Title
Essays on Labor Markets and Inequality

Author
Gouin-Bonenfant, Emilien

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Labor Markets and Inequality

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Émilien Gouin-Bonenfant

Committee in charge:

Professor James D. Hamilton, Chair
Professor David Lagakos
Professor Tommaso Porzio
Professor Natalia Ramondo
Professor Allan Timmermann
Professor Alexis Akira Toda

2019

The dissertation of Émilien Gouin-Bonenfant is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

2019

DEDICATION

I dedicate this dissertation to my family. To Gilles, Édith, Mathilde, and Anne: thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my thesis committee members for their continuing guidance, support, and training.

Chapter 1, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien. “Productivity Dispersion, Between-Firm Competition, and the Labor Share”. The dissertation author was the sole investigator and author of this material.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien; Toda, Alexis Akira. “Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality”. The dissertation author was a primary author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien. “Earnings Dynamics and Inequality Within the Firm”. The dissertation author was the sole investigator and author of this material.

VITA

2013	B. Sc. in Economics, Université de Montréal
2019	Ph. D. in Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays on Labor Markets and Inequality

by

Émilien Gouin-Bonenfant

Doctor of Philosophy in Economics

University of California San Diego, 2019

Professor James D. Hamilton, Chair

In the first chapter, I propose a tractable model of the labor share that emphasizes the interaction between labor market imperfections and productivity dispersion. I bring the model to the data using an administrative dataset covering the universe of firms in Canada. As in the data, most firms have a high labor share, yet the aggregate labor share is low due to the disproportionate effect of a small fraction of large, extremely productive "superstar firms". I find that a rise in the dispersion of productivity across firms leads to a decline of the aggregate labor share in favor of firm profit. The mechanism is that productivity dispersion effectively shields high-productivity firms from wage competition. Reduced-form evidence from cross-country and cross-industry data supports both the prediction and the mechanism. Through the lens of the model, rising

productivity dispersion has caused the U.S. labor share to decline starting around 1990.

In the second chapter, we propose a new, systematic approach for analyzing and solving heterogeneous-agent models with fat-tailed wealth distributions. Our approach exploits the asymptotic linearity of policy functions and the analytical characterization of the Pareto exponent to make the solution algorithm more transparent, efficient, and accurate with zero additional computational cost. As an application, we solve a heterogeneous-agent model that features persistent earnings and investment risk, borrowing constraint, portfolio decision, and endogenous Pareto-tailed wealth distribution. We find that a wealth tax is a "lose-lose" policy: the introduction of a 1% wealth tax (with extra tax revenue used as consumption rebate) decreases wage by 6.5%, welfare (in consumption equivalent) by 7.7%, and total tax revenue by 0.72%.

In the third chapter, I propose a model of earnings dynamics and inequality within the firm. The model combines a production hierarchy with a "rank order tournament" promotion scheme. I motivate the theory by documenting three sets of facts using proprietary personnel data. First, most of inequality within the firm is between hierarchy levels rather than within. Second, there is on average very little upward mobility within the firm. Third, there is important heterogeneity in earnings *trajectories*. The model provides a positive theory of these facts and sheds light on the determinants of inequality within the firm.

Chapter 1

Productivity Dispersion, Between-Firm Competition, and the Labor Share

1.1 Introduction

For most of the twentieth century, worker compensation in the U.S. grew one-for-one with labor productivity. This observation has been enshrined as one of the stylized facts of economic growth (Kaldor (1961)). However, starting around 1990 worker compensation has stagnated relative to labor productivity, leading to a decline of the aggregate labor share (Elsby, Hobijn, and Şahin (2013), Karabarbounis and Neiman (2014)). Over that same period, there has been a dramatic rise in the dispersion of firm productivity (Barth et al. (2016), Andrews, Criscuolo, and Gal (2016), Berlingieri, Blanchenay, and Criscuolo (2017)). In other words, recent productivity gains have been extremely concentrated among some firms rather than being broad-based.

In this paper, I argue that the rise in productivity dispersion is in fact a core part of why the aggregate labor share has declined. In a nutshell, my argument is that since recent productivity gains have been concentrated, firms experiencing high productivity growth have been shielded from wage competition by firms whose productivity has lagged. As a result, the increase in

productivity dispersion has caused the aggregate labor share to decline in favor of firm profits.

To formalize this argument, I build a tractable model of the labor share that aggregates up from the firm level. To guide my modelling choices, I begin by summarizing the distribution of labor shares across firms. I draw on a new dataset covering the universe of firms in Canada and focus on three features of the data. First, within narrow industries, there are large and persistent differences in labor shares across firms. This suggests that it is not simply technology that determines a firm's labor share. Second, in any given year, a large fraction of firms have a labor share above one. This is not just a handful of firms, but close to 20 percent of firms, which is far too large a group to write off as simple measurement error. Third, labor shares at the firm are strongly decreasing in firm output. This means that firms who contribute the most to GDP have the lowest labor shares. These facts are not only a feature of the Canadian economy, they have all been previously documented in some form using U.S. data (Autor et al. (2017), Kehrig and Vincent (2017)). I will argue that any quantitative theory of the labor share aggregating from the firm level must be confronted with these empirical regularities.

My model builds on two ideas. First, there are search frictions in the labor market and firms set wages (as in Burdett and Mortensen (1998)). Labor market imperfections will generate firm-specific gaps between wages and labor productivity, thus implying differences in labor shares between firms operating the same technology. Second, firms face persistent productivity shocks and grow over time by accumulating workers (as in Hopenhayn and Rogerson (1993)). Introducing firm dynamics and an endogenous exit decision to the model will generate time-series variation in firm-level labor shares and rationalize the existence of negative profits (i.e. labor shares above one).

In the model, firms choose how much capital to rent, what wages to offer, and when to exit. In equilibirum, high productivity firms offer high wages in order to poach workers from lower paying firms and therefore *grow* faster. The labor share of high-productivity firms is low, reflecting the fact that they are able to recruit and retain workers despite paying them

wages below their marginal productivity. In other words, these firms exert significant monopsony power in the labor market. In contrast, low-productivity firms offer low wages, grow slower, and have a high labor share. Firm-level productivity is persistent but not permanent, so the model endogenously generates "superstar firms", who emerge by experiencing extended periods of fast growth. The model thus explains the rise and fall of large firms and, in a stationary equilibrium, generates an extremely skewed firm size distribution that obeys Zipf's law. While there is no one-to-one relationship between employment and productivity, larger firms are on average more productive. Hence, the aggregate labor share disproportionately reflects the low labor share of large, highly-productive firms.

I then bring the model to the data using a unique firm-level, panel dataset containing employment and tax records for the universe of firms in Canada over the 2000-2015 period. The dataset is ideal as it allows me to measure value-added and worker compensation at the firm-level using the same methodology as in the National Accounts. Despite being very parsimonious, the model matches a number of features of the data that were not targeted. For example, the model generates a large interdecile range of labor shares of 1.06 (0.88 in the data) even though all firms operate the same Cobb-Douglas technology. The aggregate labor share is 0.66 (0.65 in data) while the average (unweighted) firm labor share is 0.97 (0.88 in data). The reason why the average labor share is so high is that 36 percent of firms have labor share above one (17 percent in data). The difference between the average and aggregate labor share is equal to the covariance between labor share and output share of -0.32 (-0.22 in data). Finally, the model predicts that value-added is concentrated within high-productivity, low-labor-share firms: the top decile of firms ranked by labor productivity account for 53 percent of output (42 percent in data) and have a labor share of 0.54 (0.44 in data).

Equipped with the calibrated model, I quantify the effect of an increase in the dispersion of productivity across firms. I feed in an increase in the variance of (log) labor productivity of 0.3, consistent with the measurements from Barth et al. (2016) over the 1977-2007 period. In response

to the rise in productivity dispersion, the model predicts a large labor share decline of 2.3%, which is roughly $2/3$ of the corporate labor share decline over the 1977-2007 period (3.5%). The mechanism is that productivity dispersion effectively weakens labor market competition. High productivity firms become increasingly shielded from wage competition as their competitors move further away on the productivity spectrum. As a result, the gap between the wage they pay and their productivity increases, so their profit margins increase. In contrast, the labor share of firms at the bottom of the productivity distribution *increases*. The reason is that the marginal exiting firm is now willing to absorb larger losses as she expects higher profit margins in the future when productivity reverts. Hence, the high productivity dispersion economy exhibits a higher concentration of value-added and a more polarized distribution of firm labor shares (i.e., a "winner takes most" environment).

The micro predictions of the model are almost exactly what has been documented by Autor et al. (2017) and Kehrig and Vincent (2017) using U.S. firm-level data. In particular, the model predicts that in response to a productivity dispersion shock (1) the aggregate labor share decreases, (2) the decline is driven by a reallocation of value-added towards firms with a low and declining labor share, (3) output concentration increases, and (4) employment concentration remains unchanged.

Finally, I test the prediction of the model that an increase in productivity dispersion is associated with a decline of the aggregate labor share. I use cross-country data from 7 OECD countries over the 2001-2011 period as well as cross-industry data from 69 Canadian industries (3-digit NAICS) over the 2000-2015 period. In both cases, I measure productivity dispersion as the interdecile range of labor productivity (value-added per worker) in the cross-section of firms. I estimate cross-sectional and fixed effects regressions models. In all cases, the estimated coefficients range from -0.05 to -0.20 and are statistically significant. Using the Canadian cross-industry data, I also test the mechanism. I estimate the effect of productivity dispersion on the labor share and output share of firms ranked according to their productivity. As predicted by the

model, a productivity dispersion shock is associated with a polarization of firm labor shares as well as a reallocation of output shares from low to high productivity firms.

Related literature. Many studies document a large decline of the labor share across many of countries starting around 1980 (Blanchard, Nordhaus, and Phelps (1997), Elsby, Hobijn, and Şahin (2013), Karabarbounis and Neiman (2014), Dao et al. (2017)). The explanation I proposerising productivity dispersiondoes not explain all of the decline but is consistent with both microeconomic evidence and aggregate data. To the best of my knowledge, my paper is the first to link the rise in productivity dispersion to the decline of the labor share.

While it is well-known that there is a large dispersion in firm productivity (Syverson (2011)), a less appreciated fact is that productivity dispersion has increased dramatically over time. Kehrig (2015) shows that the (cross-sectional) standard deviation of TFP nearly doubled in the U.S. manufacturing sector from the 1970s to the 2000s. Barth et al. (2016) documents a similar trend, where the variance of log revenue per worker has increased within all industries (SIC 1-digit) over 1977-2007. The increase in productivity dispersion is broad-based within the US economy and holds within both young and mature firms as well as within high-tech and non-tech sectors (see Decker et al. (2018)). It also seems to be a global phenomenon. Andrews, Criscuolo, and Gal (2016) use Orbis data covering 24 OECD countries over the period 1997-2014 period and document a strong divergence in productivity growth between firms at the "frontier firms" (defined as the top 5% of firms in terms of productivity) and non-frontier firms. They show that labor productivity grew faster at frontier firms and that the gap reflects divergence in TFP, not capital deepening. Similarly, Berlingieri, Blanchenay, and Criscuolo (2017) use harmonized business register data covering a subset of OECD countries and document a broad-based increase of productivity dispersion, which holds for both labor productivity and TFP.

Existing explanations for the decline of the labor share include a rise of outsourcing along with changes in labor market institutions (Elsby, Hobijn, and Şahin (2013)) as well as capital-labor

substitution caused by a decline in the relative price of capital equipment (Karabarbounis and Neiman (2014)). A more recent set of papers argue that product market power has increased, leading to a rising importance of profits in aggregate income. Barkai (2016) uses U.S. National Accounts data to impute aggregate payments to capital equipment and finds that the decline of the labor share was not compensated by a rise of the capital share, but rather a rise of the profit share. Eggertsson, Robbins, and Wold (2018) interpret historical data through the lens of DSGE model and argue that increased markups are needed to account for a set of secular trends in the U.S. economy, including the decline of the labor share. De Loecker and Eeckhout (2017) estimate firm-level price markups using Compustat data and find that markups have increased substantially starting around 1980.

Kehrig and Vincent (2017) and Autor et al. (2017) present firm-level evidence that puts restrictions on the set of explanations for the decline of the labor share that are consistent with the data. Kehrig and Vincent (2017) focus on the manufacturing sector over the 1967-2012 period. They find that the aggregate labor share decline was driven by a *reallocation* of value-added towards firms with a low (and declining) labor share, but that there was limited reallocation of inputs (labor and capital). They also show that low-labor-shares firms are mostly firms with high TFP, not firms with high capital intensity. Autor et al. (2017) use data covering most sectors over the 1982-2012 period. They find that the labor share decline was driven by reallocation of market shares, as opposed to a broad-based decline of firm-level labor shares. They also show that industries that saw the largest increases in sales concentration also saw the largest declines in their labor shares.

Finally, Hartman-Glaser, Lustig, and Zhang (2018) study the link between the volatility of sales and the aggregate capital share in a firm dynamics model where firms (risk neutral) insure workers (risk averse). They find that an increase in firm-level risk generates an increase in the aggregate capital share. My paper differs in that I study the effect of a rise in productivity dispersion (a cross-sectional moment) while they study the effect of a rise in volatility (a time-

series moment). Moreover, the feature that generates a link between productivity dispersion and aggregate labor share in my model is the presence of search frictions—which grants firms monopsony power—as opposed to risk aversion.

My paper also relates to the literature on labor market monopsony. The theoretical underpinning of the *new monopsony literature* is the equilibrium search model developed in Burdett and Mortensen (1998) (henceforth BM).[1] Recent applications of the BM model to study the wage-setting behavior of firms include: Meghir, Narita, and Robin (2015), who study the role of informal labor markets; Engbom and Moser (2017), who study the effect of the large increase of the minimum wage in Brazil, and Heise and Porzio (2018), who study the role of firms and location preferences in shaping spatial wage gaps. My model differs in that it incorporates firm dynamics (endogenous entry, exit, firm growth and productivity shocks). There is a limited amount of work on firm dynamics models with search frictions in the labor market. Notable exceptions include Elsby and Michaels (2013), Kaas and Kircher (2015), Gavazza, Mongey, and Violante (2018) and Coles and Mortensen (2016). The main difference with the models in the first three papers is that they do not model job-to-job flows (no on-the-job search). On-the-job search is central element of my model: it is the worker's threat of quitting to a higher paying firm that forces firms to offer wages above the value of nonemployment.

Coles and Mortensen (2016) (henceforth CM) allow for job-to-job flows, but they do not model endogenous firm entry and exit. They assume that the lower bound of firm productivity is high enough that every firm makes positive profits in every date and state, so firms never wish to exit. In contrast, I do not restrict the firm productivity process and model firm exit explicitly (i.e. firms solve an optimal stopping time problem). Hence, my model generates negative firm profits in equilibrium, which is a pervasive feature of the data. Moreover, CM model the hiring process as costly but unrelated to the wage-setting decision. Instead, I model the matching process exactly as in BM (random matching with exogenous meeting rates), which implies that high-wage firms

[1] see Ashenfelter, Farber, and Ransom (2010) and Manning (2011) for literature reviews

grow faster due to higher poaching flows. The wage-setting decision of firms is tightly related to this "poaching incentive".

I leverage the insight from CM that the firm problem in the BM model is tractable out-of-steady-state when the production technology exhibits constant-returns-to-scale (CRS). I build on this result and show that, with exogenous meeting rates and a CRS matching technology, the allocation of workers to firms in a stationary equilibrium can be characterized analytically. In particular, I provide necessary conditions for the existence of a stationary equilibrium (Assumption 3) and a closed-form solution for the employment-weighted productivity distribution and unemployment rate (Proposition 3). These results are crucial for my application, as they allow me to tractably aggregate the model and prove a number of results analytically. Finally, I leverage theoretical results on power laws (Beare and Toda (2017a)) and a novel solution method for heterogeneous-agent models with fat-tails (Gouin-Bonenfant and Toda (2018)) to numerically solve the joint-distribution of productivity, employment and labor share.

1.2 Motivating facts

The aggregate labor share LS is the sum of firm-level labor shares LS_i weighted by their output share Y_i/Y.

$$LS = \sum_i \underbrace{(Y_i/Y)}_{\text{output share}} \times \underbrace{LS_i}_{\text{labor share}} \tag{1.1}$$

Recent evidence highlights the fact that the decline of the U.S. labor share was driven by a reallocation of output shares toward firms with low labor shares, rather than a uniform decline of firm labor shares (Kehrig and Vincent (2017), Autor et al. (2017)). Hence, to be able to interpret movements of the aggregate labor share through the lens of a model, one needs to work with a model that can speak to the firm-level heterogeneity in the data.

From Equation 1.1, we can see that, to be consistent with both firm-level and aggregate

data, a model needs to match (1) the distribution of labor shares, (2) the distribution of output shares, and (3) the dependence between labor shares and output shares. It is well known that output shares are distributed according to Zipf's law (Axtell (2001a)), but little is know regarding the distribution of firm labor shares, especially outside of the manufacturing sector. In this section, I focus on three facts that summarize the distribution of firm-level labor shares and their dependence with output shares.

I use microdata data from the *National Accounts Longitudinal Microdata File* (NALMF), which is produced by Statistics Canada by merging administrative data from different sources. The NALMF contains de-identified data covering the universe of private sector employers in Canada over the 2000-2015 period. The unit of observation is an enterprise, which is an entity larger than an establishment but smaller than a (consolidated) corporation.[2] I restrict the main sample to the private corporate sector excluding: Agriculture, Mining, Utilities, Education, and Health (NAICS 11, 21, 22, 61, and 62). Agriculture and Mining are excluded due to data limitations while Utilities, Education, and Health are excluded due to the fact that these sectors are dominated by public entities in Canada. Finally, I restrict the sample to firm-year observations with more than 5 employees. The labor share of firm i in year t is defined as

$$LS_{i,t} = \frac{\text{worker compensation}_{i,t}}{\text{worker compensation}_{i,t} + \text{gross profits}_{i,t}}$$

In Appendix A.2.1, I describe in detail how I construct the variables and then validate against aggregate data. The final sample contains 3,084,182 firm-year observations, so I do not report standard errors when presenting cross-sectional averages.

