less cyclical than the rest, which speaks against financial frictions. We define size using employment, and extend their empirical specification by including the interactions of size and age. We find strong effects of size on cyclicality, which just happen to have opposite signs for young and old firms and hence offset each other in the aggregate, highlighting the importance of studying firm age and size together over the business cycle. Dinlersoz et al. (2024) merge balance sheet data from Compustat and Orbis into the US Longitudinal Business Database (LBD) to study finance for both private and public firms. They find results highly complementary to our own, including that highly levered young-small were particularly affected during the financial crisis. We differ from their paper by additionally studying heterogeneous RTS and its interactions with finance, both empirically and theoretically.

Finally, we relate to the previously-cited literature on production function estimation, and the consequences of heterogeneity in RTS. Gao and Kehrig (2021) estimate RTS at the industry level and show that industries with larger average firm size have higher RTS. Smirnyagin (2023) also estimates RTS at the industry level, and shows that the entry rate of high RTS firms is more pro-cyclical. Relative to these papers, we estimate RTS at the firm level and document that larger firms have higher RTS, an approach also followed by Hubmer et al. (2024), and develop new macroeconomic insights.[6]

[4]In the US there are around 4,000 publicly traded firms (Gupta et al., 2021) in the population of over 5 million firms. Further recent contributions using Compustat include Duygan-Bump et al. (2015), Jungherr et al. (2022), Jeenas (2019), and Grob and Züllig (2024). Another literature, exemplified by Chodorow-Reich (2014), uses match firm-bank relationships to identify financial shocks.

[5]For recent papers going beyond public data see Alder et al. (2023) (France), Bahaj et al. (2022) (UK), Cao et al. (2024) (Norway), Castillo-Martinez and Bornstein (2024) (Orbis, Europe), and Ferreira et al. (2023) (Portugal).

[6]Gavazza et al. (2018) is an early contribution leveraging RTS heterogeneity to generate firm size differences in a heterogeneous firm model, which also includes financial shocks. Relative to their paper we provide direct evidence of RTS heterogeneity and emphasize general equilibrium effects from wages, interest rates, and capital prices, while they build a search and matching model and emphasize general equilibrium effects from labor market slack and recruitment effort. Smirnyagin (2023) and Hubmer et al. (2024) also build models investigating the interactions between financial frictions and RTS. The former studies business cycles and how fewer high RTS firms enter in recessions, also employing a Khan and Thomas (2013) style model. RTS are calibrated using the firm age distribution, while we directly calibrate RTS from our estimates. The latter studies steady states and how heterogeneous RTS amplifies the costs of financial frictions in a model

The rest of this paper is organised as follows. In Section 2 we discuss the data and present our empirical results about cyclicality. In Section 3 we provide a simple analytical framework to build intuitions on the interactions between RTS and finance. In Section 4 we return to the data to provide supporting evidence on heterogeneity in RTS and the role of financial frictions. Section 5 develops our quantitative model. Finally, in Section 6 we conclude and discuss the implications of our results for future research.

2 Measuring cyclicality in the data

2.1 Data

Our dataset covers firms in Denmark between 2001 and 2019 at an annual frequency. It contains the universe of of Danish firms across all sizes and ages. In order to analyse firm outcomes and financial balance sheet data together, we merge two datasets ("data registers") provided by Statistics Denmark (DST): the FIRE dataset ("Regnskabsstatistikken"), which broadly contains data on accounting variables, is merged with the FIRM dataset ("Firmastatistik"), containing data regarding economic, employment and accounting information at company level. The quality of this data is generally believed to be very high, as Statistics Denmark is a government agency, and most of the variables we use are originally collected by Denmark's tax authority, SKAT.[7] Additionally, DST also runs independent checks on the datasets. Individual firms are identified by a unique number that is generated at the time of registration. The merging of the datasets is done using this identifier, and thus provides exact matches. More information on data itself and the cleaning process is provided in Appendix A.1.

Our cleaned dataset is an unbalanced panel capturing employer firms in Denmark. Our baseline sample is firm-year observations containing both valid accounting (e.g. sales, employment) and balance sheet (e.g. debt, assets) data. Due to the availability of balance sheet data our baseline sample starts in 2001 and so runs from 2001 to 2019 (effectively 2002 to 2019, since our main empirical specifications use growth rates and require the existence of lagged data). Moreover, while balance sheet data is available for firms of all sizes, small firms are sampled less than large firms, making our sample stratified. Nonetheless,

of entrepreneurial choice. We study both steady states and business cycle dynamics, and i) demonstrate how heterogeneous RTS allows the model to match our new cyclicality results along the size distribution, and ii) emphasise the role of unconstrained high RTS firms in crowding in following a financial shock, making factor prices less responsive, and hence amplifying missing generation effects.

[7] Sales, assets, liabilities, investment and information about employment based on payroll.

we still observe firms of all sizes and ages, both publicly listed and privately owned, with positive probability in our dataset.[8] This makes our dataset uniquely suited to studying the role of financial frictions across the whole distribution of firms, especially at younger and smaller firms that are not featured in datasets like Compustat. We exclude banks and other financial corporations (not to distort financial variables) as well as non-profit, charities and government controlled companies such as public hospitals. Our baseline dataset contains roughly 2 million firm-year observations. In terms terms of employment, it covers between 1.5M to close to 2.2M workers every year, which corresponds to 90-100% private employment.[9]

Given the start and end date of the underlying registers one might reasonably worry whether our results might be overweighting the role of finance due to the financial crisis. While it would be theoretically possible to extend our sample by using alternative datasets that cover different time periods, we believe that there was enough other variation in the Danish business cycle that other shocks are also well represented. According to the OECD[10], our sample covers the following business cycle turning points: troughs in 2003M7, 2009M7, and 2014M4, and peaks in 2006M7, 2011M4, and 2019M6. Denmark thus experienced at least three recessions in our sample: the early 2000s recession, the global financial crisis and the Eurocrisis. This means that while certainly important, the financial crisis is not the only recession in our dataset driving variation in aggregate GDP.

Key variables. To measure firm how firms react to the business cycles, we focus on sales and employment. Sales ("Omsætning") are based either on balance sheet information or on VAT declarations. Employment is measured in hours scalled by annual full time (roughly 1900 hours per year). Our measure of debt contains both short and long-term liabilities. Specifically, beyond short and long-term debt ("Anden langfristet gæld" and "Andenkortfristet gæld"), our measure of debt also includes provisions ("Hensættelser") — unknown obligations such as deferred tax or pension obligations— and long (maturity beyond 1 year) and short-term debt to suppliers ("Langfristet/Kortfristet gæld til leverandører").[11]

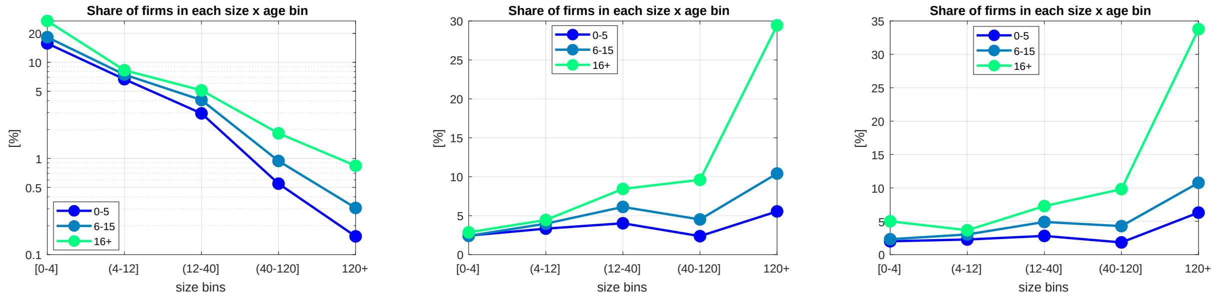
[8]The stratified sampling reduces the frequency of observations for small firms, but this should just reduce the power of our empirical results, rather than altering the point estimates we find, for this group of firms. This is supported by the fact that our main cyclicality results for employment and sales do not change when only including firm-year observations with balance sheet data. See Figure 22 in Appendix B.2.2.

[9]We get higher coverage in the secnd half of the sample, see Figure 18 in Appendix A.2.

[10]See OECD turning points.

[11]The inclusion of these nontraditional liabilities in the definition of debt is the likely culprit behind the leverage being higher than what is usually reported for firms in other countries.

Figure 1: Share of firms across age and size bins



(a) number of firms (b) employment (c) sales

Note: Fraction of observations in each joint age-size bin. Lines correspond to age bins and x-axis to size bins. Panel (a), number of firms, uses a logarithmic scale.

For assets, we consider a combination of intangible ("Immaterielle anlægsaktiver"), tangible ("Materielle anlægsaktiver") and financial assets ("finansielle anlægsaktiver"). Financial assets are both short term (such as cash) and long term.

When estimating firms' production functions, we use the expenditures on intermediates from FIRE, sectoral level PPI and dissagregated input-output tables.[12] We additionally use data on a firm's sector of operation so the results are not driven by differences between sectors. We use DB07 classification at 36 sector grouping to control for sectoral trends and fluctuations and the highest level of dissagration available when it comes to computing the sectoral price indeces for purposes of RTS estimation.[13]

Size and age groups. We define firm size by its lagged employment (headcount). The firm size measure thus changes as the firm grows or shrinks as it ages and is hit by shocks. We sort firms into bins based on five headcount thresholds (0-4], (4-12], (12-40], (40-120] and (120+) of size across the population of firms active that year. At the same time, we also sort firms into three age groups: 0-5 (the young), 6-15 (the mature), and 15+ years (the old). Firm age is measured in our data from the moment the firm is registered.[14] This notion of age is thus the true age since foundation of the firm, which distinguishes us from other datasets which can only measure age since, for example, the firm was publicly listed on stock markets.

[12]We extract prices and quantities of intermediates from the expenditures by combining sectoral input-output matrices with sectoral price indices, see Appendix C.2.2.

[13]For details, see DST website.

[14]Given that it takes very little time to start a new firm in Denmark, there is not a large need to formally register the firm long before the firm becomes economically active. Nonetheless, to align with the notion of an economically active firm in standard models, we assign firms age 0 and include them in our dataset only from the first year that they register positive employment or sales.

The firm shares are captured in Figure 1. Due to the well-known skew in the firm size distribution, a large share of total employment is done by a small number of large firms. For example, the largest firm size bin consists of roughly 1.3% of the firm population, out of which 0.84% belong to the oldest, 0.3% to the middle-aged and 0.16% to the youngest age group. At the same time, these firms contribute 45% of aggregate employment and 51% of aggregate sales. Despite being smaller, young firms also contribute disproportionally to job growth (older firms are on average shrinking), see Figure 20 in Appendix A.2. The number of firms within each age group changes over time (see Figure 17) and cyclical fluctuations in entry create swings in cohort size that propagate over the age distribution. More on the moments of the data and the timeseries can be seen in Appendix A.2.

Table 1: Averages of variables of interest by age and size

	Age groups			Size groups				
	0-5	6-15	16+	[0-4]	(4-12]	(12-40]	(40-120]	120+
N	575	685	947	1365	497	268	73	29
Employment	7.8	11.1	17.8	1.9	5.6	16.4	53.0	375.3
	536	580	724	987	496	268	73	29
Sales	18765	26990	47712	5321	12442	39259	144040	1109402
	470	543	723	1022	420	223	64	27
Assets	21691	34200	83217	12386	20517	32113	136890	1289130
	374	404	462	634	332	199	59	26
Debt	12283	18435	43428	5957	10517	18154	72493	696463
	372	401	457	630	329	197	58	25
Equity	6688	11501	29885	4596	8493	14156	57368	554581
	464	511	613	886	396	233	64	25
Equity< 0	0.3	0.2	0.3	0.4	0.1	0.1	0.1	0.0
	372	401	457	630	329	197	58	25
D/A	0.43	0.43	0.36	0.29	0.48	0.65	0.79	0.86
	575	685	947	1365	497	268	73	29

Note: "N" shows the total number of firm observations in thousands. Sales, assets, balance sheet debt, and net worth (assets-debt) in thousands of DKK,[15] negative Net worth in percent. Not all variables are reported every year, the grey numbers show the number of non-missing observations in thousands. Reported numbers are the average values within a bin. Debt/assets (D/A = leverage) and net worth winsorized at 99.5th percentile. Sample includes only incumbent firms that do not exit in the current period. The fraction of firms with negative equity reported as percentage.

2.2 Estimation framework

To study the intricate interplay between firm size and age we allow for interactions between size and age bins. Therefore, the effect of being old, for example, is allowed to be different for small and large firms, extending the framework of Crouzet and Mehrotra (2020). Using the definition of groups from the previous section, we run various regressions with a set of dummies controlling for the interaction of size and age:

$$x_{i,t} = \sum_j \sum_k (\alpha_{j,k} + \beta_{j,k} y_t) \mathbb{1}_{i \in I_t^j} \mathbb{1}_{i \in A(k)} + \sum_l (\gamma_l + \delta_l y_t) \mathbb{1}_{i \in S(l)}, \tag{1}$$

where $x_{i,t}$ is a variable of interest, such as the level or growth rate of employment at firm i.[16]. GDP growth rate $y_t \equiv \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$ is collected from the DST National accounts. The indices j, k, and l index firm size bins, firm age bins, and firm sectors respectively.[17] $\mathbb{1}_{i \in I_t^j}$ is an indicator variable for firm i being in size group j at time t (and similarly $\mathbb{1}_{i \in A(k)}$ for age and $\mathbb{1}_{i \in S(l)}$ for sector).

Depending on the variable of interest, we focus on the level coefficients (α) or the cyclicality coefficients (β). Since we include sector controls, these capture the within-sector average level and cyclicality of each age-size group respectively.[18] We take the old-large firm bin as the base group, so confidence intervals capture the null that a given age-size bin coefficient is equal to that of old-large firms. We present most results graphically, plotting coefficients across size bins and grouping the age bins into a lines of age-specific colour. In plots we add a common shifter coefficient that captures the unconditional cyclicality of the base group, indicated by the horizontal black line. The regression table can be found in Appendix B.2.1. Our baseline results exclude firms that are either entering or exiting at the given period, and we investigate entry and exit in later results.

2.3 Firm cyclicality

In Figure 2 we present our baseline cyclicality results from (1) for firm-level employment and sales growth, focusing only on non-exiting incumbents.[19] Cyclicality of both employment and sales follows a similar pattern: average cyclicality is negatively correlated across size bins for firms in the youngest age (0-5) bin, but it increases across with size for firms in the oldest age (16+) bin. Accordingly, young firms are typically more cyclical than old firms

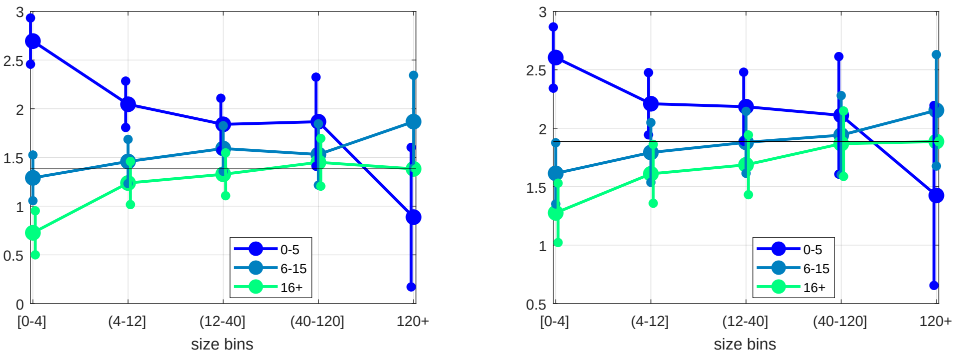
[16]When computing firm-level growth rates of we use the normalized growth rate $x_{i,t} = \frac{X_{i,t} - X_{i,t-1}}{0.5 * (X_{i,t} + X_{i,t})}$. See Haltiwanger et al. (2013) for a discussion.

[17]We use the Danish 36 sector industrial classification DB07, based on NACE rev.2

[18]Sector controls are important since sectors can be differentially exposed to the cycle (Abraham and Katz, 1986) which would bias coefficients if the sectors also differ in their composition across age and size. A similar logic suggests that the level coefficients should estimated in a regression controlling for the cycle if the age-size composition of firms changes over the cycle.

[19]For the results that include entry and exit, see Appenix B.2.5, for alternative size definitions see Figure 23

Figure 2: Cyclicality



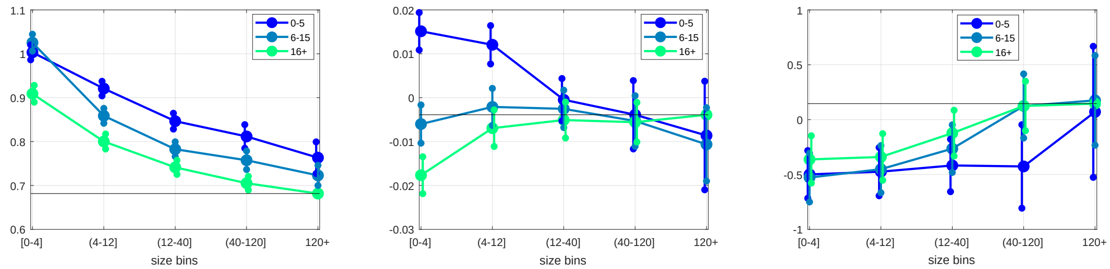
(a) Employment (b) Sales

Note: This figure present cyclicality coefficients $\beta_{j,k}$ shifted as described in Section 2.2. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms. The underlying regression table is available in Table 2 in Appendix B.2.

even conditioning on size, but this gap disappears for sufficiently large firms. In the rest of this paper we shorten this and similar statements about average cyclicality across bins to "cyclicality decreases with size for young firms and increases with size for old firms", which we use without any causal interpretation. The results for the middle 6-15 workers bin are located mostly in between the two extreme age groups, but relatively closer to the oldest firms. The difference between cyclicality of young and old firms is statistically significant for the first three size bins for employment and two bins for sales, and this general pattern is robust to different specifications. The confidence intervals become very wide for cyclicality for the largest firms, especially for the young, which is a direct consequence of much smaller number of firms that belong to these bins.

To place our results in context with the literature, it is interesting to compare our full joint age-size cyclicality results to simpler specifications which only investigate the role of size or age on cyclicality independently. We run these specifications, and present the results in Figure 24, Appendix B.2.4. If we regress age and size without the interaction, we find that younger firms are more cyclical than old firms (which is consistent with the general view of the literature) and larger firms are more cyclical than small, but the differences are not statistically significant within age groups. If we regress size alone, we do not observe any statistically significant pattern, consistent with the mixed results in the literature. Including the joint age-size interactions provides a natural explanation of this result: size alone does not predict cyclicality particularly strongly, since the relationship between size

Figure 3: Average levels, growth rates and cyclicality of leverage



(a) level (b) growth rate (c) cyclicality (growth rate)

Note: Leverage is defined as the ration of Debt to Assets (DA) and it is winsorised at 99.5%. "Level" panel (a) plots $\alpha_{j,k}$ coefficients respectively from regression (1) with the *level of leverage* as the left-had side variable. "Growth rate" panel (b) and "cyclicality" panel (c) show $\alpha_{j,k}$ and β_{lk} coefficients from regression (1) with the *growth rate* of leverage.

and cyclicality has opposite signs for young and old firms, which roughly cancel out on average. This finding could help reconcile the conflicting results about cyclicality by firm size discussed in the introduction.

2.4 Levels, growth rate and cyclicality of firm finance variables

In this section we turn to financial variables, analyzing the level, growth, and cyclicality of firm-level leverage, defined as debt over assets,[20] by joint age-size bin. We estimate (1) for both the level and growth rate of leverage, and present results in Figure 3. Starting with the average levels of leverage (panel a), we find that younger firms are on average more leveraged than older firms across all size groups. These firms are the likely candidates for being the most constrained, both because they already have the most (in relative terms) debt and the shortest track-record with lenders. At the same size, across all ages, average leverage is lower in the larger firm size bins.

Consistent with the declining level of leverage by age, the growth rate of leverage (panel b) is negative for most firm groups, although often statistically hard to distinguish from zero.[21] The exception is young firms in the smallest two size bins whose leverage is growing. This is suggestive evidence that they have high and increasing financial needs, and might therefore be more prone to financial frictions. Turning to the cyclicality results, leverage (panel c) seems to be countercyclical for almost all age-size groups with the exception of large firms of all ages, where the point estimate appears acyclical, but with very wide

[20] For the results on net worth, see Appendix B.4.

[21] Declining leverage with age can also reflect selection if high leverage firms exit more, consistent with our findings in Section 4.3.

confidence bands.

3 A simple model of finance and heterogeneous RTS

In this section we set up a simple model of financial frictions where firms differ in their returns to scale. The goal is to understand what mechanisms a model might need in order to match the joint size-age distribution of cyclicality we described in the previous section, and develop intuitions about the interactions between financial frictions and heterogeneous RTS.

3.1 Environment

We consider a model of heterogeneous firms cast in continuous time $t \in [0, \infty)$, although much of the analysis is static Firms produce with a Leontief production function $y_t = a_t z \min\{k_t, l_t\}^\eta$ in capital k_t and labor l_t. a_t is an aggregate productivity shock, z is permanent firm-level productivity, and $\eta \in (0, 1)$ measures the firm's returns to scale. Firms purchase and sell capital at price one and it depreciates at rate δ, and labor is paid the wage w_t. Firms borrow and save at a constant risk free rate r.

For the moment, suppose that at time t there is a mass M of active firms with joint distribution $G_t(n, z, \eta)$ over their net worth, n, and production parameters (z, η). We suppress the firm subscript $i \in [0, M]$. Firms borrow in risk-free debt b_t, and their balance sheet gives net worth n_t as assets less liabilities: $n_t = k_t - b_t$. We define leverage as debt over assets, $\lambda_t = d_t / k_t$, and let $\phi_t = k_t / n_t$ denote assets over net worth. Firms cannot raise equity, but they can borrow up to an exogenous and potentially time-varying borrowing limit expressed as a leverage constraint $\lambda_t \leq \bar{\lambda}_t$, so that $k_t \leq \bar{\phi}_t n_t$ where $\bar{\phi}_t \equiv 1/(1 - \bar{\lambda}_t)$. The Leontief structure simplifies the problem as firms simply set labor equal to capital: $l_t = k_t$, giving sales as $y_t = a_t z k_t^\eta$.

The simplicity of the setup, in particular there is no time to build assumption, means we can solve for the firm's optimal capital choice without reference to a full dynamic programming problem. For firms who are currently financially unconstrained, arbitrage between production and investing in the risk free bond means their optimal capital $k_t^u(z, \eta)$ satisfies the first order condition $\eta a_t z k_t^{\eta-1} = \delta + r + w_t$. For firms who are currently constrained, their optimal capital $k_t^c(n)$ is trivially equal to $\bar{\phi}_t n_t$. Combining these gives the capital policy function $k_t(n,z,\eta)$ as

$$k_t(n,z,\eta) = \min\left\{k^c(n), k_t^u(z,\eta)\right\} \text{ where } k_t^c(n) = \bar{\phi}_t n, \;\; k_t^u(z,\eta) = \left(\frac{\eta a_t z}{\delta + r + w_t}\right)^{\frac{1}{1-\eta}}. \quad (2)$$

Let $\bar{n}_t(z,\eta) \equiv k_t^u(z,\eta)/\bar{\phi}_t$ denote the level of net worth required to afford the unconstrained level of capital. To close the model, we assume that firms are owned by a representative household who also consumes and supplies labor. The household has instantaneous utility function over consumption, C, and labor supply, L, of $U(C,L) = C - \chi L^{1+\eta_L}$ and discount rate ρ. This gives a constant equilibrium interest rate of $r = \rho$. The household's labor supply condition gives the equilibrium relationship between wage and labor as $w_t = \chi(1+\eta_L)L_t^{\eta_L}$, where $1/\eta_L$ is the Frisch elasticity of labour supply. Integrating across firms yields total output, Y_t, and other aggregates. We relegate derivations to Appendix D.

3.2 Cyclicality in response to productivity and financial shocks

Our empirical results measured cyclicality by firm age and size averaging over two decades of data, and so our empirical results most likely reflect an average across several kinds of aggregate shocks. We consider a recession with a joint aggregate productivity, a_t,[22] and borrowing constraint, $\bar{\phi}_t$, shock. We measure cyclicality as the contemporaneous response of a firm's capital to the shocks.[23] Taking the total derivative of (2) in logs gives

$$d\log k_t^c(n) = d\log\bar{\phi}_t \quad (3)$$

$$d\log k_t^u(z,\eta) = \frac{1}{1-\eta}\left(d\log a_t - \frac{w_t}{\delta + r + w_t} d\log w_t\right) \quad (4)$$

where $d\log\bar{\phi}_t$ and $d\log a_t$ are the aggregate shocks and $d\log w_t$ the equilibrium response of wages. Both shocks increase aggregate output ($dY_t/\bar{\phi}_t > 0$, $dY_t/a_t > 0$) allowing us to discuss cyclicality.

While simple, these equations give three useful insights. Firstly, the capital choice of financially constrained firms, $d\log k_t^c(n)$, responds only to the financial shock, $d\log\bar{\phi}_t$, and in a procyclical manner. Secondly, the capital choice of unconstrained firms, $d\log k_t^u(z,\eta)$, does not directly respond to the financial shock. Instead, they directly respond procyclically to the productivity shock, $d\log a_t$, and may indirectly respond countercyclically to

[22]The productivity shock can be considered a stand in for other unmodelled shocks, such as demand or factor price shocks, which move profits or costs in an equivalent way.

[23]This measures the cyclicality at the moment the shock hits, and ignores later dynamics, but provides intuition. In the quantitative model we measure cyclicality using the same regressions as in the data.

both shocks if they introduce procyclical wage adjustments, $d \log w_t$.[24] Finally, the responsiveness of unconstrained firms depends on their RTS, η, with higher RTS firms more responsive to both productivity shocks and wages.

3.3 Mapping the model to the data

Under what conditions can this simple model be consistent with our results on cyclicality by joint age-size bin? Consider mapping financial constraints to age and size in the model, specifically thinking about the behavior of entrants and how they accumulate net worth. We assume that firms receive an initial equity injection at birth giving them initial net worth n_0. This along with their permanent productivity and returns to scale (z, η) is drawn from a CDF $G^e(n_0, z, \eta)$. Firms enter at exogenous rate μ_0, exit at exogenous rate ζ, and retain all profits, only paying out a final dividend when they exogenously exit. We provide a formal discussion in Appendix D, and simply state here the intuitive results that: 1) firms become less financially constrained as they age and accumulate net worth, and 2) under a minimal condition on productivity, optimal firm size when unconstrained is increasing in RTS, for a given productivity level. We consider a typical recession experiment, and identify three conditions needed for the model to match the data.

Requirement 1: Financial constraints bind for young firms. In Figure 2 we showed that firms become less cyclical with age, even controlling for firm size. Because net worth is the only variable in the model which fundamentally changes with age, for the model to generate differences in cyclicality by age group requires financial frictions to bind for young firms. Age matters is because firms need time to grow out of financial constraints. This happens if some firms are born with initial net worth draws $n_0 < \bar{n}_t(z, \eta)$. Indeed, if all firms were born rich enough to be unconstrained, then all firms would be able to choose the unconstrained capital level $k_t^u(z, \eta)$ even when young, and there would be no differences in cyclicality by age, once size is controlled for.[25]

Requirement 2: Existence of both real and financial shocks. Building on the first requirement, the existence of financially constrained firms is necessary but not sufficient to generate our results by firm age. We find that both young and old firms react procycli-

[24] Since $w_t = \chi(1 + \eta_L) L_t^{\eta_L}$, the wage is procyclical as long as L_t is, unless $\eta_L = 0$ which induces wages to be endogenously acyclical.

[25] It is important that we can control for firm size in order to make this statement, otherwise changing composition of firms by size as we move along the age distribution could be a candidate for explaining cyclicality differences by age.

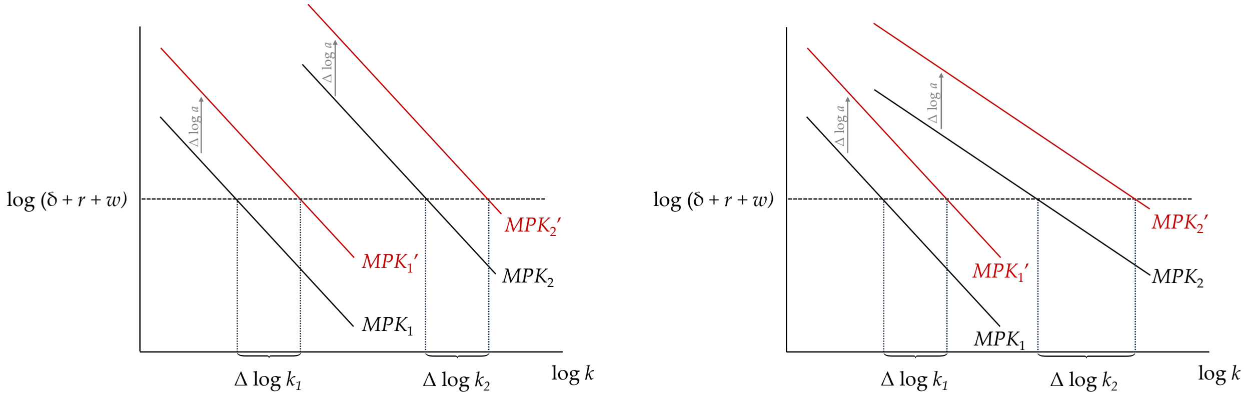
cally to the cycle, with young firms more cyclical than old. Since young firms are driven by the financial shock and old firms by the real shock, this requires that both shocks hit in our typical recession experiment.[26] This also demonstrates that the firm-level responses by firm age can be used to identify the aggregate shocks hitting the economy: the relative response of young versus old firms identifies the relative importance of financial versus real shocks, through the lens of the model. Since financial constraints make firms smaller, among young firms we would also expect larger firms to be less financially constrained and hence less cyclical, as we saw in the data.