[2]The full definition from *Statistics Canada*'s website is "Enterprise refers to the highest level of the Business Register statistical hierarchy and is associated with a complete set of financial statements. The enterprise, as a statistical unit, is defined as the organizational unit of a business that directs and controls the allocation of resources relating to its domestic operations, and for which consolidated financial and balance sheet accounts are maintained from which international transactions, an international investment position and a consolidated financial position for the unit can be derived. It corresponds to the institutional unit as defined for the System of National Accounts."

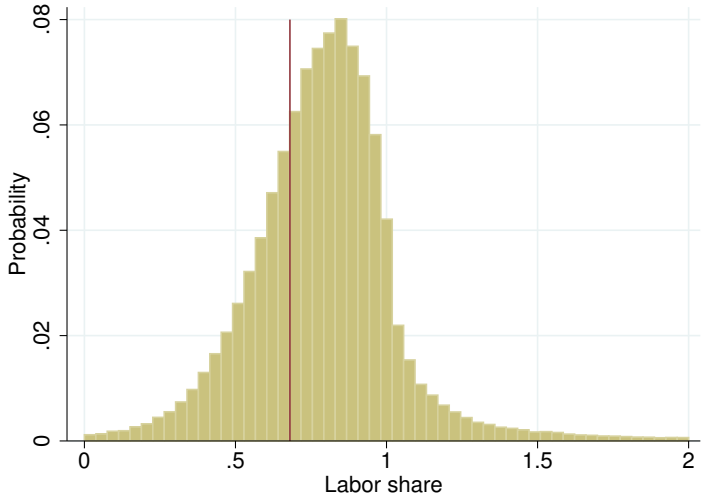

Figure 1.1: Distribution of firm labor shares.

1.2.1 Fact #1: There are large and persistent differences in labor shares across firms

Figure 1.1 presents the distribution of firm labor shares for the year 2015. A striking fact is that behind the aggregate labor share lies a lot of heterogeneity. The vertical red line represents the aggregate labor share of 0.67. Those large labor share differentials also hold within industry. To emphasize this point, Figure 1.2 presents the average labor share of firms by labor share deciles. I construct the deciles by sorting firms within 2-digit NAICS industry and year bin, and then pooling all industries and years together. The dispersion in firm labor shares within industry is large. For instance, firms in the second decile have a labor share of 0.6 on average while those in the ninth decile have a labor share of almost one. The fact that these difference hold within an industry suggests that there are factors other than technology that generate dispersion in firm labor shares.

These labor share differentials are also quite persistent. Figure 1.3 plots the empirical transition probability matrix for labor share deciles. For example, the $(1,2)$ entry contains the empirical frequency at which a firm in the first labor share decile moves to the second decile

from one year to the next. If differences in labor shares were only due to noise, all entries would be equal to 0.1. I find that the persistence in the data is high, which is exemplified by the fact that values along the diagonal are dark. The probability that a firm's labor share remains within the same or an adjacent decile from one year to the next is between 0.6 and 0.8 depending on the initial decile. For example, the probability that a firm in the fifth decile moves to either the fourth, fifth, or sixth decile in the following year is 0.63. Kehrig and Vincent (2017) have also documented large and persistent labor share differentials in the U.S. manufacturing sector.

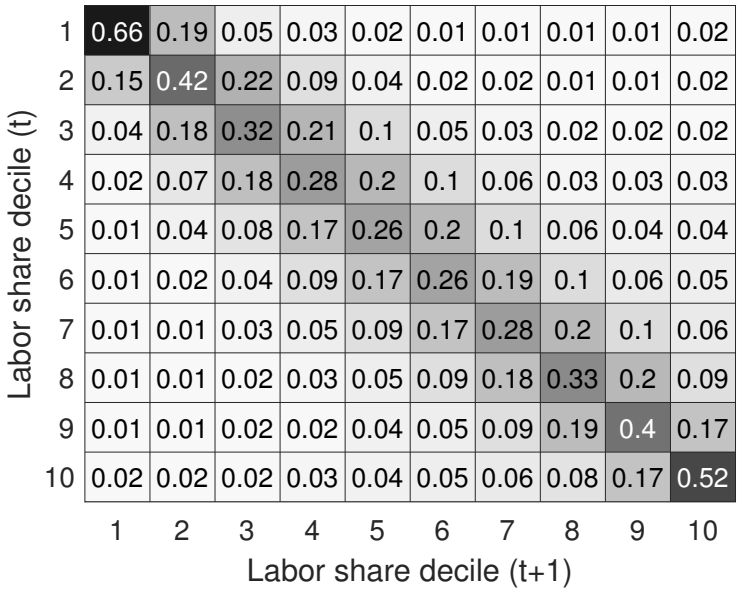

Figure 1.2: Transition matrix.

1.2.2 Fact #2: In any given year, a large fraction of firms have a labor share above one

Notice from Figure 1.1 that the distribution of labor shares has a long upper tail. In fact, 17% of firms have a labor share larger than one. A labor share above one means that worker compensation exceeds value-added, and thus implies that gross profits are negative. While the aggregate labor share is always below one, in any given year many firms will make losses. This

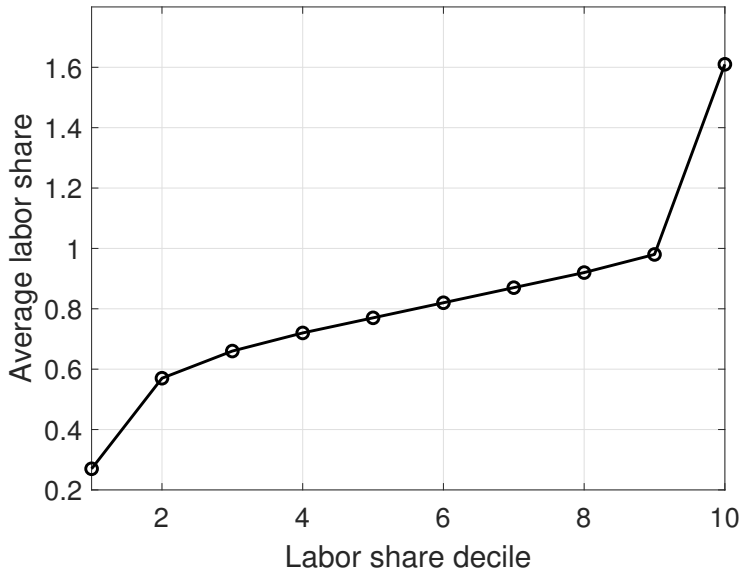

Figure 1.3: Labor share by decile.

finding is not only true for the year 2015, Figure 1.2) shows that the average labor share for the top two deciles of firms are around 1 and 1.6 respectively. So labor shares above one are a pervasive feature of the data, not just the result measurement errors or a feature of the Canadian data.[3] Truncating the labor share distribution at one would imply throwing away nearly one out of five observation and would bias the aggregate labor share downwards.

1.2.3 Fact #3: There is a strong negative relationship between firm labor share and firm output

Figure 1.4 contains a scatterplot of firm-level output (in millions of Canadian dollars) and labor share for the year 2015. The data is smoothed (locally weighted scatterplot smoothing with bandwidth of 0.8) in order to maintain confidentiality. A clear relationship emerges where firms with high output tend to have a lower labor share. To formalize this point, consider the following decomposition of the aggregate labor share in the spirit of Olley and Pakes (1996), which follows

[3]For instance, Figure 1 in Kehrig and Vincent (2017) shows that the distribution of firm labor shares in the U.S. manufacturing sector has significant mass above one.

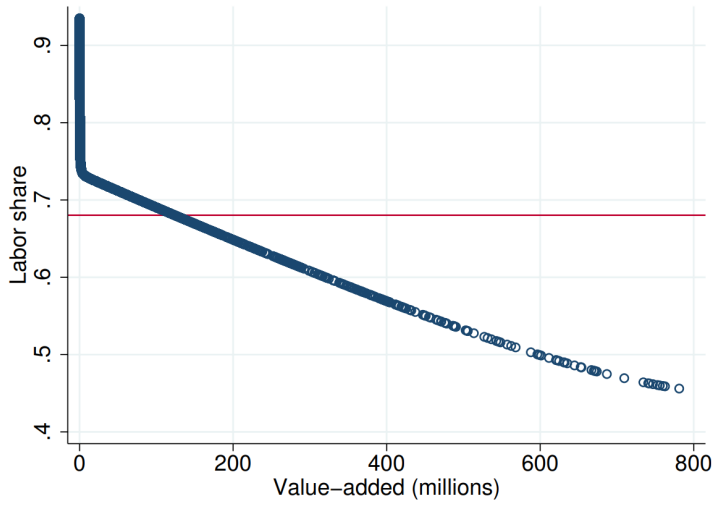

Figure 1.4: Smoothed scatter plot of firm-level labor share and value-added.

directly from Equation 1.1.

$$\underbrace{LS}_{\substack{\text{aggregate}\\ \text{labor share}}} = \underbrace{\bar{LS}}_{\substack{\text{average}\\ \text{labor share}}} + \underbrace{cov\left(LS_i\,,\, Y_i/Y\right).}_{\substack{\text{covariance between}\\ \text{labor share and output share}}} \tag{1.2}$$

The aggregate labor share can thus be expressed as the sum of the average (unweighted) firm labor share across and the covariance between labor share and output share. Table 1.1 presents the results of the decomposition. To account for differences across industries and years, I compute each of the three components in Equation 1.2 separately within industry-year bin (NAICS 2-digit). I then average industries within a year using industry value-added weights and then average over all years. The aggregate labor share is 0.65 yet the average (unweighted) firm labor share is much higher at 0.88. The reason why the gap between average and aggregate labor share is so large is that firms with higher output tend to have a lower labor share. The covariance between firm-level output share and labor share is -0.23. The negative association between firm-level labor share and output share had previously been documented in the case of the U.S. manufacturing sector (Kehrig and Vincent (2017)), the Taiwanese manufacturing sector (Edmond, Midrigan, and Xu

(2015)), and most other sectors of the U.S. economy (Autor et al. (2017)).

Table 1.1: Aggregate labor share decomposition.

Component	Value
Aggregate labor share	0.65
Average (unweighted) labor share	0.88
Covariance between labor share and output share	-0.23

Notes. The three components are defined in Equation 1.2. Each component is computed within each 2-digit NAICS industry and then averaged using value-added weights. The results are 2000-2015 averages.

1.3 Model

1.3.1 Preferences and production

Time is continuous and an interval $[t, t+1)$ represents a year. The economy is populated by an endogenous measure F of heterogeneous firms and a unit measure of identical, infinitely-lived workers. Firms and workers are risk neutral and discount the future at rate $r > 0$. Firms compete for workers by posting wage contracts w and rent capital in a perfectly competitive market at user cost $R \equiv r + \delta$, where δ is the depreciation rate of capital. Firms can change the wage policy at any time and do not precommit to future wages. For simplicity, I assume that firms are required to pay all of their workers the same wage w. Workers can be either employed or unemployed and are available to receive job offers in both states. The flow value of unemployment is exogenous and given by $b > 0$. Firms produce an homogeneous good and differ only in their total factor productivity (TFP) level $z \geq 0$. The production function is Cobb-Douglas with constant-returns-to-scale $zK^{\alpha}N^{1-\alpha}$, where $N \geq 0$ denotes the measure of workers a firm employs and $K \geq 0$ denotes its stock of capital.

1.3.2 Matching technology

At exogenous rate $\mu \geq 0$, an unemployed worker meets an entrepreneur (a potential firm). The entrepreneur then draws a productivity level z from the distribution Γ_0 and decides whether or not to enter. There is free entry, so entrepreneurs enter whenever expected discounted profits are positive. Without loss of generality, entry will follow a threshold rule where the entrepreneur enters if and only if $z \geq z_l$. The threshold z_l is an equilibrium object which I will characterize later. I now make the following assumption regarding the source distribution (CDF) Γ_0.

Assumption 1. *The cumulative distribution function* $\Gamma_0 : [0, +\infty) \to [0, 1]$ *is differentiable and satisfies*

$$\Gamma_0'(z) > 0, \quad \int_0^\infty z\Gamma_0'(z)dz = 1, \quad \int_0^\infty z^{\frac{1}{1-\alpha}}\Gamma_0'(z)dz < \infty.$$

The first condition ensures that there is positive mass everywhere on $[0, \infty)$, the second condition is merely a normalization, and the third condition ensures that aggregate output is finite.

Firm productivity changes over time. At rate χ, continuing firms draw a new productivity level from Γ_0. This particular process implies that firm productivity z remains constant for a stochastic duration, and then jumps to a new level that is independent from the previous one. The optimal behavior of firms will be characterized later, but the decision rule will be to exit whenever z falls below an endogenous threshold z_l. Notice that I am using the same notation z_l for the entry and exit thresholds. I will later show (Lemma 2) that those two objects are equal in equilibrium. Firms decide when to exit. When they exit, their workers are sent into unemployment. To simplify notation, define the distribution Γ_0 truncated at z_l by $\Gamma(z)$, the rate at which an unemployed workers meets an entering firm χ_e, the firm exit rate by χ_x, and the arrival rate of productivity shocks χ_s by

$$\Gamma(z) \equiv \frac{\Gamma_0(z) - \Gamma_0(z_l)}{1 - \Gamma_0(z_l)}, \quad \chi_e \equiv \mu(1 - \Gamma_0(z_l)\big), \quad \chi_x \equiv \chi\Gamma_0(z_l), \quad \chi_s \equiv \chi\big(1 - \Gamma_0(z_l)\big). \tag{1.3}$$

The law of motion for the measure of active firms F is thus given by

$$\dot{F} = \underbrace{\chi_e u}_{\text{firm entry}} - \underbrace{\chi_x F}_{\text{firm exit}} . \tag{1.4}$$

I now describe how continuing firms meet new workers and grow. A firm with a measure N of workers meets a new worker at rate λN, where $\lambda > 0$ is a technological parameter. I assume random matching, meaning that the identity of the workers a firm meets is randomly drawn from the population, independently of their current status (employed or unemployed).[4] Of those workers, a fraction u will be unemployed while the remainder $1-u$ will be employed. The rate at which workers (employed or unemployed) receive a job offer from a continuing firm is given by $\lambda(1-u)$. This is because each continuing firm meets λN workers at every instant, so the measure of meetings is given by λ times aggregate employment $1-u$. When workers meet firms, I assume that they do not observe the firm's productivity z.

Assumption 2. *The wage paid by a firm is public information but its productivity z is private.*

The solution that I study is a *Pure Strategy Bayesian Equilibrium*. The equilibrium is *Bayesian* given that workers need to form expectations about the future path of wages at different firms in order to compute continuation values. The *pure strategy* part prevents agents from randomizing their actions. I will focus on size-invariant equilibria in which wages are increasing in firm productivity (as in Coles and Mortensen (2016)). The focus on size-invariant equilibria means that a firm with productivity z and size N offers a wage $w(z)$ that does not depend on its size N. This is not very restrictive since both the production technology and matching technology exhibit constant-returns-to-scale. The only reason why firms would offer size-dependent wages would be if workers' beliefs induced them to behave in that manner. The focus on equilibria in

[4] To give some context, in Burdett and Mortensen (1998) a firm meets a worker at a *constant* rate λ. Since the meeting rate does not grow with a firm's size N, it means that the matching technology exhibits decreasing-returns-to-scale. The only departure from *BM* in the labor market module of my model is that I assume that the matching technology exhibits constant-returns-to-scale.

which wages are increasing in firm productivity means that the equilibrium policy function of the firm satisfies $w(z') > w(z)$ whenever $z' > z$. This is not restrictive, since the case $w(z') \leq w(z)$ would not be consistent with an equilibrium.

To simplify the construction of the equilibrium, I will impose from now on that the workers' beliefs are consistent with a size-invariant equilibrium in which wages are increasing in firm productivity. I Denote the belief function, which maps productivity to wages, by $\widehat{w}(z)$. The interpretation is that a workers believe that firms with productivity z will pay $\widehat{w}(z)$ with probability one. Given the focus on equilibria where $w(z)$ is monotone, the equilibrium belief function must satisfy $z' > z \implies \widehat{w}(z') > \widehat{w}(z)$.

1.3.3 Worker and firm problems

The only decision made by workers is whether or not to accept a job offer. To do so, workers need to compute the present value of accepting and rejecting a job offer paying. The value of unemployment U and the value of working at a firm paying $W(w)$ are the solutions to the following system of equations.

$$rU = b + \underbrace{\chi_e \int \max\Big\{W(\widehat{w}(z)) - U, 0\Big\} d\Gamma(z)}_{\text{job offer from entrant}} + \underbrace{\lambda(1-u) \int \max\Big\{W(w') - U, 0\Big\} d\widetilde{P}(w')}_{\text{job offer from continuing firm}} \tag{1.5}$$

$$\begin{aligned} rW(w) = w &+ \underbrace{\chi_s \int \Big(W(\widehat{w}(z)) - W(w)\Big) d\Gamma(z)}_{\text{wage change due to productivity shock}} \\ &+ \underbrace{\lambda(1-u) \int \max\Big\{W(w') - W(w), 0\Big\} d\widetilde{P}(w')}_{\text{job offer from continuing firm}} + \underbrace{\chi_x \Big(U - W(w)\Big)}_{\text{job destruction}} \end{aligned} \tag{1.6}$$

$$W(w) \geq U \tag{1.7}$$

Given that workers can quit to unemployment, the value of being employed $W(w)$ must weakly exceed the value of unemployment U (Equation 1.7). The distribution $\widetilde{P}$ represents the distribution of wages in the population of workers. Since firms meet workers proportionally to their employment, the distribution of wage offers coming from continuing firms is precisely $\widetilde{P}$. The optimal job acceptance rule is characterized in the following lemma.

Proposition 1. *Employed workers accept any wage-increasing job offer. Unemployed workers accept any job offer above a an endogenously-determined reservation wage w_r given by*

$$w_r = b + (\mu - \chi) \int_{z_l}^{\infty} \max \Big\{ W(\widehat{w}(z)) - U, 0 \Big\} \Gamma_0'(z) dz \tag{1.8}$$

Proof. See Appendix A.1.1 □

By definition, active firms pay wages above w_r, otherwise all their workers would have quit to unemployment. Therefore, the law of motion for the unemployment rate can be expressed as

$$\dot{u} = \underbrace{(1-u)\chi_x}_{\substack{\text{Unemployment} \\ \text{inflows}}} - \underbrace{u\Big(\chi_e + \lambda(1-u)\Big)}_{\substack{\text{Unemployment} \\ \text{outflows}}}. \tag{1.9}$$

Firm growth. Equipped with the worker's optimal job offer acceptance rule, I now characterize the link between firm-level wage and employment growth. The instantaneous change in employment at a firm of size N paying $w \geq b$ is given by

$$\dot{N} = \widetilde{g}(w) N, \tag{1.10}$$

where the employment growth function $\widetilde{g}(w)$ is an endogenous object which depends on the wage distribution $\widetilde{P}(w)$. Notice that $\widetilde{g}$ depends on the current wage policy w but not on size N. This result is due to the fact that meeting rates are linear in size and that the worker's acceptance rule depends only on the current wage (Proposition 1). A simple expression can be obtained for the

employment growth function $\widetilde{g}$ *conditional on survival*

$$\widetilde{g}(w) \equiv \underbrace{\lambda u + \lambda(1-u)\widetilde{P}(w)}_{\text{hiring rate}} - \underbrace{\lambda(1-u)\big(1-\widetilde{P}(w)\big)}_{\text{separation rate}} \quad \forall w \geq w_r \tag{1.11}$$

Since meeting rates are exogenous, the hiring rate depends only on the fraction of job offers that are accepted. As a result of the random matching assumption, a fraction u of meetings are with unemployed workers and a fraction $1-u$ are with employed workers. Unemployed workers accept all job offers above w_r, while employed workers require a wage increase to quit (Proposition 1). The measure of employed workers at firms paying less than w is equal to $\widetilde{P}(w)$, so the hiring rate from employment is simply given by $\lambda(1-u)\widetilde{P}(w)$. Firms paying higher wages thus have a higher hiring rate from employment as they are able to poach workers higher up in the wage distribution.

Separations occur only due to quits. At rate $\lambda(1-u)$, a worker receives a competing job offer. Since firms meet workers proportionally to their size, the distribution of offered wages is equal to the distribution of wages in the population $\widetilde{P}(w)$. The rate at which a firm paying w loses its workers to competitors is thus given by $\lambda(1-u)(1-\widetilde{P}(w))$. Hence, high-wage firms face a lower separation rate due to the fact that a smaller fraction of their workers get poached by competing firms. The following Lemma characterizes the employment growth function.

Lemma 1. *The growth rate function $\widetilde{g}(w)$ is weakly increasing and bounded from above by λ.*

Proof. The results are obtained directly from Equation 1.11, using the fact that $\widetilde{P}(w) \to 1$ □

The firm problem is to choose a sequence of wage rates and levels of capital to rent $\{w_s, K_s\}_{s=0}^{T}$ as well as an "exit time" T to maximize the expected flow of discounted profits.[5] Let

[5]The notation can be confusing but the exit time T is not chosen at time zero, but rather is a function of the full history of the idiosyncratic shocks.

$v(z,N)$ denote the value of a firm with productivity z and employment N.

$$v(z,N) = \max_{\{w_s,K_s\}_{s=0}^{T},\, T\geq 0} \mathbb{E}_0 \int_0^T e^{-rs}\Big(z_s K_s^{\alpha} N_s^{1-\alpha} - w_s N_s - RK_s\Big)ds \tag{1.12}$$

$$N_0 = N, \quad z_0 = z \qquad \text{(initial condition)}$$

$$w_s \geq w_r \qquad \text{(participation constraint)}$$

$$dN_s = \widetilde{g}(w_s)N_s ds \qquad \text{(law of motion for labor)}$$

$$dz_s = d\mathcal{J}_s(z_s' - z_s) \qquad \text{(productivity process)}$$

$\{\mathcal{J}_s\}_{s=0}^{\infty}$ denotes a Poisson process with intensity χ and $\{z_s'\}_{s=0}^{\infty}$ is a sequence of independent draws from Γ_0. Due to the exit decision, the value function does not obey a standard *Hamilton-Jacobi-Bellman* (HJB) equation. It is obtained as the solution to the following variational inequality of the obstacle type.[6]

$$\min\Bigg\{ rv(z,N) - \max_{\substack{w\geq b\\ K\geq 0}}\Big\{ zK^{\alpha}N^{1-\alpha} - wN - RK + \frac{\partial}{\partial N}v(z,N)\widetilde{g}(w)N\Big\} \tag{1.13}$$
$$-\chi\Big(\int v(z',N)\Gamma_0(dz') - v(z,N)\Big)\,,\; v(z,N)\Bigg\} = 0.$$

Basically, Equation 1.13 says that either the value function obeys a regular HJB equation or it is equal to zero. Before going further, I make a parametric assumption to ensure that the value function is well-defined.

Assumption 3. *The meeting rate λ and rate of productivity resets χ satisfy the following inequality.*

$$0 < \lambda < \chi$$

The requirement that $\lambda > 0$ ensures that at least some firms grow over time (see Lemma 1),

[6]For a discussion on variational inequalities in economics, and in particular to solve optimal stopping time problems in firm dynamics models, I refer the reader to Achdou et al. (2014).

while $\lambda < \chi$ ensures that firms do not grow too fast. The case $\lambda \geq \chi$ would imply that the most productive firm would "take over" the whole economy.

Lemma 2. *The value function homogeneous of degree one in N, which means that* $v(z,N) = v(z,1)N$. *Firm entry and exit follows a threshold rule, so that firms exit whenever* $z < z_l$ *and entrepreneurs enter whenever* $z \geq z_l$.

Proof. See Appendix A.1.2 □

From now on, I denote the value of a firm with productivity z and a unit measure of worker by $v(z) \equiv v(z,1)$ and the stock of capital per worker by $k \equiv K/N$. Over the range $z \geq z_l$, the value function $v(z)$ obeys the following HJB equation

$$rv(z) = \max_{\substack{w \geq b \\ k \geq 0}} \Big\{ \underbrace{zk^{\alpha} - w - Rk}_{\text{profits}} + \underbrace{v(z)\widetilde{g}(w)}_{\text{growth}} \Big\} + \underbrace{\chi\Big(\int v(z')\Gamma_0(dz') - v(z) \Big)}_{\text{productivity shocks}}. \tag{1.14}$$

Over the range $z < z_l$, we have that $v(z) = 0$.[7] From Equation 1.14, we can see that the annuity value of a firm $rv(z)$ depends not only on current profits $zk^{\alpha} - w - Rk$ but also on future growth prospects and accounts for the risk of productivity shocks. A corollary of the homogeneity result (Lemma 2) is that the wage function $w(z)$ and the optimal the stock of capital per worker $k(z)$ are

[7] More generally, the value function can be characterized as the solution to a *Linear Complementary Problem*.

$$v(z)\Big(rv(z) - \max\Big\{ zk^{\alpha} - w - Rk + v(z)\widetilde{g}(w) \Big\} - \chi\Big(\int v(z')\Gamma_0(dz') - v(z) \Big) \Big) = 0 \tag{1.15}$$

$$rv(z) - \max\Big\{ zk^{\alpha} - w - Rk + v(z)\widetilde{g}(w) \Big\} - \chi\Big(\int v(z')\Gamma_0(dz') - v(z) \Big) \geq 0 \tag{1.16}$$

$$v(z) \geq 0 \tag{1.17}$$

The exit region is characterized by a set $\mathcal{X} = \{z : v(z) = 0\}$. I use this representation of the problem for the numerical solution (see Appendix A.1.9).

also size-independent. From the first-order conditions, we have

$$\underbrace{z\alpha k(z)^{\alpha-1}}_{\substack{\text{marginal product}\\ \text{of capital}}} = \underbrace{R}_{\substack{\text{User cost}\\ \text{of capital}}} \tag{1.18}$$

$$\underbrace{1}_{\substack{\text{marginal increase}\\ \text{in wage bill}}} = \underbrace{v(z)\widetilde{g}'\big(w(z)\big).}_{\substack{\text{marginal increase in}\\ \text{value of the firm}}} \tag{1.19}$$

The first-order condition for capital is standard and implies that firms rent capital up to the point where the marginal product of capital is equal to its user cost. The first-order condition for wages captures the core trade-off faced by a wage-setting firm in a dynamic environment. When a firm offers a high wage, it makes lower per-employee profit but grows faster, which increases the future value of the firm. I denote the equilibrium growth rate of employment by $g(z) \equiv \widetilde{g}(w(z))$. The following proposition provides closed-form expressions for the policy functions.

Proposition 2. *Capital intensity $k(z)$, labor productivity $LP(z)$ and the wage schedule $w(z)$ are increasing in z and are given by*

$$k(z) = \big(\alpha/R\big)^{\frac{1}{1-\alpha}} z^{\frac{1}{1-\alpha}} \tag{1.20}$$

$$LP(z) = \big(\alpha/R\big)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}} \tag{1.21}$$

$$w(z) = w_r + \int_{z_l}^{z} v(\zeta) g'(\zeta) d\zeta. \tag{1.22}$$

Proof. The expressions for $k(z)$ and $LP(z)$ are obtained directly from the first-order condition 1.18. For the derivation of $w(z)$, see Appendix A.1.3. □

Given that employment is predetermined while capital per worker is flexible, high z firms rent more capital per worker (see Equation 1.20). Equation 1.22 highlights the fact that workers earn a wage above their reservations wage and that the premium $w(z) - w_r$ is higher at more

productive firms. Finally, the firm-level labor share $LS(z)$ is given by

$$LS(z) \equiv \frac{w(z)}{LP(z)} = \frac{w_r + \int_{z_l}^{z} v(\zeta) g'(\zeta) d\zeta}{(\alpha/R)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}}}. \tag{1.23}$$

1.3.4 Equilibrium and characterizations

The analysis thus far is in partial equilibrium since the employment growth function $\widetilde{g}$ depends on the endogenously-determined wage distribution $\widetilde{P}$ through Equation 1.11. The wage distribution is central to my analysis. It fully summarizes the amount of wage competition that firms face. Instead of deriving a law of motion of the wage distribution, I focus on the employment-weighted productivity distribution $P(z) \equiv \widetilde{P}(w(z))$, as it will be easier to characterize. The law of motion for $P(z)$ is given by

$$\dot{P}(z) = \underbrace{(1-u)\lambda P(z)(P(z)-1)}_{\text{Job-to-job flows}} + \underbrace{\frac{u}{1-u}\chi_e \Gamma(z) + u\lambda P(z)}_{\text{Employment inflows}} - \underbrace{\chi_x P(z)}_{\substack{\text{Employment} \\ \text{outflows}}} + \underbrace{\chi_s(\Gamma(z) - P(z))}_{\text{Productivity shocks}}. \tag{1.24}$$

In Appendix A.1.4, I provide a derivation of the law of motion. Conditional on $P(z)$, the equilibrium employment growth function $g(z)$ can be computed using the definition of $\widetilde{g}(w)$ (Equation 1.11).

$$g(z) = \lambda u + \lambda(1-u)P(z) - \lambda(1-u)(1-P(z)) \tag{1.25}$$

I now define the equilibrium concept.

A *Stationary Bayesian Equilibrium* consists of a productivity distribution $\Gamma(z)$, rates (χ_e, χ_s, χ_x), a value function $v(z)$, a productivity threshold z_l, policy functions $w(z)$, $k(z)$, $g(z)$, a reservation wage w_r, an unemployment rate u, and an employment-weighted productivity distribution $P(z)$ that satisfy the following conditions:

1. The truncated productivity distribution $\Gamma(z)$ and rates χ_e, χ_s, χ_x are determined by the optimal entry and exit of firms (Equation 1.3);

2. The value function $v(z)$ and productivity threshold z_l solve the optimal stopping time problem of the firm (Equations 1.14);

3. The policy functions $k(z)$, $w(z)$, and $g(z)$ solve the firm problem (Equations 1.20, 1.22, and 1.25);

4. The reservation wage solve the worker problem (Equation 1.8);

5. The unemployment rate and employment-weighted productivity distribution $P(z)$ are stationary solutions to their respective laws of motion (Equations 1.9 and 1.24);

6. Workers beliefs are correct, meaning that $\widehat{w}(z) = w(z)$.

In a stationary equilibrium the measure of active firms is $F = \frac{\chi_e}{\chi_x} u$ (follows from Equation 1.4). Notice that F does not appear in the equilibrium definition. The reason is that, due to constant-returns-to-scale, large firms are merely scaled replicas of small firms. Hence, the measure of firms is irrelevant. What matters is the allocation of workers to firms. The following proposition provides closed-form expressions for the equilibrium unemployment rate u and the employment-weighted productivity distribution $P(z)$.

Proposition 3. *Conditional on the productivity threshold z_l, there exists a unique pair of unemployment rate u and employment-weighted distribution $P(z)$ that are stationary solutions to their respective laws of motion (Equations 1.9 and 1.24). They are given by*

$$u = \frac{\lambda + \chi_e + \chi_x - \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x}}{2\lambda} \tag{1.26}$$

$$P(z) = \frac{\lambda(1-u) + \chi - \lambda u - \sqrt{\big(\lambda(1-u) + \chi - \lambda u\big)^2 - 4\lambda(1-u)(\chi - \lambda u)\Gamma(z)}}{2(1-u)\lambda} \tag{1.27}$$

Proof. See Appendix A.1.4. The existence of a solution is guaranteed by Assumption 3. □

A corollary of Proposition 3 is that the distribution of workers across firm productivity ranks is invariant to the underlying productivity distribution Γ. For example, the measure of workers employed at firms below the median productivity level $\Gamma^{-1}(1/2)$ is given by $P(\Gamma^{-1}(1/2))$. Notice from Equation 1.27 that $P(\Gamma^{-1}(1/2))$ depends only on search frictions (χ, λ, u), and in particular *does not* depend on the productivity distribution Γ. Before taking the model to the data, I discuss the theoretical mechanisms that shape the equilibrium relationship between firm-level productivity z, size N, and wages w.

1.3.5 Pass-through of productivity to wages

How do wages offered relate to firm productivity? Recall that the only incentive for firms to offer wages above the reservation wage w_r is to increase the growth rate of employment. This incentive is captured by the first-order condition for wages evaluated along the equilibrium path (Equation 1.19)

$$w'(z) = v(z)g'(z). \tag{1.28}$$

Equation 1.28 states that the pass-through of firm productivity to wages is equal to the increase in the growth rate $g'(z)$ multiplied by the present value of rents accruing to the firm $v(z)$. The interpretation of $w'(z)$ as a pass-through comes from the fact that if the productivity of a firm increases from z to $z+\Delta$, its wage will increase from $w(z)$ to approximately $w(z)+w'(z)\Delta$. I now use the expression for $g(z)$ (Equation 1.25) to unpack the growth effect $g'(z)$.

$$w'(z) = \underbrace{v(z)}_{\text{value effect}} \times \underbrace{2\lambda(1-u)}_{\text{search frictions}} \times \underbrace{P'(z)}_{\text{local competition}} \tag{1.29}$$

Equation 1.29 highlights the importance of wage competition in shaping the pass-through of productivity to wages. The "local competition" term $P'(z)$ represents the density of employment at z. If the density of employment at z is high, a firm faces strong incentives to increase its

wage so that it can poach workers from firms with similar productivity. If, on the other hand, a firm operates at a productivity level where the density of employment is low, it faces weak incentives to increase its wage. Optimal wage-setting behavior in the model is thus intimately tied to between-firm competition for workers. Notice that the pass-through of productivity to wages is stronger when search frictions are less severe (λ is high) and when the labor market is tight (u is low). Using this characterization of $w'(z)$, I now establish two properties of the pass-through of labor productivity $LP(z)$ to wages.

Proposition 4. *The pass-through of labor productivity to wages* $\frac{dw}{dLP}(z)$ *satisfies two properties:*

1. *It is non-negative*

$$\frac{dw}{dLP}(z) \geq 0 \quad \forall z \geq 0$$

2. *It converges to zero*

$$\lim_{z \to \infty} \frac{dw}{dLP}(z) = 0$$

Proof. See Appendix A.1.5 □

The main takeaway from Proposition 4 is that very productive firms face weak incentives to pass productivity gains to their workers in the form of higher wages—a prediction shared by standard wage posting models as Burdett and Mortensen (1998). The reason is that they face very few competitors with similar levels of productivity. Increasing wages does not increase employment growth much as they already are the highest paying firms, so they face no incentive to do so. In the language of monopsony theory, high-wage firms face a (locally) inelastic labor supply curve, so their wage mark-down (i.e. gap between wage and marginal productivity) is high.

Figure 1.5 illustrates the relationship between labor productivity and wages in the calibrated model (see Section 1.4.1 for a description of the calibration strategy). We can see that wages increase with labor productivity, but the relationship becomes very weak in the upper range.