Requirement 3: Size is positively correlated with RTS. The final requirement concerns the behavior of old firms, and therefore again exploits that we control for joint age-size bins. Among old firms, Figure 2 showed that larger firms were more cyclical than small. This is especially true among firms aged 16 years and over, who are plausibly financially unconstrained through the lens of the model. If this is the case, then all sufficiently old firms choose the unconstrained capital level $k_t^u(z,\eta)$ and have cyclicality given by (4), driven by the productivity shock. This immediately shows that if all firms have the same returns to scale, η, then all old firms would have the same cyclicality and the model cannot match the data. Differences in productivity, z, lead to differences in firm size, but do not affect how reactive firms are to shocks. Cyclicality is instead proportional to $\frac{1}{1-\eta}$, showing that firms with higher returns to scale are more responsive to an aggregate productivity shock. If large firms have higher returns to scale — which needs to be verified empirically — then this simple mechanism provides a natural explanation for our empirical finding.

A simple partial equilibrium experiment (setting $d\log w_t = 0$) explains the intuition. Figure 4 plots the optimal capital choice for financially unconstrained firms as the intersection of the marginal product of capital, $\eta a_t z k_t^{\eta-1}$, and input costs, $\delta + r + w_t$, both in logs. We compare a small and a large firm, labelled 1 and 2, where in panel (a) their size difference is driven by productivity $z_2 > z_1$, and in panel (b) by returns to scale $\eta_2 > \eta_1$. In logs, the marginal product line has slope $\eta_i - 1 < 0$, and hence is shallower for firms with higher RTS. An increase in aggregate productivity a_t shifts the log-MPK line upwards by the same amount for all firms. This leads to equal increases in optimal capital for the two firms in panel (a), since their log-MPK lines have the same slope. In contrast, in panel (b) we see that this leads to a larger increase in optimal capital for firm 2, since their larger size

[26] Or alternatively, endogenous collateral price movements in response to other shocks which trigger financial accelerator effects, which are subsumed into the financial shock in our model.

Figure 4: Higher returns to scale increases the cyclicality of unconstrained firms



(a) Size driven by productivity

(b) Size driven by returns to scale

Note: Response of unconstrained firms to a rise in aggregate productivity, $\Delta \log a$, in partial equilibrium. The MPK line gives $\log \eta + \log a + \log z + (\eta - 1) \log k$.

is accompanied with a flatter log-MPK line.

Summary and implications of heterogeneous RTS for financial frictions. We stress that while the three requirements above are needed for our simple model to match the data, other features may also do the same. Our goal is not to rule all other explanations out — they could sit alongside our mechanisms— but to develop requirements within the class of models we study. We verify that these requirements are met in our data in the next section. The presence of RTS heterogeneity has implications for financial frictions models over-and-above just matching our new cyclicality facts, which we then develop in the simple model (Appendix D.2) and in our quantitative model (Section 5).

4 Empirical supportive evidence

In this section we provide supportive evidence for the requirements and mechanisms proposed in the previous chapter. First, we directly estimate returns to scale in the micro data and find that larger firms have higher returns to scale (Requirement 3) and this is the case even when we only look future size among currently the youngest or the smallest firms. We then investigate financial frictions (Requirement 1) especially with the emphasis on the firm life-cycle. We show that there is a large heterogeneity in net worth among entrants, and entering net worth is strongly predictive about future survival rates. We provide further evidence by directly studying how cyclicality varies with leverage and RTS, and document patterns of firm growth over the lifecycle. Taking together, there are many firms that are born with low levels of net worth, which makes them more vulnerable to shocks.

4.1 Empirical estimates of returns to scale

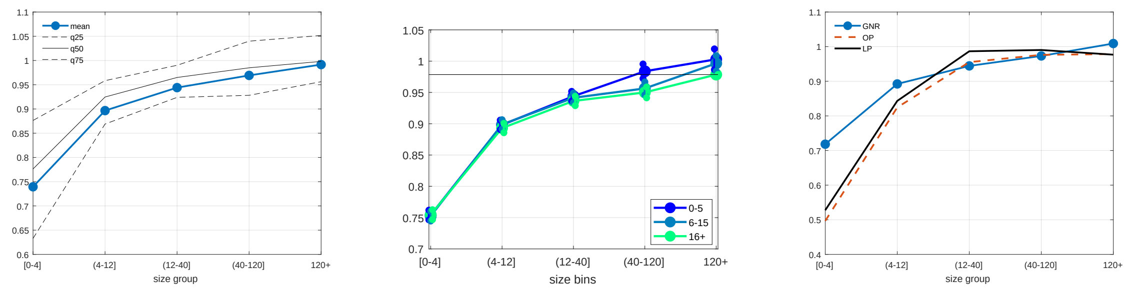
Our RTS estimates are based on production function estimation techniques. Since firm inputs are an endogenous choice of the firm, standard approaches in this literature (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015, henceforth OP, LP, ACF) use proxy variables, such as investment or intermediate inputs, to address endogeneity bias.[27] Our baseline results are obtained using a state of the art production function estimation method proposed by Gandhi et al. (2020) (henceforth GNR) that extends the existing methodology by proposing a new way to exploit the cross-equation restrictions between the first-order condition for the intermediate inputs and the production function.

Estimation methodology. Production functions are typically estimated at the sectoral level, imposing an assumption that all firms in a given sector have common production function parameters and hence (in the case of Cobb Douglas) RTS. Since our cyclicality results are within-sector, we require specifications which allow firms even within the same sector to potentially have different RTS, which we address in two ways. Firstly, the Gandhi et al. (2020) approach estimates a translog production function where RTS depend on the firm's current inputs are not therefore not constant even for fixed coefficients. Secondly, motivated by our desire to see if RTS differ by firm size, we estimate separate production functions for firms based on the maximum firm size they achieve during the sample.[28] We estimate 180 separate production functions: one for each of 5 maximum size partitions in 36 sectors. Beyond that, our procedure is standard and our baseline results feature a gross output translog production function with capital, labor and intermediate inputs. The sample features over 79 thousands firms and roughly 380 thousands firm-year observations. We estimate production functions using revenue or value added data, and so our estimates should be interpreted as returns to scale in revenue, not physical quantity produced. For more details on the estimation procedure, robustness using other estimation methods, and

[27] Production function estimation is challenging because of long-recognized endogeneity problems (Marschak and Andrews, 1944; Griliches and Mairesse, 1995). If firms respond to shocks that are unobserved by the econometrician, this introduces correlation between flexible inputs and the error term.

[28] By focusing on maximum size we aim to identify firms' long run optimal size (e.g. after financial frictions have been overcome in our model) which is a better measure of their underlying technology than their current size. This approach obviously delivers downward biased estimates of true maximum size, especially for young firms who exit young. For this reason we only use firms that survive for at least 4 years in the estimation. This partition also makes sure that every firm is featured in exactly one estimation, unlike if we were to split firms by their current size.

Figure 5: Returns to scale by firm size



(a) distribution of RTS (b) size and age effects (c) alternative methods

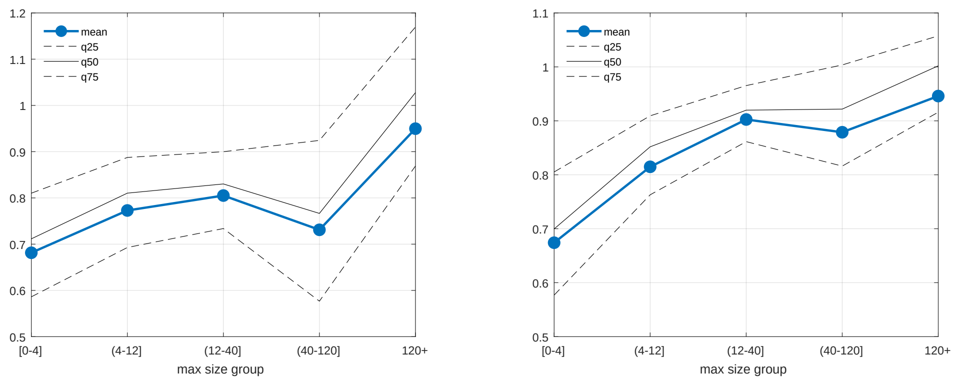
Note: Panel (a) depicts the distribution of estimated returns to scale across all industries using GNR. Panel (b) shows the estimated values of $\alpha_{j,k}$ from regression equation (1) when using the estimated RTS from GNR as the right hand side variable. Panel (c) plots average RTS by firm size for the baseline (GNR, blue with ball markers), OP (solid black line), and LP (dashed red line) methods. All three methods suggest that larger firms have higher returns to scale.

additional results (including results on productivity), see Appendix C.

Results. In Figure 5 we present our baseline RTS estimates in two ways. In panel (a) we present the inter-quartile range, median and mean of the estimated RTS across all firms in all sectors, grouping firms by their current size. In panel (b) we use equation (1) to see how average RTS varies across size and age within sectors. We find a positive and statistically significant relationship between firm size and RTS, and the results are very similar both in the full cross section of firms (panel a) and also within sectors (panel b). Average RTS are around 0.75 for the smallest firms, and above 0.95 (i.e. close to constant RTS) for the largest firms, with the median slightly higher than the mean in the full cross section. These RTS differences are economically significant, as demonstrated by the formula for the optimal choice of unconstrained capital in (2) from our toy model. All else equal, in response to a 1% rise in aggregate productivity, a firm with returns to scale of $\eta = 0.75$ would expand their capital by $\frac{1}{1-0.75} \times 1\% = 4\%$, while a firm with returns to scale of $\eta = 0.95$ would expand by $\frac{1}{1-0.95} \times 1\% = 20\%$, a five times higher elasticity. In panel (c) we show that the average differences by size are even larger using the established OP and LP methods with a Cobb-Douglas production function, where we again estimate separate production functions for each sector and max-size group.

Since firms may be born small and take several years to grow to their optimal size, an important question is whether RTS is related to a firm's current size or their underlying optimal size. In Figure 6 we focus only on firms who are currently small (panel a) or currently

Figure 6: Returns to scale of small or young firms (sorted by eventual size)



(a) only currently smallest size bin (b) only currently youngest age bin

Note: This panel shows the distribution of net worth per workers across eventual maximum firm size bin. Panel (a) restricts the sample to only firms currently in the smallest size (0-5 workers) bin, whereas panel (b) restricts the sample only to firms currently in the youngest age (0-5 years) bin.

young (panel b). We plot the distributions of RTS for these firms against the maximum size these firms achieve during their observed lifetime. We see that while small or young, the firms that eventually grow large have higher returns to scale, with a quantitative pattern similar to in the sample of all firms. We interpret this finding as RTS being a permanent feature of firms production technology. This supports models where firms have to commit to their production process when they enter (Sedláček and Sterk, 2017; Smirnyagin, 2023) as we assume in this paper.

Exploring hetoregeneity in RTS is an active research area. Our results are in line with Gao and Kehrig (2021) and Smirnyagin (2023), who find a positive relation between size and RTS on industry level data. Using Canadian and US firm-level data, Hubmer et al. (2024) also find a positive relationship between size and returns to scale using GNR, although their definition of size is different (which might explain differences in estimated curvature of returns to scale with respect to size). For more detailed discussion see Appendix C.5.

4.2 The effect of Finance and Returns to Scale on Firms Cyclicality

Now that we have established the link between firm size and RTS, we turn to exploring its relation with cyclicality, estimated together with finance, measured by firm leverage.

Consider the following regression specification:

$$x_{i,t} = \sum_m (\omega_m^{DA} + \psi_m^{DA} y_t) \mathbb{1}_{i \in DA(m)} + \sum_m (\omega_m^{rts} + \psi_m^{rts} y_t) \mathbb{1}_{i \in RTS(m)} + \sum_l (\gamma_l + \delta_l y_t) \mathbb{1}_{i \in S(l)}, \quad (5)$$

where we consider both leverage and returns to scale terciles, $m = 1, 2, 3$. This regression is a variation on (1), and controls for sector as before, but replaces the size and age controls with leverage and RTS dummies (without interaction), where the coefficients capture cyclicality differences across the leverage (via ψ_m^{DA} coefficient) and RTS (ψ_m^{rts}) distributions. We run this regression on our whole sample, and additionally on two subsamples that only cover young or old firms respectively. The results are reported in Figure 7 and the underlying regression table can be found in Table 3 in Appendix B.3. The results for employment and sales are not identical but qualitatively point in the same direction, especially when comparing the lowest versus the highest tercile, and leverage seems more strongly related to employment cyclicality, while RTS more so for sales.

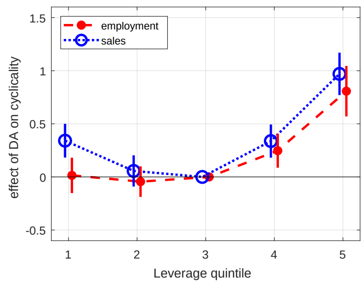
Figure 7 shows that cyclicality is increasing in both leverage and RTS.[29] Since we previously showed that leverage is decreasing in age and RTS is increasing in size, this is supportive of our story that young firms are more cyclical due to financial constraints (Requirement 1) and large firms due to high RTS (Requirement 3). Splitting the regression by age provides further support. Firstly, the effect of leverage on cyclicality is around twice as strong for young than old firms, consistent with the idea that finance matters less for old firms since they have time to outgrow their financial constraints. Secondly, the effect of RTS on cyclicality is more positive for old firms than young, consistent with our model, where binding financial constraints render RTS less relevant for the cyclicality of young firms.[30,31]

[29] Specifically, we compare the first and third terciles for this statement. For leverage, there is a small decrease from the first to second terciles, but this is not statistically significant in either of the age subsamples.

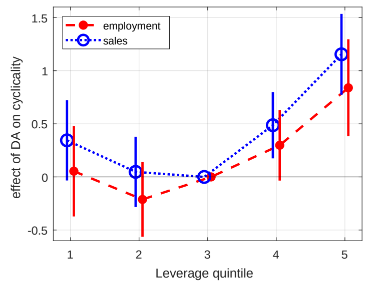
[30] As an alternative to regression with tercile dummies, Table 4 in Appendix B.3 reports the size of cyclicality coefficients if we treat the tercile indicators as continuous variables. Just as in Figure 7, the RTS is positive and significant only among old firms and leverage being significant among firms of all ages, but the coefficient being larger among the young.

[31] These results have comparable quantitative magnitudes to our cyclicality results in Figure 2, suggesting that these channels are quantitatively relevant. E.g. for RTS, the cyclicality difference from low to high RTS for old firms is around 0.5, which is reasonable fraction of the cyclicality difference by size for old firms in the data (0.5-0.75). The cyclicality difference from low to high leverage for young firms is about 0.5-0.75, which is a reasonable fraction of the excess cyclicality of young firms which is around 1-1.5.

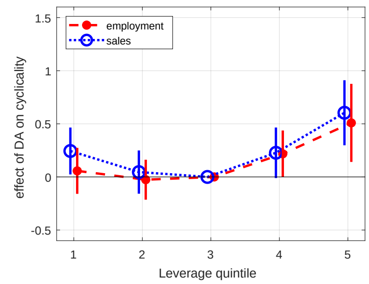
Figure 7: The effect of leverage and returns to scale on cyclicality



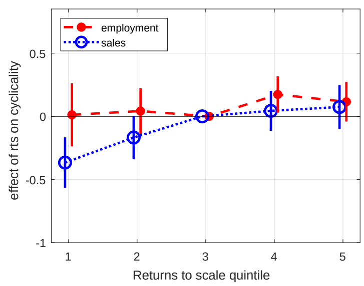
(a) leverage, all firms



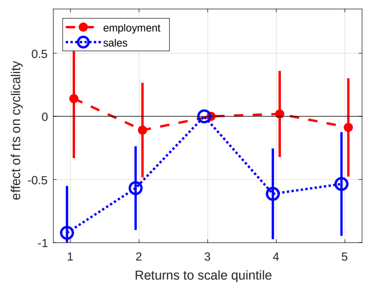
(b) leverage, young firms



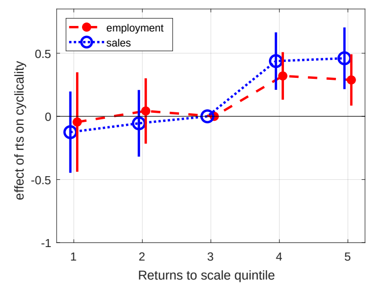
(c) leverage, old firms



(d) RTS, all firms



(e) RTS, young firms



(f) RTS, old firms

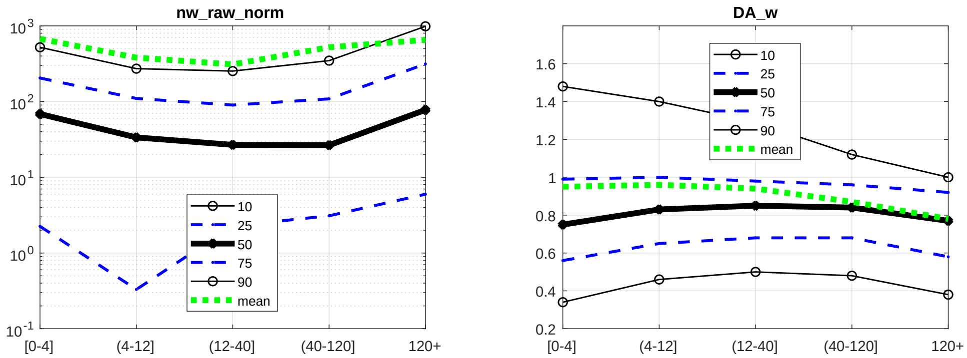
Note: The panels show the coefficients ψ^{DA} and ψ^{rts} from regression (5) estimated on three samples; full, only young (0-5) and only old firms(15+). The vertical line capture 95% confidence interval. The middle quintile of leverage or rts used as the base group.

4.3 Financial situation at entry and its implications

Distribution of net worth at entry. In this section we provide descriptive statistics on the distribution of financial variables among entrant firms. Splitting by firm size, we document the distributions of leverage and net worth per worker for entrant firms in Figure 8.[32] Four observations stand out. First, there is wide heterogeneity in the financial position of entrants measured via net worth or leverage, even controlling for firm size bin. The interquartile range and the median of both net worth per worker and leverage at entry are relatively flat across the firm size distribution, with the IQR typically spanning firms with debt to asset ratios near 1 (i.e. near zero net worth) or as low as 0.6. Second, there is a very heavy right tail of net worth: the average is close to the 90th percentile. This is not the case for average leverage which is located between the 50th and 75th percentile across the whole size distribution. Third, our data also contain a nontrivial number of firms that have negative net worth (leverage above one) which is something that the model will not be able to replicate. Fourth, although there is strong trend in the absolute net worth (see

[32]For the splices over the distribution for ages 5, 10, 15 and 20 see Appendix B.4 in Figure 27.

Figure 8: Net worth at entry



(a) net worth per worker (b) leverage

Note: This figure shows the distribution of net worth per workers and leverage among entrants. Legend: solid and dashed lines: corresponding percentiles, dotted line: mean. Net worth per worker is measured in thousands of DKK.

Table 1), this is not the case for net worth per worker.

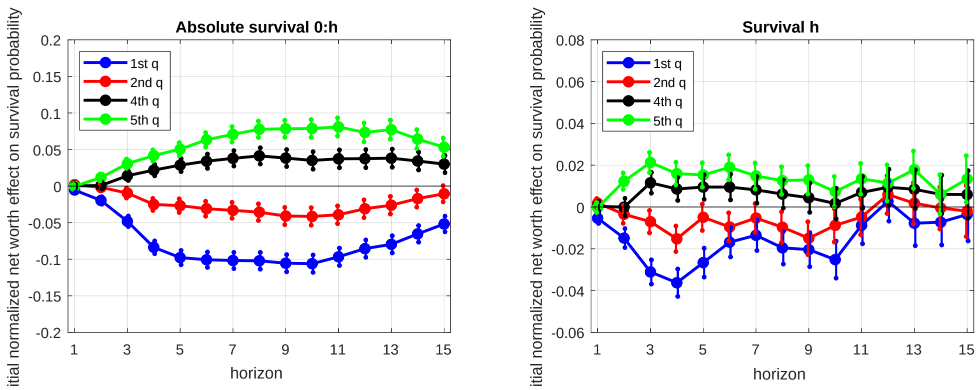
Our conclusion is that among firms of all sizes there is significant dispersion of starting net worth. If this is not perfectly related to the quality of the firm, this opens the door to some firms being born financially constrained and needing time to reach their optimal size.

Net worth at entry and survival odds. Do the aforementioned differences in entry net worth translate into differences in real outcomes for firms? We show that they do, by documenting that the net worth the firms start with is predictive of their future survival odds. Specifically, we run a series of linear projections with indicator variables capturing firm survival status at age h with the quintile j of net worth per worker *at entry*:

$$S_{i,t+h} = \alpha_h + \sum_{j=1}^{5} \beta_{j,h} \mathbb{1}_{i \in N(j)} + X_{i,t} + \varepsilon_{i,t,h} \qquad \forall h = 1, 2, ..., 15 \tag{6}$$

In this regression we include only firms who entered at time t, and the indicator $S_{i,t+h}$ equals 1 if the firm is still operating at age h (i.e. time $t+h$) and zero otherwise. $X_{i,t}$ contains sector, entering size, and time fixed effects. The coefficient of interest is $\beta_{j,h}$ which, in the absence of other controls, would simply measure the fraction of firms in group j who survive up to age h. With controls, $\beta_{j,h}$ measures the effect of initial net worth on the probability of surviving up to age h relative to the excluded base group $j = 3$. We plot the results from this regression in Figure 9(a). In panel (b) we break down the cumulative survival probability into the marginal effects of surviving from age $h-1$ to h, by rerunning each regression h additionally only including observations where $S_{i,t+h-1} = 1$. The full

Figure 9: Starting with higher net worth significnatly increases odds of surviving



(a) total effect until horizon h (b) marginal effect at horizon h

Note: This figure shows the coefficients $\beta_{j,h}$ from regression equation (6). Vertical lines (offset for better visibility when overlaying) show 95% convidence interval.

regression table is in Appendix B.5 in Tables 5 and 6.

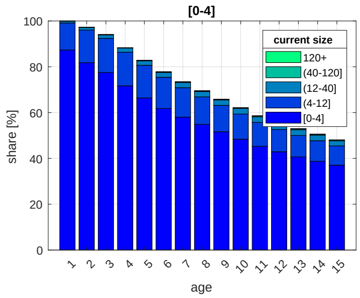
We find the effect of starting net worth to be important and it explains a large share of variation in exits (R^2 of over 30% by age 10 in the regression). The cumulative effect seems to be largest around age 10 where we see a nearly 20p.p. difference between the survival rate of firms that started with the highest versus lowest normalized net worth.[33] This difference is initially small but rapidly increases up to around age 10. After age 10 the cumulative effects stop growing reflected in the marginal effects mostly losing statistical significance. These results suggest that finance, measured via net worth, matters especially early on in a firm's life. Even if a firm is promising in the sense of large productivity or returns to scale, if it starts with low starting net worth, it faces higher chances of exit and therefore never realizing its potential.

4.4 Persistence of firm size

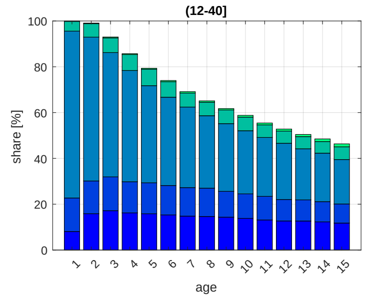
We postulated that financial frictions slow down firms' convergence to their optimal size. In our model, firms born with low net worth will be financially constrained and hence

[33]The regression (6) is subject to the usual *age-period-cohort* identification issue so the time fixed effect in $X_{i,t}$ and the age effect in α_h are not interpretable as such without additional assumptions. Importantly, this identification problem does not affect the interpretation of $\beta_{j,h}$, since the time fixed effect is not interacted with initial net worth bin. If we assume additionally that there are no cohort effects, we can interpret the coefficient α_h as the average survival rate at age h. The estimated difference between in marginal survival rate between the richest and the poorest firms at the time of entry at horizon 5 (15) is 6 pp, relative to the marginal survival odds of 95% at age 5 (96% at age 15). At the same time, the estimated gap total odds of surviving more than 5 (15) years is 17 pp (13 pp) relative to the surviving odds average of 75% (26%) on average.

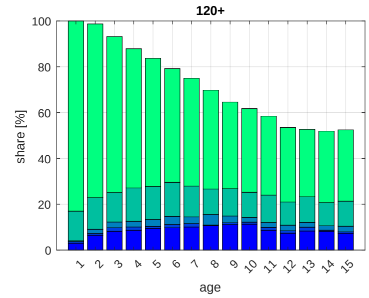
Figure 10: Firm size distribution according to their current/entry size bin



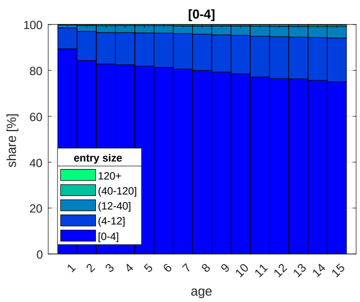
(a) *started* in the bottom bin



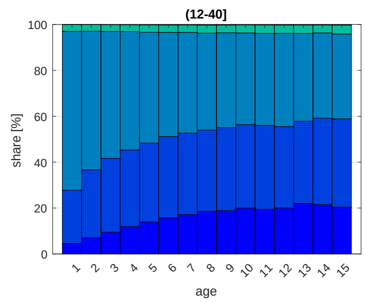
(b) *started* in the middle bin



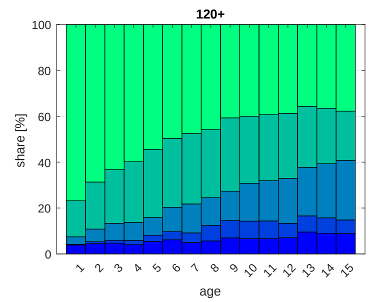
(c) *started* in the top bin



(d) *currently* in the bottom bin



(e) *currently* in the middle bin



(f) *currently* in the top bin

Note: Note: The first row panels show the *starting* size distribution depending on the current size bin, whereas the second row shows the *current* size distribution based on starting size bin. In the first row, the white gap until 100 captures the fraction of firms that exited.

small when young, and then grow towards their optimal size with age. In this section, we provide evidence on how many firms that start small grow to become very large. To do so, Figure 10 presents the size transition rates in two ways. In the top row, we sort the firms by their size bin at entry and track what size bins they belong to (or whether they have exited) at each age up to age 15. In the bottom row, we sort firms based on their current size at each age and go backwards to see what size bin they belonged to at the point of their entry.

Unsurprisingly, firm size is very persistent: conditioning on survival, almost 80% of firms born into the smallest size bin remain in the smallest size bin at age 15 (panel a). Similarly, around 50% and 60% of firms born in the middle and top bins find themselves in the same bin at age 15 (panels b and c). Outside of this persistence, firms grow and shrink, moving into the higher and lower bins. Surprisingly, starting firm size has a relatively small effect on long term survival: while marginally just below 50 percent of firms that started in the smallest size bin survive to age 15, this number is marginally above 50 percent for the firms that started in the largest size bin.[34]

[34]This is not the case if we look at *the current* size bin, instead of *the starting* size bin, see Appendix B.6. For example, firms that start small and make it to the larger size bins are less likely to exit compared to the firms

However, given the small number of large firms, this persistence masks an important fact: a large share of large firms were born small. We can see this by switching perspective and looking from current firm size backwards. A large share of the firms that currently find themselves in the largest bins came from other bins, including the smallest. In fact, less than 40% of 15 years old firms that are currently in the largest firms size bin actually started in the largest bin, and almost 10 percent started in the smallest bin (panel f). Similarly, nearly 60% of 15 year old firms in the middle bin started in the two smallest bins (panel e). If this growth driven by firms overcoming financial frictions it implies costs of financial frictions, and we use this data as an untargeted check in our model.

5 Quantitative model

In this section we build our quantitative model, which builds on classic heterogeneous-firm financial frictions models, such as Khan and Thomas (2013).[35] We extend this class of models and discipline it using our new facts on the heterogeneity in returns to scale across the size distribution and firm-level data on the financial position of entrant firms. We use the model to investigate cyclicality along the age-size distribution, and the costs of financial frictions in steady state and following financial shocks.

5.1 Model setup

The model features a continuum of heterogeneous firms, with both ex-ante and ex-post heterogeneity. There is a representative consumer, who owns firms and supplies labor. The key extension that we use to match our new stylized facts is different returns to scale across firms, combined with rich heterogeneity in the financial position of firms at birth. The model is set in continuous time $t \in [0, \infty)$ with an infinite horizon. We focus on the case without aggregate uncertainty, and conduct business-cycle experiments using unanticipated one-time shocks. The model is presented in steady state, for expositional simplicity, and we therefore drop the time subscript, t, in most of what follows.

Firm types. A continuum of heterogeneous firms $i \in [0, G]$ are our firms of interest. These firms produce a homogeneous final good using capital and labor. This homogeneous good is the numeraire, with price normalized to one, which is used for consumption and

that were born large.

[35]For examples of other work building on this framework, see Jo and Senga (2019), Ottonello and Winberry (2020), and the references therein.

investment. The mass G of firms arises via endogenous entry and exit. Firms have both ex-ante and ex-post heterogeneity, are owned by the representative household, and discount the future at the interest rate, r.

At birth, firms draw a permanent "size type" $s = \{1, 2, ..., S\}$, which determines features which we wish to relate to firm size. Specifically, their returns to scale, η_s, depends on this size type, as well as a permanent component of their physical productivity, which we label z_s^S. We allow for idiosyncratic shocks to firm productivity, z_j^J with $j = \{1, ..., J\}$ denoting discrete productivity levels with transition rates from j to j' of $\pi_{j,j'}^J$. All firms share the common production function $y = z \min\left\{k, \frac{l}{\alpha}\right\}^{\eta_s}$ where $z \equiv z_s^S z_j^J$ denotes overall productivity, which combines the permanent and idiosyncratic components. Firms have Leontief production functions in capital and labor, with labor share determined by α.[36] If all firms had $\eta_s = 1$ then all firms would have constant returns to scale in production, and $\eta_s < 0$ denotes decreasing returns to scale. Firms pay a stochastic fixed cost, which we discuss further below.

Firm choices. Firms make decisions both on the extensive (entry/exit) and intensive (how much to produce) margins. At the firm level, all factors of production can be adjusted freely without cost. We are in continuous time, and there is no time to build for capital. It is convenient to first optimise labor for a given level of capital. The Leontief production function gives the solution simply as $l(k) = \alpha k$, and $\pi(k, s, j) = z_s^S z_j^J k^{\eta_s} - \alpha w k$ denotes revenue less the wage bill. A firm's capital stock evolves through a standard accumulation equation. Given investment i per unit of time and depreciation rate δ we have: $\dot{k} = i - \delta k$. One unit of investment costs p_K units of final good. Old capital and investment are perfect substitutes for firms, so capital also trades at the price p_K.

Firms can borrow using a risk-free short-term bond b with interest rate r. We will consider endogenous firm exit but firms will pay off their debt when exiting, so never default and so pay only the risk free rate. They face a borrowing constraint which limits the amount they can borrow according to the amount they can post as collateral: $b \leq \lambda p_K k$, where recall that k is a firm's physical capital. The parameter λ controls the tightness of

[36]The use of a Leontief production function is helpful in matching the wide size distribution in the data, when combined with financial frictions which directly affect the purchase of capital only, and not labor. With a Cobb Douglas production function, a financially constrained firm heavily substitutes from capital to labor while young. By ruling this out, the Leontief production function forces firms to maintain a fixed capital-labor ratio, so that financial frictions directly affect both capital and labor equally. This helps keep firms of size type s in the associated group (0-5 employees and so on) that they are designed to match.

the leverage limit, with smaller λ making the constraint tighter. In the business cycle experiments, we allow λ to evolve as an aggregate financial shock. A firm's net worth, n is defined as its assets less its liabilities: $n = p_K k - b$. Combining this with the leverage limit gives $k \leq \frac{n}{p_K(1-\lambda)}$. Define a firm's assets over net worth, ϕ, as $\phi \equiv p_K k/n$. Combining this with the leverage limit expresses the borrowing constraint as: $\phi \leq \bar{\phi}$, where $\bar{\phi} \equiv \frac{1}{1-\lambda}$ is the exogenous borrowing limit. Absent the arrival of a stochastic fixed cost shock, a firm's net worth evolves according to

$$\dot{n} = \left(\frac{\pi(k,s,j)}{k} - (\delta + r)p_K\right)k + rn - d \tag{7}$$

where the first term is the net return on leveraged investment, and d denotes the dividend payout flow. We assume that firms cannot raise equity at all after the moment of birth, and so impose $d \geq 0$.

Firm HJB. The firm's problem can be stated recursively using a Hamilton Jacobi Bellman (HJB) equation. Optimized firm value, $v(n,s,j)$, can be expressed as

$$\begin{aligned} rv(n,s,j) = \max_{0 \leq p_K k \leq \bar{\phi} n, d \geq 0} & \, d + \zeta\left(n - v(n,s,j)\right) + \sum_{j'} \pi^J_{j,j'}\left(v(n,s,j') - v(n,s,j)\right) \\ & + v_n(n,s,j)\left(\left(\frac{\pi(k,s,j)}{k} - (\delta + r)p_K\right)k + rn - d\right) \\ & + \alpha_\omega\left(\int_{\omega=0}^{\infty} \max\left\{v(n-\omega,s,j), n\right\} dF(\omega,s) - v(n,s,j)\right) \end{aligned} \tag{8}$$

Firms maximize the discounted sum of dividends, d. The v_n term is the drift in net worth, which depends on the capital choice and dividend payout. The ζ term captures exogenous firm exit, where firms pay out their net worth as a final dividend, and the j' term captures transitions across idiosyncratic productivity states. The α_ω term is captures the stochastic fixed cost shock driving endogenous firm exit, which we discuss further below.

Investment policy. The firm investment policy in this setting can be expressed as an unconstrained optimal capital stock, which firms will achieve only if they are financially unconstrained. The first order condition with respect to capital gives

$$v_n(n,s,j)\left(\pi_k(k,s,j) - (\delta + r)p_K\right) = \mu_k,$$

where $\mu_k \geq 0$ is the multiplier on the borrowing constraint. If a firm hits its borrowing constraint then we know that $k = \bar{\phi} n/p_K$. If a firm is rich enough to be unconstrained, then $\mu_k = 0$ and the capital FOC gives us $\frac{\pi_k(k,s,j)}{p_K} = \delta + r$. This gives the unconstrained investment policy if unconstrained, $k^u(s,j)$, which an analytical solution of the same form as the toy model from Section 3. The overall investment policy can then be simply expressed as $k(n,s,j) = \min\{\bar{\phi}n/p_K, k^u(s,j)\}$.

Dividend policy. The first order condition with respect to dividends gives $\mu_d = v_n(n,s,j) - 1$, where $\mu_d \geq 0$ is the multiplier on paying positive dividends. The firm optimally pays zero dividends whenever the marginal value of keeping net worth inside the firm is larger than one, and only pays out dividends when $v_n(n,s,j) \leq 1$. Given that firms discount the future at the interest rate, and there is always the possibility that the fixed cost shock can drive you to be financially constrained in the future, we prove in Appendix E.1 that $v_n(n,s,j) > 1$ at all times in our model, so firms never pay out dividends voluntarily ($d(n,s,j) = 0$) and only do so when they exit.

Firm exit and entry. All firms face an exogenous exit rate ζ. When firms exit in this way, they pay out their remaining net worth, n, as a final dividend. Beyond the common exit shock, there is additional endogenous firm exit driven by a stochastic fixed cost shock. At rate α_ω firms have to pay a stochastic fixed cost $\omega \in [0,\infty)$ in order to continue producing. If they do not pay, they must immediately and permanently exit. The fixed cost is paid as a purchase of the final good, and is drawn from a CDF $F(\omega,s)$. Notice that the distribution is allowed to depend on s, so that high size type firms might pay higher fixed costs, for example. Notice that this fixed cost is not a flow (i.e. you do not pay ωdt) but rather a stock cost, which will be paid out of net worth. So when a fixed cost shock arrives, if the firm chooses to pay it and continue operating their net worth jumps from n to $n' = n - \omega$. If they choose not to pay it they must exit and pay their net worth n out as a final dividend. Firms optimally choose whether to exit or not by comparing the value of continuing to the value of exiting.[37] As long as v is increasing in n, there is a threshold $\bar{\omega}$ where the firm would exit for any $\omega \geq \bar{\omega}$, defined by $v(n-\bar{\omega},s,g,j) = n$. This defines threshold $\bar{\omega}(n,s,j)$ where for draws of ω between 0 and $\bar{\omega}(n,s,j)$ the firm optimally remains active but with a reduced net worth $n - \omega$. For draws of ω greater than $\bar{\omega}(n,s,j)$ the firm optimally exits and pays out n as a final dividend.

For a firm with state (n,s,j) their total exit rate is $EX(n,s,j) = \zeta + \alpha_\omega\left(1 - F(\bar{\omega}(n,s,j),s)\right)$, where the first term is exogenous exit and the second is endogenous exit. This formulation

[37] Jo and Senga (2019) use a similar structure to drive firm exit, but where the cost is paid as a utility cost which does not directly affect the firm's resources. In our formulation, the fixed cost is a true resource cost which worsens the firm's financial position if they choose to pay it and not exit.

allows firm exit to depend on net worth, which allows the model to endogenously generate exit rates which decline with firm age as firms accumulate net worth.

Firm entry is endogenous and allowed to vary independently for each size type. There is a large fixed mass M_s^e of potential entrants of each size type s at each instant of time. Each potential entrant draws an entry cost ξ from a CDF $G_s^e(\xi)$. They decide to enter after observing ξ but before knowing their draw of their initial net worth n_0 or idiosyncratic shock j. Thus, the ex-ante value of entry, excluding the entry cost, is $v^e(s) = E\left[v(n_0, s, 1, j) - n_0\right]$, where j is drawn from its ergodic distribution, π_j^J. Initial net worth is drawn from a distribution with CDF $F^e(n_0, s)$ and PDF $f^e(n_0, s)$. This distribution will be crucial in disciplining the model, as it determines how financially constrained firms are. A potential entrant enters if $v^e(s) - \xi \geq 0$, giving the total entry flow into each size type as $\mu_0(s) = M_s^e G_s^e(v^e(s))$. The flow of entrants of any type (n_0, s, j) is equal to $f^e(n_0, s)\pi_j^J \mu_0(s)$, and the total entry flow is $\mu_0 = \sum_s \mu_0(s)$.

Closing the model. Given the solution to the firm problem, we can simulate the endogenous firm distribution in steady state, where we denote the ergodic distribution $G(n, s, j)$, or in transition experiments. We can then calculate aggregates such as output and employment, and moments of the firm size and age distribution. We close the model by specifying how the prices that firms face (real wage, interest rate, and capital prices) are determined in a general equilibrium setting.

We assume that the representative household has GHH instantaneous utility function over consumption, c, and labor supply, L, of $U(C, L) = \frac{1}{1-\eta_C}\left(C - \chi L^{1+\eta_L}\right)^{1-\eta_C}$ and discount rate ρ. The household's labor supply condition gives labor supply as a function only of the wage: $L = (w/(\chi(1+\eta_L)))^{1/\eta_L}$, where $1/\eta_L$ is the Frisch elasticity of labour supply and χ the labor disutility shifter. We suppose that investment goods are produced by a representative investment good producing firm subject to quadratic adjustment costs. This gives the equilibrium capital price as $p_K = 1 + \psi_K\left(\frac{I}{K} - \delta\right)$, where I is aggregate investment and K aggregate capital. This formulation normalises capital prices to one in any steady state, while they deviate from one if the aggregate investment rate changes during transitions, where ψ_K controls the strength of adjustment costs. Since Denmark is a small economy with a trade to GDP ratio of over 100%, we use a small open economy (SOE) framework as our baseline. We assume a constant exogenous world real interest rate of $r_w = \rho$, and since this is a one good model the equilibrium interest rate in Denmark takes the same value $r = r_w = \rho$. We also consider a closed economy version of the model, where the interest rate is still $r = \rho$ in steady state but is determined by the household's Euler equation during transitions. Aggregate output is the sum over firms $Y \equiv \int y(n,s,j)dG(n,s,j)$, and goods market clearing gives $Y = C + I + FC + AC + EC + NX$ where FC is spending on the stochastic fixed cost, AC spending on investment adjustment costs, EC on entry costs, and NX are net exports. See Appendix E.1 for details and the definition of equilibrium.

Calibration. The key novelty of our calibration is that we discipline our model using our new facts on how returns to scale vary across the firm size distribution, and on the distribution of initial net worth across firms. Apart from these tweaks, we purposefully aim for a standard calibration of a heterogeneous-firm financial frictions model, in the spirit of Khan and Thomas (2013), and in the interests of space we relegate most calibration details to Appendix E.2. We discuss here the novel features of our calibration: heterogeneous RTS and disciplining the entrant net worth distribution.

We assume $S = 5$ permanent size types in the model, corresponding to the five size bins in our empirical work. Each size type s is designed to make up the majority of firms in a given size bin in steady state, with $s = 1$ corresponding to the 0-4 employee bin, $s = 2$ to the 4-12 employee bin, and so on. We set the RTS parameter η_i for each size type directly from our data, using our estimated RTS for each size group from Figure 5(a). Accordingly we set $\eta_1 = 0.75$, $\eta_2 = 0.875$, $\eta_3 = 0.925$, $\eta_4 = 0.95$, $\eta_5 = 0.97$.[38] We choose the size-type specific productivity shifters z_s^S to match the employment share of each associated size bin. We calibrate the entry process to set the relative flow of entrants into each size type in steady state, $\mu_0(s)/\mu_0$, to match the share of firms in each size bin in the data.

We calibrate the initial financial position of entrants to achieve a match to two key features of the distribution of leverage across age 0 firms from our data in Figure 8(b): Firstly, the data shows that the leverage distribution is relatively similar across entrants in the different size bins, motivating a simple parameterization with common parameters across size types s. Secondly, the data shows that the distribution of leverage across entrants within the same size bin is wide, motivating a wide initial net worth draw distribution. We parameterize the distribution for initial net worth draws as log normal with common standard deviation σ_e and size-type-specific mean: $n_0 \sim \text{Lognormal}(\mu_e \bar{n}_s, \sigma_e^2)$. We assume that the mean net worth draw is a common fraction, $\mu_e \in (0,1)$, of the net worth each size-type

[38]Note that some of our estimates gave slightly increasing returns to scale, which is incompatible with our model which requires decreasing returns to scale to have well defined firm sizes. Since not all firms in each size bin are of the associated size type, the average returns to scale in bin s differ slightly from η_s, but this difference is small in practice, as shown in Figure 33(c).

needs to become financially unconstrained, $\bar{n}_s$.[39] We set the leverage constraint to allow a maximum debt to asset ratio of 0.7. We follow Khan and Thomas (2013) and choose the initial net worth fraction μ_e to match the share of aggregate employment in young firms (the age 0-5 bin) in the data. We set $\sigma_e = 1.25$, which generates a wide distribution of initial leverage. These parameters generate a distribution of leverage roughly comparable to the data, as shown in Figure 37.[40]

The model is able to replicate important moments from the data, above the targeted age and size distributions, as discussed in Appendix E.2. The model generates higher exit rates for low net worth entrants, as in the data, and replicates well the fact that a large share of large old firms were born in smaller size bins.

For comparison purposes, we also calibrate a more standard model where all firms are assumed to have the same returns to scale of $\eta_s = 0.85$. This value is in line with the values in typical heterogeneous firm models (Khan and Thomas, 2013; Winberry, 2021) and so allows us to see how heterogeneous RTS alters the basic model.[41] We refer to the two calibrations as our "baseline" and "common RTS" calibrations respectively.

Cyclicality. In Appendix E.3 we show that the model is able to replicate our cyclicality results from Section 2.3. We simulate a recession meant to emulate the average of the recessions in our empirical sample. In response to a combined financial and TFP shock, the model matches that young firms are more cyclical than old firms, and that cyclicality is increasing in size among old firms, due to the heterogeneity in RTS. We also compare the firm-level responses to TFP and financial shocks and discuss how firm-level data identify the relative size of different aggregate shocks.

Overview of the experiments. In the following two sections, our two main experiments

[39]Specifically, for each s, $\bar{n}_s$ is the level of n needed to afford the unconstrained capital choice at the highest idiosyncratic productivity draw: $\bar{n}_s = k^u(s, J)/\bar{\phi}$.

[40]We aim for a good match studying the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles of the data. The 90th and 95th percentiles of DA in the data are above 1, and hence incompatible with our model. The model does well between the 5th and 50th percentiles, but struggles to match the 75th percentile. This is because the median firm is financially constrained in the model, so all firms from the median and above have the same level of leverage. In practice, matching the data exactly is challenging because even very similar firms in the data can hold quite different financial positions, for reasons outside of our model.

[41]We recalibrate the other parameters, and the calibration procedure for these other parameters is identical, with two exceptions. Firstly, we fix the standard deviation of firm-level shocks at the value from our main calibration. Recalibrating this value led to issues with the calibration, since small firms are more responsive to the shock in the common RTS model and fall out of their targeted bin during the calibration. Secondly, this counterfactual model is not able to match the employment share of age 0-5 firms, since firms outgrow their financial constraints too quickly. We therefore instead choose the parameter μ_e to match the same employment share of age 0 firms as in our baseline calibration.

investigate the aggregate effects of financial frictions in our model, and how they are altered by their interaction with heterogeneous RTS. We show how financial frictions interact with heterogenous returns to scale. In particular, the results will vary by the degree to which prices adjust, which will change how much resources the firms that are close to constant returns are able to absorb.

5.2 Steady state costs of financial frictions

We begin with an illustrative steady state experiment, where we permanently tighten the borrowing constraint by 50%. While introducing heterogeneous RTS does not change the direction of the effect of financial frictions in the aggregate, it does alter the adjustment of prices vs. quantities, resulting in the aggregate contribution of different firm types shrinking or expanding in equilibrium. The magnitude of these differences is driven by the responsiveness of prices, which in steady state are controlled by the labor supply elasticity η_L. In the next paragraphs we describe in detail the results for our baseline elasticity $\eta_L = 0.3$, but we also show how the aggregate response of key variables differs for a range of elasticities, plotting the results in Figure 11 for both the baseline and common RTS models.[42]

Allowing heterogeneous RTS leads to quantitatively important changes in how financial frictions distort the economy. On the one hand, the model with heterogeneous RTS sees a larger fall in productivity, measured as output per worker,[43] which nearly doubles relative to the common RTS model (panel c). This reflects big changes in the allocation of resources across firms by age and size, which we discuss further below. On the other hand, these changes actually dampen the total fall in output and employment relative to the common RTS model (panels a and b), reducing the overall impact of financial frictions on aggregate welfare.[44]

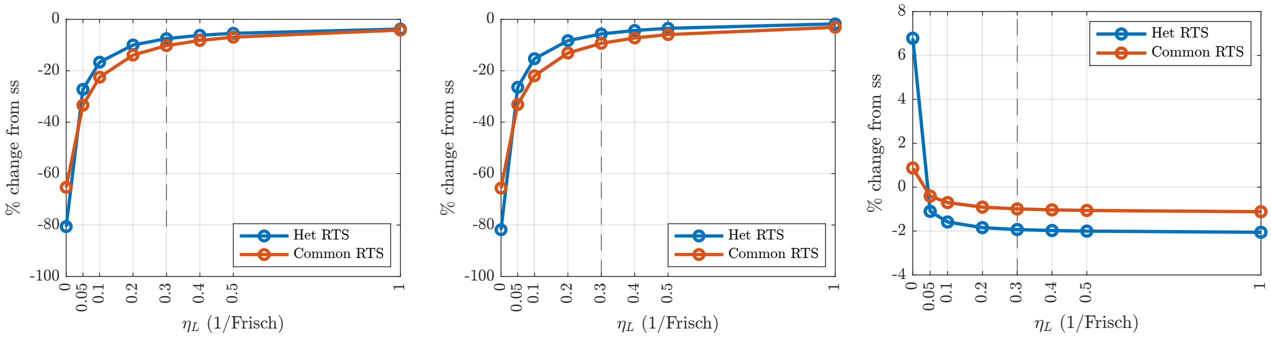
Strikingly, allowing for heterogeneous RTS more than doubles the fall in the number of firms following the tightened financial constraint (panel d), and this is mostly due to the larger entry fall (panel e). Since output falls less than the number of firms, output per firm rises, implying that some firms must have expanded to take advantage of the disruption

[42]Recall that the interest rate and capital price are always fixed at ρ and 1 in any steady state, so the wage is the only price which can adjust. When varying η_L we recalibrate χ to maintain $w_{ss} = 1$, so that varying η_L does not affect the steady state of the calibrated model.

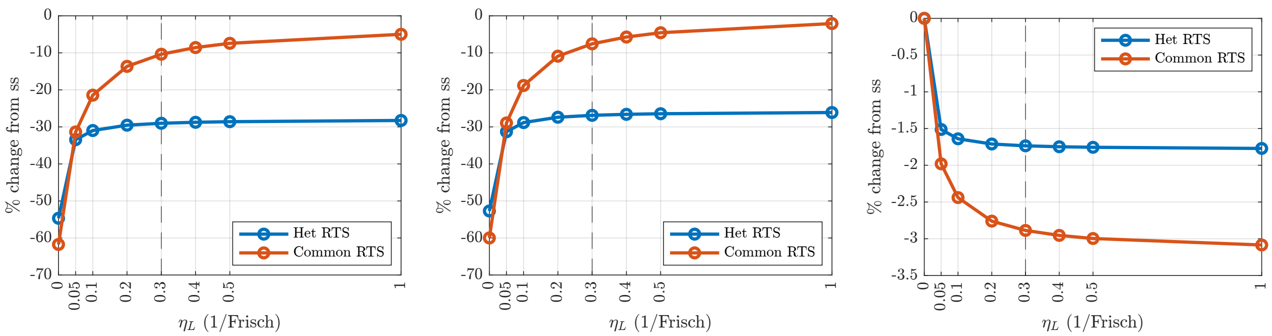
[43]With our Leontief production function assumption the firm-level and aggregate capital-labor ratios are fixed, so we use output per worker as our measure of aggregate productivity.

[44]Indeed, when $\eta_L = 0.3$, measured in consumption equivalents the welfare decline is 3.2% in the common RTS model and only 2.5% in the baseline model.

Figure 11: Aggregate effect of permanent financial tightening



(a) Output (b) Employment (c) Productivity (Y/L)



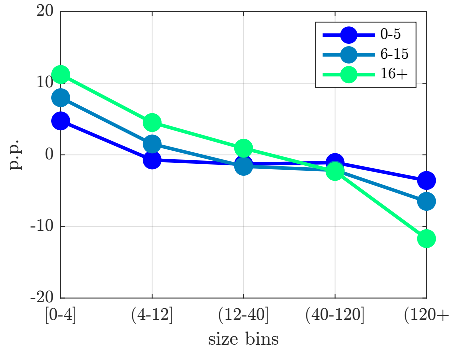
(d) No. firms (e) Entry (f) Wage

Note: Change in aggregates following a permanent 50% tightening of borrowing constraint. Blue line is for our baseline model, and red the model with common RTS. x-axis traces different elasticities of labor supply, with $\eta_L = 0.3$ the baseline value.

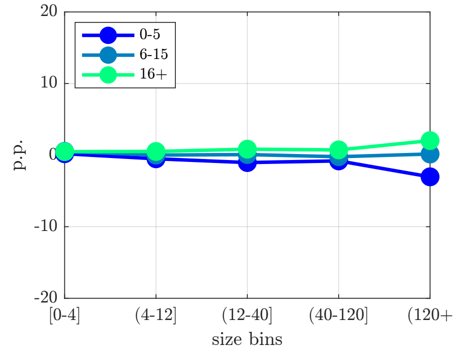
caused by the financial tightening. This is induced by the fall in the aggregate wage (panel f), which raises profits and the optimal size of financially unconstrained firms. Three key ideas, one partial equilibrium and two general equilibrium, explain the aggregate response in our model.

Firstly, holding factor prices fixed, heterogeneous RTS make large *constrained* firms more responsive to financial frictions. We see this in Figure 12, which plots the change in the sales share of each age-size group following the financial tightening, either allowing the wage to adjust ($\eta_L = 0.3$), or holding the wage fixed ($\eta_L = 0$). For a fixed wage, tightening the constraint leads to a significant shift in sales away from large firms and towards small firms in our model (panel a), which is absent with common RTS (panel b). This is despite our assumption that firms of all size types started with the same net worth (on average) relative to the amount needed to be unconstrained, μ_e. Starting from the same net worth, it takes high RTS firms longer to outgrow their financial constraints since their profit margins are lower, making them more sensitive to a financial shock. We demonstrate this effect analytically in Appendix D.2.

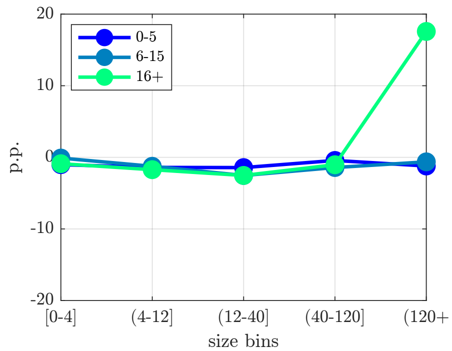
Figure 12: Effect of permanent financial tightening on sales shares



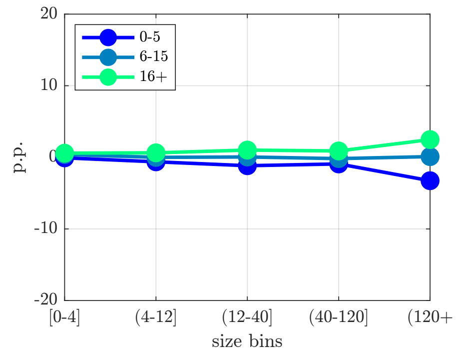
(a) Het RTS (fixed wage)



(b) Common RTS (fixed wage)



(c) Het RTS



(d) Common RTS

Note: Change in fraction of firms in each joint age-size bin following a permanent 50% tightening of borrowing constraint. Each line is a different age group. Panels (a) and (b) hold the wage constant (by setting $\eta_L = 0$) and panels (c) and (d) allow the wage to adjust ($\eta_L = 0.3$).

Secondly, this effect is overturned in general equilibrium in our model, mainly due to the behavior of large *unconstrained* firms. As employment falls the wage falls, which raises profit margins for constrained firms and optimal firm size for unconstrained firms. This effect is most pronounced for large firms with high RTS, who are the most responsive to changes in factor prices. This causes a significant increase in size at large, old (and hence unconstrained) firms, leading to a nearly 20pp increase in their sales share (panel c). In other words, with heterogeneous RTS, large unconstrained firms "crowd in" after a financial tightening. We demonstrate this effect analytically in Appendix D.2 for our analytical model where r is fixed: in the limit where there exists even a single unconstrained firm with constant RTS, aggregate employment is completely unresponsive to the financial shock. This effect is effectively absent in the model with common decreasing RTS (panel d), because large firms in that model have lower RTS and hence are unresponsive to changes in factor prices.

Finally, because large unconstrained firms crowd in so much after a financial tightening with heterogeneous RTS, this amplifies missing generation effects. Specifically, since high RTS firms are so responsive to wage changes, a smaller fall in wages is needed to equilibrate the labor market than in the common RTS model. For $\eta_L = 0.3$, the required wage fall is only half as large (Figure 11 panel f) and since wages are endogenously less responsive to shocks this creates problems for entrant firms. Entrants have a high need for finance, and with a smaller decline in wages, the value of being an entrant firm falls more after the financial shock, which is why entry falls so much more in our model than in the common RTS model. To see that it is wages driving the larger entry fall in our model, Figure 11(f) shows that that for fixed wages ($\eta_L = 0$) the decline in entry is instead smaller in our model than the common RTS model. With fixed wages the general equilibrium forces are absent, which also flips the relative magnitude of the output and productivity (panels a and c) effects between the heterogeneous and common RTS models.

In summary, allowing for heterogeneous RTS dampens the steady state costs of financial frictions for aggregate output, but increases the amount of misallocation and amplifies missing generation effects. Put together, the underlying mechanisms show the importance of properly understanding i) the general equilibrium role of factor prices, and ii) how responsive firms are to those factor prices, when modelling financial shocks in heterogeneous firm models. Winberry (2021) emphasised the role of factor prices, and a contribution of our paper is to emphasise that understanding firm-level responsiveness may be equally important.

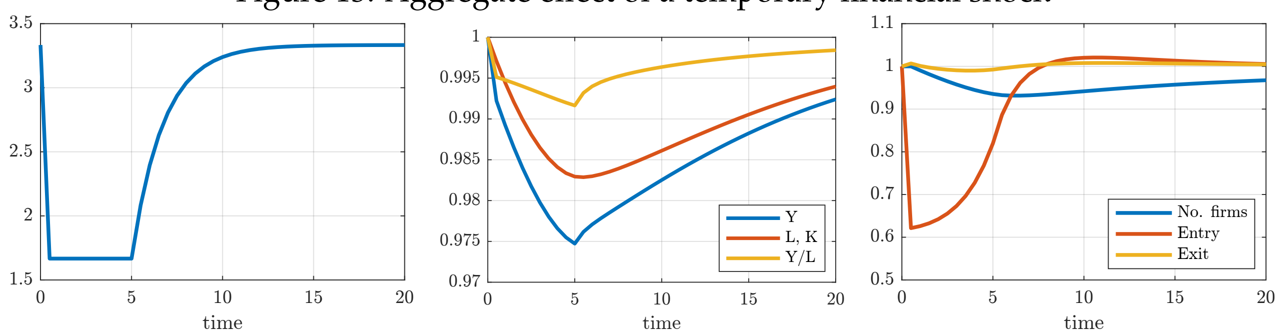
5.3 Propagation of financial crises

The second experiment uses the model to assess the costs of a temporary financial shock. To learn more about the interaction of finance and RTS with firm size and age, we present the how variables of interest respond, first in aggregate but then split across different criteria (young vs old, small vs large, constrained vs unconstrained). To provide the most detailed insight into the micro-level responses, we conduct a cohort by cohort exercise (Figure 15). All these exercises document the large heterogeneity in responses and show how financial crises lead to low entry resulting in slow recoveries. Like in the steady state experiment, introducing heterogeneous RTS does not change direction of the responses to shocks, but amplifies the micro differences: in response to a temporary financial shock currently constrained firms shrink and unconstrained firms expand, and this gap gets larger in the presence of heterogeneous RTS.

We consider an unanticipated shock, and at time 0 agents learn that the borrowing constraint will follow a path $\{\bar{\phi}_t\}$ which eventually converges back to the calibrated steady state value. We tighten the borrowing constraint by 50% for five years, and then allow it to converge smoothly back to the steady state. All prices adjust in general equilibrium during the transition, with the shock and key aggregates plotted in Figure 13, and prices and further aggregates in Figure 40. This setup following the spirit of the exercise in Khan and Thomas (2013), which was designed to mimic the Great Recession.

Figure 13 shows that the financial shock causes a large recession with a peak output fall of 2.5%, due to a fall in both input use and a nearly 1% fall in productivity. The fall in output is very persistent, recovering only half of its value in the decade after the shock has receded. At the same time, firm entry falls by nearly 40%, leading to a large and persistent decline in the number of firms. The overall firm exit rate actually falls in impact, but we show later that this reflects composition effects and the firm exit rate actually rises for most cohorts.