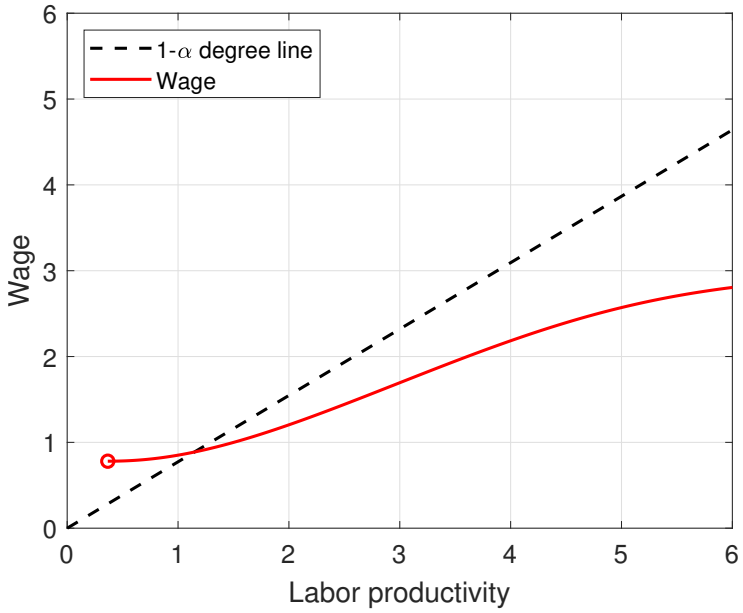

Figure 1.5: Equilibrium wage schedule in the calibrated model.

The dashed line represents a share $1-\alpha$ of labor productivity, which would be the wage in a frictionless, competitive model.[8] Figure 1.6 contains the resulting relationship between labor share and labor productivity. Notice that some firms have a higher labor share than than $1-\alpha$, which means that after payments to capital are taken into account, they make negative profits. In contrast, high-productivity firms have a labor share below $1-\alpha$.

In the calibrated model, the labor share is everywhere decreasing. This is not always the case, yet I can prove the following result.

Proposition 5. *There exists a productivity threshold $\widetilde{z}$ such that the labor share becomes decreasing in firm productivity.*

$$\exists\, \widetilde{z} \,:\, \forall z > \widetilde{z},\ \frac{d}{dz} LS(z) < 0$$

Proof. See Appendix A.1.5. □

[8] With perfect competition, wages are equal to the marginal product of labor $w = (1-\alpha)LP$, which implies that the labor should be $1-\alpha$ at every firm.

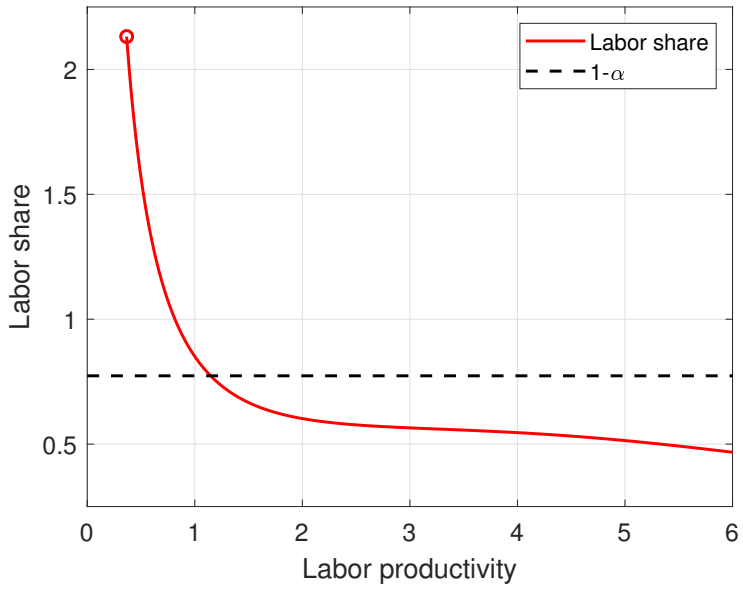

Figure 1.6: Equilibrium labor share schedule in the calibrated model.

1.3.6 Joint-distribution of productivity and employment

In the model, firm decisions are size-invariant. Yet, in a stationary equilibrium firm productivity and employment size are *dependent* random variables. While there is no one-to-one relationship between productivity z and size N, I prove that *on average*, more productive firms are larger.

Proposition 6. *The expected value of firm employment N conditional on productivity z is an increasing function of z. It is given by*

$$\mathbb{E}(N|z) = \frac{1-u}{u}\frac{\chi_x}{\chi_e}\frac{\chi-\lambda u}{\sqrt{\left(\lambda(1-u)+\chi-\lambda u\right)^2-4\lambda(1-u)(\chi-\lambda u)\Gamma(z)}}.$$

Proof. See Appendix A.1.6 □

The reason why productivity and size are contemporaneously correlated is that more productive firms pay higher wages and thus grow faster, and thus tend to be larger. The dependence between size and productivity in a stationary equilibrium is fully summarized by the joint density

$\varphi(z,N)$, which can be obtained as the solution to the following *Kolmogorov Forward Equation*

$$0 = \underbrace{\chi_s\Big(\frac{\partial}{\partial z}\Gamma(z)\int_{z_l}^{\infty}\varphi(z',N)dz' - \varphi(z,N)\Big)}_{\text{productivity shocks}} - \underbrace{g(z)\frac{\partial}{\partial N}\Big(\varphi(z,N)N\Big)}_{\text{firm growth}} + \underbrace{\chi_x\Big(\frac{\partial}{\partial z}\Gamma(z)\psi(N-1) - \varphi(z,N)\Big)}_{\text{firm turnover}}, \tag{1.30}$$

where $\psi(N)$ denotes the *Dirac Delta* function. The following proposition characterizes the upper tail of the firm size distribution.

Proposition 7. *The firm size distribution has a Pareto upper tail, meaning that there exists a unique* ζ *such that*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N>n) = -\zeta,$$

Moreover, the Pareto exponent satisfies $\eta > 1$.

Proof. The equilibrium firm growth process falls into the class of models covered by Theorem 3.1 in Beare and Toda (2017a). The result that $\zeta > 1$ follows from Assumption 3. □

The stationary density $\varphi(z,N)$ does not admit a closed-form solution, so I use the "Pareto Extrapolation" method developed in Gouin-Bonenfant and Toda (2018) to numerically solve φ over a finite grid and extrapolate the tails off the grid using the analytical characterization of the Pareto exponent ζ (see Appendix A.1.9). One well-known empirical regularity regarding the firm size distribution is Zipf's law, which states that the size distribution of firms obeys a power law with Pareto exponent close to (but above) one (see Gabaix (1999) and Axtell (2001a)). Proposition 7 states that the stationary firm size distribution has a Pareto upper tail with exponent *above* one. Whether or not the model delivers Zipf's law is ultimately a quantitative question.

Aggregating the model. The aggregate labor share is obtained by integrating firm-level labor shares weighted by their output share against the stationary distribution.

$$LS = \int\int \underbrace{\omega(z,N)}_{\text{output share}} \times \underbrace{LS(z)}_{\text{labor share}} \times \underbrace{\varphi(dz,dN)}_{\text{stationary distribution}} \tag{1.31}$$

I denote by $\omega(z,N)$ the output share of a firm with employment N and productivity z

$$\omega(z,N) = \frac{N \times LP(z)}{Y},$$

where Y is aggregate output.[9] Equation 1.31 highlights the fact that the aggregate labor share is shaped by rich interactions between firms in the labor market. In general, changes in the distribution of TFP draws Γ_0 will affect the aggregate labor share LS by changing the relative productivity differentials of firms throughout the productivity distribution, and therefore their incentives.

Proposition 8. *If TFP increases by a factor $\pi > 0$ for all firms and the value of unemployment b increases by a factor $\pi^{\frac{1}{1-\alpha}}$, then the aggregate labor share LS remains unchanged.*

Proof. See Appendix A.1.7 □

1.3.7 Interpreting differences in firm-level TFP

Before I go to the data, I want to emphasize the fact that I am being agnostic about the reasons why firms differ in their TFP. In the data, I can not observe prices separately from quantities, so the measure of labor productivity that I will construct is revenue-based. Hence,

[9] Using the employment-weighted productivity distribution $P(z)$, the labor share can be expressed as

$$LS = \int \omega(z,1) \times LS(z) \times P(dz)$$

.

differences in labor productivity can be due to differences in prices and/or differences in physical productivity.

I now give an example of a micro foundation of TFP in the model that arises purely from differences in prices. So far, I have assumed that firms produce an homogeneous good whose price is normalized to one. I now relax this assumption. Suppose that firms all have the same productivity $z = 1$ but produce differentiated goods. Let firms be indexed by $\omega \in [0, F]$. A representative consumer derives utility from a bundle of goods according to the following utility function

$$\mathcal{U}(c) = \int_0^F \beta(\omega) c(\omega) d\omega,$$

where c and β are infinite-dimensional non-negative vectors that represent respectively the amount of each variety that is being consumed and the preference weights.

Suppose that the price of each variety $p(\omega)$ is determined in a centralized market and that the representative consumer maximizes utility. For the market to clear, it must be the case that $p(\omega) = \beta(\omega)$ for all $\omega \in [0, F]$. The revenue production function of firm ω is thus given by $zK^{\alpha}N^{1-\alpha}$, where $z = \beta(\omega)$. Hence, we can reinterpret TFP differences in the model—both between firms and within firm over time—as arising purely from changing consumer preferences rather than from physical productivity. The ultimate question remains the same: when does firm (revenue) productivity translates into wages?

1.4 Quantifying the model

1.4.1 Calibration strategy

First, I model the distribution of TFP draws Γ_0 as a Gamma distribution with mean normalized to one. The distribution is thus characterized by a single shape parameter $\eta > 0$ which determines the thickness of the upper tail of distribution. The resulting PDF is defined over $[0, \infty)$

and given by

$$\Gamma_0'(z) = \frac{\eta^\eta}{G(\eta)} z^{\eta-1} e^{-z\eta}, \tag{1.32}$$

where $G(\eta)$ is the Gamma function. Lower values of $\eta > 0$ are thus associated with both higher variance and skewness. Second, I set the rate at which an unemployed worker meets an entrepreneur to $\mu = \chi$. Since the measure of active firms F is irrelevant for the labor share, the difference μ and χ only affects the reservations wage and the unemployment rate. By setting $\mu = \chi$, the reservation wage w_r becomes equal to the value of unemployment b (see Equation 1.8). I will therefore identify b directly. The model has seven parameters. I assign values to the first three (r, δ, α) and jointly calibrate the last four (λ, χ, η, b).

I set the net interest rate r to 3%, which is the difference between the effective business borrowing rate in Canada over the 2000-2008 period of as measured by the Bank of Canada[10] (5.3%) and the average realized CPI inflation over that same period (2.3%). I obtain the depreciation rate δ by using industry-level data from Statistics Canada. For the set of industries covered in the main sample, the average depreciation rate over the 2000-2015 is 17.6% (Table 031-0006). The Cobb-Douglas exponent α is set to 0.185 as to imply, conditional on $R \equiv r + \delta$, a capital-to-output ratio of 1.10 as measured in the main sample. I find a similar capital-output ratio of 1.02 using public data from Canadian National Accounts for the covered industries (Statistics Canada Tables 031-0006 and 383-0063). Table 1.2 summarizes these choices.

The remaining parameters $\theta \equiv \left(\lambda, \chi, \eta, b\right)'$ are jointly calibrated. I target moments from the data and choose the set of parameters that minimizes the distance between the model-implied moments and the data.[11] To ensure that Assumption 3 is satisfied, I restrict the parameters space

[10] "The effective interest rate for businesses is a weighted-average borrowing rate for new lending to non-financial businesses, estimated as a function of bank and market interest rates. The weights are derived from business credit data. The business effective rate is a function of: short-term commercial paper and bankers acceptance rates, with terms of one and three months; the bank prime business lending rate, which is adjusted for movements in bank funding costs, so as to estimate the effective bank prime lending rate faced by new borrowers; and longer term borrowing rates, approximated using Merrill Lynch bond indices, which include both investment and non-investment grade companies (non-financial)."

[11] See Gourieroux, Monfort, and Renault (1993) for a theoretical treatment of the indirect inference method

Table 1.2: Assigned parameters.

Parameter	Symbol	Value	Target	Source
Discount factor	r	0.030	Real interest rate of 3%	Bank of Canada
Depreciation rate	δ	0.176	Depreciation rate of 17.6%	Statistics Canada
Capital share	α	0.227	Capital output ratio of 1.10	Statistics Canada

to Θ, which is defined by

$$\Theta = \mathbb{R}^4_+ \cap \Big\{ \theta \mid 0 < \lambda < \chi \Big\}. \tag{1.33}$$

I choose 4 moments (statistics to be precise): the job reallocation rate, the autocorrelation of labor productivity, the interdecile range of labor productivity, and the unemployment rate. While there is no one-to-one mapping between the parameters and the moments, I now provide an informal mapping. The meeting rate λ determines the pace at which firms meet new workers, and therefore directly affects the job reallocation rate. The rate of productivity resets χ directly affects the autocorrelation of labor productivity at the firm. The shape parameter η, which determines the variance and skewness of the distribution of TFP draws, will determine the interdecile range of labor productivity. Finally, the flow value of nonemployment b will determine the productivity threshold z_l, and therefore the firm exit rate χ_x. Since all unemployment inflows comes from firm exit, b will determine the unemployment rate. Table 1.3 summarizes these relationships.

Table 1.3: Informal mapping between calibrated parameters and moments.

Parameter	Symbol	Moment
Meeting rate	λ	Job reallocation rate
Rate of productivity resets	χ	Autocorrelation of labor productivity
Shape parameter of TFP distribution	η	Interdecile range of labor productivity
Flow value of nonemployment	b	Unemployment rate

I take the annual unemployment rate and job reallocation rate directly from Canadian public data and are average them over the 2000-2015 period (Tables 282-0087 and 527-0001). Job reallocation is defined as the average of the job creation rate and the job destruction rate.

The autocorrelation of labor productivity is obtained by estimating ρ in the following regression model using the Canadian microdata

$$\log LP_{i,t} = c + \mu_t + \gamma_j + \rho \log LP_{i,t-1} + \eta_{i,t}, \quad \mathbb{E}_{i,t}\eta_{i,t} = 0, \tag{1.34}$$

where $LP_{i,t}$ represent the labor productivity (value-added per worker) at firm i in period t and μ_t, γ_j are year and 3-digit NAICS industry fixed-effects. I obtain a coefficient of 0.816.

The interdecile range of labor productivity is also computed using the microdata. First, I remove year and industry (2-digit NAICS) fixed effects to firm-level log labor productivity. I then compute the difference the 90th and 10th percentiles of the (residualized) log labor productivity in the pooled 2000-2015 sample.

Denote the vector of statistics in the data by $\widehat{\Lambda}$ and the vector of the statistics in the model evaluated at θ by $\Lambda(\theta)$. For each value of $\theta \in \Theta$, I solve the model numerically (see Appendix A.1.9 for a description of the solution method) and compute the model-implied statistics $\Lambda(\theta)$. I then minimize the following quadratic form

$$\min_{\theta \in \bar{\Theta}} \left(\Lambda(\theta) - \widehat{\Lambda}\right)' \left(\Lambda(\theta) - \widehat{\Lambda}\right). \tag{1.35}$$

Since Θ is not compact, I implement the numerical optimization problem by using its closure $\bar{\Theta} \equiv \Theta \cup \left\{\theta \mid 0 \leq \lambda \leq \chi\right\}$. I then verify that the minimizer θ^* is in Θ. I obtain the parameter estimate $\left(\lambda, \chi, \eta, b\right)' = \left(0.2095, 0.2295, 2.9518, 0.7806\right)'$. Even though the parameter b is identified mainly through the unemployment rate, which arguably is not very informative about the value of nonemployment, its estimated value is fairly close to standard values in the literature. The average wage in the economy is 1.2186, which means that b is equal to 64% of the average wage which is close (but higher) than the widely used value of 0.4 in the labor search literature.[12]

Table 1.4 contains the targeted statistics in the model and in the data. One thing to notice

[12]Note that Hagedorn and Manovskii (2008) propose an calibration strategy that yields a value of $b = 0.955$.

is that, even though there are as many parameters as moments, the fit is not perfect. This can be explained by the fact that the equations mapping the structural parameters to the statistics are non-linear, so that an exact solution $\beta(\theta^*) = \widehat{\beta}$ need not exist. Nevertheless, the only statistic that the model does not match satisfactorily is the job reallocation, and the value in the model is 7.5% which is not too far from the 10.5% found in the data.[13]

Table 1.4: Targeted moments.

Statistic	Data	Model	BM
Job reallocation rate	0.105	0.075	0
Autocorrelation of labor productivity	0.816	0.807	1
Interdecile range of labor productivity	1.544	1.551	1.551
Unemployment rate	0.071	0.086	0.086

Notes. The "Model" column present the model-implied statistics while the "BM" column presents the model-implied statistics in a Burdett and Mortensen (1998) model.

To contrast the predictions of my model with those of the classic model of Burdett and Mortensen (1998) (BM), I calibrate a plain-vanilla BM model augmented with a capital as a factor of production. Since the model is well-known, I relegate the derivation to the Appendix (See Appendix A.1.8). For the calibration, I use the two free parameters of the model to exactly match the unemployment rate and interdecile range of labor productivity in the full model. From Table 1.4, we can see that the static nature of the *BM* model restricts the productivity process to be fully persistent (firm productivity does not change through time) and job reallocation to be zero (firms do not grow or shrink over time).

1.4.2 Non-targeted moments

I now assess the ability of the calibrated model to match the data along several non-targeted dimensions, paying particular attention to the facts documented in Section 1.2.

[13] In Appendix A.2.3, I provide a detailed decomposition of the job flows in the model and in the data (job creation, destruction; by entrants, continuers, exiters). The model generates too low job reallocation amongst continuing firms.

Consider the decomposition of the aggregate labor share LS described in Equation 1.2.

$$\underbrace{LS}_{\substack{\text{aggregate}\\ \text{labor share}}} = \underbrace{\bar{LS}}_{\substack{\text{average}\\ \text{labor share}}} + \underbrace{cov\big(LS_i\,,\,Y_i/Y\big)}_{\substack{\text{covariance between}\\ \text{labor share and output share}}}.$$

Table 1.5: Aggregate labor share decomposition (medvel versus data).