Figure 13: Aggregate effect of a temporary financial shock



(a) Financial shock ($\bar{\phi}$) (b) Output and productivity (c) No. firms, entry, exit

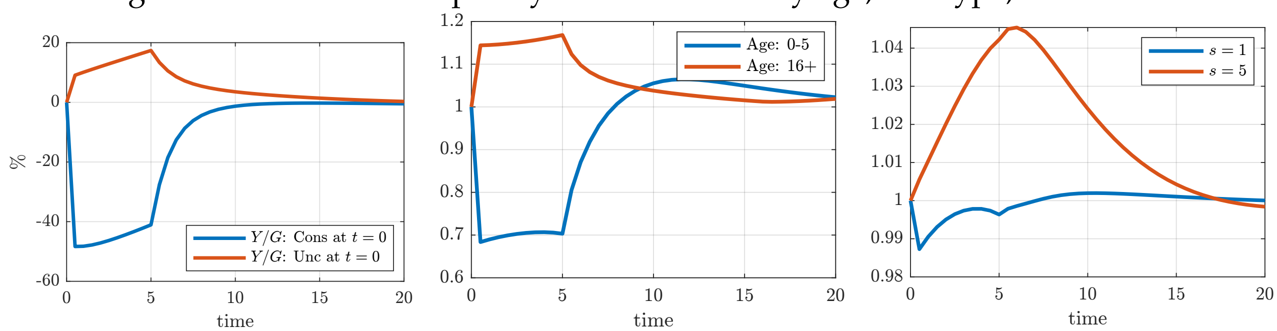
Note: Response of aggregates to a temporary 50% tightening of the borrowing constraint, whose path is given in panel (a). Panels (b) and (c) break down output and the number of firms respectively.

The decline in input use is driven by the financial shock directly reducing the ability of young, constrained firms to purchase capital. To demonstrate this, in Figure 14(a) we study the response of firms who were alive at the time the shock hit, and compare the evolution of their aggregate output per firm relative to a world where the shock did not hit. We categorise firms into those financially constrained or unconstrained the moment before the aggregate shock hit, and find a nearly 50% fall in output per firm for initially constrained firms, offset by a nearly 20% rise in output per firm from initially unconstrained firms.[45]

[45] An alternative is to simply look at the total output per firm of constrained and unconstrained firms at each period of the simulation. However, the number of firms who are financially constrained mechanically

Financially unconstrained firms respond to the shock by expanding, since wages and interest rates fall. This offsets the fall in aggregate output, but the large change in the allocation of inputs across firms is what causes measured productivity to fall. Since constrained firms also tend to be younger, these mechanics are also reflected in the firm age distribution (panel b), where young firms shrink and older firms expand. Exactly as in response to the permanent financial shock, it is the *large* old, unconstrained firms who are able to expand the most in response to the declining factor prices, due to their high RTS. This is reflected in the changes in the firm size distribution, and in panel (c) we see that output per firm rises for the high size type ($s = 5$) firms while it falls for the low size type ($s = 1$) firms.

Figure 14: Effect of a temporary financial shock by age, size type, and finance



(a) Y/G by constraint status (b) Output per firm by age (c) Output per firm by size type

Note: Response of average output per firm (Y/G) of various firm groups to a temporary 50% tightening of the borrowing constraint. Panels (b) and (c) plot averages by current firm age and firm size type respectively. Panel (a) plots only firms alive at the moment the shock hit, split by whether they were financially constrained or not.

Why does the economy respond so persistently to a temporary financial shock? This is driven entirely by cohort effects: there are fewer firms born in the years immediately after the shock, and the firms that are born never manage to outgrow the disadvantage of being born during a financial crisis. We document this in detail in Figure 15. In panels (a) and (b) we study the total output produced by different cohorts of firms at each point in time, where the x-axis is calendar time following the shock. The black dashed line plots total output for comparison, and each solid line tracks one cohort of firms from the year they are born until they reach 20 years old. We plot the deviation of the cohort's total output from its usual lifecycle, so a negative value means that this cohort is producing less output than they usually would at that age, in the absence of the shock. In panel (a) we plot cohorts of firms who were alive when the shock hit, and in panel (b) we plot the cohorts born in

shoots up when the constraint is tightened, and the resulting composition effect masks the underlying dynamics. Our approach avoids this problem.

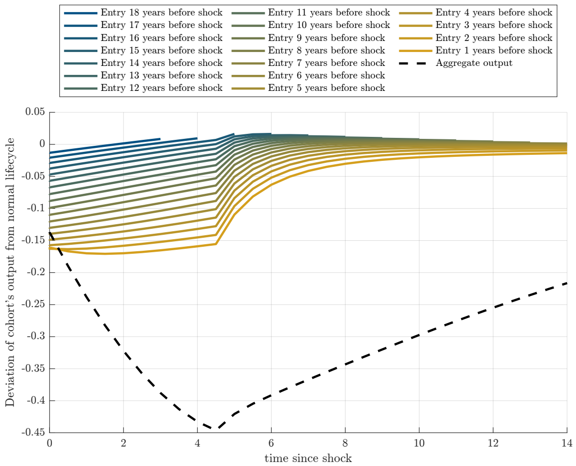
or after the year the shock first hit. We visualise cohorts across the two panels with older cohorts in blue merging to yellow and then red for the younger cohorts. In panel (b) firms have age 0 in the first time they appear in the plot, and then age along with calendar time as we move to the right. In panel (a) all firms are alive at the time of the shock.

Cohorts which were old when the shock hit (panel a, blue lines) contract their total output the least, or even expand, following the shock, because these tend to be financially unconstrained firms. These are the firms who crowd in and dampen the aggregate output response to the shock. In contrast, firms who were young when the shock hit (panel a, yellow lines) or born shortly after (panel b, yellow lines) contract total output the most following the shock, because they are young and financially constrained. Importantly, these effects are persistent: the total output in young cohorts remains below their usual lifecycle amount for the entire 14 year period plotted, despite the shock having faded. Thus, the financial shock has a permanent scarring effect on these generations, which is why the recovery from the crisis is so slow. Eventually cohorts born seven years after the shock hit (panel b, red lines) have higher total output than their usual lifecycle, as they benefit from the shock fading and lower factor prices, and help drive the recovery.

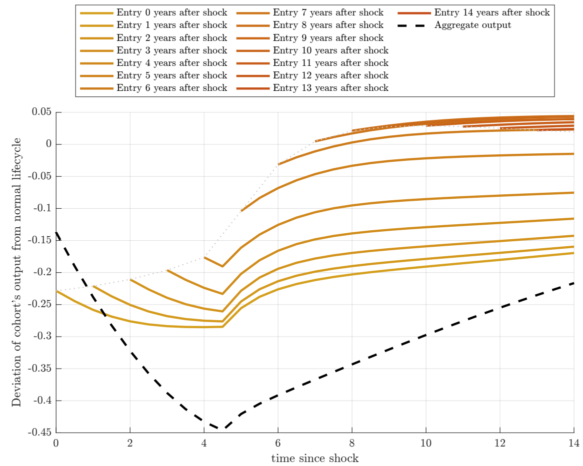
In panels (c) to (f) we decompose the total output changes into output per firm and the total number of firms. Panels (g) and (h) plot the exit rates of each cohort, again all relative to their usual lifecycle. For firms alive at the time the shock hits, the changes in the number of firms are minimal (panel e) because while exit rates do change they do so relatively little (panel g). It is instead changes in output per firm (panel c) which drive the response of total output for these firms. For firms born when or after the shock hits, we see important movements in both output per firm (panel d) and the number of firms (panel f). The cohorts born in the seven years after the shock hits see both smaller output per firm and a smaller number of firms. The smaller number of firms is driven mainly by the large decline in firm entry, but the rise in firm exit for these cohorts (panel h) also plays a role. These deficits persist over the whole lifecycle of the cohorts, even after the shock has receded. Thus, we see that the persistent fall in output in the model is driven by both a missing generation of young firms, and the persistent scarring of the young firms that did enter in the years just before and following the shock.

How does heterogeneity in RTS affect the transmission of the financial shock? To answer this question we compute the response of the economy with common RTS to the same shock, and compare it to our baseline economy in Figure 16. The results and intuitions mir-

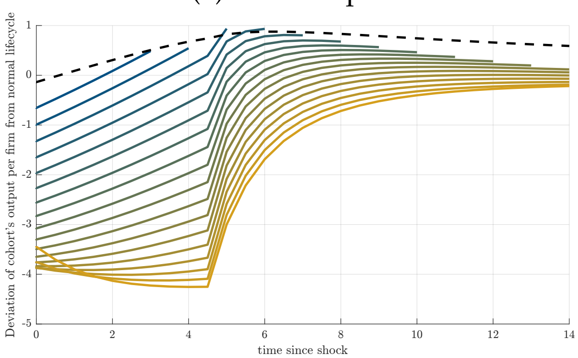
Figure 15: Cohort analysis: effect of a temporary financial shock



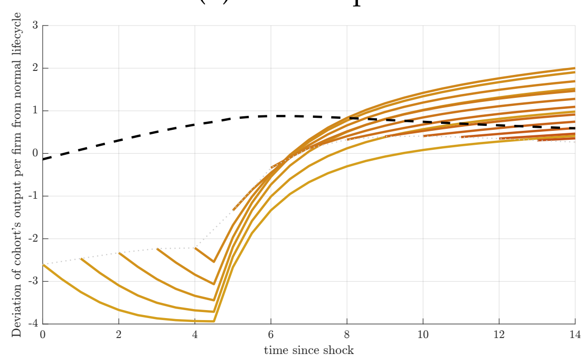
(a) Total output



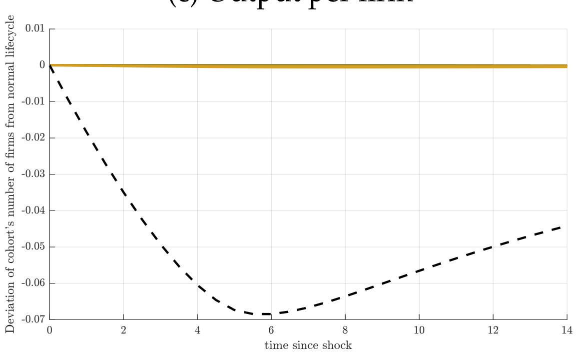
(b) Total output



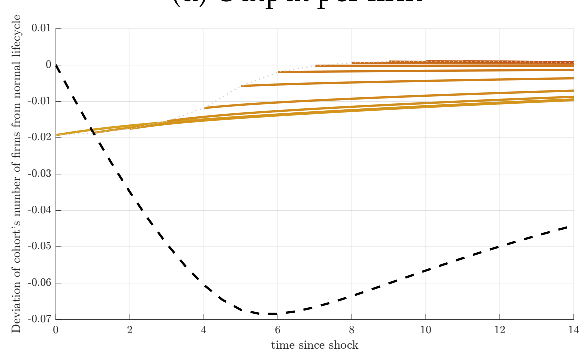
(c) Output per firm



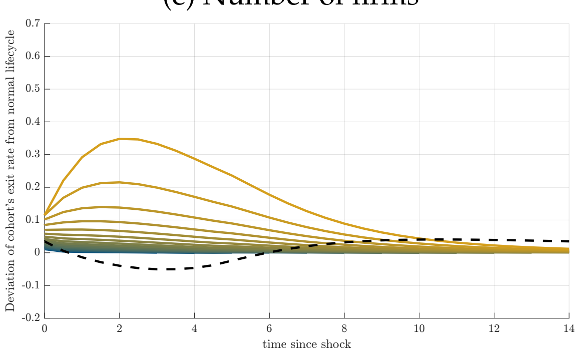
(d) Output per firm



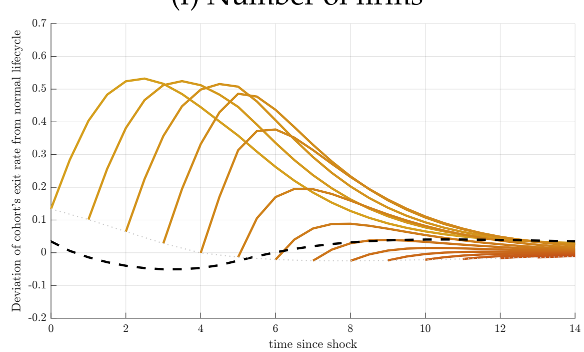
(e) Number of firms



(f) Number of firms



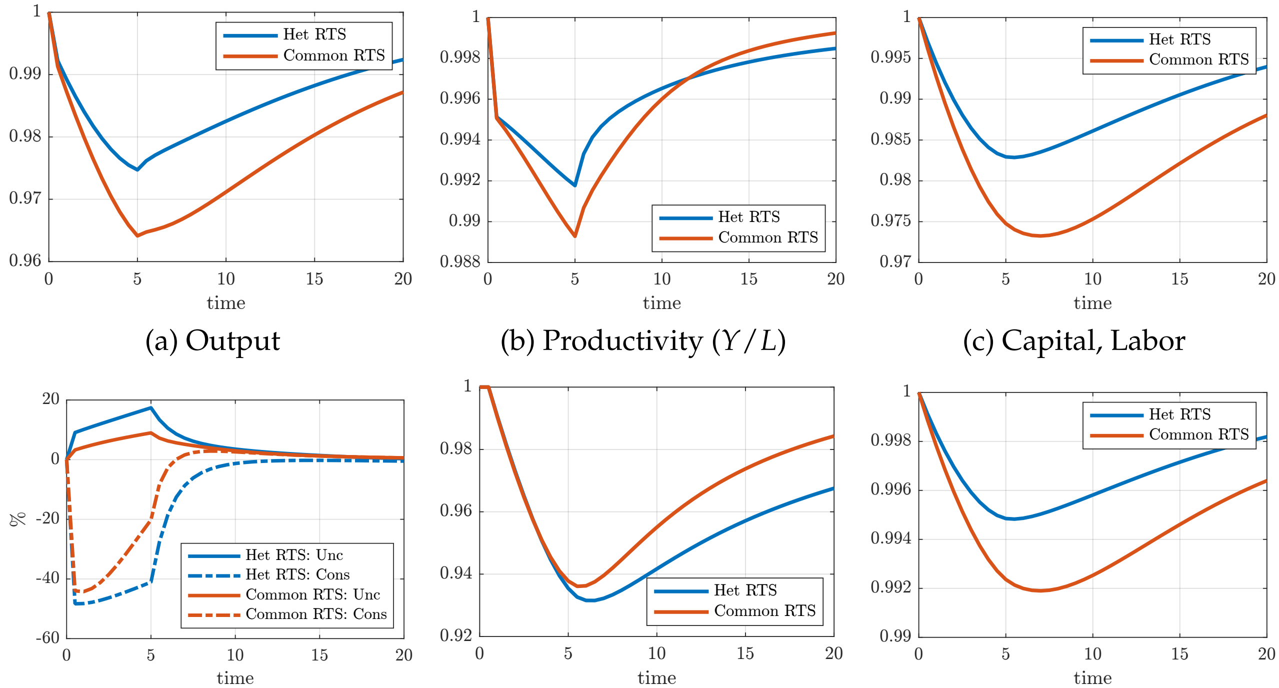
(g) Exit rate



(h) Exit rate

Note: Response of to a temporary 50% tightening of the borrowing constraint by firm cohort. Each line tracks a different cohort, with the value at time t giving the value of that variable (as deviation from their usual lifecycle) for that cohort at calendar time t. Left panels plot firms alive before the shock hit and right panels firms born afterwards.

Figure 16: Temporary financial shock: comparison to homogenous RTS model



(a) Output (b) Productivity (Y/L) (c) Capital, Labor

(d) Y/G by constraint status (e) No firms (f) Wage

Note: Response of economy to a temporary 50% tightening of the borrowing constraint in the baseline (blue line) common RTS (red line) models. Panel (d) plots only firms alive at the moment the shock hit, split by whether they were financially constrained or not.

ror exactly those from the permanent financial tightening experiment, and so we restate them only briefly here. Heterogeneity in RTS actually dampens the response of aggregate output to a financial shock (panel a), while amplifying the fall in entry and the number of firms (panel e). This leads to a more persistent fall in productivity (panel b). The dampened output response is because unconstrained old large firms are able to crowd in more when they have high RTS, and so offset more of the financial shock. This is shown in panel (d), where the offsetting response of initially unconstrained firms to the shock is more than twice as large in our model than the common RTS model. At the same time, heterogeneity in RTS amplifies the direct effect of the financial shock, as we see that the contraction in output per worker of initially constrained firms is much larger and more persistent in our model, as is the decline in entry and the number of firms. The key mechanism here is again that young cohorts suffer from the increased ability to crowd in that having large, high RTS firms in the model brings. Since the factor prices need to fall less to restore equilibrium (see e.g. the wage in panel f) young cohorts do not benefit from low factor prices following the shock. While heterogeneity in RTS is therefore good for overall GDP following a financial shock, it indirectly amplifies missing generation and scarring effects on young firms.

6 Conclusion

In this paper, we present a new fact about firm cyclicality using high quality registry data from the universe of Danish firms: among young firms, cyclicality decreases with firm size, while the opposite is true for old firms. We propose and test that two channels can explain this heterogeneity. Firstly, young firms are more likely to be financially constrained, making them more cyclical than old firms who have had more time to accumulate financial resources. Secondly, among older firms, larger firms are more cyclical because their higher returns to scale make them more sensitive to shocks even in the absence of financial frictions. These conjectures are confirmed using balance sheet data and production function estimation, where we provide direct evidence that leverage is decreasing in firm age, and RTS are increasing in firm size. Leverage is a stronger predictor of cyclicality for young firms than old, while the opposite is true for RTS. At the same time, financial frictions can leave scars even among older firms, as we show that many old large firms are born small, and the financial situation of entrants affects their odds of surviving into old age.

We build a quantitative heterogeneous firm model and show that the introduction of heterogeneity in RTS greatly changes the implications of financial frictions in this class of models. Typical heterogeneous firm models assume decreasing RTS, and given that we estimate large firms to be very close to constant RTS, it is perhaps not surprising that introducing even a small number of such firms changes how an otherwise standard model behaves. Many of our results stem from the fact that high RTS make firms much more sensitive to shocks or changes in factor prices. To the extent that this could even exaggerate the sensitivity of firms to temporary shocks, this elevates further the case for studying and modelling firm-level adjustment costs (Cooper and Haltiwanger, 2006), something we abstract from in the current study. On the other hand, our framework could have implications for other topics such as the transmission of monetary policy (Ottonello and Winberry, 2020) or recent declines in business dynamism (Pugsley and Şahin, 2019; Decker et al., 2020), where changes in the distribution of RTS could affect the responsiveness of firms to their environment.

References

Abraham, K. G. and Katz, L. F. (1986). Cyclical unemployment: Sectoral shifts or aggregate disturbances? *Journal of Political Economy*, 94(3):507–522.

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2021). Income and Wealth

Distribution in Macroeconomics: A Continuous-Time Approach. *The Review of Economic Studies*, 89(1):45–86.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Alder, M., Coimbra, N., and Szczerbowicz, U. (2023). Corporate Debt Structure and Heterogeneous Monetary Policy Transmission. Technical report.

Andersen, S. G. and Rozsypal, F. (2021). What can old firms tell us about the effect of age on firm size. working paper.

Bahaj, S., Foulis, A., Pinter, G., and Surico, P. (2022). Employment and the residential collateral channel of monetary policy. *Journal of Monetary Economics*, 131:26–44.

Begenau, J. and Salomao, J. (2018). Firm Financing over the Business Cycle. *The Review of Financial Studies*, 32(4):1235–1274.

Bernanke, B. and Gertler, M. (1989). Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, 79(1):14–31.

Bernanke, B., Gertler, M., and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1, Part C, chapter 21, pages 1341–1393. Elsevier, 1 edition.

Brinca, P., Chari, V. V., Kehoe, P. J., and McGrattan, E. (2016). Accounting for Business Cycles. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 1013–1063. Elsevier.

Cao, J., Hegna, T., Holm, M. B., Juelsrud, R., König, T., and Riiser, M. (2024). The Investment Channel of Monetary Policy: Evidence from Norway.

Casiraghi, M., McGregor, T., and Palazzo, B. (2021). Young Firms and Monetary Policy Transmission. IMF Working Papers 2021/063, International Monetary Fund.

Castillo-Martinez, L. and Bornstein, G. (2024). Firm Exit and Financial Frictions.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2012). Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities. *NBER Macroeconomics Annual*, 27:1–56.

Chodorow-Reich, G. (2014). The employment effects of credit market disruptions: Firm-level evidence from the 2008-09 financial crisis. *Quarterly Journal of Economics*, 129(1):1–59. Lead article.

Cloyne, J., Ferreira, C., Froemel, M., and Surico, P. (2023). Monetary policy, corporate finance and investment. *Journal of the European Economic Association*, (1911).

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

Correia, S. (2016). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report. Working Paper.

Covas, F. and Haan, W. J. D. (2011). The Cyclical Behavior of Debt and Equity Finance. *American Economic Review*, 101(2):877–899.

Crouzet, N. (2017). Aggregate Implications of Corporate Debt Choices. *The Review of Economic Studies*, 85(3):1635–1682.

Crouzet, N. and Mehrotra, N. R. (2020). Small and large firms over the business cycle. *American Economic Review*, 110(11):3549–3601.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing Business Dynamism and Productivity: Shocks versus Responsiveness. *American Economic Review*, 110(12):3952–3990.

Dinlersoz, E., Kalemli-Ozcan, S., Hyatt, H., and Penciakova, V. (2024). Leverage over the Life Cycle and Implications for Firm Growth and Shock Responsiveness. NBER Working Papers 25226, National Bureau of Economic Research, Inc.

Drechsel, T. (2023). Earnings-Based Borrowing Constraints and Macroeconomic Fluctuations. *AEJ: Macroeconomics, forthcoming*, (pdr141).

Duygan-Bump, B., Levkov, A., and Montoriol-Garriga, J. (2015). Financing constraints and unemployment: Evidence from the great recession. *Journal of Monetary Economics*, 75(C):89–105.

Díez, F. J., Fan, J., and Villegas-Sánchez, C. (2021). Global declining competition? *Journal of International Economics*, 132:103492.

Ferreira, M. H., Haber, T., and Rorig, C. (2023). Financial Constraints and Firm Size: Micro-Evidence and Aggregate Implications. *SSRN Electronic Journal*.

Fort, T. C., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2013). How Firms Respond to Business Cycles: The Role of Firm Age and Firm Size. *IMF Economic Review*, 61(3):520–559.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016.

Gao, W. and Kehrig, M. (2021). Returns to scale, productivity and competition: Empirical evidence from u.s. manufacturing and construction establishments.

Gavazza, A., Mongey, S., and Violante, G. L. (2018). Aggregate recruiting intensity. *American Economic Review*, 108(8):2088–2127.

Gertler, M. and Gilchrist, S. (1994). Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms*. *The Quarterly Journal of Economics*, 109(2):309–340.

Giroud, X. and Mueller, H. M. (2017). Firm leverage, consumer demand, and employment losses during the great recession. *The Quarterly Journal of Economics*, 132(1):271–316.

Griliches, Z. and Mairesse, J. (1995). Production functions: the search for identification. Technical report, NBER working paper series, wp No. 5067.

Grob, V. and Züllig, G. (2024). Corporate leverage and the effects of monetary policy on investment: A reconciliation of micro and macro elasticities. *SNB working paper series*.

Gupta, V., Koller, T., , and Stumpner, P. (2021). Reports of corporates'demise have been-greatly exaggerated. *McKinsey & Company Strategy & Corporate Finance Practice*.

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

Hubmer, J., Chan, M., Ozkan, S., and amd Guangbin Hong, S. S. (2024). Scalable vs. productive technologies.

Jeenas, P. (2019). Firm balance sheet liquidity, monetary policy shocks, and investment dynamics. Technical report, WP.

Jensen, T. L., Lando, D., and Medhat, M. (2017). Cyclicality and Firm Size in Private Firm Defaults. *International Journal of Central Banking*, 13(4):97–145.

Jermann, U. and Quadrini, V. (2012). Macroeconomic effects of financial shocks. *American Economic Review*, 102(1):238–71.

Jo, I. H. and Senga, T. (2019). Aggregate consequences of credit subsidy policies: Firm dynamics and misallocation. *Review of Economic Dynamics*, 32:68–93.

Jungherr, J., Meier, M., Reinelt, T., and Schott, I. (2022). Corporate Debt Maturity Matters For Monetary Policy. Technical report.

Khan, A. and Thomas, J. (2013). Credit shocks and aggregate fluctuations in an economy with production heterogeneity. *Journal of Political Economy*, 121(6):1055–1107.

Kiyotaki, N. and Moore, J. (1997). Credit Cycles. *Journal of Political Economy*, 105(2):211–248.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12(3/4):143–205.

Mian, A. and Sufi, A. (2014). What explains the 2007–2009 drop in employment? *Econometrica*, 82(6):2197–2223.

Moscarini, G. and Postel-Vinay, F. (2012). The contribution of large and small employers to job creation in times of high and low unemployment. *American Economic Review*, 102(6):2509–39.

Nikolov, B., Schmid, L., and Steri, R. (2018). The Sources of Financing Constraints. Swiss Finance Institute Research Paper Series 18-74, Swiss Finance Institute.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Ottonello, P. and Winberry, T. (2020). Financial heterogeneity and the investment channel of monetary policy. *Econometrica*, 88(6):2473–2502.

Poeschl, J. (2023). Corporate debt maturity and investment over the business cycle. *European Economic Review*, 152:104348.

Pugsley, B. W. and Ṣahin, A. (2019). Grown-up Business Cycles. *The Review of Financial Studies*, 32(3):1102–1147.

Reinhart, C. M. and Rogoff, K. S. (2014). Recovery from Financial Crises: Evidence from 100 Episodes. *American Economic Review*, 104(5):50–55.

Rovigatti, G. and Mollisi, V. (2016). PRODEST: Stata module for production function estimation based on the control function approach. Statistical Software Components, Boston College Department of Economics.

Sedláček, P. and Sterk, V. (2017). The growth potential of startups over the business cycle. *American Economic Review*, 107(10):3182–3210.

Sharpe, S. A. (1994). Financial Market Imperfections, Firm Leverage, and the Cyclicality of Employment. *American Economic Review*, 84(4):1060–1074.

Siemer, M. (2019). Employment effects of financial constraints during the great recession. *The Review of Economics and Statistics*, 101(1):16–29.

Smirnyagin, V. (2023). Returns to scale, firm entry, and the business cycle. *Journal of Monetary Economics*, 134:118–134.

Sufi, A. and Taylor, A. M. (2022). Financial crises: A survey. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 6 of *Handbook of International Economics: International Macroeconomics, Volume 6*, pages 291–340. Elsevier.

Winberry, T. (2021). Lumpy investment, business cycles, and stimulus policy. *American Economic Review*, 111(1):364–96.

ONLINE APPENDIX

A Data construction appendix

A.1 Additional information on dataset building process

We use the following datasets provided by Statistics Denmark (DST). All data is at the yearly frequency.

- FIGT ("Gammel Firmastatistik") + FIGF ("Gammel firmastatistik regnskabsdata", 1992-1999), FIRM ("Firmastatistik", 1999-2019): General firm-level data for sales and employment.
- FIRE ("Regnskabsstatistikken", 2001-2019): Firm-level accounting data such as assets or debt of non-financial corporations. Information in FIRE comes from *either a survey done by DST* with a rotating sample of approximately 9000 firms, *or directly from tax authority SKAT*. The sampling of firms in the survey depends on their size: firms with more than 50 workers are always included, 20-49 are included for 5 years every 10 years, firms with 10-19 workers are included every 2 years every 10 years

Overall, we have universal coverage of Danish firms regarding employment and sales as well as financial variables for the period starting from 2001 until 2019. Due to the stratified sampling of balance sheet data from FIRE, for smaller firms we do not have information for every firm in every year, but we do have data that has positive coverage even up to the smallest firms.

Subject to some minimal threshold on economic activity,[46] all firms are legally obliged to report data to SKAT or DST, which are then collected in these databases. We drop all observations that we deem as inactive by our definition, i.e. firms that provide no information about employment, sales, value added, or profits.

We also drop all firms that never in their life employ more than one worker.[47] Finally, we also drop firms listed as non-profits as well as entities controlled by government at any level. In our baseline exercises, we include only firms that do not exit in the current or the next year. We thus do not separately investigate the role of firm entry or exit in driving cyclicality.

Sometimes, information about a particular variable for a given firm is missing in the aforementioned registers. This is more likely for for financial rather than real variables, for smaller firms and for firms in the process of exiting. The year of exit also causes problems for variables that measure stock at a given point in time, rather than annual average. For these reasons, we only consider observations for firms that are not exiting in a given year when estimating the cyclicality of variables that is based on growth rates. For indicator of entry and exit, we do include the year of entry and exit in the regression samples.

Firm-level growth outcomes are defined by the normalized growth rates suggested by Haltiwanger et al. (2013): for any firm-level variable $x_{i,t}$, we measure growth from $t-1$ to

[46] In most situations, firms that report employment that corresponds to less than 0.5 full-time workers are considered inactive by DST, but still present in our data.

[47] We do this to eliminate sole proprietorship firms and also firms that exist due to tax optimization purposes.

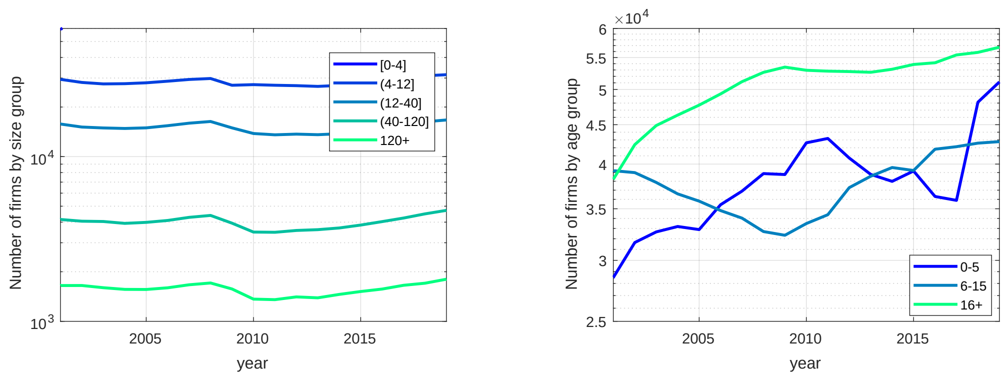
t as

$$\hat{g}_{x_{i,t}} \equiv \frac{x_{i,t} - x_{i,t-1}}{\frac{1}{2}(x_{i,t} + x_{i,t-1})},$$

where i indexes firms and t years. As discussed by Haltiwanger et al. (2013), this growth rate, which uses the average of the current and past value as the denominator, rather than just the past value, is more robust and typically has better properties in firm-level data. With this definition, the growth rate of any variable at the year of entry is 2 and it is equal to -2 at the year when firms exit.

A.2 Basic data

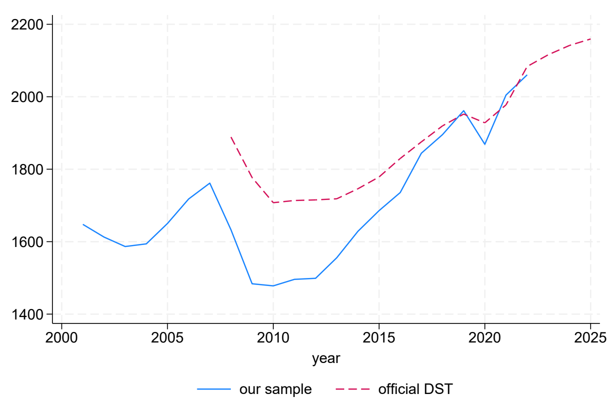
Figure 17: Number of firms in size and age bins over time



(a) Number of firms in each size group (b) Number of firms in each age group

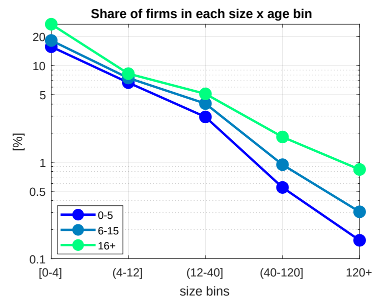
Note: This figure includes entering and exiting firms.

Figure 18: Employment coverage

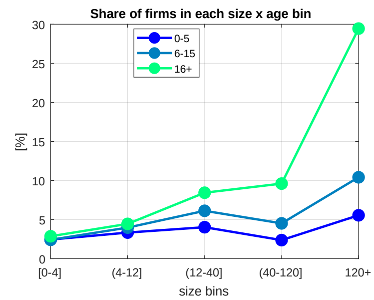


Note: This figure shows the comparison in coverage of people employed in our data (blue line) and the official Statistics Denmark reported number of people working in private companies (aggregated from monthly to anual frequency by averaging). Y-axis in thousands of workers.

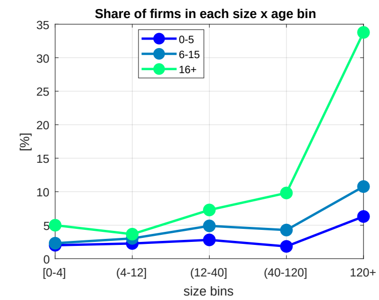
Figure 19: Share of firms across age and size bins



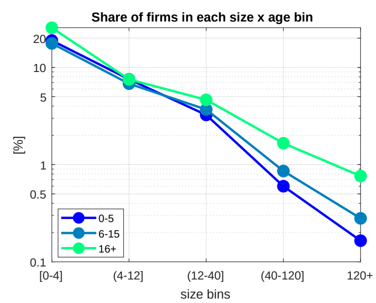
(a) firm share, excluding entry/exit



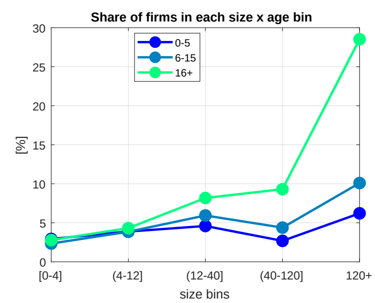
(b) employment share, excluding entry/exit



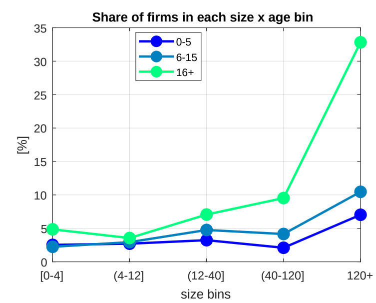
(c) sales share, excluding entry/exit



(d) firm share, including entry/exit



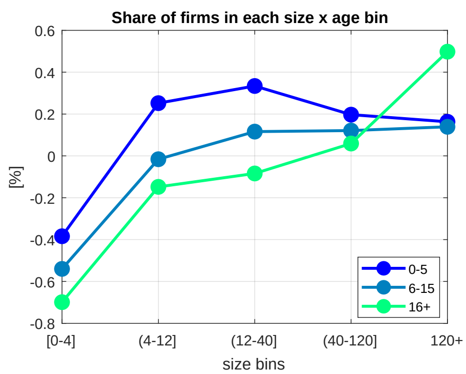
(e) employment share, including entry/exit



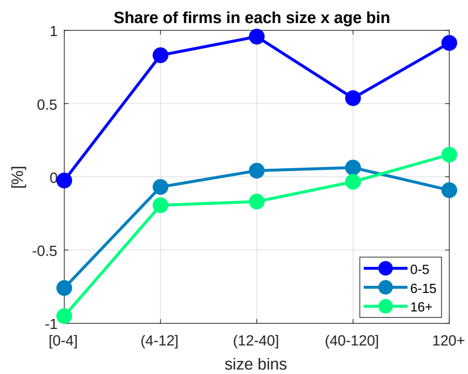
(f) sales share, including entry/exit

Note: Fraction of observations in each joint age-size bin (panels (a,d) or contribution to the aggregate employment (b,e) or sales (c,f)). First row shows the baseline results that exlude entry and exit, second row shows the contribution including entry and exit (by construction, firms in their exiting year have zero inputs and production). Panels (a) and (d) have log scale. Lines correspond to age bins and x-axis to size bins.

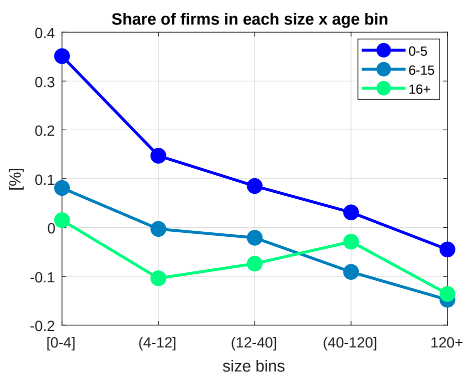
Figure 20: Contribution to job creation/destruction



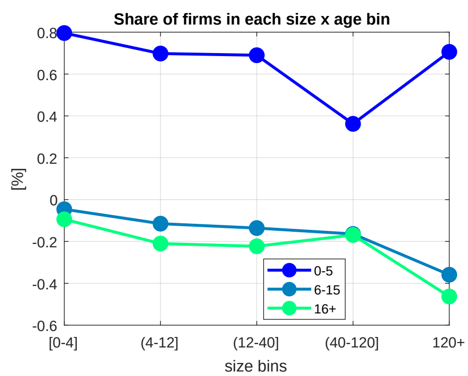
(a) current size, excluding entry/exit



(b) current size, including entry/exit



(c) start size, excluding entry/exit

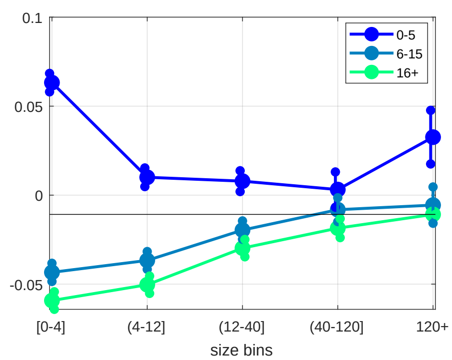


(d) start size, including entry/exit

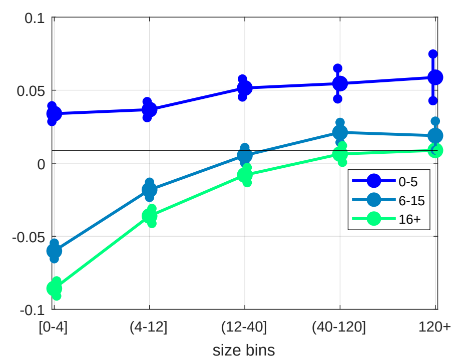
Note: Job creation defined as net change in employment for incumbent firms or as the current employment for the entrants. The left column reports the results excluding entrants and exiters (the baseline sample), the right column includes exiters and entrants. The first row uses the current firm size bin as the sorting variable on the x-axis, the second row uses the size at the time of entry as the sorting variable.

Note about large entrants. Who are the young firms that enter with size that places them already into the largest size bin? We provide information about the firm demographics in Section 4.4. In our sample an entry of a firm is recorded when a new firm registration number is established so we are not able to exclude the possibility that an exiting firms for whatever reason changes its registration number is treated as a new firm. However, we can show despite this, young large firms still behave on average differently than older large firms. For example, they grow much faster both in terms of employment and sales, as Figure 21 shows. In fact, across all size bins larger than five workers, the young firms are much more similar to each other in terms of their growth rates than to older firms. This suggests to us that mischaracterising some existing firms or mergers as entrants is not the driving the average behavior of large entrants.

Figure 21: Growth rates



(a) Employment



(b) Sales

Note: This figure present level coefficients $\alpha_{j,k}$ from regression equation (1) with growth rate of emplyoment (panel (a)) or sales (panel (b)) as the left hand side variable, shifted as described in Section 2.2. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms. The difference between this Figure and Figure 2 is in exclusion of all firms that do not report financial variables in a given year.

B Empirical results appendix

B.1 Regression details

Specifically when using growth rate of for example employment, for each size bin, α_{jk} captures the marginal effect on the average growth rate of firms of being in that size bin. For cyclicality, we are interested in the β_{jk} parameters, which capture how the firm-level growth rates, $\hat{g}_{x_{i,t}}$, are differently related to the aggregate growth rate, y_t. The interpretation of β_{jk} is that a 1pp increase in aggregate growth is on average associated with a "β_{jk}"pp increase in firm-level growth for firms in size group j and age group k, on top of any additional effects captured by sector-specific cyclicality. Thus, the β_{jk} captures the cyclicalities of each firm age-size group. Similarly, δ_l coefficients control the cyclicalities of the different sectors, to strip out the potentially differing average cyclicality of different industries.

The practical implementation of (1) is done via REGHDFE[48] package in STATA. The coefficients α and β are computed relative to the base group, which is the oldest-largest firms. To reconstruct the cyclicality of this group, we run the regression

$$\hat{g}_{x_{i,t}} = \sum_{l} (\gamma_l + \delta_l y_t) \mathbb{1}_{i \in S(l)}, \tag{9}$$

only for firms in the base group and use the average cyclicality across all sectors as a level shifter for all coefficients obtained from regression equation (1).

It is also important to remember that the regression coefficients with sectoral level and cyclicality controls do not necessarily look identical to the average values of any variable over the size x age bin distribution. This is because firms in any given sector are not necessarily unifirmly distributed across all size x age bins. For this reason, we prefer the regression setting to report the average effect of age and size. However, sometimes it is also useful to look at the distribution of specific firm characteristic, such as estimated returns to scale in Figure 5 in the cross section, without controlling for sectoral differences.

B.2 Additional cyclicality results

B.2.1 Regression table for the baseline results

[48] For details, see Correia (2016).

Table 2: Cyclicality regression coefficients

	(1) employment	(2) employment	(3) sales	(4) sales
[0-4] × y	-0.66***		-0.61***	
	(-5.69)		(-4.71)	
(4-12] × y	-0.15		-0.28*	
	(-1.29)		(-2.15)	
(12-40] × y	-0.06		-0.20	
	(-0.49)		(-1.53)	
(40-120] × y	0.07		-0.02	
	(0.54)		(-0.13)	
0-5 × y	-0.50		-0.46	
	(-1.36)		(-1.17)	
6-15 × y	0.48*		0.27	
	(2.00)		(1.10)	
[0-4] × 0-5 × y	2.46***		1.79***	
	(6.61)		(4.49)	
[0-4] × 6-15 × y	0.08		0.07	
	(0.31)		(0.28)	
(4-12] × 0-5 × y	1.30***		1.06**	
	(3.50)		(2.65)	
(4-12] × 6-15 × y	-0.26		-0.08	
	(-1.05)		(-0.33)	
(12-40] × 0-5 × y	1.01**		0.96*	
	(2.67)		(2.36)	
(12-40] × 6-15 × y	-0.22		-0.07	
	(-0.87)		(-0.29)	
(40-120] × 0-5 × y	0.91*		0.70	
	(2.14)		(1.54)	
(40-120] × 6-15 × y	-0.40		-0.19	
	(-1.44)		(-0.68)	
y		1.38***		1.89***
		(13.11)		(15.60)
Observations	1807893	18525	1569709	17064
Adjusted R^2	0.023	0.034	0.022	0.026
reg type	Size x Age	shifter	Size x Age	shifter

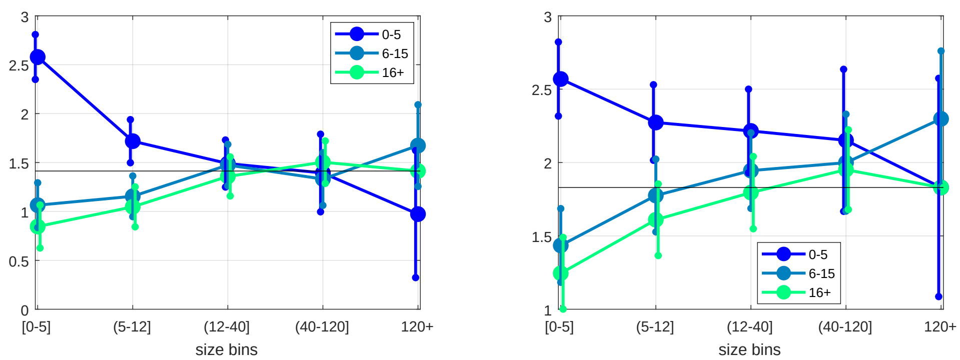
t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This Table present cyclicality coefficients $\beta_{j,k}$ from regression equation (1) as well as the shifter that corresponds to the average cyclicality of base group, which are the largest oldest firms (average across all sectors).

B.2.2 Results for the sample limited to firms reporting financial variables

Figure 22: Cyclicality, subsample with firms reporting financial variables

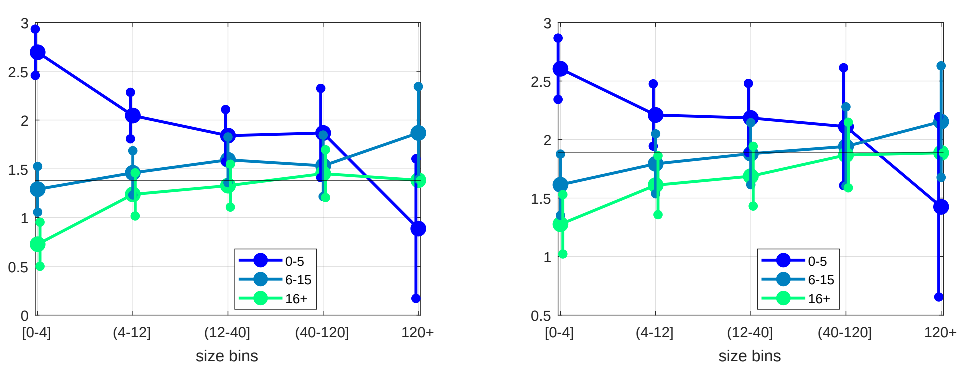


(a) Employment (b) Sales

Note: This figure present cyclicality coefficients $\beta_{j,k}$ shifted as described in Section 2.2. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms. The difference between this Figure and Figure 2 is in exclusion of all firms that do not report financial variables in a given year.

B.2.3 Cyclicality with assets based size definition

Figure 23: Cyclicality when size is defined by assets

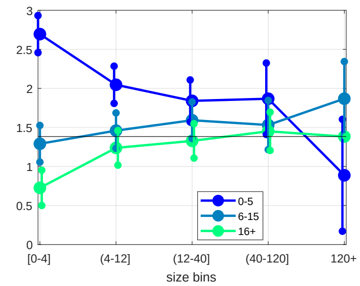


(a) Employment (b) Sales

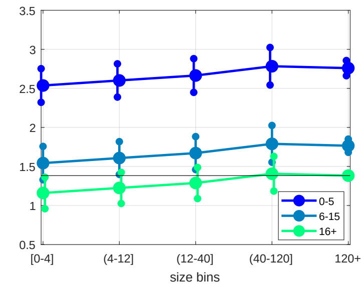
Note: This figure present cyclicality coefficients $\beta_{j,k}$ shifted as described in Section 2.2. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms.

B.2.4 Alternative specification of age and size interaction

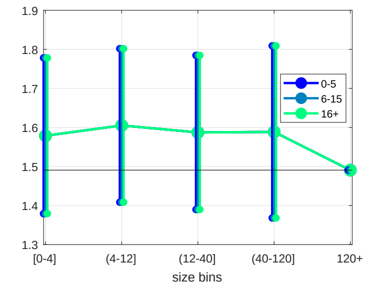
Figure 24: Alternative age specifications and the cyclicality of employment and sales



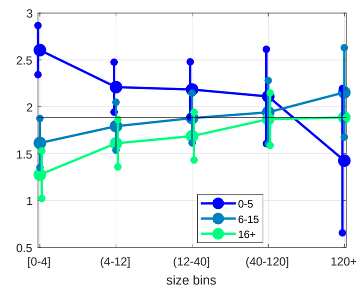
(a) employment, baseline



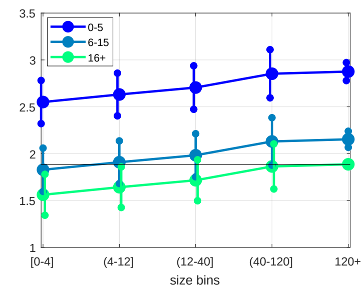
(b) employment, additive age



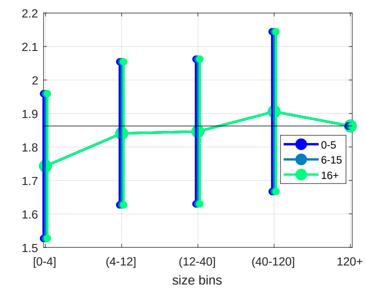
(c) employment, no age



(d) sales, baseline



(e) sales, additive age



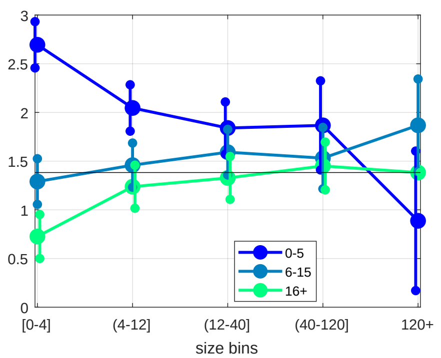
(f) sales, no age

Note: This figure present cyclicality coefficients $\beta_{j,k}$ (shifted as described in Section 2.2) for different specifications of age and size interaction in equation (1). The left column shows the baseline (and and size interacted, middle column shows the results where age and size enter additively and the right column shows the results without any age control. In the additive specification (middle column), the age effect has to be the same for any size group. The estimated age effects are driven much more by the small firms, because there are more numerous than old firms. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms.

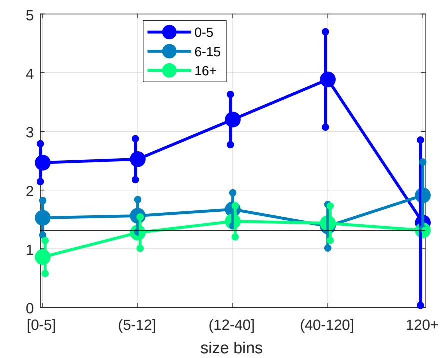
B.2.5 Cyclicality with entry and exit

For the baseline results, we exclude both entry and exit. In that sense, the reported cyclicality of employment and sales only capture the intensive margin of cyclical adjustment and we report the cyclicality of exit separately. However, the cyclicality to be computed even in the year of entry (where $\hat{g}_{it} = 2$) or exit ($\hat{g}_{it} = -2$). These results are following the logic of Siemer (2019). Results with inclusing entry and exit are reported in Figure 25.

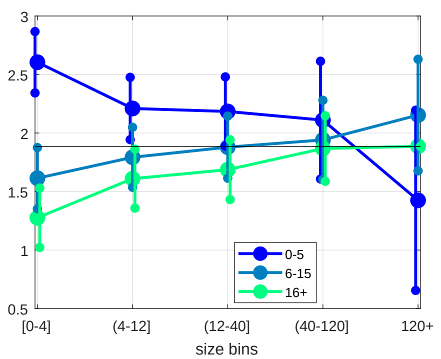
Figure 25: Cyclicality result: baseline vs sample with entry/exit



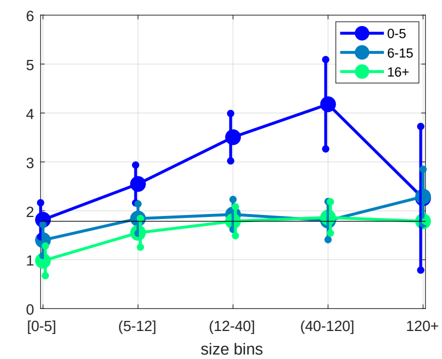
(a) Employment - baseline: no entry/exit



(b) Employment - with entry/exit



(c) Sales - baseline: no entry/exit



(d) Sales - with entry/exit

Note: This figure present cyclicality coefficients $\beta_{j,k}$ (shifted as described in Section 2.2) for baseline (left column) and including entrants and exiting firms in the sample (right column).

The largest change from the baseline results is the increased cyclicality of young firms in the third and fourth size bin (12-40 and 40-120). This suggest that these firms have more pro-cyclical entry and countercyclical exit than other firm types.

B.3 RTS and Leverage regression details

Table 3: RTS vs Leverage regression

	(1) employment	(2) employment	(3) employment	(4) sales	(5) sales	(6) sales
1st terc rts × y	0.00195	0.0232	-0.0639	-0.234**	-0.461**	-0.198
	(0.02)	(0.14)	(-0.48)	(-3.11)	(-3.21)	(-1.66)
3rd terc rts × y	0.0959	-0.0856	0.256**	0.0465	-0.309*	0.302**
	(1.61)	(-0.59)	(3.27)	(0.70)	(-1.97)	(3.28)
1st terc DA × y	-0.0336	-0.0416	-0.0390	0.127*	0.158	0.0674
	(-0.54)	(-0.27)	(-0.48)	(2.08)	(1.10)	(0.80)
3rd terc DA × y	0.535***	0.604***	0.310**	0.620***	0.825***	0.340**
	(6.85)	(3.91)	(2.72)	(8.68)	(6.02)	(3.01)
Observations	379100	104001	145647	357495	98480	137356
adj-r2	0.0171	0.0175	0.0215	0.0282	0.0298	0.0345
Firm age	all	young	old	all	young	old

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This regression table presents the results of regression (5). The base groups in the regresion is always the middle tercile of RTS or leverge.

Table 4: Slope effect of leverage and returns to scale on cyclicality

	(1) employment	(2) employment	(3) employment	(4) sales	(5) sales	(6) sales
y × rts_var	-0.0250	-0.0798	0.0538	0.0342	0.0400	0.0858*
	(-0.98)	(-1.49)	(1.45)	(1.53)	(0.86)	(2.54)
y × DA_var	0.150***	0.189***	0.0733*	0.101***	0.170***	0.0460
	(6.07)	(3.49)	(2.10)	(4.70)	(3.73)	(1.44)
Observations	376570	105084	142683	355216	99546	134602
adj-r2	0.0154	0.0158	0.0187	0.0265	0.0285	0.0316
Firm age	all	young	old	all	young	old

t statistics in parentheses

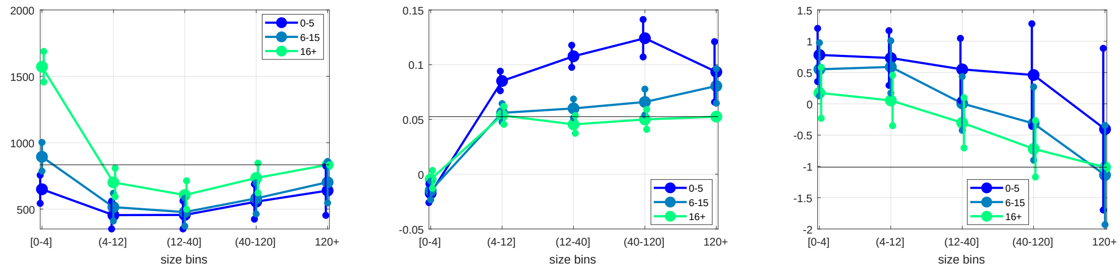
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This regression table presents the results of regression (5) but then the indicator variables of RTS and leverage quintiles are treated as continuous variables. As the bin numbers are treated as continuous variables, we only interpret these results as indicating whether cyclicality increases or decreases with returns to scale of the amount of leverage.

B.4 Net worth

Level and cyclicality.

Figure 26: Average levels, growth rates and cyclicality of net worth per worker

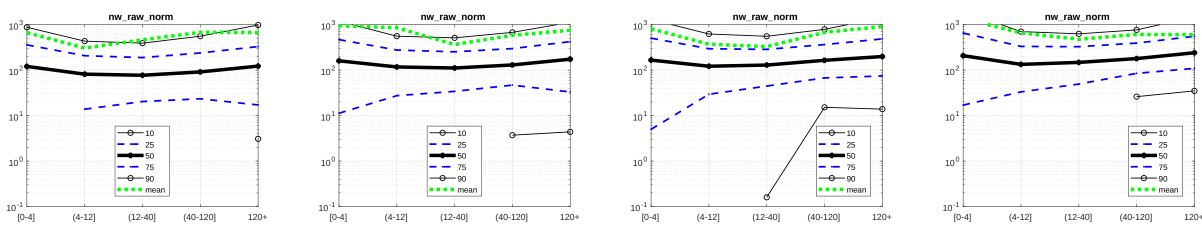


(a) level (b) growth rate (c) cyclicality of growth rate

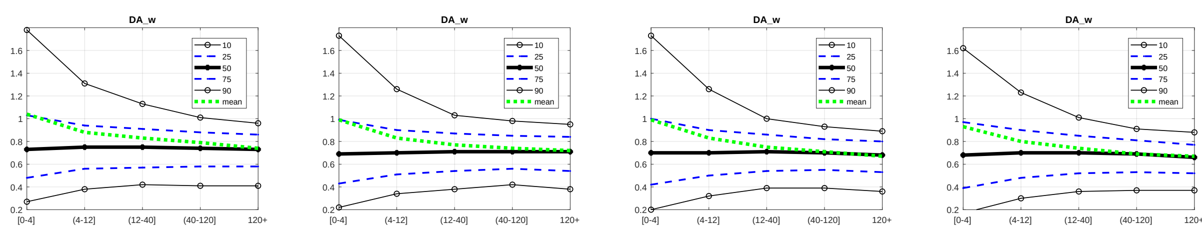
Note: This figure present cyclicality coefficients $\alpha_{j,k}$ and $\beta_{j,k}$ from regression (1) with net worth per worker (level or growth rate), shifted as described in Section 2.2. Coefficients belonging to the same age group are connected by colored lines. Vertical lines show 95% confidence intervals corresponding to H_0 of $\beta_{j,k} = \beta_{oldest,largest}$. The sample only includes incumbent non-exiting firms.

Distribution at different ages. Here we report the distribution of net worth per worker similar to the Figure 8, but with additional ages: 5, 10 15 and 20. Net worth per worker is very dispersed: the average is often close to the 90th percentile.

Figure 27: Net worth per worker at age 0,5,10,15 and 20.



(a) nw/w, age 5 (b) nw/w, age 10 (c) nw/w, age 15 (d) nw/w, age 20



(e) leverage, age 5 (f) leverage, 10 (g) leverage, 15 (h) leverage, 20

Note: Net worth is in thousand of krone per worker. More detailed legend description: thick dotted line: median, medium thinkness line: 25/75 percentiles, think balled line: 10/90th percentiles,: dotted green line: mean. Missing values for 10th percentile of net worth per worker indicate a negative value.

B.5 Net worth at entry and odds of surviving

Table 5: Net worth and survival, total over $0 - h$, coefficients from regression equation (6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	sur_1	sur_2	sur_3	sur_4	sur_5	sur_7	sur_9	sur_11	sur_13	sur_15
1.nw_n	-0.00539***	-0.0197***	-0.0482***	-0.0837***	-0.0978***	-0.102***	-0.105***	-0.0967***	-0.0796***	-0.0518***
	(-4.28)	(-7.60)	(-13.14)	(-18.79)	(-19.63)	(-18.07)	(-17.68)	(-15.97)	(-13.35)	(-9.56)
2.nw_n	0.00185	-0.00178	-0.00954**	-0.0251***	-0.0268***	-0.0332***	-0.0409***	-0.0392***	-0.0261***	-0.0105
	(1.67)	(-0.73)	(-2.76)	(-5.94)	(-5.61)	(-5.98)	(-6.83)	(-6.34)	(-4.25)	(-1.86)
4.nw_n	0.000340	0.000802	0.0141***	0.0218***	0.0287***	0.0379***	0.0383***	0.0373***	0.0382***	0.0301***
	(0.29)	(0.33)	(4.31)	(5.55)	(6.35)	(7.09)	(6.50)	(6.02)	(6.05)	(5.06)
5.nw_n	-0.000279	0.0118***	0.0306***	0.0418***	0.0505***	0.0708***	0.0784***	0.0809***	0.0772***	0.0532***
	(-0.23)	(5.09)	(9.49)	(10.67)	(11.11)	(13.14)	(13.10)	(12.72)	(11.71)	(8.33)
_cons	0.988***	0.946***	0.887***	0.817***	0.763***	0.660***	0.566***	0.472***	0.365***	0.253***
	(1199.98)	(557.31)	(370.76)	(284.10)	(232.33)	(171.23)	(135.04)	(108.15)	(83.16)	(61.72)
N	87199	81968	76775	66098	62991	57292	52205	47406	42029	37040
adj-r2	0.00998	0.0735	0.126	0.261	0.244	0.243	0.255	0.280	0.307	0.354
Sector FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Age 0 size FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Note: sur_*X* gives the survival odds *from entry until* horizon *X* and nw_w stands for *net worth per worker* with 1 being the lowest quintile and 5 being the highest quintile. Middle quintile used as the base group.