Component	Data	Model	BM
Aggregate labor share	0.653	0.655	0.538
Average (unweighted) labor share	0.884	0.970	0.542
Covariance between labor share and output share	-0.231	-0.315	-0.004

Notes. The "Data" columns is taken from Table 1.1; the "Model" column present the model-implied statistics using the calibrated model; the "BM" column presents the model-implied statistics in the calibrated Burdett and Mortensen (1998) model.

I now reproduce Table 1.1 and include model-implied moments (see Table 1.5). The model generates an aggregate labor share of 0.66, which is nearly identical to the one in the data. The average (unweighted) firm-level labor share in the model is 0.97, which is close but higher than in the data (0.88). The resulting covariance between labor share and output share is negative (fact #3), and stands at -0.32 in the model (-0.23 in the data). Overall, the fit is very good, which is surprising given that the calibration procedure did not use data on the labor share. In contrasts, the BM model generates a counterfactually low aggregate labor share (0.54) and a near zero correlation between firm-level labor share and output share.

Both in the model and in the data, the average firm has a high labor share, yet the most important firms—those with high output shares—tend to have a lower labor share. I now provide a non parametric decomposition of the aggregate labor share. Denote by LS_d and ω_d the labor share and output share within labor productivity decile d. The aggregate labor share can be expressed as

$$LS = \sum_{d=1}^{10} \omega_d LS_d. \tag{1.36}$$

Figures 1.7 and 1.8 and plots the labor share and output shares within each labor productivity decile, both in the model and in the data.[14] To provide a benchmark, I include an horizontal line at $1-\alpha$ in Figure 1.7, which represents the labor share that would prevail in a frictionless, competitive model.

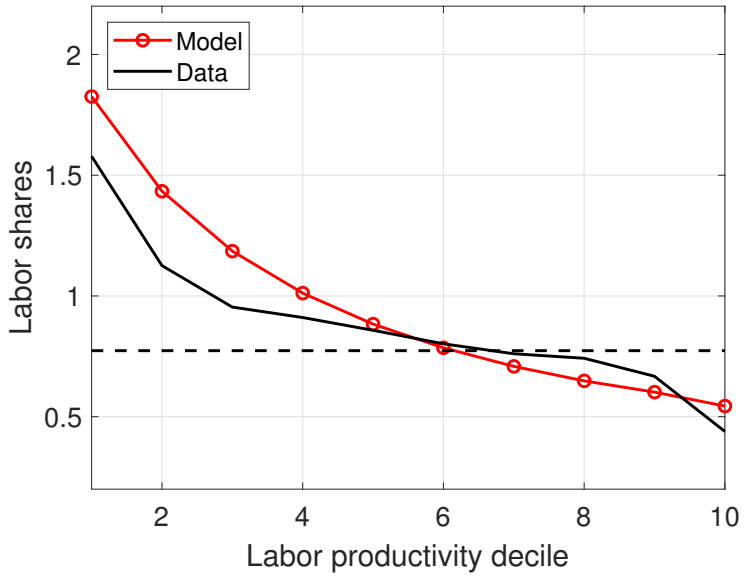

Figure 1.7: Labor shares.

First, notice that there are large differences in labor shares across firms (fact #1). In the model, the difference between the labor share at the 10th decile and the 1st decile is 1.28 (1.14 in the data). A large fraction of firms have a labor share above above one (fact #2). In the model, the bottom 4 deciles of firms have a labor share above one (bottom 2 deciles in the data) while the bottom 5 deciles have a labor share above $1-\alpha$ (bottom 5 deciles in the data). In contrast, the top decile of firms have a low labor share of roughly 54% (44% in the data).

The reason why the aggregate labor share is disproportionately determined by the low labor share of top firms can be understood from Figure 1.8. Both in the model and in the data,

[14]To construct the productivity deciles in the data, I first sort firm-year observations by labor productivity within industry-year bins (2-digit NAICS) and then group them into 10 deciles. I then compute the measure of interest (labor share, output share, etc.) within each decile and average over industries and years (see Appendix A.2.4 for details). The advantage of doing things in this manner is that (1) each productivity decile contains the same industry mix as the overall economy and (2) the aggregate labor share *LS* has an exact decomposition, both in the model and in the data.

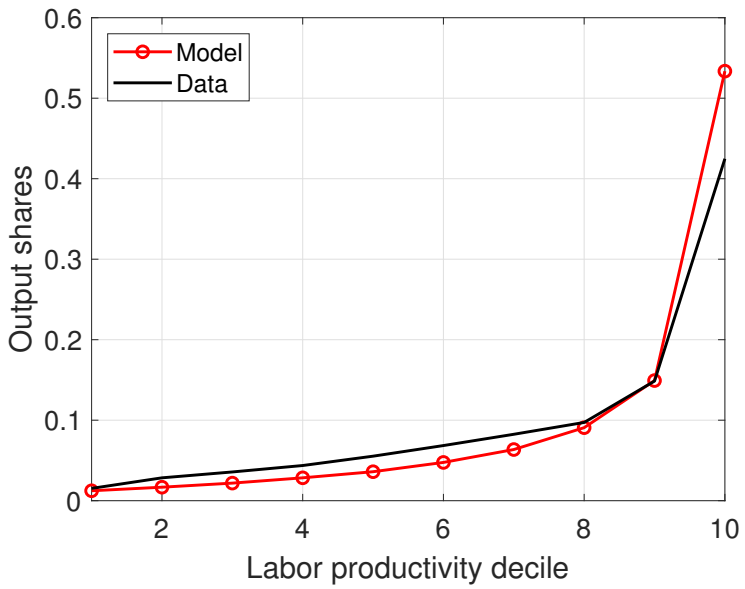

Figure 1.8: Output shares.

value-added is extremely concentrated within high-productivity firms. In the model, the top decile of firms account for 53% of of value-added (43% in the data), while the bottom 5 deciles accounts for a mere 12% (18% in the data). The negative covariance between labor share and output share documented earlier can thus explained by the fact that, in a stationary equilibrium, labor shares are decreasing in firm productivity while output shares are increasing in firm productivity. Firms in the top deciles achieve achieve high output shares through a combination of employing more workers (Proposition 6) and being more productive.

Table 1.5 highlights the fact that the *BM* model—which is essentially a static version of my model—generates predictions that are not in line with the data. Most importantly, the aggregate labor share is too low (0.54 versus 0.65 in the data) due to the fact that firm profits are too high. In both model, the share of value-added going to payments to capital is α, which implies that the share of firm profits is 11% in my model and 23% in the BM model. For comparison, Eggertsson, Robbins, and Wold (2018) and Barkai (2016) estimate a profit share of roughly 0.12 in the U.S. over the 2000-2015 period.

Figures 1.9 and 1.10 reproduce the previous analysis in the context of the *BM* model.

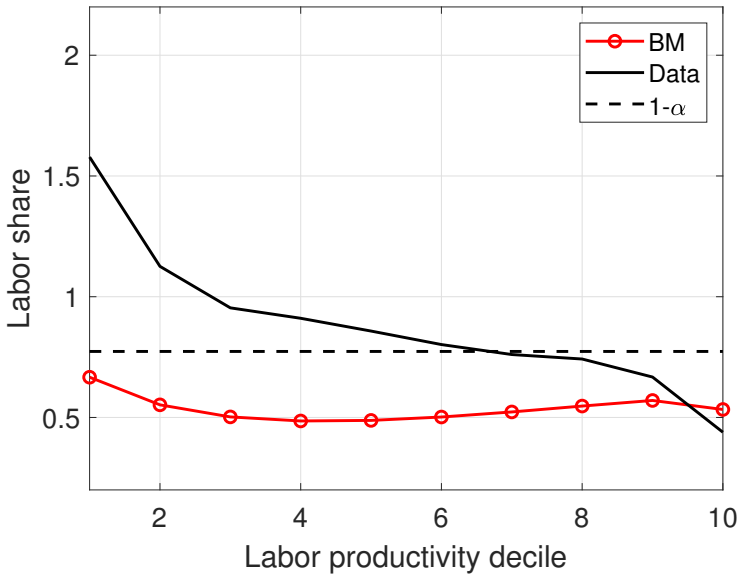

Figure 1.9: Labor shares

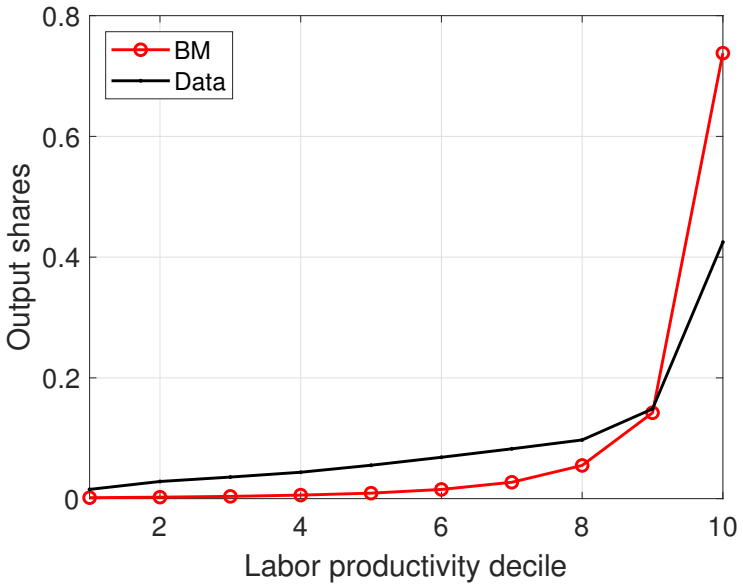

Figure 1.10: Output shares

Notice that firm-level labor shares are uniformly below $1-\alpha$. This is because firms that select into entry only if they make positive profits. In my model, flow profits can be either positive or negative. The reason is that firm productivity changes over time and some firms may choose to continue operating despite being temporarily unprofitable in order to maintain they scale. Hence, the *selection margin* accounts for the fact that the BM model predicts such a low labor share.

Another counterfactual prediction of the BM model is that value-added is too concentrated (see Figure 1.10). The top decile of firms account for 74% of aggregate value-added (43% in the model). This is because firm productivity does not change over time, so high productivity firms end up hiring most workers.

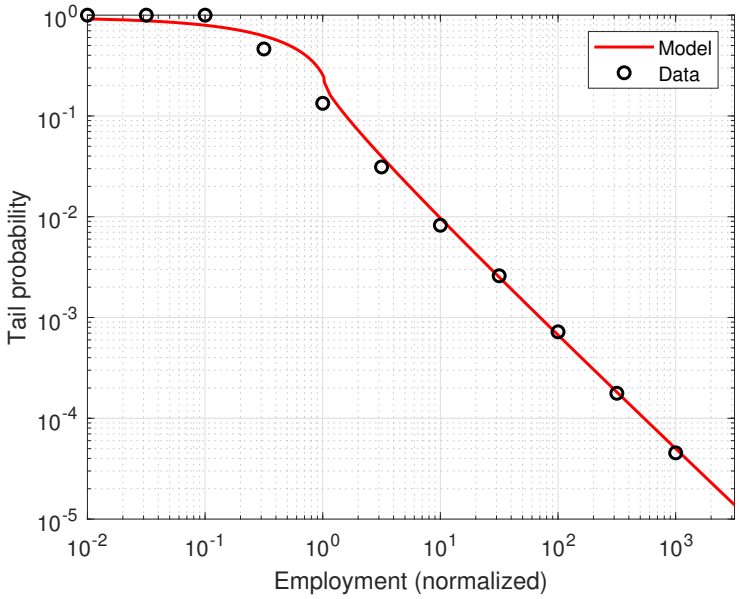

Figure 1.11: Complementary CDF of normalized firm size (firm employment over average firm employment)

In the data, it is well-known that the firm size distribution is well approximated by Zipf's law (i.e. a Pareto distribution with exponent slightly above one). Figure 1.11 contains the complementary CDF of normalized firm size (firm employment over average firm employment) in the model and in the data. Despite being a non targeted moment, the fit is exceptional. The reason is that the model generates a Pareto upper tail (Proposition 7) and in the calibrated model the Pareto exponent is 1.08 (see Appendix A.1.9 for details), which is close to the value of 1.06 estimated by Dixon and Rollin (2012) using Canadian data, and Axtell (2001a) using U.S. data. The fact that the model endogenously generate a Pareto exponent close to one is consistent with the findings in Toda (2016), who shows that Zipf's law robustly arises in models with a constant-returns-to-scale technology and a fixed supply of labor.

An advantage of working with a constant-returns-to-scale firm dynamics model is that it can easily replicate the extreme concentration of employment and value-added in the data without having to compromise on the fit of the productivity distribution. In fact, the firm-size distribution is invariant to the productivity distribution (see remark after Proposition 3), so the model would generate Zipf's law for any value of η. The model endogenously generates "superstar firms", who emerge by experiencing extended periods of fast growth. Since firm-level productivity is persistent but not permanent, the model provides a natural explanation for the rise and fall of large firms.

1.4.3 Model validation

In the model, the reason why high-output-share firms tend to have a lower labor share is entirely due to the fact that they tend to be more productive and pay wages significantly below labor productivity. I now test this prediction. First, I estimate the relationship between firm-level labor share and output in a regression framework. I estimate the following model

$$\log LS_{i,t} = \mu_t + \gamma_j + \beta_Y \log Y_{i,t} + \varepsilon_{i,t},$$

where μ_t and γ_j are year and industry (3-digit NAICS) fixed effects. The estimated coefficient $\widehat{\beta}_Y = -0.112$ is negative and highly statistically significant (column 1 of Table 1.6). Is the covariance between output share and labor share driven by the fact that the capital share $R\frac{K}{Y}$ is higher at larger firms? Recall that in the model, the capital share is equal to α at every firm, so cross-sectional differences in the capital-output ratio play no role in driving the relationship between firm-level output and the labor share. This needs not be true in the data, so I add the capital-output ratio as a separate regressor in the regression

$$\log LS_{i,t} = \mu_t + \gamma_j + \beta_Y \log Y_{i,t} + \beta_{K/Y} \log(K/Y)_{i,t} + \varepsilon_{i,t}.$$

Column 2 of Table 1.6 reports the estimated coefficients. The estimated relationship between labor share and output barely changes after controlling for K/Y (-0.113 versus -0.112). Hence, high-output firms tend to have a lower labor share, irrespective of their capital-output ratio. Consistently with economic theory, the coefficient on the capital-output ratio is negative ($\widehat{\beta}_{K/Y} = -0.018$), meaning that more capital-intensive firms tend to have a lower labor share. But notice that adding the capital-output ratio as a regressor barely improves the fraction of variance explained by the model; the R^2 increases from 0.120 to 0.125.

I now decompose the effect of output Y on the labor share. Notice that the logarithm of firm output is the sum of number of the logarithms of employment N and labor productivity LP. I now include (log) employment and labor productivity separately

$$\log LS_{i,t} = \mu_t + \gamma_j + \beta_{LP} \log LP_{i,t} + \beta_N \log N_{i,t} + \varepsilon_{i,t}.$$

Column 3 of Table 1.6 reports the results. In the model, the labor share is size-invariant conditional on firm productivity (Proposition 2). Consistently with this prediction, the estimated coefficient on labor productivity is large significant ($\widehat{\beta}_{LP} = -0.314$) but the coefficient on size is very small ($\widehat{\beta}_N = 0.009$). The interpretation is that high-productivity firms tend to have a lower labor share, but after controlling for labor productivity, size is mostly irrelevant and if anything predicts a *higher* labor share.

The market imperfection that allows firms in the model to earn profits is the presence of search frictions. It is well known that a wage-setting firm will exert monopsony power in the labor market when there are search frictions (Bontemps, Robin, and Berg (2000)). The key difference between competitive and monopsonistic labor market models relates to the elasticity of the labor supply curve faced by a firm. In a competitive environment, firms face a perfectly elastic labor supply curve, which means that they can hire any amount of labor at the going wage rate. In contrast, monopsonists face an upward-sloping labor supply curve, which means that they must

Table 1.6: Labor share and output at the firm-level.

Labor share ($\log LS_{i,t}$)	(1)	(2)	(3)
Value-added ($\log Y_{i,t}$)	-0.112***	-0.113***	
	(0.001)	(0.001)	
Capital-output ratio ($\log (K/Y)_{i,t}$)		-0.018***	
		(0.000)	
Size ($\log N_{i,t}$)			0.009***
			(0.001)
Labor productivity ($\log LP_{i,t}$)			-0.313***
			(0.002)
Industry fixed effects	✓	✓	✓
Year fixed effects	✓	✓	✓
Sample size	3,084,182	3,084,182	3,084,182
R^2	0.120	0.125	0.273

Notes. Standard errors in parentheses clustered at the firm-level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

increase wages in order to recruit additional workers. In my model, the relationship is captured by Equation 1.9, which implies that firm-level employment growth over a time period $[t, t+1)$ is related to the offered wage through

$$\log N_{t+1} - \log N_t = \int_t^{t+1} \widetilde{g}(w_s) ds,$$

where the employment growth function $\widetilde{g}(w)$ is increasing in w (Lemma 1). I now test this prediction in the data by estimating the following equation

$$\log N_{i,t+1} - \log N_{i,t} = \mu_t + \gamma_i + \theta \log w_{i,t} + \varepsilon_{i,t},$$

where $N_{i,t}$ and $w_{i,t}$ denote the employment and average wage of firm i in year t and the terms

μ_t and γ_i are year and firm fixed effects. I estimate a slope of $\widehat{\theta} = 0.213$ (see Table 1.7). In the calibrated model, the population regression coefficient is $\theta = 0.368$ which is larger but of the same order of magnitude.[15] The interpretation is that when a firm increases its wage by 10%, its annual growth rate of employment increase by 2.1 percentage points (3.7 in the model).

Table 1.7: Employment growth and wage at the firm-level.

Employment growth ($\log N_{i,t+1} - \log N_{i,t}$)	
Logarithm of average wage ($\log w_{i,t}$)	0.212***
	(0.001)
Firm fixed effects	✓
Year fixed effects	✓
Sample size	1,167,691
R^2	0.275

Notes. Standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). The sample, which contains annual data, is restricted to observations for which there is no missing values at time $t+1$ up to $t+5$.

To further validate the model, I estimate the "firm size wage premium"—in the tradition of Brown and Medoff (1989)—by estimating the following cross-sectional regression

$$\log w_{i,t} = \mu_t + \gamma_j + \beta_N \log N_{i,t} + \varepsilon_{i,t},$$

where $w_{i,t}$ and $N_{i,t}$ denote the average wage and employment at firm i in year t. The terms μ_t and γ_j are year and industry (3-digit NAICS) fixed effects. The point I want to make is that, despite the fact that the wage schedule is size-invariant, the model still generates a cross-sectional relationship between wage and size. I estimate a firm size wage premium $\widehat{\beta}$ of 0.030 (see Table

[15] I compute the model-implied regression coefficient as $\theta = \frac{cov(g(z), \log w(z))}{var(\log(w(z)))}$, where $z \sim \Gamma$.

1.8), which is close to the model-implied regression coefficient of 0.042.[16] The interpretation is that a 10% increase in a firm's size is associated with a 0.3% (0.42% in the model) increase in the average wage. The positive relationship between size and wage in the model is explained by the fact that high productivity firms tend to be larger (Proposition 6) and pay higher wages (Proposition 4).

Table 1.8: Firm size wage premium.

Logarithm of average wage ($\log w_t$)	
Logarithm of firm size ($\log N_t$)	0.030***
	(0.001)
Year fixed effects	✓
Industry fixed effects	✓
Sample size	3,084,182
R^2	0.390

Notes. Standard errors clustered at the firm level in parentheses (*** $p < 0.01$, ** $p < 0.054$, * $p < 0.1$)

1.5 Productivity dispersion and the labor share

1.5.1 Quantifying the effect of productivity dispersion

I now quantify the effect of an exogenous and permanent increase in productivity dispersion using the calibrated model. To do so, I feed in an exogenous increase in productivity dispersion of the same magnitude as the U.S. economy experienced over the 1977-2007 period. I use the measure of productivity dispersion constructed by Barth et al. (2016) (their Table 3). They use data from the Census Bureaus Economic Census files and compute the variance of log

[16] I compute the model-implied regression coefficient as $\beta_N = \frac{cov(\log N, \log w(z))}{var(\log N)}$, where $(z,N) \sim \varphi$.

revenue per worker in the cross-section of establishments.[17] The data is based on quinquennial censuses of establishments, so they report values for the years 1977, 1982, 1987, 1992, 1997, 2002, and 2007. They find an increase of 0.311 over the full sample, almost all of which occurs in the post 2000 period. While the measure does not account for industry differences, they provide measures for 8 broad industries, and the average increase with in industries is nearly identical an stands at 0.297.

To generate a "productivity dispersion shock" in the model, I decrease the parameter η, which determines the shape of the distribution of TFP draws Γ_0. Recall that the distribution was parametrized in a way to ensure that the mean is equal to one (Equation 1.32), so the change in η represents a *mean-preserving spread* (see Figure 1.15). I choose $\eta = 1.6176$, which implies a 0.3 increase in the cross-sectional variance of (log) revenue. I also increase the flow value of unemployment to $b = 0.8304$ in order to keep the firm exit rate constant before and after the shock. If we thinks of b as an unemployment benefit, then increasing b is consistent with unemployment benefits increasing with GDP.

The model experiment consists in comparing the stationary equilibrium of the model before and after the shock. Table 1.9 contains the response of several variables to the productivity dispersion shock. First, the aggregate labor share declines by 2.26 percentage point, from 0.655 to 0.632. Interestingly, the average (unweighted) labor share moves in the opposite direction, increasing from 0.970 to 1.265. The employment share of the top decile of firms does not change. This is not surprising in light of the corollary of Proposition 3, which states that the allocation of workers to firm productivity ranks is invariant to the underlying distribution Γ. Given that the productivity gap between the bottom and top decile of firms increases and the employment shares remain unchanged, the output share of the top decile of firms increases by 6.3 percentage points

[17] The authors caution that "The Economic Census expanded in scope over the 19772002 period, but the Business Register and LBD covered all industries throughout. As a check, we calculated the variance of revenues per worker restricted to industries where in each year total industry employment in the economic census is greater or equal to 90% of total industry employment in the LBD. The variance trend is very similar for 1977, the variance is 0.945; for 1982, it is 0.965; for 1987, it is 0.991; for 1992, it is 1.036; and for 1997, it is 1.111"

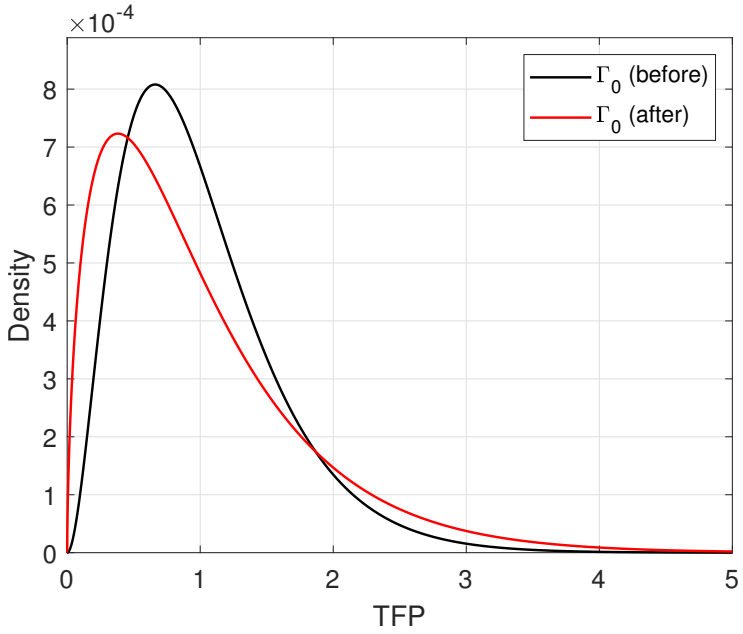

Figure 1.12: Distribution of productivity dispersion Γ_0 before and after the shock.

from 0.540 to 0.597.

Table 1.9: Model response to the productivity dispersion shock.

Variable	Before	After	Change (%)
Aggregate labor share	0.655	0.632	-2.26
Average (unweighted) labor share	0.970	1.265	$+29.59$
Output share of the top decile	0.540	0.597	$+3.10$
Employment share of the top decile	0.285	0.285	0

Notes. The employment/output shares of the top decile are defined as the share of employment/output of the top 10% of firms ranked by productivity.

To highlight the forces at play, I now study the impact of the productivity dispersion shock amongst different segments of firms. Figures 1.13 and 1.14 plots the response of labor shares and output shares in the model for all productivity deciles. The positive response of the average (unweighted) labor share can be explained by the fact that, for all but the top two deciles, firm labor shares increase. The *adjustment* of firm labor shares is more pronounced for low productivity firms, which leads to an increase in the dispersion of firm labor shares. Turning to

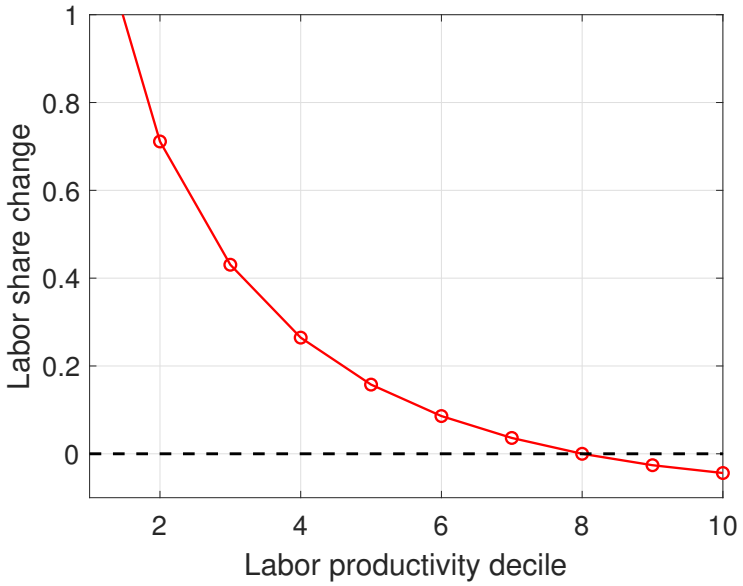

Figure 1.13: Adjustment of labor shares.

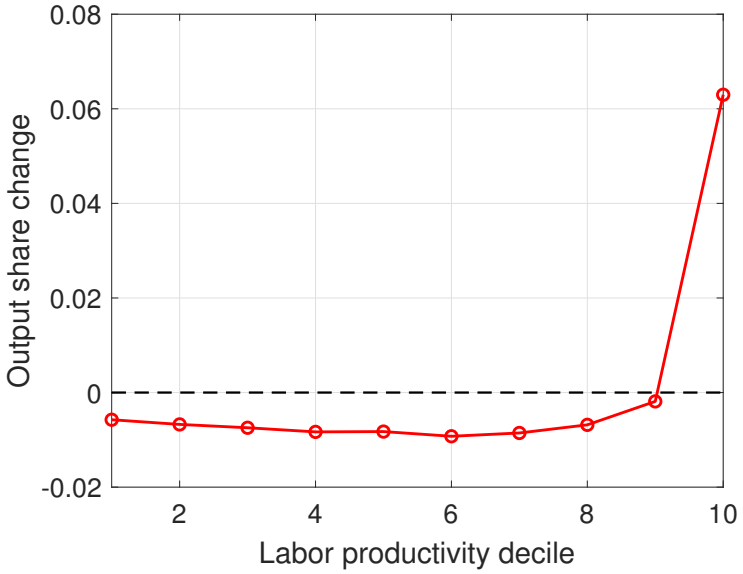

Figure 1.14: Reallocation of output shares.

Figure 1.14, we see that the bottom 9 deciles of firms lose output shares at the expense of the top decile. Having established that high productivity firms have lower labor shares, the *reallocation* of output shares towards low-labor-share firms mechanically decreases the aggregate labor share.

To quantify the relative importance of the "adjustment" and "reallocation" components of the decline of the aggregate labor share implied by the productivity dispersion shock, I use a

traditional shift-share decomposition as in Baily, Hulten, and Campbell (1992).[18] Denote by x_d the value of a variable within productivity decile d. Using the exact decomposition from Equation 1.36, the change in the aggregate labor share can be expressed as

$$\Delta LS = \underbrace{\sum \omega_d \Delta LS_d}_{\text{adjustment}} + \underbrace{\sum \Delta\omega_d LS_d}_{\text{reallocation}} + \underbrace{\sum \Delta\omega_d \Delta LS_d}_{\text{directed reallocation}}. \tag{1.37}$$

The adjustment measures the effect of changing the labor shares LS_d while keeping the labor shares ω_d constant. The reallocation component measures the effect of changing the output shares ω_d while keeping the labor shares LS_d constant. The third term, which I interpret as "directed reallocation", accounts for the comovements between *changes* in labor shares and output shares. For example, if there is reallocation of output shares towards firms with a *declining* labor share, then the directed reallocation term will be negative. Table 1.10 decomposes the model-implied aggregate labor share change into the contribution of three components Equation 1.37. The contribution of the "reallocation" and "directed reallocation" components are the largest at -1.15 and -0.26 percentage points respectively. The adjustment component works in the opposite direction and accounts for a positive contribution of 0.52 percentage points.

Table 1.10: Shift-share decomposition of the aggregate labor share response in the model.

Component	Contribution (%)
Adjustment	+2.91
Reallocation	−2.93
Directed reallocation	−2.25
Total	−2.26

Notes. The three components are defined in Equation 1.37.

The model predicts that, in response to a productivity dispersion of the same magnitude

[18] Baily, Hulten, and Campbell (1992) decompose the data at the establishment-level while I decompose it at the productivity decile level. Since the ranking of firms in terms of productivity remains unchanged after the productivity dispersion shock, these two approaches are equivalent.

as what has been estimated in the U.S. over 1977-2007, the aggregate labor share declines by 2.26 percentage points. To compare this number with the data, I first compute the U.S. corporate sector labor share as the ratio of "compensation of employees" to "gross value-added" using BEA Table 1.14. The labor share exhibits cyclicality (Nekarda and Ramey (2013)), so I also estimate a cubic trend over the full sample (1929-2017). Over 1977-2007, the trend labor share decline has been 3.62 percentage points. Hence, through the lens of the model, rising productivity dispersion explains $2.26/3.62 \approx 63\%$ of the labor share decline, making it a primary contributor.

Figure 1.5.1 contains the time series of productivity dispersion and the labor share. As we can see, in the first half of the sample (1977-1992), the increase in productivity dispersion—the variance of (log) revenue per worker—has been limited ($+0.07$) and the trend labor decline has been only 0.4 percentage point. Over the second half of the sample (1992-2007), the increase in productivity dispersion has been large ($+0.24$) and the labor share declined by 3.2 percentage points.

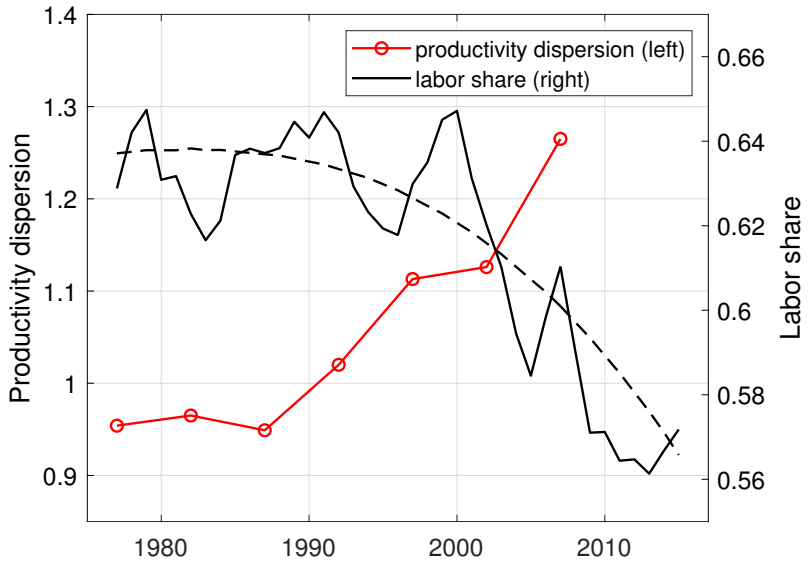

Figure 1.15: U.S. corporate sector labor share and productivity dispersion.

In addition to a decline of the aggregate labor share, the model generates micro-level predictions that can be confronted with the data. First, the decline of the labor share in the model is overwhelmingly driven by the reallocation of output shares towards firms with a low (and

declining) labor share, and is partially offset by an *increase* in the labor share of most firms (Table 1.10). This is precisely what Kehrig and Vincent (2017) document in the U.S. manufacturing sector. They show that over the 1967-2012 period, the aggregate labor share declined by 4.5 percentage point per decade, while the median labor share increased by 0.7 percentage point per decade. They reconcile this divergence by showing that low-labor-share firms have gained output shares over time. Autor et al. (2017) document a similar pattern within most industries.

Second, the model predicts that, in response to a productivity dispersion shock, the concentration of output increases but employment concentration remains constant (Table 1.9). Consistent with this prediction, Kehrig and Vincent (2017) find that there has been a massive reallocation of output towards low-labor-share firms, but that there was limited reallocation of employment. At the aggregate level, there has been a secular increase in sales concentration in the U.S. over the past two decades (Grullon, Larkin, and Michaely (2017)), which coincided with the decline of the labor share. The negative comovement between labor share and concentration is often interpreted as evidence in favor of rising monopoly power (Barkai (2016)), but my in my model the two series move in opposite direction as the result of an increase in productivity dispersion.

1.5.2 Empirical evidence

I now leverage cross-country and cross-industry data to test the prediction that there is a negative relationship between productivity dispersion and the labor share. The model predicts that the *level* of productivity dispersion should be negatively correlated with the *level* labor share. But it also predicts that *changes* in productivity dispersion should translate in *changes* in the labor share. I will use different data sources—which I will describe shortly—and estimate the

following regressions models.

$$LS_{j,t} = \mu_t + \beta_{\text{OLS}}\, PD_{j,t} + \varepsilon_{j,t} \tag{1.38}$$

$$LS_{j,t} = \mu_t + \gamma_j + \beta_{\text{FE}}\, PD_{j,t} + \varepsilon_{j,t} \tag{1.39}$$

$$\Delta^k LS_{j,t} = \mu_t + \gamma_j + \beta_{\text{LD}}\, \Delta^k PD_{j,t} + \varepsilon_{j,t+k}, \tag{1.40}$$

PD denotes the interdecile range of labor productivity across firms, *LS* is the labor share, μ_t are year fixed-effects, and γ_j is an industry and/or country fixed effects. The first specification (Equation 1.38), which I call "OLS", uses cross-sectional variation in the *levels* for identification. The second specification (Equation 1.39), from now on fixed-effects or "FE", uses variation in the *level* of *LS* and *PD* within a unit j over time for identification. For the last specification (Equation 1.40), from now on long differences or "LD", I use non-overlapping k-year periods to calculate the changes Δ, so the identification comes from the low frequency comovement of *LS* and *PD* for identification.