Table 6: Net worth and survival, marginal at horizon h , coefficients from (6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	sur2_1	sur2_2	sur2_3	sur2_4	sur2_5	sur2_7	sur2_9	sur2_11	sur2_13	sur2_15
1.nw_n	-0.00539***	-0.0149***	-0.0311***	-0.0362***	-0.0266***	-0.0135***	-0.0204***	-0.00876	-0.00776	-0.00357
	(-4.28)	(-6.42)	(-10.51)	(-10.84)	(-7.55)	(-3.64)	(-4.90)	(-1.95)	(-1.42)	(-0.55)
2.nw_n	0.00185	-0.00348	-0.00703*	-0.0152***	-0.00491	-0.00529	-0.0149***	-0.00480	0.00181	-0.00198
	(1.67)	(-1.59)	(-2.57)	(-4.89)	(-1.53)	(-1.50)	(-3.72)	(-1.11)	(0.36)	(-0.32)
4.nw_n	0.000340	-0.0000722	0.0116***	0.00861**	0.00954**	0.00818*	0.00444	0.00695	0.00848	0.00599
	(0.29)	(-0.03)	(4.58)	(3.08)	(3.18)	(2.52)	(1.25)	(1.78)	(1.80)	(1.03)
5.nw_n	-0.000279	0.0123***	0.0213***	0.0160***	0.0153***	0.0148***	0.0130***	0.0135***	0.0178***	0.0134*
	(-0.23)	(6.14)	(8.67)	(5.79)	(5.13)	(4.68)	(3.74)	(3.58)	(3.92)	(2.36)
_cons	0.988***	0.958***	0.941***	0.939***	0.944***	0.949***	0.954***	0.956***	0.955***	0.953***
	(1199.98)	(633.46)	(500.77)	(455.45)	(426.40)	(394.22)	(367.63)	(329.62)	(269.65)	(219.99)
N	87199	80857	72163	57143	50217	39313	30537	22971	15852	9761
adj-r2	0.00998	0.0691	0.0820	0.151	0.0556	0.0497	0.0435	0.0771	0.0565	0.202
Sector FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Time FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Age 0 size FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

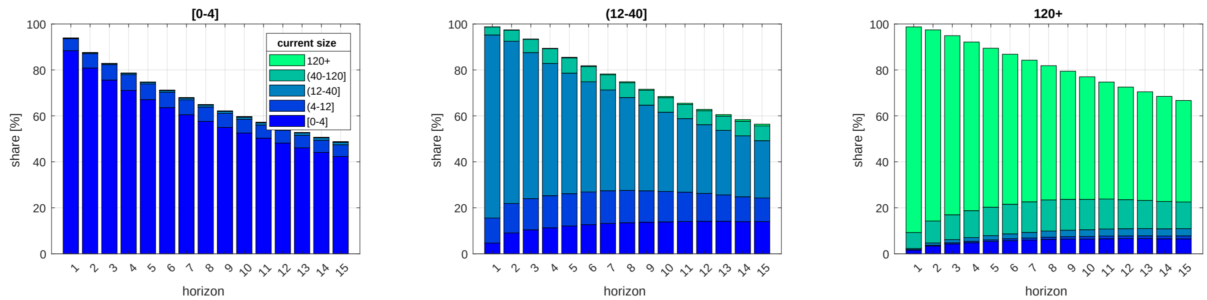
t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Note: sur_*X* gives the survival *at* horizon *X* and nw_w stands for *net worth per worker* with 1 being the lowest quintile and 5 being the highest quintile. Middle quintile used as the base group.

B.6 Demographics

Figure 10 presented the relationship between the current size and the starting size bin (both ways). Additionally, here we also present the transitional odds when ignoring firm age.

Figure 28: Unconditional transition rates based on the current size bin

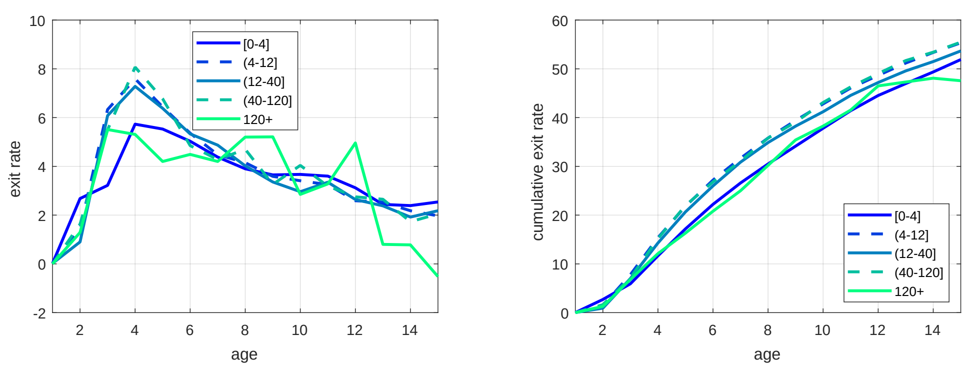


(a) *started* in the bottom bin (b) *started* in the middle bin (c) *started* in the top bin

Note: This figure shows the transition odds based on the current size of a firm going x years to the future. Unlike in the first row of figure 10, here we do not limit ourselves only to current entrants.

Conditional on *starting* size, exit rates follow similar pattern, spiking at age between 3-5 and falling afterwards. This means, that unconditionally on age, size *at the time of entry* does not predict exit rate very strongly (because of our definition of exit, all firms exit with zero workers and hence at the point of exit they are in size bin 1).

Figure 29: Exit rates conditional on entry size bin



(a) per period (b) cumulative

Note: This figure shows implied exit rates conditional on size at the point of entry, computed from the data underlying Figure 10. Note that the exit year is the year *after* the final positive sales or employment is observed. For example, a firm that enters (age 0) in March 2005 and is active for 9 months until December 2005, would be recorded as exited in 2006 when aged 1.

C Returns to scale estimation appendix

Here we present the differences in estimation methodologies. Specifically, we compare our baseline GNR to OP and to LP with ACF. There are two specification differences between our implementations of the baseline and the two alternative methods. The latter two use value added instead of sales as the left-hand side variable (and therefore intermediates do not enter as a factor of production) and they are estimated using standard Cobb-Douglas production functions (unlike GNR where we use a translog production function).

The general pattern in returns to scale is consistent across all specifications: RTS are increasing in size with the largest difference being between the first two size groups. The results for productivity are not the same between the baseline GNR and OP/LP. While it is true that the largest firms are always the most productive, the methods do not agree on the gradient of productivity with respect to size among the smallest firm bins. However, this discrepancy is much smaller in the cross-section of the results than when the size and age effects are estimating in a regression setting with additional sectoral fixed effects.

C.1 Estimation Methodology

Production function estimation is challenging because of long-recognized endogeneity problems (Marschak and Andrews, 1944; Griliches and Mairesse, 1995): it is at least plausible that firms are able to respond to shocks that are unobservable by the econometrician. If so, firms are likely to choose to use more of the flexible inputs at favorable times. For the econometrician, however, this introduces correlation between some of the inputs and the error terms, leading to endogeneity bias.

The methods such as OP, LP, ACF, and GNR recognize this bias and use different insights from the firm problem to overcome it. The basic insight is that some factors such as capital are likely fixed in the short term and so they cannot respond to current productivity shocks. In contrast, other inputs (labor for OP and LP, intermediates for GNR) are flexible and hence it is possible for a firm to change their usage to respond to a shock. Furthermore, the economic problem of a profit maximizing firm is used to derive a additional set of moments. In particular, if a firms is able to observe a part of the productivity realization, it would react with its investment (as in OP) or with intermediate inputs (LP and GNR). Given this insight and the assumptions about the stochastic process of firm productivity, one can derive conditions that can be used to identify the parameters of the production function.

Table 7: Overview of methods used to estimate returns to scale

method	output variable	fixed variables	free	proxy	prod function
Gandhi et al. (2020)	sales	capital, labor	intermediates	intermediates	translog
Olley and Pakes (1996)	VA	capital	labor	investment	C-D
Levinsohn and Petrin (2003)	VA	capital	labor	intermediates	C-D

Implementation. For OP and LP we use`prodest` package (Rovigatti and Mollisi, 2016). For GNR we base our estimation on their replication package (Gandhi et al., 2020). To follow the standard setting, we use translog production function for GNR estimated on sales and Cobb-Douglas for OP and LP with value added as the left-hand side variable.

C.2 RTS Data construction

When estimating value added production functions, the following data is used. For the OP method, the "proxy" variable is investment, whereas for the LP method, the proxy variable is intermediate inputs. For the "free" variable, labor, full-time equivalent employees are used.

C.2.1 Variables

Inputs. Capital is always a "state" variable and it is constructed by the perpetual inventory method using investments and the bookkeeping value of capital. Specifically, an initial capital level is determined as the highest of either reported capital or investments divided by an assumed depreciation rate of 10%. For subsequent years, capital is determined by the highest of either reported capital or the depreciated capital determined in the previous year plus investments. If there are any gaps in the series of any variable used in the production function estimation, the capital series is re-based with a new firm identifier. We use the full time equivalent number of workers (the main cyclicality variable) to measure firm employment input.

Intermediates. Intermediates are measured as "Purchase of raw materials, consumables, finished products and packaging" minus the change in stock of inventories from FIRE register provided by Statistics Denmark. Specifically, the variable definitions change over time, so use `khr-dlg` for years prior 2004, `khre+kvv-dlg` for 2004 to 2016 and `frhe+fvv` from 2017 on-wards. To compute the real use of inventories, we deflate the nominal values by sector-specific producer price index for inputs obtained from IO data collected by Statistics Denmark.

Sectoral deflation is applied to value added, intermediates, investments, and capital using sector-level producer price index (PPI) data obtained from Statistics Denmark. Sectoral PPI data is published for varying levels of classification specificity. If several levels of PPI data are published for the sector of a given firm, the level with the highest frequency is used. If more levels have the same frequency, the most specific level is used. For the baseline results using GNR methodology, we construct the intermediate share as the share of nominal intermediate inputs relative to nominal sales.

C.2.2 Sectoral prices

First, we introduce some notation. There are N sectors, in each sector there are N_s firms. Firms are indexed by i, so that $\forall i = 1, \ldots, N_s$ in sector s we have

$$sales_{i,t} = p_{i,t} y_{i,t},$$
$$interm_{i,t} = \phi_{i,t} m_{i,t}.$$

sales and (expenditures on) intermediate inputs *interm* are directly observable in the micro data. However, we are interested in the physical m to properly estimate the production function and hence returns to scale. Sales and intermediates can be further split by the trading partner (firm j):

$$sales_{i,t} = p_{i,t}y_{i,t} = \sum_j p_{i,j,t}y_{i,j,t},$$
$$interm_{i,t} = \phi_{i,t}m_{i,t} = \sum_j \phi_{i,j,t}m_{i,j,t}.$$

Note that $p_{i,j,t} = \phi_{j,i,t}$ and $y_{i,j,t} = m_{j,i,t}$ by definition. In aggregate, we observe producer price indeces for each sector p_t^s. Second, IO table captures the value of the flow goods purchased in sector s_{dest} coming from s_{source} in a given year:

$$f_{s_{source},s_{dest}} = f(s,d) = \sum_{i \in dest, j \in source} \phi_{i,j,t}m_{i,j,t} = \sum_{i \in dest, j \in source} p_{j,i,t}y_{j,i,t}$$

Now we introduce two assumptions about pricing behavior and the use of intermediates, which we use to reverse engineer intermediate input quantities.

Assumption 1 *All firms in a given sector sell output at the same price, regardless of the buyer:* $\forall i \in$ *sector* s *and* $\forall j \in d : p_{i,j,t} = p_t(s)$ *and* $\phi_{j,i,t} = \phi_t(s)$.

Assumption 2 *All firms in a given sector use the intermediates in the same proportion.*

Using the first assumption, all firms that are buying intermediate inputs of sector s are buying it at the same price, hence

$$f_t(s,d) = \phi_t(s) \sum_{i \in dest, j \in source} m_{i,j,t} = p_t(s) \sum_{i \in dest, j \in source} y_{j,i,t}$$

if $p(s)$ is the producer price index associated with production of y^s for intermediate use, we have $y(s,d) = m(s,d)$ and $\phi(s) = p(s)$, so the previous equation can be simplified to:

$$f(s,d) = p(s)m(s,d).$$

Note that two out of the three terms are observable in the data: $f(s,d)$ is observable from IO tables, and p_t^s is the producer price in sector s.Let's define the total expenditures on intermediates in sector d as $f(d) = \sum_s f(s,d)$. By Assumption 2, the share of resources on intermediates from particular sector is the same in the aggregate and on firm level:

$$\frac{f(s,d)}{f(d)} = \frac{\phi(s)m_i(s,d)}{\sum_s \phi(s)m_i(s,d)}$$

however, note that $\sum_s \phi(s)m_i(s,d)$ is the value of intermediate inputs which is observed. So this equation can be re-organised to get

$$m_i(s,d) = interm_{i,t}\frac{f(s,d)}{f(d)}\frac{1}{p(s)}$$

which gives the (physical) amount of the intermediate output of type s used by a firm that uses $interm_{i,t}$ in sector d.

The total (physical) use of intermediates is then

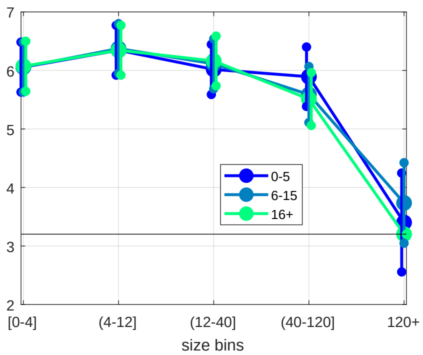
$$m_{i,t} = \sum_s m_i(s,d) = \frac{interm_{i,t}}{f(d)} \sum_s \frac{f_t(s,d)}{p_t(s)}$$

and all the terms on the right hand side of the equation are observable.

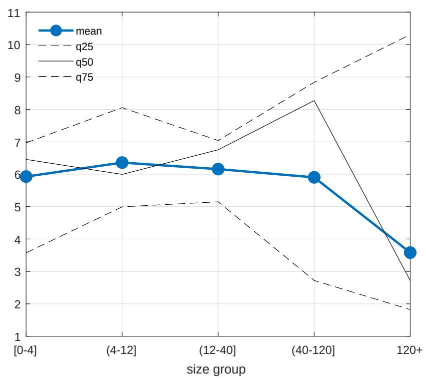
C.3 Production function estimation, results for productivity

While we find that RTS are increasing with firm size, the relationship between size and productivity is non-monotonic. It is initially increasing, but may even decline with size for large firms as shown in Figure 30.

Figure 30: Productivity by current size



(a) Size and age effects, productivity



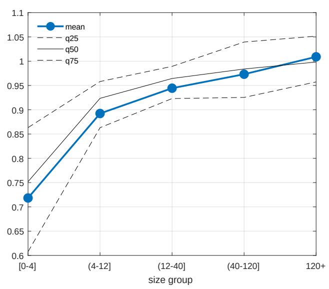
(b) Distribution of estimated productivity

Note: Firm level productivity is computed as the residual from production function estimation. Panel (a) shows the age and size effect from regression (1), panel (b) shows the median, mean and the interquartile range of estimated productivity across the size bins.

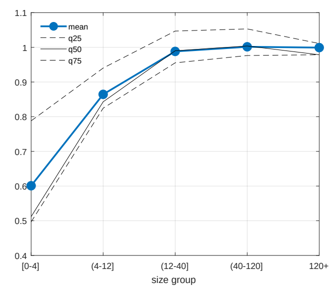
C.4 Alternative partitioning of the firm sample when estimating rts

Partitioning the sample is an important step in estimating the production function as within each "cell", the production function coefficients will be shared by all firms. Furthermore, in the case of Cobb-Douglas production function, $Y(L,K) = AL^{\beta}K^{\alpha}$, the returns to scale $RTS = \alpha + \beta$ will be also the same for all firms within the cell. For translog production functions, $\log Y = \alpha_0 \log A + \sum_{i=1}^{N} \alpha_i \log X_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} \log X_i \log X_j$, the returns to scale are not constant and change with the level of inputs X. This implies that for Cobb-Douglas production function estimated at 2 digit sectoral classification, with no further partitioning, any gradient with respect to size can only come from the changes in sectoral composition across the size distribution. As discussed in the main text, this is why we additionally partition our estimates by maximum-size groups within each sector. Even with a Cobb-Douglas production function and focusing within a single sector this approach would allow us to investigate differences in RTS across firm size groups.

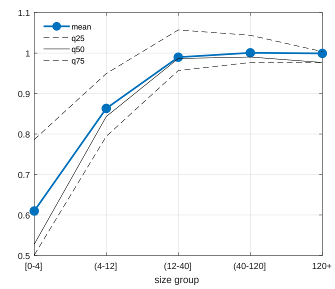
Figure 31: Returns to scale by current size by method of estimation



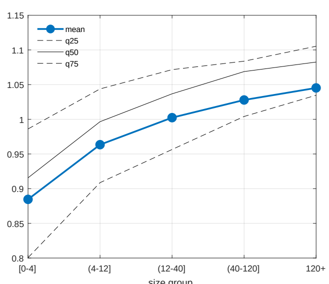
(a) GNR, max (baseline)



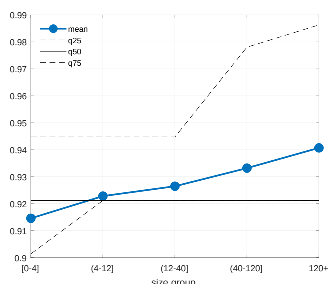
(b) OP, max



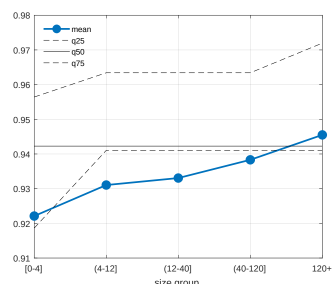
(c) LP, max



(d) GNR, all



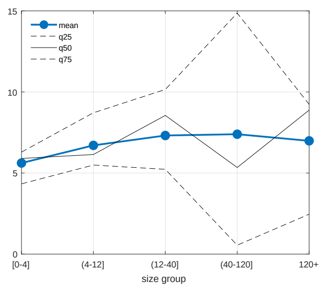
(e) OP, all



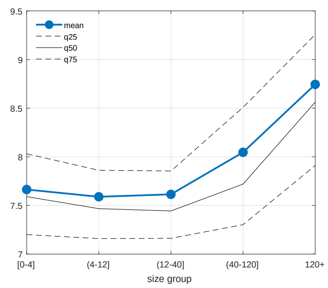
(f) LP, all

Note: This figure shows RTS estimates by size for three estimation methods (rows: GNR-baseline, OP and LP) and two sample partitioning: by max size (first row - baseline) and no additional partitioning (beyond 2 digit sectors).

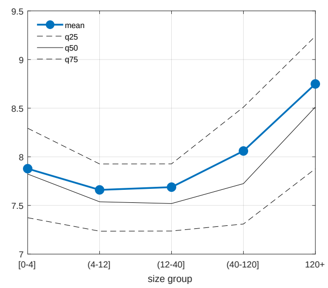
Figure 32: Productivity by current size by method of estimation



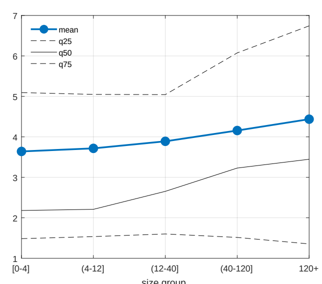
(a) GNR, max (baseline)



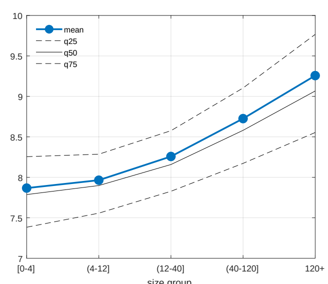
(b) OP, max



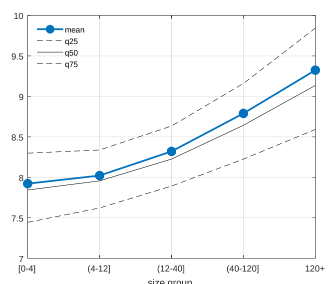
(c) LP, max



(d) GNR, all



(e) OP, all



(f) LP, all

Note: This figure shows productivity estimates by size for three estimation methods (rows: GNR-baseline, OP and LP) and two sample partitioning: by max size (first row - baseline) and no additional partitioning (beyond 2 digit sectors). The level difference between GNR and OP/LP is given by a difference in the underlying production function.

C.5 Relation to existing empirical findings about RTS heterogeneity

A recent literature, some emerging since we submitted the paper, provides new complementary evidence on heterogeneity in returns to scale. In an early contribution, Gao and Kehrig (2021) estimate RTS at the industry level and show that larger industries have higher RTS. Smirnyagin (2023) also shows that larger industries have higher RTS using US manufacturing data, and provides evidence that the entry rate of firms in higher RTS industries is more procyclical. Relative to these approaches we provide firm-level RTS estimates using the Gandhi et al. (2020) approach and/or OP/LP/ACF additionally running the estimation splitting the firms by their maximum achieved size. Hubmer et al. (2024) estimate RTS at the firm level also using the approach of Gandhi et al. (2020). We go beyond Gandhi et al. (2020) by also partitioning the sample by the maximum size a firm achieves during its lifetime. Including our own, all these papers find that large firms (or industries with higher firm size) have higher RTS, which is reassuring validation of our empirical finding.

Another way to interpret heterogeneous RTS is as a heterogeneous *demand elasticity*. In particular, with a CES demand curve it is well-known that the demand elasticity acts isomorphically to RTS in the production function, if one focuses only on data for sales (and not real quantities) as we do. In our model, we could assume that all firms have constant RTS in production but heterogeneous demand elasticities which drive differing levels of RTS in revenue. This would mean that small firms charge higher *markups* than large firms, based on the usual result that inelastic demand leads to higher markups: recall the standard static result that the optimal markup is equal to $\mu = \epsilon/(1-\epsilon) = 1/\theta$. This interpretation allows us to use data on average markups by firm size to interpret our finding of more decreasing returns to scale at small firms. If markups are higher at small firms than large firms, this would provide additional support. Indeed, this appears to be the case in the data. Díez et al. (2021) compute markups for both private and public firms using the global Orbis dataset, for a set of firms accounting for 70% of global GDP. They find that there is a U-shaped relationship between markups and firm size, and that markups are decreasing with firm size for most of the size distribution: *"Contrary to common wisdom, we find that, unconditionally, smaller firms have higher markups even within narrowly defined industries—only when we focus on very large firms we do find a positive relation... markups first decrease with firm size and only when a (fairly large) size threshold is reached, markups start increasing with firm size"* (p2).

D Analytical model appendix

D.1 Model details: Firm size, firm age and life-cycle dynamics

Firms receive an initial equity injection at birth giving them initial net worth n_0. This along with their permanent productivity and returns to scale (z, η) is drawn from a CDF $G^e(n_0, z, \eta)$. Firms enter at exogenous rate μ_0 and exit at exogenous rate ζ, so the total mass of firms M is fixed at $M = \mu_0/\zeta$, and we suppress the firm subscript $i \in [0, M]$.

Net worth evolves according to

$$\dot{n}_t = a_t z k_t^{\eta} - (\delta + r + w_t)k_t + r n_t - d_t \tag{10}$$

where $\frac{dn_t}{dt} \equiv \dot{n}_t$, and $d_t \geq 0$ is the firm's dividend payout.[49] To finish characterizing the firm's problem requires specifying a dividend policy. A full model of dividend payout requires specifying the firm's dynamic problem (as we do in the quantitative section) as the decision to pay out dividends depends on the relative value to the firm of retaining funds inside the firm versus paying them out. We simplify the problem by assuming that firms pay out dividends once their net worth becomes sufficiently large. Specifically, if their leverage falls below some level $\underline{\phi}$ (implying net worth rises above $\underline{n}_t(z, \eta) \equiv k_t^u(z, \eta)/\underline{\phi}$) we assume they pay out dividends to immediately return leverage to this target level.[50]

Firm age. We begin with a brief description of the firm life-cycle in this model, focusing on a world without aggregate shocks where $a_t = 1$, $\bar{\phi}_t = \bar{\phi}$, and the wage takes a fixed value $w_t = w$. By assumption, firm-level productivity and returns to scale do not change as firms age, so the life-cycle is determined entirely by the evolution of net worth, n_t. In this section we index firm age by co-opting the time index t, with firms born at age $t = 0$. We discuss the firm life-cycle for a generic firm, who is assumed to be born with sufficiently low initial net worth so as to be financially constrained at birth: $n_0 < \bar{n}(z, \eta)$.

The firm life-cycle is characterized by three regions. In the first region, from birth until some age t_u, the firm cannot afford the unconstrained level of capital. Leverage is at the borrowing limit, and as the firm accumulates retained earnings their net worth and capital grow. The second region is from age t_u to t_d. They enter it once their net worth reaches $\bar{n}(z, \eta)$ and the firm can afford the unconstrained level of capital, so their capital stops growing. As they continue to earn profits and retain earnings, their net worth continues to increase which allows them to begin lowering their leverage ratio. This continues until they lower leverage to the point they start paying dividends, at leverage $\underline{\phi}$ and net worth $\underline{n}$. This happens at age t_d, defining the end of the second region and beginning of the third. In this region the firm pays dividends, and net worth stops increasing.

---

[49] To derive this, consider a discrete time model with period length Δt. Between periods t and $t + \Delta t$ net worth evolves according to $n_{t+\Delta t} = a_t z k_t^{\eta} \Delta t - w_t k_t \Delta t + (1 - \delta \Delta t)k_t - (1 + r\Delta t)b_t - d_t \Delta t$, where the production flow, interest rate, and depreciation rate are all scaled with the length of a period. Combining this with the balance sheet $n_t = k_t - b_t$ gives $n_{t+\Delta t} - n_t = a_t z k_t^{\eta} \Delta t - w_t k_t \Delta t - (\delta + r)\Delta t k_t + r \Delta t n_t - d_t \Delta t$. Diving by Δt and taking the limit as $\Delta t \to 0$ gives (10).

[50] This dividend policy can be shown to be optimal under certain model assumptions. Following the logic of the minimum savings policy in Khan and Thomas (2013), if firms discount the future at the risk free rate, there exists a level of net worth below which they do not pay dividends, and above which they are indifferent about paying dividends. In our simple model, if $(a_t, \bar{\phi}_t)$ were constant firms would be indifferent about paying out dividends the moment their leverage falls below $\bar{\phi}$. In general, as long as $\underline{\phi}$ is chosen to be within the true minimum saving policy region, this policy is optimal.

What this emphasizes in a stark way is how firm age is a helpful proxy for the firm's financial position. However, not all young firms are necessarily financially constrained, since firms born with initial net worth $n_0 > \bar{n}(z,\eta)$ can jump immediately to their optimal capital. This highlights the importance of directly measuring, net worth, leverage, and firm growth to investigate the role of financial frictions, even among young firms. In particular, if a young firm has high and declining leverage, and their size increases as they age, this is consistent with a binding borrowing constraint for that firm in our model.

D.2 Proofs

Effect of a_t and $\bar{\phi}_t$ on aggregate output. Aggregate output and labor are:[51]

$$Y_t = \int_{unc} a_t z \left(\frac{\eta a_t z}{\delta + r + w_t}\right)^{\frac{\eta}{1-\eta}} dG_t(n,z,\eta) + \int_{cons} a_t z(\bar{\phi}_t n)^{\eta} dG_t(n,z,\eta) \tag{11}$$

$$L_t = \int_{unc} \left(\frac{\eta a_t z}{\delta + r + w_t}\right)^{\frac{1}{1-\eta}} dG_t(n,z,\eta) + \int_{cons} (\bar{\phi}_t n) dG_t(n,z,\eta) \tag{12}$$

Equilibrium is given by (12) and the wage equation $w_t = \chi(1+\eta_L)L_t^{\eta_L}$, and given w_t we can then solve for Y_t from (11). Start with the response to a productivity shock. Differentiating (12) and the wage equation gives

$$\frac{dL_t}{da_t} = L'(a_t) = \frac{(\delta + r + w_t)/a_t - w'(a_t)}{\delta + r + w_t} X^u \tag{13}$$

$$w'(a_t) = \eta_L \chi(1+\eta_L)L'(a_t)L_t^{\eta_L - 1} \tag{14}$$

where $X^u = \int_{unc} \frac{1}{1-\eta}\left(\frac{\eta a_t z}{\delta + r + w_t}\right)^{\frac{1}{1-\eta}} dG_t(n,z,\eta) > 0$. Combine:

$$L'(a_t) = \frac{(\delta + r + w_t)/a_t - \eta_L \chi(1+\eta_L)L'(a_t)L_t^{\eta_L - 1}}{\delta + r + w_t} X^u \tag{15}$$

Rearranging establishes that $L'(a_t) > 0$. If $\eta_L > 0$ then $w'(a_t) > 0$, while if $\eta_L = 0$ then $w'(a_t) = 0$. Differentiating (11) gives

$$\frac{dY_t}{da_t} = Y'(a_t) = \frac{Y_t}{a_t} + a_t \frac{(\delta + r + w_t)/a_t - w'(a_t)}{\delta + r + w_t} Z^u \tag{16}$$

where $Z^u = \int_{unc} z \frac{\eta}{1-\eta}\left(\frac{\eta a_t z}{\delta + r + w_t}\right)^{\frac{1}{1-\eta}} dG_t(n,z,\eta) > 0$. Since $L'(a_t) > 0$, (13) implies that $(\delta + r + w_t)/a_t - w'(a_t) > 0$. Inspecting (16), this implies that $Y'(a_t) > 0$.

Moving on to the financial shock, equivalent steps establish that $Y'(\bar{\phi}_t) > 0$, $L'(\bar{\phi}_t) > 0$, and that if $\eta_L > 0$ then $w'(\bar{\phi}_t) > 0$, while if $\eta_L = 0$ then $w'(\bar{\phi}_t) = 0$.

Conditions under which unconstrained size is increasing in RTS. To ease notation, consider a steady state without aggregate shocks where $a_t = 1$ and we drop the t subscript. For firms old enough to be financially unconstrained, their size, measured as input use k,

[51] Aggregate output is $Y_t = \int a_t z k_t(n,z,\eta)^{\eta} dG_t(n,z,\eta)$ and we split the integral using *cons* and *unc* to refer to constrained and unconstrained firms. Note that since at the margin of being constrained or unconstrained $k_t^u(z,\eta) = k_t^c(n)$, the endpoint terms in the Leibniz rule for any derivatives cancel out.

is unconstrained optimum $k^u(z,\eta)$. Clearly, higher productivity leads to larger size: Differentiating (2) gives $k^u_z(z,\eta) = \frac{1}{1-\eta} z^{\frac{\eta}{1-\eta}} (\eta/(\delta+r+w))^{\frac{1}{1-\eta}} > 0$.

What is the effect of returns to scale, η, on size? Intuitively, firms with higher returns to scale should choose a larger optimal size, since their profitability falls less as they expand. This turns out to be true, but only under a minimal assumption that firms are sufficiently productive to begin with. A sufficient condition for size to be increasing in returns to scale (i.e. $k^u_\eta(z,\eta) > 0$) is that $z \geq \bar{z}$, where $\bar{z} \equiv \delta + r + w$. This can be seen most clearly by taking the semi-elasticity $\frac{\partial \log k^u(z,\eta)}{\partial \eta} = \frac{1}{\eta(1-\eta)} + \frac{1}{(1-\eta)^2}(\log \eta + \log z - \log(\delta + r + w))$, which shows that the effect of RTS on size is actually ambiguous. To build intuition, take the limiting case of CRS, where $\eta = 1$ and static profit is $(z - \delta - r - w)k$. Then we know that firms with $z > \bar{z}$ would choose infinite optimal size $k^u \to \infty$ while firms with $z < \bar{z}$ would shrink to $k = 0$. Hence if we start from $\eta < 1$, where all firms have finite optimal size, and move to $\eta = 1$ then $z > \bar{z}$ firms will expand to infinity and $z < \bar{z}$ will shrink to zero, demonstrating the ambiguous effect of RTS on optimal size. Inspecting the semi-elasticity shows that $z \geq \bar{z}$ is sufficient for $k^u_\eta(z,\eta) > 0$.

High RTS firms take longer to outgrow borrowing constraints. We continue to work in an aggregate steady state, and prove the following proposition. Consider firms with the same unconstrained optimal size, $k^u(z,\eta)$, but different RTS, η. Suppose they are currently financially constrained and have net worth n_t a fraction $\omega_t \equiv n_t/\bar{n}(z,\eta) < 1$ of the amount required to be financially unconstrained. The higher a firm's RTS, the slower the current growth rate of its net worth, and hence capital.

Since $k^u(z,\eta) = \left(\frac{\eta z}{\delta+r+w}\right)^{\frac{1}{1-\eta}}$, firms with different RTS but same $k^u(z,\eta)$ must have different productivities, z. Their net worth grows according to (10). Use the definitions of ω_t and $\bar{n}(z,\eta)$ to replace n_t with $n_t = \omega_t k^u(z,\eta)/\bar{\phi}$ and replace z using the formula for $k^u(z,\eta)$ to yield

$$\frac{\dot{n}_t}{n_t} = \left(\frac{1}{\eta \omega_t^{1-\eta}} - 1\right)(\delta + r + w)\bar{\phi} + r > 0 \tag{17}$$

where net worth growth is positive because $\omega_t < 1$ and $\eta < 1$. Most importantly, holding the current net worth deficit ω_t fixed, $d(\frac{\dot{n}_t}{n_t})/d\eta < 0$. This means that if we compare two firms with the same optimal size and current net worth, net worth grows more slowly for the higher RTS firm. This is because higher RTS reduces the excess profits made when firms are below optimal scale. One can also interpret higher RTS as being equivalent to more elastic demand for the firm's products, and hence lower markups. If one models large firms as having higher RTS, this does not say that net worth for large firms should grow more slowly than for small firms, since we hold size constant in the exercise. Instead, it says that large firms will grow more slowly than if you instead modelled large firms as having the same RTS as small firms but higher productivity.

Aggregate labor unresponsive to financial shock if there exist unconstrained CRS firms. Suppose that there exist firms with constant RTS, $\eta = 1$, in the model. If these firms are financially unconstrained their optimal capital choice has the usual "bang bang" solution: $k^c_t(z,1) = \infty$ if $a_t z > \delta + r + w_t$, $k^c_t(z,1) = 0$ if $a_t z < \delta + r + w_t$, and indifferent about any level of capital (i.e. $k^c_t(z,1)$ indeterminate) if $a_t z = \delta + r + w_t$. Assume that all firms with

$\eta = 1$ have the same z. If a positive mass of these firms is financially unconstrained, then they pin down the wage as $w_t = a_t z - \delta - r$. The household's labor supply condition then pins down aggregate labor as $w_t = \chi(1+\eta_L)L_t^{\eta_L} \implies L_t = ((a_t z - \delta - r)/(\chi(1+\eta_L)))^{\frac{1}{\eta_L}}$. Notice that L_t depends only on the productivity shock, a_t, and not on the distribution of resources across firms or the financial shock $\bar{\phi}_t$. Intuitively, if the financial constraint is tightened, the unconstrained CRS firms can perfectly expand to keep labor fixed at desired labor supply given the fixed wage. This equilibrium is valid as long as the group of CRS firms collectively has enough net worth to finance the gap between L_t and the labor demanded by constrained firms.

E Quantitative model appendix

E.1 Further model details

Investment good firms. A representative capital producing firm purchases the final good and converts it into capital which it sells to firms at price $p_{K,t}$. In particular, to create I_t units of new capital the capital producer must purchase I_t units of output and then pay an additional cost $\frac{\psi_K}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t$ denominated in the final good. This is a quadratic cost of deviating the investment rate away from the steady state investment rate, $I/K = \delta$, which is scaled by the size of the capital stock. ψ_K controls the degree of adjustment costs. The capital producing firm's maximization problem can be stated as:

$$\pi_t^K = \max_{I_t} p_{K,t} I_t - I_t - \frac{\psi_K}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t \tag{18}$$

The first order condition for investment implies that the equilibrium capital price is linear in the investment rate:

$$p_{K,t} = 1 + \psi_K \left(\frac{I_t}{K_t} - \delta\right) \tag{19}$$

Note that by choosing to have the quadratic cost paid for investment rates away from steady state, we ensure that the capital price is normalized to $p_K = 1$ in steady state. For a given level of investment, the total resources spent on adjustment costs are $AC_t = \frac{\psi_K}{2}\left(\frac{I_t}{K_t} - \delta\right)^2 K_t$.

Entry cost distribution. We parameterize the entry cost CDF as $G_s^e(\xi) = a_{e,s}\xi^{b_e}$ on support $\xi \in [0, a_{e,s}^{-1/b_e}]$, giving PDF $g_s^e(\xi) = b_e a_{e,s}\xi^{b_e - 1}$. As long as the entry values satisfy $v_s^e \in [0, a_{e,s}^{-1/b_e}]$ at all times, $a_{e,s}$ and M_s^e are not separately identified, and we can directly estimate m_s^e. We follow this approach, and implicitly assume that $a_{e,s}$ is chosen such this condition holds in the neighborhood of the steady state. The total flow of entry costs incurred by firms who choose to enter is then

$$EC = \sum_s M_s^e \int_{\xi=0}^{v_s^e} \xi b_e a_{e,s} \xi^{b_e-1} d\xi = \sum_s m_s^e b_e \int_{\xi=0}^{v_s^e} \xi^{b_e} d\xi = \sum_s \frac{m_s^e b_e (v_s^e)^{b_e+1}}{b_e+1} \tag{20}$$

Aggregate fixed cost spending. This is the sum of the fixed cost paid by all shocked firms who decide not to shut down: $FC = \int_{n,s,g,j} \alpha_\omega \int_{\omega=0}^{\bar{\omega}(n,s,g,j)} \omega dF(\omega, s) dG(n, s, g, j)$.

Proof that firms do not pay positive dividends until exit. Differentiate the value function (8) with respect to n and apply the envelope condition. The Lagrange multiplier on the borrowing constraint $p_K k \leq \bar{\phi} n$ implies that $v_n(n, s, j) > 1$ whenever the borrowing constraint is binding. Moreover, if the borrowing constraint is not binding and expected never to bind again then $v_n(n, s, j) = 1$. However, inspection of the equations reveals that the borrowing constraint may always bind in the future for a bad enough sequence of fixed cost shocks ω, which would lower n back into the region where the borrowing constraint binds again. Even if this happens with very low probability it drives $v_n(n, s, j) > 1$ at all times and firms optimally never pay dividends.

Representative household problem. The representative household owns all firms in the economy, as well as consuming and supplying labor. Firms discount the future using the household's stochastic discount factor. Since we consider only steady states and deterministic transitions to unanticipated shocks ("MIT shocks") the stochastic discount factor is simply equal to the risk free interest rate r_t. Here we formally derive the household optimality conditions determining the equilibrium interest rate and wage for the closed economy case. Let B_t denote household saving in a risk free bond with interest rate r_t. The household's budget constraint reads $\dot{B}_t = w_t L_t + r_t B_t + D_t - C_t$, where D_t is the household's net receipts from firm ownership. In a world without aggregate uncertainty, we can state the household's problem recursively as

$$\rho W(B, t) = \max_{C,L} \frac{1}{1 - \eta_C} \left(C - \chi L^{1+\eta_L} \right)^{1-\eta_C} + W_B(B, t) \left(w_t L + r_t B + D_t - C \right) + W_t(B, t) \tag{21}$$

where we allow the value function to depend directly on time t to allow for arbitrary dynamics of equilibrium prices. $W(B, t)$ is the household value function and $W_B(B, t)$ and $W_t(B, t)$ derivatives with respect to its first and second argument. The first order conditions with respect to consumption and labor yield $\left(C_t - \chi L_t^{1+\eta_L} \right)^{-\eta_C} = W_B(B, t)$ and $(1 + \eta_L) \chi L_t^{\eta_L} \left(C_t - \chi L_t^{1+\eta_L} \right)^{-\eta_C} = w_t W_B(B, t)$, where $C_t \equiv C(B, t)$ and $L_t \equiv L(B, t)$ are shorthand for the optimized consumption and labor choice. Combining these gives optimal labor as $L_t = (w_t / (\chi(1 + \eta_L)))^{1/\eta_L}$. Conversely put, the equilibrium wage satisfies

$$w_t = (1 + \eta_L) \chi L_t^{\eta_L} \tag{22}$$

Let $\dot{C}_t$ and $\dot{L}_t$ be the total derivatives of the consumption and labor functions with respect to time, i.e. $\dot{C}_t \equiv C_t(B, t) + \dot{B}_t C_B(B, t)$ and $\dot{L}_t \equiv L_t(B, t) + \dot{B}_t L_B(B, t)$, where $\dot{B}_t \equiv \dot{B}(B, t)$ is the optimized asset drift. To derive the equilibrium interest rate, first differentiate (21) with respect to B to yield

$$\rho W_B(B, t) = W_{BB}(B, t) \dot{B}_t + r_t W_B(B, t) + W_{tB}(B, t) \tag{23}$$

Totally differentiate $\left(C(B, t) - \chi L(B, t)^{1+\eta_L} \right)^{-\eta_C} = W_B(B, t)$ with respect to t to yield

$$-\eta_C \left(\dot{C}_t - (1 + \eta_L) \chi L(B, t)^{\eta_L} \dot{L}_t \right) \left(C(B, t) - \chi L(B, t)^{1+\eta_L} \right)^{-\eta_C - 1} = W_{Bt}(B, t) + \dot{B}_t W_{BB}(B, t)$$

Using $W_{Bt}(B, t) = W_{tB}(B, t)$ and combining the two equations yields the continuous time

Euler equation for consumption under GHH preferences:

$$r_t = \rho + \eta_C \frac{\dot{C}_t - (1+\eta_L)\chi L_t^{\eta_L} \dot{L}_t}{C_t - \chi L_t^{1+\eta_L}} \tag{24}$$

Definition of equilibrium. The definition of equilibrium is standard. We state first the definition in the stationary equilibrium without aggregate shocks in a closed economy (i.e. $NX_t = 0$). This is given by prices (r, w, p_K) such that

1. The representative household and investment good firm optimality conditions hold. This implies that aggregate prices satisfy (19), (22), and (24) in steady state, giving $r = \rho$, $p_K = 1$, and $w = (1+\eta_L)\chi L^{\eta_L}$.

2. Firms optimize given prices, solving the problem in (8), giving policy functions including $k(n,s,j)$, $l(n,s,j)$, and $d(n,s,j)$, and value function $v(n,s,j)$. This induces a firm entry flow of $\mu_0 = M^e G^e(v^e)$, and an ergodic distribution $G(n,s,j)$ over firms.

3. Prices clear the goods, labor, and capital markets. In particular, labor demand equals labor supply such that $L = \int l(n,s,j)dG(n,s,j)$. Aggregate capital and output are $K = \int k(n,s,j)dG(n,s,j)$ and $Y = \int y(n,s,j)dG(n,s,j)$, $I = \delta K$, and goods market clearing gives $Y = C + I + FC + AC + EC$.

For aggregate transition experiments the definition is similar. We now solve for price sequences $\{r_t, w_t, p_{K,t}\}$ which are determined via the dynamic equations (19), (22), and (24) which depend on the paths for aggregate variables. Consumption satisfies goods market clearing at all dates, such that $Y_t = C_t + I_t + FC_t + AC_t + EC_t$, which determines the equilibrium interest rate. The wage clears the labor market clears at all dates, and the capital price is consistent with the first order condition of the investment goods firms. The firm problem is identical except that firm value is a function of time, $v_t(n,s,j)$, and firms account for the time-paths of prices in their continuation values. For the open economy case the definition is similar except that the interest rate takes the exogenous world rate at all times and net exports NX_t adjust to clear the goods market.[52]

E.2 Calibration

Calibration procedure. Parameters are either pre-set to a known value, or chosen to exactly hit one moment using an associated parameter. We use an iterative updating scheme, and stop once all moments are hit with 10% tolerance or less. There are 34 parameters of the model, which are given in Table 11, with each associated moment given in the Source column. We start by describing our relatively standard parameters. We take one unit of time to be one year. We set the interest rate r to a 4% annual real interest rate. The capital depreciation rate δ is set to a 10% annual rate.[53] We choose the labor to capital ratio α to

[52] A helpful property of GHH in our context is that we can solve for the equilibrium path of wages, labor, and output (and all firm-level objects) without reference to the path for consumption and hence the split of output between consumption and net exports. If desired, the consumption path can be solved as follows. The household's consumption path still satisfies the Euler equation, with the initial level C_0 determined by the economy's intertemporal budget constraint given the time-0 financial position relative to the rest of the world. The level of net exports NX_t then clears the goods market at all dates.

[53] Since one unit of time is one year, we convert the annual rate X into the continuous time flow x using $x = -\log(1 - X)$.

generate a labour share in total firm costs of 2/3. The labor supply disutility χ is chosen to normalize the steady state wage to one. The labor supply elasticity η_L is set to 0.3, which implies a Frisch elasticity of 3.33 in line with the typical values used in macro calibrations.[54] In our baseline exercises choose ψ_K to set an elasticity of the capital price to the investment rate of 0.25 following Brinca et al. (2016) and Ottonello and Winberry (2020). For some exercises we keep capital prices fixed ($\psi_K = 0$) or experiment with different values.

We specify the firms' idiosyncratic shock process $\pi^J_{j,j'}$ as follows. Shocks arrive on average once per year, and when drawn new values come from an AR(1) process discretised with $J = 2$ nodes.[55] To remain close to Khan and Thomas (2013), we fix the annual autocorrelation at their value of 0.659, which is the range of usual values, and normalize the mean to one. We choose the standard deviation to match the standard deviation of log employment among large, old firms in our dataset. More details are given below.

We calibrate the exit process to generate the higher exit rate of young firms in the data. The arrival rate of the stochastic fixed cost shock, α_ω, is chosen to match the exit rate of firms at age 0 relative to at age 16+. We use our own data and data from Andersen and Rozsypal (2021) and calculate this ratio to be around 2. We use the exogenous exit rate ζ to target an overall average exit rate of 5% per year, in line with the data. We seek a simple parameterization of the exit cost function, $F(\omega, s)$. We let $\omega = \kappa_s \tilde{\omega}$, where $\tilde{\omega}$ is drawn from an exponential distribution with unit variance. The size-type specific scaling parameters κ_s are chosen so that sufficiently wealthy — and hence old — firms of each size type effectively only exit at the exogenous rate ζ.[56]

We make a functional form assumption on the entry cost draw distribution $G^e_s(\xi)$ so that it takes the constant elasticity form $G^e_s(\xi) = a_{e,s}\xi^{b_e}$. b_e controls the elasticity of entry to firm value, and is assumed to be equal for all size types to reduce our degrees of freedom. We set b_e to ensure a sensible response of firm entry to aggregate shocks, and discuss this further in our counterfactual experiments. The entry flow to each type is therefore $\mu_0(s) = m^e_s(v^e(s))^{b_e}$, where we subsume M^e_s and $a_{e,s}$ into the hyper-parameter $m^e_s = M^e_s a_{e,s}$.[57]. We set $b_e = 7$, which generates a reasonable decline in entry of around 30% in response to a negative shock to borrowing constraint in both the steady state and business cycle financial shock experiments.

The remaining parameters relate to the novel features of our study: heterogeneity in returns to scale and disciplining the initial net worth of entrant firms, and were discussed in the main text. We assume $S = 5$ permanent size types in the model, corresponding to the five size bins in our empirical work. We use the m^e_s parameters to set the relative flow of entrants into each size type, $\mu_0(s)/\mu_0$, to match the share of firms in each associated size bin in the data, and normalize the number of firms in steady state to one. The net worth distribution of entrants is chosen to provide a match to our data.

Estimation of the idiosyncratic shock process. Our procedure broadly follows that of

[54] Chetty et al. (2012) provide a meta analysis of values of the Frisch elasticity, finding an average 3.31 for papers studying the macro elasticity.

[55] Specifically $\pi^J_{j,j'} = \alpha_J \tilde{\pi}^J_{j,j'}$ where $\alpha_J = 1$ is the arrival rate of a new draw, and $\tilde{\pi}^J_{j,j'}$ is the discretised transition matrix of the AR(1) process.

[56] Specifically, we choose κ_s so that, for each s, firms with very high net worth only exit with 1% probability in the event that the stochastic fixed cost shock arrives.

[57] As long as $G^e_s(v^e(s))$ takes interior values in the steady state and during simulations, M^e_s and $a_{e,s}$ are not separately identified, as we discuss further in the appendix.

Khan and Thomas (2013). Firstly, the autocorrelation of idiosyncratic shocks is known to be hard to estimate, and Khan and Thomas (2013) choose an annual autocorrelation of 0.659, which we do too. Since we are in continuous time, we first specify that firms draw a new value of their idiosyncratic shock on average once a year, and that when they do it is drawn from a discretised AR(1) process with autocorrelation $\rho^I = 0.659$, mean $\mu^I = 1$, and unconditional standard deviation σ^I. This leaves the standard deviation to calibrate, which we choose to match the standard deviation of within firm employment over time. Specifically, we regress employment in our data on firm and time fixed effects, and compute the standard deviation of the residual, which gives a measure of how much employment changes over time within a firm. To avoid issues of life-cycle growth, we do so only for large and old firms in the data (120+ employees, age 16+ years), and do so for old unconstrained firms from the largest size type in the model. We time-aggregate the simulated data to form yearly employment, and compute the standard deviation, adjusting σ^I to match the standard deviation of 22.6% in the data.

Numerical solution and simulation procedure. We solve the model using continuous time numerical methods which draw on Achdou et al. (2021). We use their finite difference methods, and discretize the state variable n with a grid of 300 nodes, with a different grid for each of the five size types since their net worth distributions are very different.[58] Ergodic distributions and the aggregate simulations are calculated using the grid based simulation procedure that forms part of the Achdou et al. (2021) method.

To solve transition paths of the economy to aggregate shocks, use the grid-based simulation approach of the Achdou et al. (2021) method, iterating over guesses of aggregate price paths until the economy converges to the true transition path. This ensures an accurate solution to our transition experiments, which does not rely on simulated data from a finite number of firms.

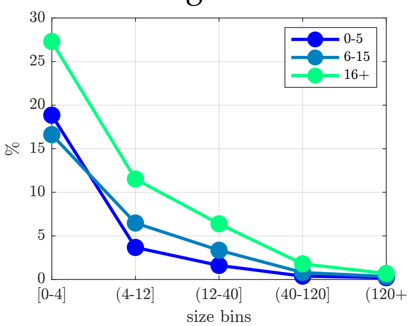
When replicating our cyclicality regressions on model-simulated data, we construct time-aggregated yearly data in such a way as to be comparable to our Danish data source. We construct a panel of 100,000 firms, accounting for entry and exit, which we simulate in response to the aggregate shocks. The policy functions of these firms are the policies solved for exactly during the grid-based transition experiment. We aggregate the data up to yearly frequency to make firm-year observations, and regress this data on the growth rate of aggregate output, as done in our data work, using the same regression specification. We generate 15 years of data from the model to use for our regressions, which contains the single recession event driven by an MIT shock. Specifically, after a burn in period, we allow for 5 years of data pre shock, and then 10 years of data from the moment the shock hits and through the economic recovery.

Calibration results. The model generates a very good match to the marginal age and size distributions (by number of firms and employment) by construction. In Figure 33 we plot key features of the calibration across the joint age-size distribution. Panels (a) and (b) give the firm and sales shares in the model, which are comparable to the data from Figure 1. We directly compare the marginal age and size distributions in the model

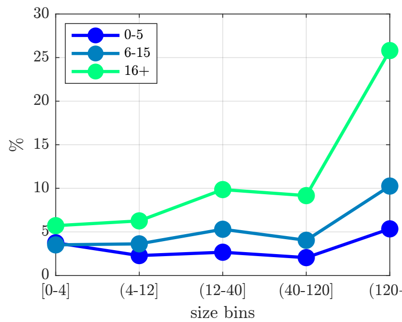
[58] For each size type, the grids run linearly from $n = 0$ to some $n_s^{max} = k^u(s, J)/\underline{\phi}$. At the maximum n, the firm is forced to pay out dividends to maintain a capital to net worth ratio of $\underline{\phi} = 1$. Recall that firms never optimally pay out dividends in this model. However, we choose a very high value of n_s^{max} and in practice firms are nearly indifferent about paying dividends at this level of net worth.

and data in Table 8 and see the fit (which was targeted for the size distribution and for the average employment of the young firms) is very good. The model underpredicts the average employment of old (16+) firms slightly, but does match that there continues to be growth in average firm size between the ages of 6-15 and 16+. The model generates an exit rate (panel f) which is decreasing in firm age, in line with the data. In Figure 39 we show that exit rates are higher along the whole life-cycle for firms born with lower initial net worth in our model, in line with the data from Figure 9, and highlighting that net worth and finance are a key feature driving exit in the model. We plot an average life-cycle in Figure 38, and the distribution of size-types across size bins in Table 10.

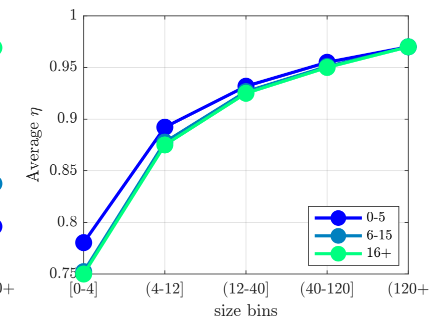
Figure 33: Firm shares and characteristics across age and size bins



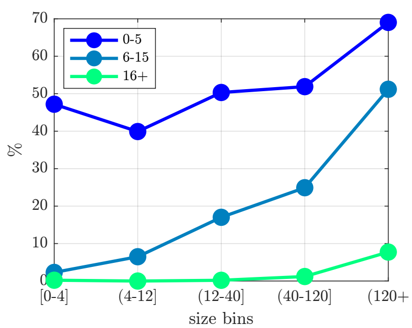
(a) Firm share



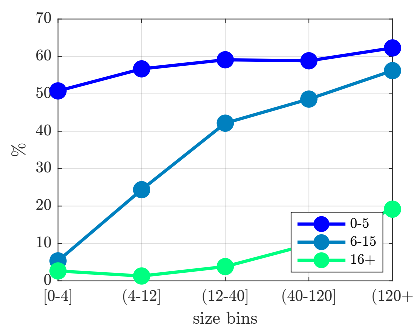
(b) Sales share



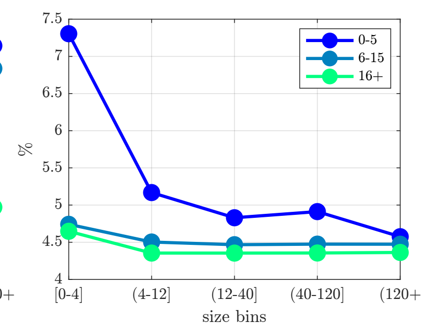
(c) Average returns to scale



(d) Fraction constrained



(e) Leverage (D/A)



(f) Exit rate

Note: Characteristics of each joint age-size bin in the model.

A key use of the model, over directly studying the data, is that it gives us a theory of financial constraints over the life-cycle. Figure 33(e) gives average leverage by age-size bin. Leverage is declining in age, and roughly constant across size groups for the youngest firms. Leverage falls less quickly with age for large firms, because they take longer to outgrow their financial constraints due to their higher RTS and lower profit margins. This is reflected in panel d, where a larger share of middle-aged large firms are financially constrained than middle aged small firms.

An untargeted success of the model is that it replicates well the firm demographics we discussed in Section 4.4. Specifically, a high fraction of old, large firms were significantly smaller when they were younger. In Table 9(a) we look at firms currently aged 15 in the model, and compute the size bin they were in at age 0. For aged 15 firms in all size bins, a significant majority of firms were in smaller size bins at age 0. Panel (b) does the same in the data, and shows that the model provides a good qualitative fit to the data. In the model, this growth is entirely driven by firms overcoming their financial frictions. For example,

Table 8: Marginal firm size and age distributions in the model and data

	Fraction of firms					Average employment				
Size	0-4	4-12	12-40	40-120	120+	0-4	4-12	12-40	40-120	120+
Model	0.63	0.22	0.11	0.03	0.01	1.83	5.78	17.07	57.22	383.35
Data	0.61	0.22	0.12	0.03	0.01	1.85	5.59	16.35	52.99	375.33

(a) Size distribution

	Fraction of firms			Average employment		
Age	0-5	6-15	16+	0-5	6-15	16+
Model	0.25	0.28	0.48	6.82	10.42	13.07
Data	0.26	0.31	0.43	7.77	11.11	17.78

(b) Age distribution

Note: Firm age and size distributions in the model and data. Size bins refer to employment and age bins to age in years since birth. Average employment refers to total employment in the bin divided by the number of firms in the bin.

the fact that firms can eventually reach size bin 120+ despite being born in any of the 12-40, 40-120, or 120+ bins is because of the heterogeneity in initial net wealth draws, n_0. Without this feature, it would be impossible for the model to generate more than one non-zero entry per column in this table.

Table 9: Firm size over the life-cycle ("looking backwards")

		Age 15				
		0-4	4-12	12-40	40-120	120+
Age 0	0-4	1.00	0.85	0.53	0.14	0.00
	4-12	0.00	0.15	0.27	0.25	0.01
	12-40	0.00	0.00	0.21	0.39	0.14
	40-120	0.00	0.00	0.00	0.22	0.33
	120+	0.00	0.00	0.00	0.00	0.52
		1.00	1.00	1.00	1.00	1.00

(a) Model

		Age 15				
		0-4	4-12	12-40	40-120	120+
Age 0	0-4	0.87	0.64	0.38	0.30	0.19
	4-12	0.11	0.31	0.35	0.21	0.10
	12-40	0.02	0.05	0.25	0.31	0.17
	40-120	0.00	0.00	0.02	0.15	0.15
	120+	0.00	0.00	0.00	0.02	0.39
		1.00	1.00	1.00	1.00	1.00

(b) Data

Note: Distributions of size bins at age 0 for firms currently age 15 in the model (panel a) and data (panel b), conditional on size bin at age 15. Blue cells highlight the diagonal.

E.3 Firm cyclicality in response to a typical business cycle

In this section we demonstrate that our calibrated model does a reasonable job of matching the cyclicality facts we documented in Figure 2. In a previous version of this paper we explicitly reverse engineered the calibration required to match these facts and showed that heterogeneous RTS and initial net worth were necessary in this model. We now instead directly calibrate RTS and initial net worth from our micro data, and show that the model continues to a good job at matching these facts. Part of this exercise involves using the model to identify the composition of the aggregate shocks hitting the economy during a hypothetical typical business cycle, meant to capture the average over the business cycles covered by our dataset. We showed analytically in Section 3 that a combination of a financial and TFP shock could match the qualitative patterns in our data when combined with heterogeneous RTS, and we now investigate that quantitatively.