The OECD has developed a project called *MultiProd* that seeks to provide harmonized cross-country data on productivity and wage dispersion. The data they collect is obtained by running a standardized routine on individual country's production surveys and business registers. The variables that I will use (productivity dispersion) is taken directly from the replication material of Berlingieri et al. [2017]. As I did with the Canadian microdata, they measure productivity dispersion as the logarithm of the interdecile range of labor productivity in the cross-section of firms. The measures are computed at the NACE two-digit industry level and then averaged with employment weights to provide, for each country-year, a measure for the manufacturing sector and for non-financial services sectors. I merge the productivity dispersion data with publicly-available data on the labor share.[19] I obtain a balanced panel of 7 countries (Denmark,

[19]The data on the labor share by sector is obtained from EUKLEMS project (Denmark, Finland, France and Italy), WORLDKLEMS (Japan), and directly from national statistics institutes (Norway and New Zealand). For a description of the methodology underlying the EUKLEMS and WORLDKLEMS data, see respectively O'Mahony and Timmer (2009) and Jorgenson (2012).

Finland, France, Italy, Japan, Norway and New Zealand) and two industry groups (manufacturing and services) for the 2001-2011 period. One year of data was missing for France and Japan, so I imputed values using linear interpolation.

Figure 1.16 plots the cross-sectional relationship between productivity dispersion and the labor share at the beginning of the sample in 2001. There is a clear negative relationship, where country-industry observations with higher productivity dispersion tend to have a lower labor share. Figure 1.17 plots the same variables, but in 10 year differences from 2001 to 2011. Again, the relationship is negative. In Table 1.11, present the results in regression form for specifications 1.38 through 1.40. Given the small cross-section ($N = 14$), I conduct hypothesis ($\beta = 0$) by doing a permutation tests. Usual methods such as cluster-robust standard errors would not be valid. For specifications OLS and FE, the data is "block permuted" at the country-industry level to account for the autocorrelation of errors within a country-industry over time. In all specification, I obtain negative slope coefficients ranging from -0.077 (OLS) to -0.203 (10-year differences). The results are statistically significant at the 5% and 10% level for the FE and 10-year differences specifications respectively.

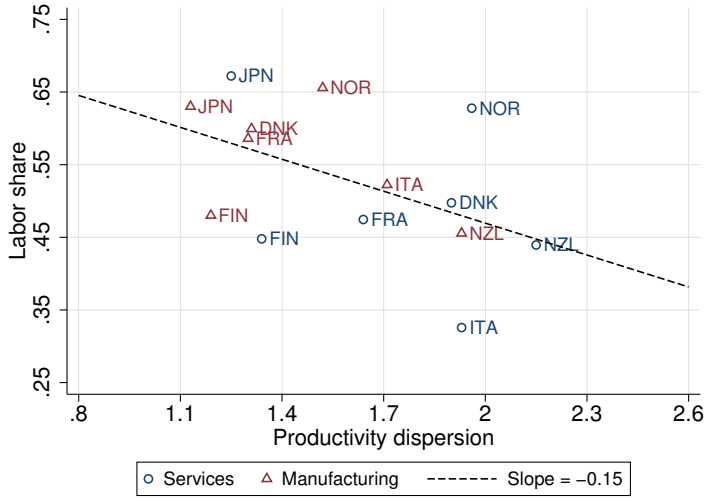

Figure 1.16: Relationship in level (2001)

I now aggregate the Canadian microdata to construct a panel dataset covering most 3-

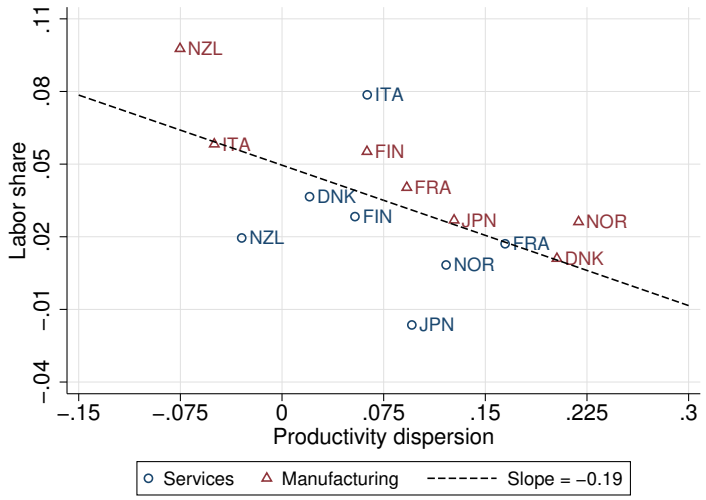

Figure 1.17: Relationship in changes (2001-2011)

Table 1.11: Productivity dispersion and the labor share: cross-country regressions.

	(1) OLS	(2) FE	(3) 5y diff.	(4) 10y diff.
Productivity dispersion	-0.077	-0.199*	-0.120	-0.203**
	($p = 0.360$)	($p = 0.096$)	($p = 0.130$)	($p = 0.029$)
Year FE	✓	✓		
Industry FE	✓	✓	✓	✓
Country FE		✓		
N	154	154	28	14
R^2	0.250	0.860	0.136	0.498

Notes. The p-values are obtained by permutation test (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). For specifications (1) and (2), the data is permuted at the country-industry. Each column contains an estimated coefficient from a separate regression of the labor share *LS* on productivity dispersion. Productivity dispersion is defined as the logarithm of the interdecile range of labor productivity in the cross-section of firms. The four specifications used are described in Equations 1.38, 1.39 and 1.40.

digit NAICS industries over the 2000-2015 period (see Appendix A.2.5 for details regarding the construction of the dataset) and repeat the previous exercise. Figures 1.18 and 1.19 plot

the relationship between productivity dispersion and the labor share for the year 2001 and the period 2001-2011 respectively. Again, there is a clear negative relationship, where industries with higher (increasing) productivity dispersion tend to have a lower (decreasing) labor share. Table 1.12 presents the regressions results. For the OLS and FE specification, I use standard errors clustered at the industry level to account for possible time-series dependence of the errors within an industry. In all specification, I obtain negative slope coefficients ranging from -0.063 (5-year differences) to -0.208 (10-year differences). The results are statistically significant at the 5% level for all specifications except for the 5-year differences. The results are remarkably similar to the cross-country exercise despite the fact that completely different datasets are used.

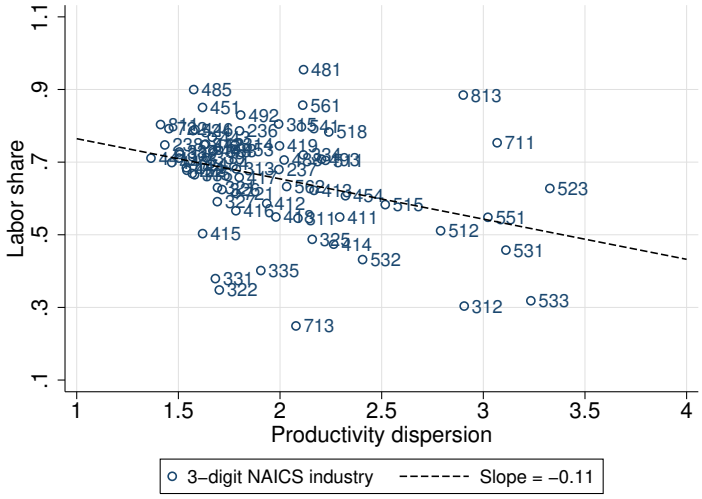

Figure 1.18: Relationship in level (2001)

So far, I have established that the labor share comoves negatively with the interdecile range (90/10 percentile ratio) of the firm productivity dispersion. I now use the cross-industry data to quantify the relative importance of right-tail dispersion (90/50 percentile ratio) and left-tail dispersion (50/10 percentile ratio). To do so, I estimate Equations 1.38, 1.39 and 1.40 but this time adding both left-tail and right-tail dispersion as two separate regressors (the results are reported in Table 1.13). A clear pattern emerges, where right-tail dispersion is associated with a lower labor share, but left-tail dispersion seems unrelated with the labor share. These findings are consistent

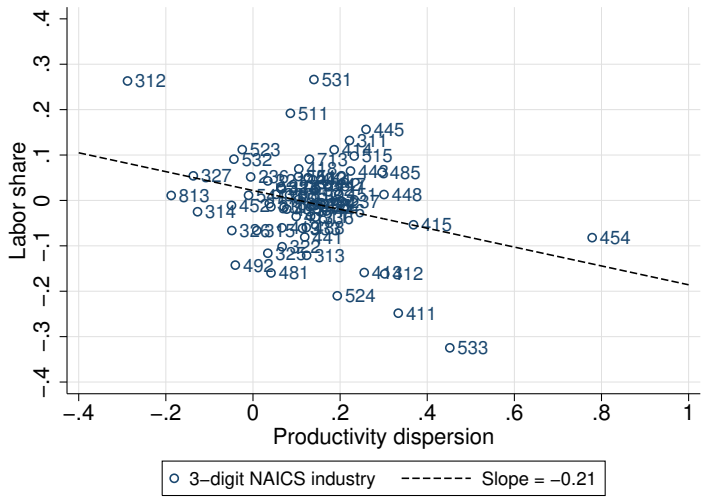

Figure 1.19: Relationship in changes (2001-2011)

Table 1.12: Productivity dispersion and the aggregate labor share: cross-industry regressions.

	(1) OLS	(2) FE	(3) 5y diff.	(4) 10y diff.
Productivity dispersion	-0.112***	-0.118**	-0.063*	-0.208**
	(0.029)	(0.047)	(0.034)	(0.079)
Year FE	✓	✓		
Industry FE		✓	✓	
N	1104	1104	207	69
R^2	0.154	0.736	0.228	0.092

Notes. Standard errors for specification (1) and (2) are clustered at the industry level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Each column contains an estimated coefficient from a separate regression of the labor share *LS* on productivity dispersion. Productivity dispersion is defined as the logarithm of the interdecile range of labor productivity in the cross-section of firms. The four specifications used are described in Equations 1.38, 1.39 and 1.40.

with my narrative, which says that it is the existence of extremely high-productivity firms that weakens labor market competition, not the existence of extremely low-productivity firms.

In the model, a rise in productivity dispersion leads to a decline of the aggregate labor share that operates through a reallocation of value-added towards high-productivity firms, as opposed

Table 1.13: Right/left-tail productivity dispersion and the aggregate labor share: cross-industry regressions.

	(1) OLS	(2) FE	(3) 5y diff.	(4) 10y diff.
Right-tail productivity dispersion	-0.445***	-0.290***	-0.203**	-0.316***
	(0.081)	(0.100)	(0.083)	(0.107)
Left-tail productivity dispersion	0.201**	-0.031	0.006	-0.103
	(0.087)	(0.078)	(0.051)	(0.105)
Year FE	✓	✓		
Industry FE		✓		
N	1104	1104	207	69
R^2	0.308	0.743	0.247	0.122

Notes. For specifications (1) and (2), the standard errors in parentheses are clustered at the industry level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Each column contains an estimated coefficient from a separate regression of the aggregate labor share *LS* on left-tail and right-tail productivity dispersion. Right-tail (left-tail) productivity dispersion is measured as the logarithm of the 90/50 (50/10) percentile ratio of labor productivity in the cross-section of firms. The four specifications used are described in Equations 1.38, 1.39 and 1.40.

to a broad-based decline of firm labor shares. In contrast, the model predicts an *increase* the average (unweighted) firm labor share (Table 1.9). I now test this prediction using the Canadian cross-industry data. First, table 1.14 presents the estimated relationship between productivity dispersion and the average firm labor share. In all specification, I obtain positive slope coefficients ranging from 0.044 (5-year differences) to 0.168 (FE). The results are statistically significant at the 5% level for all specifications except for the 5-year differences.

These results are consistent with the model experiment, in which the labor share of the bottom 80% of firms (in terms of productivity) increased while the labor share of the top 20% of firms decreased. In addition, the top 10% of firms gained output shares at the expense of the bottom 90% (see Figure 1.14). A more stringent test of the model mechanism is to estimate the response of labor shares and output shares along the productivity distribution. Specifically, for

Table 1.14: Productivity dispersion and the average labor share: cross-industry regressions.

	(1) OLS	(2) FE	(3) 5y diff.	(4) 10y diff.
Productivity dispersion	0.094***	0.168***	0.044*	0.140**
	(0.031)	(0.027)	(0.024)	(0.058)
Year FE	✓	✓		
Industry FE		✓		
N	1104	1104	207	69
R^2	0.177	0.826	0.136	0.077

Notes. For specifications (1) and (2), the standard errors in parentheses are clustered at the industry level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Each column contains an estimated coefficient from a separate regression of the average (unweighted) firm labor share $\bar{LS}$ on productivity dispersion. Productivity dispersion is measured as the logarithm of the 90/10 percentile ratio of labor productivity in the cross-section of firms. The four specifications used are described in Equations 1.38, 1.39 and 1.40.

each productivity quintile $q \in \{1,\dots,5\}$, I estimate the following fixed-effects regression

$$Y_{q,j,t} = \mu_t + \gamma_j + \beta_q PD_{j,t} + \varepsilon_{q,j,t}, \tag{1.41}$$

where μ_t is a year fixed-effect and γ_j is a 3-digit NAICS industry fixed effect. $Y_{q,j,t}$ represents the variable of interest (labor share or output share) in productivity quintile q, industry j, and year t. $PD_{j,t}$ represents the interdecile range of labor productivity within industry j in year t. I use clustered standard errors at the industry-level. The productivity quintiles are constructed by sorting firms within industry-years. Figures 1.20 and 1.21 plot the estimated coefficients with their 95% confidence intervals. As predicted by the model, the labor share of low-productivity firms increases with productivity dispersion while the opposite is true for high-productivity firms (Figure 1.20). The coefficients are imprecisely estimated, but for the two top quintiles they are negative and significant at the 5% level. Turning to the coefficients for output shares, we have

that the coefficient for the top quintile of firms is positive while it is negative or near-zero for all the other quintiles. Only the top quintile is significant at the 5% level. Taken as a whole, these findings are remarkably consistent with the transmission mechanism of productivity dispersion to aggregate labor share in the model (see Figure 1.13 and 1.14 for comparison)

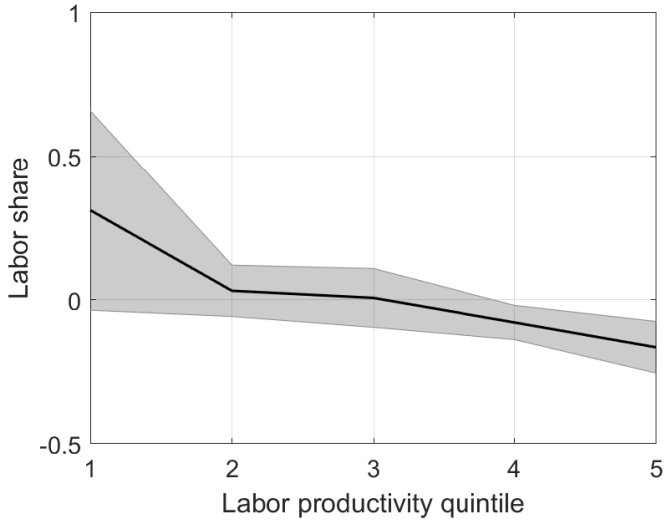

Figure 1.20: Labor shares.

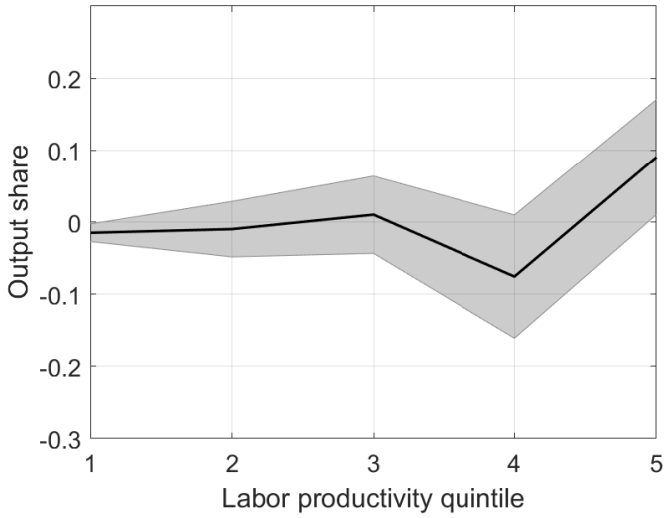

Figure 1.21: output shares.

1.6 Concluding Remarks

I present a new theory of the labor share that emphasizes the role of firm competition for workers. In my model, heterogeneous firms grow by accumulating workers and compete through wages in a frictional labor market. In equilibrium, there are firm-specific gaps between wage and productivity: low-productivity firms pay wages above productivity while high-productivity firms pay wages below productivity. The interaction between firm dynamics and labor market imperfections gives rise to an empirically relevant labor market "microstucture". In particular, value-added is concentrated within large firms who tend to be highly-productive and have a low labor share.

The main insight is that the *distribution* of firm productivity is a central determinant of the aggregate labor share. I find that productivity dispersion effectively weakens the intensity of wage competition in the labor market.

The model predicts that broad-based productivity growth translate into higher wages one-for-one. But when productivity gains are concentrated at the top of the firm productivity distribution, wages lag behind labor productivity and the labor share declines.

I find that rising productivity dispersion was in fact an important driving force behind the labor share decline in the US over the 1977-2007 period, explaining nearly $2/3$ of the decline. In my analysis, I take the rise in productivity dispersion as given. Little is known regarding the evolution of productivity dispersion over time and my paper falls short of explaining the trend increase. Future work should seek to understand the determinants of rising productivity dispersion.

1.7 Acknowledgements

Chapter 1, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien. "Productivity Dispersion, Between-Firm Competition, and the Labor

Share". The dissertation author was the sole investigator and author of this material.

Chapter 2

Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality

2.1 Introduction

Macroeconomic models increasingly incorporate heterogeneity. Doing so allows researchers to identify who gains and who loses from new policies, but also to assess how the effectiveness of policies depend on the nature of heterogeneity. Since the first generation of heterogeneous-agent models such as Huggett (1993) and Aiyagari (1994), one challenge has been to generate a realistic wealth distribution. Empirically, it is well known since Pareto (1895, 1896, 1897)'s seminal work that the wealth distribution obeys the power law: the fraction of agents with wealth w or larger decays like a power function $w^{-\zeta}$, where ζ is called the Pareto exponent. Hence, "micro-consistent" models of wealth inequality should endogenously generate fat-tailed wealth distributions.