Using the quantitative model allows us to extend the toy model analysis, and we now also allow for an endogenous capital price to investigate whether our model can also generate the cyclicality patterns from financial accelerator effects. Specifically, we consider two exercises. In the first, we hold capital prices fixed (by setting $\psi_K = 0$) and subject the economy to a temporary financial and TFP shock (both with the same persistence[59]) large enough to drive a 5% fall in output. To roughly match the fact that young firms have a coefficient of roughly 2 in our regressions, we set the financial shock to be a 10% tightening of the borrowing constraint (twice the size of the output fall) since young firms are primarily affected by the borrowing constraint. In the second exercise we instead do not consider a financial shock, and consider only a TFP shock, but allow for endogenous capital price movements by setting $\psi_K > 0$. To roughly match the cyclicality of young firms we control the strength of the financial accelerator effect by setting $\psi_K = 0.15$. In both experiments we adjust the size of the TFP shock to generate the targeted 5% output fall. Both of these exercises are carried out in the open economy model.

In Figure 34 we plot the results for the first exercise, which we take as our baseline exercise. Panels (a) to (c) show the simulated path for output, the estimated shocks, and equilibrium prices respectively. To generate the peak 5% output fall requires the combination of a 10% tightening of the borrowing constraint and 1% fall in TFP, and leads to a slightly more than 1% fall in the wage. Panel (d) shows the cyclicality coefficients for firm-level employment growth from (1) on model simulated data. We use exactly the same age and size bins as in our data, but add an extra age bin for firms aged 0-1 to zoom in on the behavior of very young firms. We see that the model captures our broad patterns correctly. Young firms are more cyclical than old firms and, among old firms, cyclicality is increasing in size. In the data cyclicality was decreasing in size for young firms in the age 0-5 bin. In the model cyclicality is actually increasing in size for age 0-5 firms, so here the model is missing the data, but we see that the model does generate a decreasing slope in size if we zoom in to the firms in the 0-1 age bin. Quantitatively, the fit to the data is quite good in several dimensions. Depending on the size bin, the gap in the cyclicality coefficient between age 0-5 and 16+ firms is around 0.5 to 1, similar to the magnitudes of the gaps in Figure 2. In the data, among age 16+ firms, cyclicality increases by around 0.75 when moving from the smallest to largest bin, and in the model this increase is similar at around 1.

In Figure 35 we present the results for the experiment with no financial shock but an endogenous capital price. In this version, a larger TFP shock of 0.94 is required, but it drives a 2.5% fall in capital prices which reduces net worth on impact and hence tightens financial constraints for young firms. This generates endogenously the fact that young firms are more cyclical (panel d) and again with a gap between young and old firms quantitatively

[59] Specifically both are continuous-time AR(1) processes. The aggregate TFP shock a_t has value $a = 1$ in steady state, is common to all firms and multiplies their firm level productivity, z. We let $\dot{a}_t = -\rho_a(a_t - a)$ and $\dot{\bar{\phi}}_t = -\rho_\phi(\bar{\phi}_t - \bar{\phi})$ and set $\rho_a = \rho_\phi = 0.3$. At time 0, a_0 and $\bar{\phi}_0$ unexpectedly jump from steady state to a new value, and then are known to deterministically converge back to steady state following these processes.

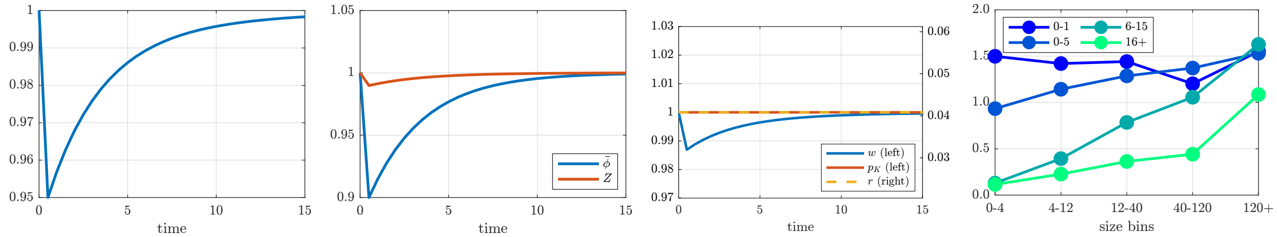
similar to the data. Among age 0-1 firms there is a slightly downwards slope of cyclicality in size. We still find an upwards slope in cyclicality in firm size among the oldest firm group, but this effect is now muted. Directly inspecting the time series of firm simulations reveals that this is due to a timing effect, and not because large firms are less cyclical in this second experiment.

Finally, in Figure 36 we repeat our first experiment but this time in a closed economy version of the model. This experiment could be interpreted as, for example, applying more to the US than Denmark, since we argued that Denmark is better modelled as a closed economy. This experiment generates results similar to the second experiment: we still find that young firms are more cyclical than old, a downwards slope of cyclicality in size for age 0-1 firms, but the positive slope in size among old firms is now muted. Directly inspecting the model simulated data reveals that this is again due to a timing effect, and not because large firms are less cyclical in this second experiment.[60]

To understand the role of firm-level data in identification of aggregate shocks, in Figure 41 we plot the impulse responses to a TFP shock and our baseline financial shock separately. The key idea here is that the aggregate responses look quite similar, despite the different nature of the shocks. It is only in the firm level responses by age (panel d) that the two shocks become more easy to distinguish, as the TFP shock causes both young and old firms to shrink, while the financial shock causes young firms to shrink and old firms to expand. Similarly, the TFP shock causes a small fall in entry (panel e) while the financial shock causes a much large fall.

[60]Endogenous interest rate or capital price movements cause unconstrained firms to smooth out their investment decisions, and so they shrink gradually in response to the shocks, and not immediately. This causes the regression to miss part of the cyclicality of large firms, since they continue shrinking even when output starts growing again as the direct effect of the TFP shock fades. In the data (not shown) we find that the employment and sales of old firms track GDP growth much more closely, suggesting that this smoothing is not entirely data-consistent. This could perhaps be rectified in the model by introducing more realistic interest rate movements (see, e.g. Winberry, 2021).

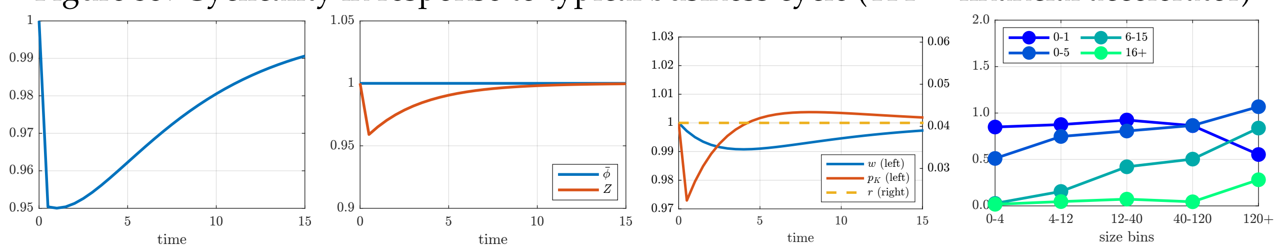
Figure 34: Cyclicality in response to typical business cycle (TFP + financial shock)



(a) Output (b) Shocks (c) Prices (d) Cyclicality

Note: Response of economy to typical business cycle experiment consisting of a TFP shock and financial shock in open economy with fixed capital price. Right column gives cyclicality regression coefficients for employment growth from (1) on model-simulated data.

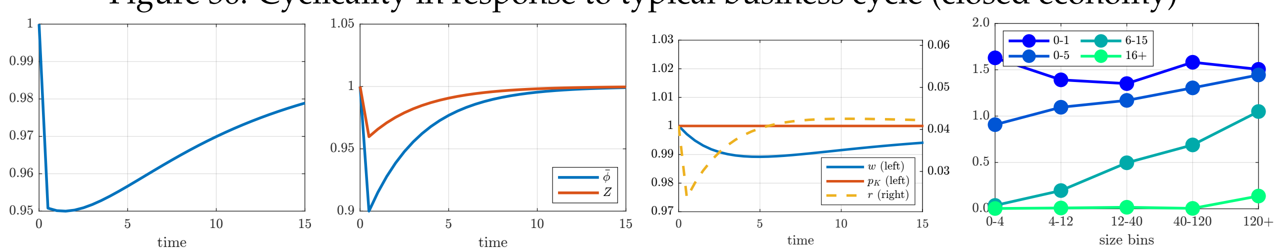
Figure 35: Cyclicality in response to typical business cycle (TFP + financial accelerator)



(a) Output (b) Shocks (c) Prices (d) Cyclicality

Note: Response of economy to typical business cycle experiment consisting of a TFP shock endogenous capital price in open economy. Right column gives cyclicality regression coefficients for employment growth from (1) on model-simulated data.

Figure 36: Cyclicality in response to typical business cycle (closed economy)



(a) Output (b) Shocks (c) Prices (d) Cyclicality

Note: Response of economy to typical business cycle experiment consisting of a TFP shock and financial shock in closed economy with fixed capital price. Right column gives cyclicality regression coefficients for employment growth from (1) on model-simulated data.

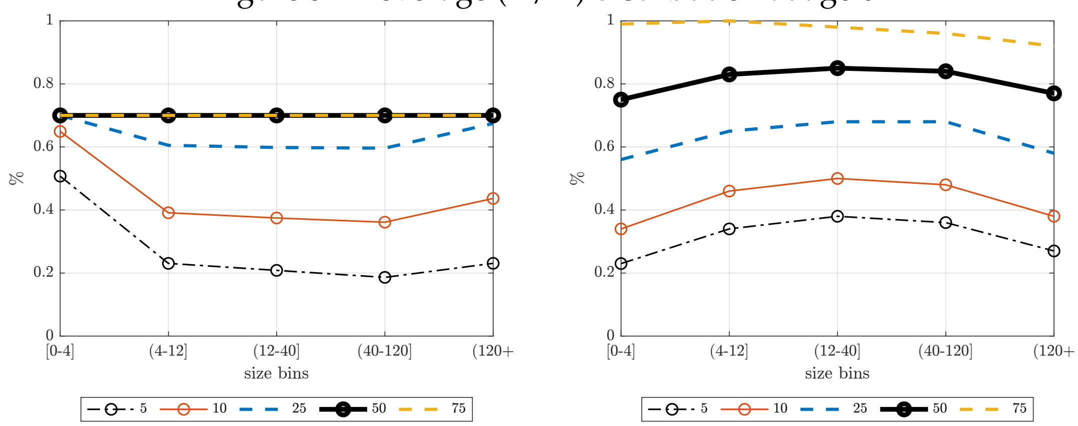
E.4 Additional model tables and figures

Table 10: Distribution of firm types s across size bins in the model

		Size type s					
		1	2	3	4	5	
Size bin	0-4	0.583	0.033	0.010	0.001	0.000	0.626
	4-12	0.000	0.202	0.013	0.002	0.000	0.217
	12-40	0.000	0.000	0.108	0.006	0.000	0.114
	40-120	0.000	0.000	0.000	0.028	0.001	0.030
	120+	0.000	0.000	0.000	0.000	0.012	0.012
		0.583	0.235	0.130	0.037	0.014	1

Note: Distribution of size types s across size bins in the calibrated steady state.

Figure 37: Leverage (D/A) distribution at age 0



(a) Model (b) Data

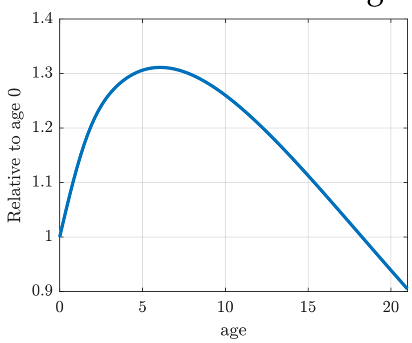
Note: Entrant leverage distribution in the model and the data, by size at age 0.

Table 11: Model Parameters and Calibration

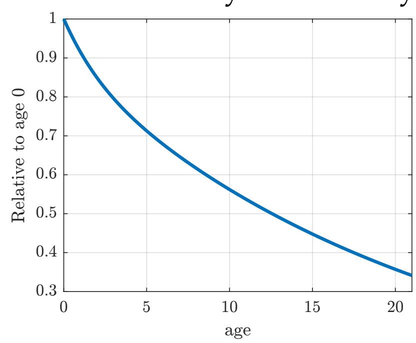
	Interpretation	Value	Source	Moment	
	Common parameters			*Data*	*Model*
r	Discount rate	0.0408	4% yearly real interest rate		
δ	Depreciation rate	0.1054	10% annual rate		
α	Labor-capital ratio in prod fun	0.2924	2/3 Labor cost share		
χ	Labor disutility shifter	0.3797	Normalise $w_{ss} = 1$		
η_L	Labor supply elasticity	0.3	Standard value		
ψ_K	Capital adjust costs	2.3728	25% elasticity (Brinca et al., 2016)		
b_e	Entry elasticity	7	Firm entry cyclicality		
ζ	Exogenous exit rate	0.0426	Average exit rate 5% per year	5%	4.96%
α_ω	Arrival rate fixed cost shock	0.0892	Exit rate age 0 / age 16	2	2.03
σ^I	Std. idiosyncratic shocks	0.0072	Std. firm-level employment	22.58%	22.58%
ρ^I	Autocorr. idiosyncratic shocks	0.6590	Khan and Thomas (2013)		
μ_e	Mean net worth fraction of entrants	0.5296	Employment share age 0-5 firms	15.44%	15.6%
σ_e	Std entrant net worth	1.25	Entrant leverage distribution	See Figure 37	
$\bar{\phi}$	S.s. collateral limit	$1/(1-0.7)$	Entrant leverage distribution	See Figure 37	
	Size-type specific parameters				
$\mu_0(1)$	S.s. entry flow $s = 1$	0.0305	Firm share 0-4 size bin	60.88%	62.64%
$\mu_0(2)$	S.s. entry flow $s = 2$	0.0116	Firm share 4-12 size bin	22.40%	21.75%
$\mu_0(3)$	S.s. entry flow $s = 3$	0.0064	Firm share 12-40 size bin	12.10%	11.39%
$\mu_0(4)$	S.s. entry flow $s = 4$	0.0018	Firm share 40-120 size bin	3.32%	2.98%
$\mu_0(5)$	S.s. entry flow $s = 5$	0.0007	Firm share 120+ size bin	1.30%	1.24%
z_1^S	Productivity for type $s = 1$	0.9270	Emp share 0-4 size bin	10.26%	10.64%
z_2^S	Productivity for type $s = 2$	0.7247	Emp share 4-12 size bin	11.39%	11.65%
z_3^S	Productivity for type $s = 3$	0.6422	Emp share 12-40 size bin	17.99%	18.01%
z_4^S	Productivity for type $s = 4$	0.6007	Emp share 40-120 size bin	15.97%	15.82%
z_5^S	Productivity for type $s = 5$	0.5626	Emp share 120+ size bin	44.39%	43.88%
η_1	Returns to scale $s = 1$	0.75	RTS estimate 0-4 size bin		
η_2	Returns to scale $s = 2$	0.875	RTS estimate 4-12 size bin		
η_3	Returns to scale $s = 3$	0.925	RTS estimate 12-40 size bin		
η_4	Returns to scale $s = 4$	0.95	RTS estimate 40-120 size bin		
η_5	Returns to scale $s = 5$	0.97	RTS estimate 120+ size bin		
κ_1	Fixed cost scaler $s = 1$	1.9381	P(endog exit) = 1% when old 0-4 size bin		
κ_2	Fixed cost scaler $s = 2$	2.5360	P(endog exit) = 1% when old 4-12 size bin		
κ_3	Fixed cost scaler $s = 3$	4.3183	P(endog exit) = 1% when old 12-40 size bin		
κ_4	Fixed cost scaler $s = 4$	9.5600	P(endog exit) = 1% when old 40-120 size bin		
κ_5	Fixed cost scaler $s = 5$	43.3127	P(endog exit) = 1% when old 120+ size bin		

Note: Parameters and calibration targets for the baseline quantitative model. The final two columns give the fit to the targeted moment for parameters updated during the calibration routine. For parameters where no number is given, either the parameter has been set outside of the calibration routine, or the parameter is set analytically and the fit is exact. We report the entry flows $\mu_0(s)$ rather than the shifters m_s^e since they are easier to interpret, and the shifters are implied by the flows via $\mu_0(s) = m_s^e(v^e(s))^{b_e}$.

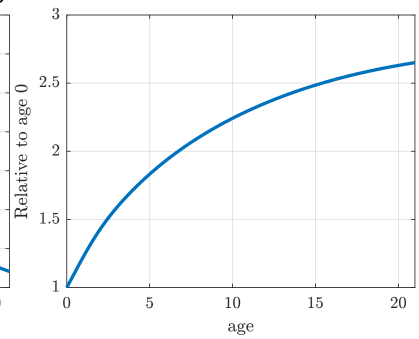
Figure 38: Firm life-cycle in steady state



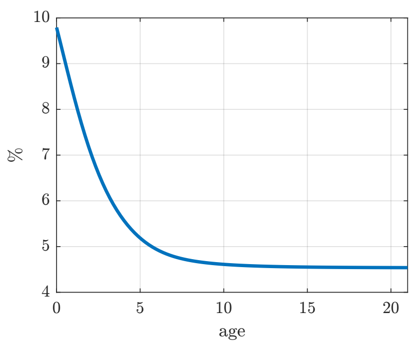
(a) Total output of cohort



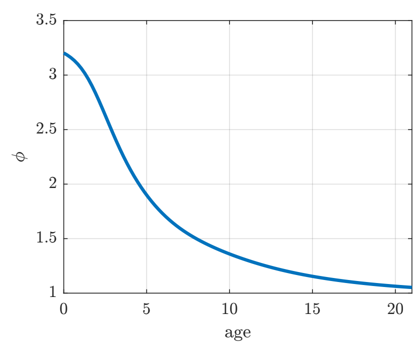
(b) No firms



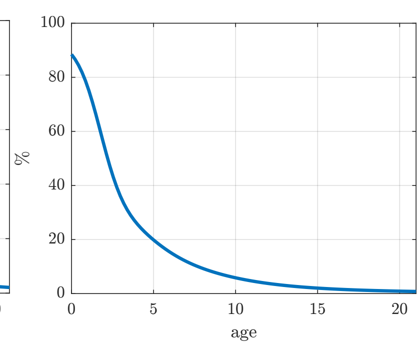
(c) Total output per firm



(d) Exit rate



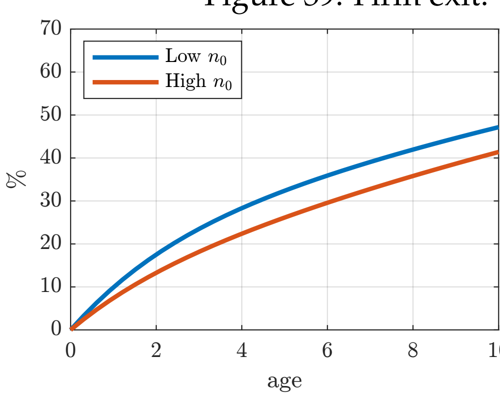
(e) Leverage



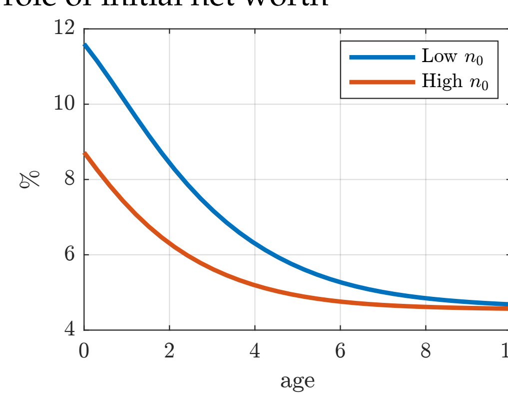
(f) Fraction constrained

Note: Firm life-cycle in the model, averaging over all firms.

Figure 39: Firm exit: the role of initial net worth



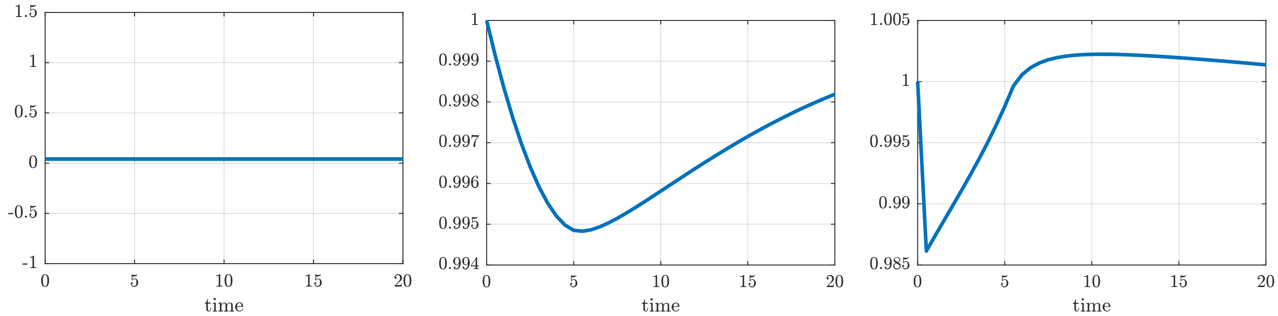
(a) Cumulative exit



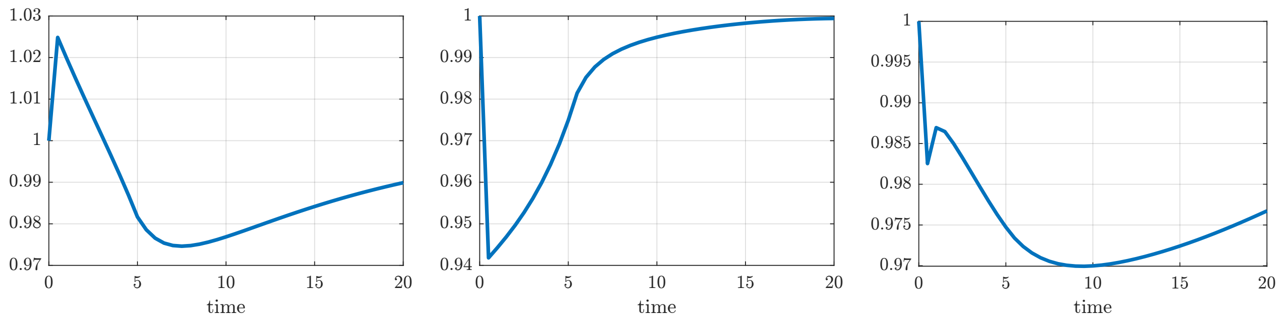
(b) Exit rate

Note: Average exit rate for firms conditioning on initial net worth. Blue (red) line is firms with below (above) median initial net worth (defined relative to their size type).

Figure 40: Aggregate effect of a temporary financial shock: further aggregates



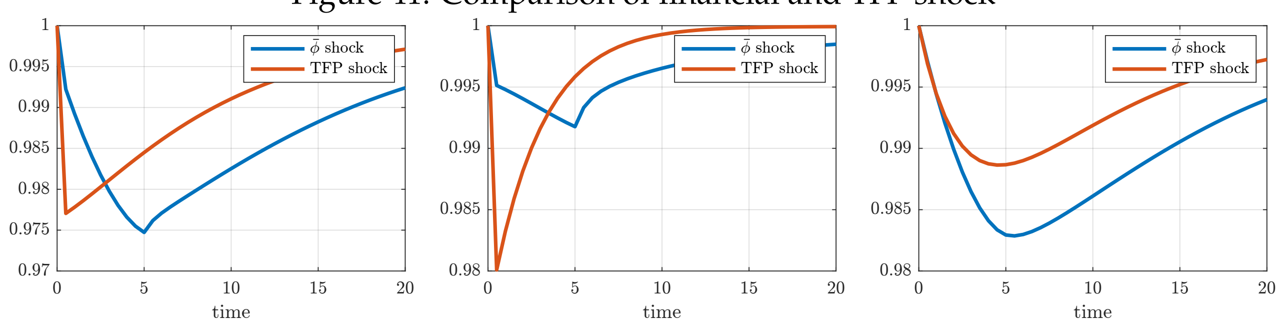
(a) Interest rate (b) Wage (c) Price of capital



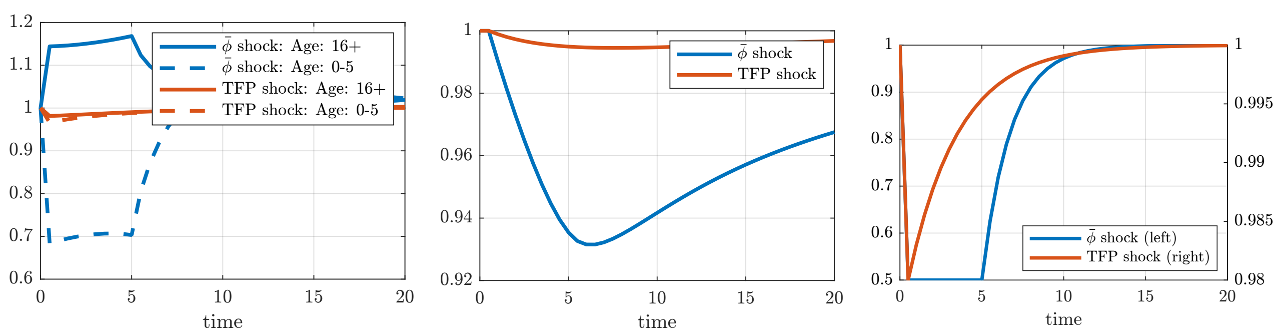
(d) Consumption (e) Investment (f) Firm net worth

Note: Response of additional aggregates to a temporary 50% tightening of the borrowing constraint.

Figure 41: Comparison of financial and TFP shock



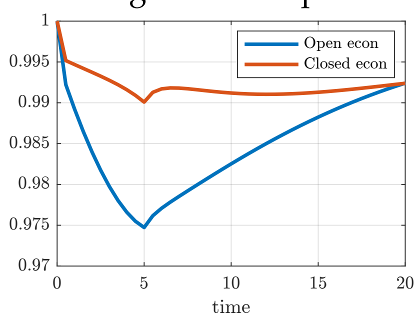
(a) Output (b) Productivity (Y/L) (c) Capital, Labor



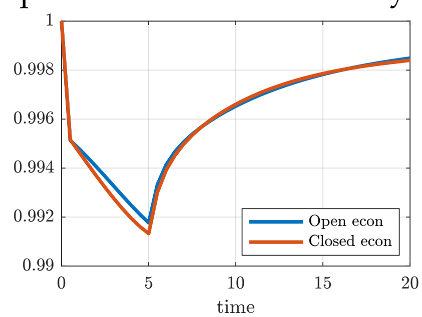
(d) Output per firm by age (e) No firms (f) Shocks

Note: Response of the model to a financial (blue line) versus TFP (red line) shock.

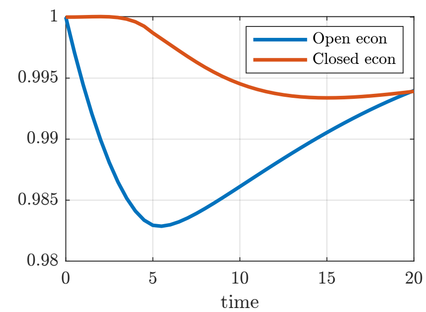
Figure 42: Response of open versus closed economy model to financial shock



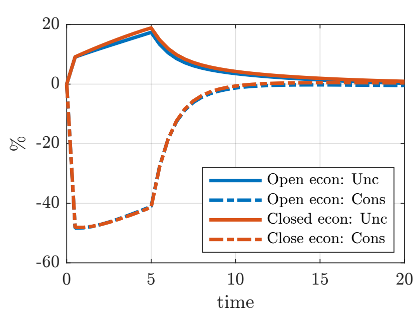
(a) Output



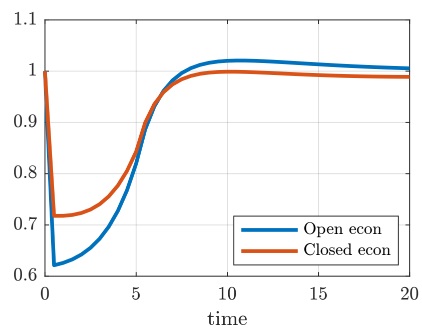
(b) Productivity (Y/L)



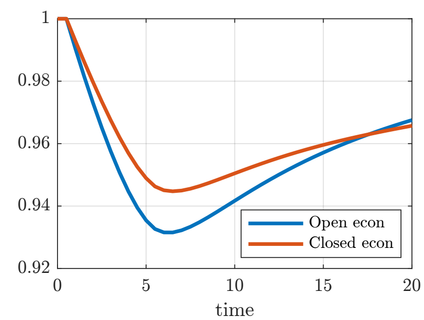
(c) Capital, Labor



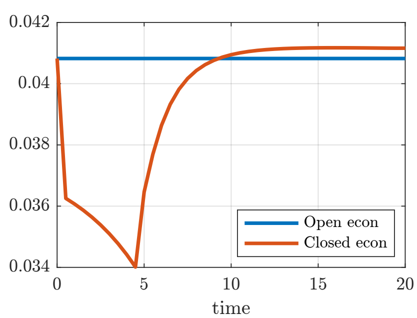
(d) Y/G by constraint status



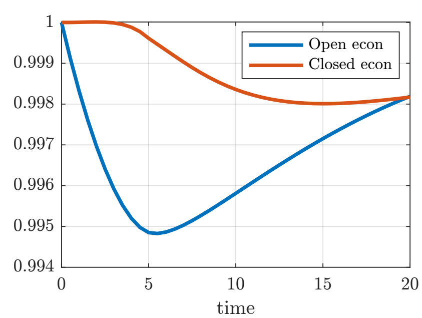
(e) Entry



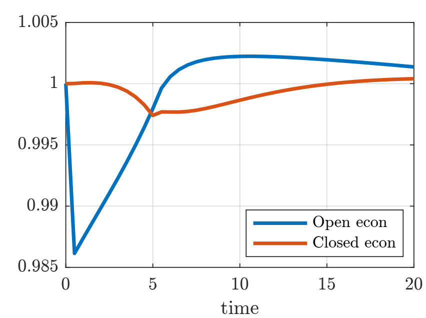
(f) No firms



(g) Interest rate



(h) Wage



(i) Capital price

Note: Response of the model to a financial shock in the open economy (blue line) and closed economy (red line) models.