We are now much closer to understanding the economic forces that determine wealth inequality, and many models have been proposed that can account for the extreme concentration

of wealth we observe in the data. Two parallel literatures have emerged. The first studies relatively simple models and provides theoretical characterizations of the power law behavior of the wealth distribution, often relying on analytical solutions. The second studies rich general equilibrium models and conducts quantitative analysis and experiments that rely heavily on numerical methods. However, there has always been the trade-off between analytical tractability and the richness of models. The former requires strong (and oftentimes unrealistic) assumptions. The latter relies on numerical methods, which are in general not well-suited for studying the tail behavior of the wealth distribution because models are commonly solved on a finite grid and hence misses the top tail *by definition*. What is lacking in the current literature is a systematic approach for analyzing and solving heterogeneous-agent models that (potentially) generate fat-tailed wealth distributions but do not admit closed-form solutions.

In this paper, we propose a simple, systematic approach for tackling heterogeneous-agent models with fat-tailed wealth distributions numerically. Our approach builds on the conventional solution algorithm but extends it with two additional steps: (i) the "asymptotic analysis" of the individual optimization problem to approximate the behavior of "wealthy" agents and compute the Pareto exponent of the wealth distribution and (ii) the "Pareto extrapolation" of the wealth distribution off the grid to compute the equilibrium and wealth distribution accurately. Our approach enables researchers to easily and accurately analyze rich heterogeneous-agent models that feature persistent earnings and investment risk, borrowing constraint, portfolio choices, recursive utility, and endogenous Pareto wealth distributions, etc.

In the "asymptotic analysis" step we solve a simplified, or "asymptotic" individual optimization problem semi-analytically. Roughly speaking, this problem ignores all additive elements and focuses on proportional elements. For example, consider the income fluctuation problem, which is a building block of models. The asymptotic problem in this case is one with no labor income (for wealthy agents labor income is negligible compared to capital income) which can be solved analytically as in Merton (1969) and Samuelson (1969). The benefit of studying

the asymptotic problem is threefold. First, its solution determines the behavior of wealthy agents, which governs the tail property of the wealth distribution. This enables researchers to determine whether the model generates a fat-tailed wealth distribution, and if so, to compute the theoretical Pareto exponent explicitly. Second, the analysis of the asymptotic problem places parametric restrictions on the equilibrium object through equilibrium considerations such as the existence of a solution to the individual optimization problem and a wealth distribution with a finite mean. This enables researchers to narrow down the set of equilibrium object and search for the equilibrium more efficiently. Third, the solution to the asymptotic problem can be used as an initial guess for solving the actual individual optimization problem, making the algorithm more efficient and stable.

In the "Pareto extrapolation" step, we extrapolate the wealth distribution off the grid using the theoretical Pareto exponent computed from the asymptotic analysis to correct for the truncation error when constructing the transition probability matrix governing the state variables and computing aggregate quantities. The benefit of Pareto extrapolation is twofold. First, it makes the solution more accurate at no additional computational cost. This is because we provide closed-form expressions for the correction terms, which depend only on the theoretical Pareto exponent and asymptotic policy functions computed in the "asymptotic analysis step". Second, and more importantly, Pareto extrapolation enables researchers to avoid making mistakes. While it is true that we can solve models to any accuracy if we use sufficiently large and fine grids and sufficiently strong computing power, with existing methods one can never be sure whether the truncation error is small enough. As an illustration, suppose some researcher says "I truncate the grid so that there is less than (say) 10^{-4} of the probability mass at the top grid point". This is not a good idea because (i) the mass at the largest grid point is severely biased downwards with existing methods, and (ii) even with mass 10^{-4} at the largest grid point, there can be substantial amount of wealth held by those agents.

In summary, our approach makes the solution and analysis of heterogeneous-agent models

with fat-tailed wealth distributions (i) more transparent (because it exploits economic theory to characterize the behavior and wealth distribution of wealthy agents), (ii) more efficient (because it narrows down the equilibrium object and uses good initial guesses), and (iii) more accurate (because it corrects for truncation errors). Furthermore, we achieve all of that with zero additional computational cost because the asymptotic analysis is semi-analytical and the correction terms are provided in closed-form. Hence, our method bridges theoretical and quantitative models of wealth inequality.

To illustrate the usefulness of our approach, we present two exercises. First, using a simple heterogeneous-agent model that admits a closed-form solution (and a Pareto wealth distribution) as a laboratory, we show that the error with existing methods can be substantial, while it is minimal with our approach. Specifically, we find that the error with our method (with an exponentially-spaced grid with 100 points) ranges from 0.13 to 0.42% depending on the choice of the largest grid point, whereas it is 2.4–26% with the usual ("Truncation") method that does not introduce correction terms. Simulation-based methods with 10,000 agents also have about 11% of error. The errors in the existing methods are especially severe when the Pareto exponent is smaller than 2, which is typical for wealth (1.5) and firm size (close to 1, Zipf's law).

Second, we develop a "Merton-Bewley-Aiyagari" model that features persistent earnings and investment risk, borrowing constraint, portfolio decision, and endogenous Pareto-tailed wealth distribution. We provide a step-by-step implementation of the Pareto extrapolation algorithm for solving the calibrated model and describe the rich general-equilibrium effects that determine the wealth shares and Pareto exponent. Using the model as a laboratory, we conduct a counterfactual experiment. We find that a wealth tax is a "lose-lose" policy: the introduction of a 1% wealth tax (with extra tax revenue used as consumption rebate) decreases wage and output by 6.5%, welfare (in consumption equivalent) by 7.7%, and total tax revenue by 0.72%. The welfare loss is uneven across the wealth distribution, the poorest households losing 7% and the richest 3%.

2.1.1 Related literature

Our paper is related to a large literature that spans across many disciplines, including quantitative macroeconomics, economic theory on consumption and portfolio choices, mathematical and statistical results on Pareto tails, and numerical analysis.

It is well known in the quantitative macroeconomics literature that idiosyncratic unemployment risk and incomplete financial markets alone are insufficient to generate a sufficiently dispersed wealth distribution (Krueger, Mitman, and Perri 2016). Recently, Stachurski and Toda (2019) have theoretically proved that in canonical models in which agents are infinitely-lived, have constant discount factors, and can invest only in a risk-free asset, the wealth distribution necessarily inherits the tail property of the income distribution. Therefore canonical heterogeneous-agent models cannot explain the wealth distribution. They also argue that introducing other ingredients such as random discount factors (Krusell and Smith 1998), idiosyncratic investment risk (Quadrini 2000; Cagetti and De Nardi 2006), and random birth/death (Carroll, Slacalek, Tokuoka, and White 2017; McKay 2017) can generate fat tails. However, because these papers are all numerical, it is not clear how to build and solve general heterogeneous-agent models that feature fat-tailed wealth distributions. Our paper contributes to the quantitative macroeconomics literature by showing the usefulness of the theoretical analysis of the asymptotic problem and providing a general solution algorithm for such models.

As mentioned in the introduction, since numerical methods are in general not well-suited for studying the tail behavior of the wealth distribution, most papers that study the power law behavior in the wealth distribution use analytical solutions. Nirei and Souma (2007) and Benhabib, Bisin, and Zhu (2011) solve growth models with idiosyncratic investment risk and use the properties of Kesten (1973) processes to obtain a Pareto wealth distribution. Moll (2014), Toda (2014), Arkolakis (2016), Benhabib, Bisin, and Zhu (2016), and Nirei and Aoki (2016) consider stochastic birth/death and obtain the double Pareto wealth distribution based on the mechanism

of Reed (2001).[1] For reviews of generative mechanisms of Pareto tails used in these papers, see Gabaix (2009). Our paper bridges this literature on power law in economics and quantitative macroeconomics by showing that the theoretical insight carries over to rich quantitative models.

Toda (2019) pointed out the usefulness of the asymptotic problem for computing the Pareto exponent in general models that admit no closed-form solutions.[2] However, he does not consider the solution algorithm for general equilibrium models with fat-tailed wealth distributions. The asymptotic linearity of consumption policies has been known for a long time since at least Huggett (1993) and Krusell and Smith (1998), among others. Benhabib, Bisin, and Zhu (2015) show the asymptotic linearity when earnings and investment risk are mutually independent and jointly i.i.d. over time and obtain a Pareto lower bound for the wealth distribution. In Appendix B.1, we argue that similar results should hold in richer models. To analytically characterize the Pareto exponent of the wealth distribution in a general Markovian environment, we apply the recent results from Beare and Toda (2017b).

Our paper is also related to the literature on solution methods for heterogeneous-agent models such as Krusell and Smith (2006), Algan, Allais, and Den Haan (2008), Reiter (2009, 2010), Den Haan (2010b, 2010a), and Algan et al. (2014), among others. In particular, we use the insight from Algan, Allais, and Den Haan (2008) and Winberry (2018), who approximate cross-sectional distributions using finite-dimensional parametric families. In our case, because economic theory suggests that the upper tail of the wealth distribution is Pareto and it is possible to compute the Pareto exponent from the solution to the asymptotic problem, we use this Pareto distribution to approximate the upper tail. Although we do not take a stance on how to deal with the rest of the distribution, we use Young (2010)'s non-stochastic simulation to compute the wealth distribution from the transition probability matrix implied by the law of motion.

[1] Other recent applications include firm dynamics (Acemoglu and Cao 2015), asset pricing (Toda and Walsh 2015; Toda and Walsh 2017), dynamics of inequality (Gabaix, Lasry, Lions, and Moll 2016; Aoki and Nirei 2017; Cao and Luo 2017; Kasa and Lei 2018), entrepreneurship (Jones and Kim 2018), and bequests (Zhu 2018).

[2] The asymptotic problem is related to the "method of moderation" in Carroll, Tokuoka, and Wu (2012), who bound the consumption policy functions from above and below by closed-form solutions to improve accuracy and stability.

The closest paper to ours in spirit is Achdou et al. (2017). They recast the model in continuous-time, which allows them to obtain a number of novel characterizations and results, including closed-form expressions for the stationary wealth distribution (in a special case) and the marginal propensity to consume of agents close to the borrowing constraint. They apply finite-difference methods and propose a fast solution algorithm that can be applied to general heterogeneous-agent models in continuous time. While our paper is different—we focus on the complications arising from fat-tailed wealth distributions—we share the same goal of bridging the gap between theoretical and quantitative work in macroeconomics.

2.2 Issues with existing algorithms

Suppose that we want to solve a Bewley (1977, 1983)-Huggett (1993)-Aiyagari (1994) model numerically when the wealth distribution could be fat-tailed. The conventional solution algorithm for heterogeneous-agent models (which we refer to as the "Truncation" method throughout the paper) combines dynamic programming over a finite grid (Blackwell 1965; Coleman 1990) with non-stochastic simulation (Young 2010). The algorithm would be roughly as follows.

1. The researcher sets up a finite grid for wealth denoted by $\mathcal{W}_N = \{w_n\}_{n=1}^N$, where N is the number of grid points and $w_1 < \cdots < w_N$. Suppose there are also other exogenous state variables (e.g., income, return on wealth, etc.), which can take S possible values indexed by $s = 1, \ldots, S$. Given the guess of the equilibrium object (e.g., interest rate, wage, etc.), we can solve the individual optimization problem on the $S \times N$ grid using dynamic programming.

2. Having solved the individual optimization problem and obtained the law of motion for individual wealth, the researcher constructs the $SN \times SN$ joint transition probability matrix Q of exogenous state and wealth. The stationary distribution $\pi \in_+^{SN}$ is obtained by solving $Q'\pi = \pi$ (so π is an eigenvector of Q' corresponding to the eigenvalue 1).

3. Finally, the researcher imposes the market clearing condition by integrating the individual decision rules (capital, labor, etc.) over the grid using the stationary distribution π to find the equilibrium objects (interest rate, wage, etc.).

There are two potential issues with this truncation method when the stationary wealth distribution is fat-tailed, both of which are related. First, consider the largest grid point w_N. This grid point in principle does not represent just the point $w = w_N$, but the entire interval $w \in [w_N, \infty)$. Therefore when we construct the transition probability from w_N to other grid points, instead of assuming that the current wealth state w is concentrated at w_N, we need to take into account that w is really distributed over the interval $[w_N, \infty)$ according to the (true) stationary distribution. Since the interval $[w_N, \infty)$ contains substantial probability mass when the wealth distribution is fat-tailed, failing to account for this will overestimate the transition probability to lower wealth states, and hence underestimate the top tail probability.

Second, suppose that we use the stationary distribution $\pi = (\pi_{sn})$ to compute aggregate quantities used in market clearing conditions. For concreteness, consider the aggregate wealth

$$W = \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} w_n. \tag{2.1}$$

The right-hand side of (2.1) essentially supposes that the top tail is concentrated on the grid point w_N, whereas in fact it is distributed over the interval $[w_N, \infty)$. Thus failing to account for this will underestimate the aggregate wealth, which affects the computation of equilibrium through market clearing conditions.

Of course, one may choose a very large truncation point w_N (say, one million times the aggregate wealth) to reduce the truncation error, but that is computationally inefficient because it will either increase the number of grid points (making the solution algorithm slower) or decrease the grid density (making the solution less accurate). One may also argue that the above two issues are specific to the particular algorithm that involves truncation, and other methods such as

simulation (Aiyagari 1994; Krusell and Smith 1998) may not be subject to those issues. As we see below, however, the situation is equally problematic. Simulation-based methods essentially use the law of large numbers to evaluate the market clearing condition. Suppose we simulate I agents and compute the sample mean of wealth $\frac{1}{I}\sum_{i=1}^{I} w_i$. The question is how fast the sample mean converges to the population mean. If the Pareto exponent ζ exceeds 2, then wealth has finite variance and we can apply the Central Limit Theorem. In this case the sample mean converges at rate $I^{1/2}$. If $\zeta < 2$ on the other hand, it is well known that the rate of convergence to the stable law is only $I^{1-1/\zeta}$.[3] Therefore solving a model accurately may require an impractically large number of agents.

As an illustration, Table 2.1 shows the order of error $I^{\max\{-1/2,1/\zeta-1\}}$ in the sample mean for various sample size I and Pareto exponent ζ.[4] If $\zeta \geq 2$ and we use 10,000 agents (the number used in Aiyagari (1994)), then the order of the error in the sample mean is $10000^{-1/2} = 1/100 = 1\%$. However, the error order is much larger if the Pareto exponent is smaller. With $\zeta = 1.5$ (a typical number for the wealth distribution according to Pareto (1897), Klass et al. (2006), and Vermeulen (2018)), the error order with 10,000 agents is 4.6%, which is substantial. If the Pareto exponent is 1.1 (a typical number for the firm size distribution, which obeys Zipf's law (Axtell 2001b)), then even with ten billion agents ($I = 10^{10}$), which is about the same order of magnitude as the world population, the error order is still 12.3%. To drive the error down to 1%, quite a modest number, the required sample size for $\zeta = 1.1$ is $I = 100^{\frac{\zeta}{\zeta-1}} = 10^{22}$ (ten sextillion), which is about the same order of magnitude as the number of stars in the universe or sand grains on earth.[5] Therefore we cannot expect to solve such models accurately using simulation.

[3] See, for example, Durrett (2010, Theorem 3.7.2) for an accessible proof. Based on this insight, Gabaix (2011) argues that a substantial fraction of aggregate fluctuations is due to idiosyncratic shocks to large firms.

[4] In Table B.3 in Appendix B.3.3, we assess the accuracy of the Aiyagari model in Section 2.4 using simulation and obtain similar results to Table 2.1.

[5] http://www.abc.net.au/science/articles/2015/08/19/4293562.htm

Table 2.1: Order of error $I^{\max\{-1/2,1/\zeta-1\}}$ in sample mean.

Sample size I	Pareto exponent ζ			
	≥ 2	1.5	1.3	1.1
$10^0 = 1$	1.00000	1.00000	1.00000	1.00000
10^2	0.10000	0.21544	0.34551	0.65793
10^4	0.01000	0.04642	0.11938	0.43288
10^6	0.00100	0.01000	0.04125	0.28480
10^8	0.00010	0.00215	0.01425	0.18738
10^{10}	0.00001	0.00046	0.00492	0.12328

2.3 The Pareto extrapolation algorithm

Our new solution algorithm, which we call the "Pareto extrapolation" method, builds on the conventional solution algorithm described in Section 2.2 but differs at several steps, most importantly when computing the stationary distribution and when aggregating individual behavior to evaluate the market clearing condition. It can be applied to solve for the stationary equilibrium of any heterogeneous-agent model, although the novel steps are needed only when the model generates a Pareto-tailed wealth distribution. In general, the stationary wealth distribution has a Pareto upper tail in models that combine homothetic preferences (e.g., additive CRRA, Epstein-Zin, etc.) with either random discount factors, stochastic returns on wealth, and/or birth and death. A byproduct of our algorithm is that it will tell whether the model generates a Pareto-tailed wealth distribution, and if so, provides an analytical characterization of the Pareto exponent.

We now provide a step-by-step description of the Pareto extrapolation method. As a leading example, we focus on a simplified version of the Krusell and Smith 1998 (henceforth KS) model without aggregate shocks, which is known to generate a Pareto-tailed wealth distribution as recently shown by Toda 2019. Since the KS model is well known, we only briefly describe the

Bellman equation for the value function of the household:

$$v_s(w) = \max_{c \geq 0} \{ \frac{c^{1-\gamma}}{1-\gamma} + \beta_s (1-p) E[v_{s'}(w')|s] \}, \tag{2.2a}$$

$$w' = R(w-c) + y_{s'}, \tag{2.2b}$$

$$w' \geq w. \tag{2.2c}$$

Here $s = 1, \ldots, S$ denotes Markov states that evolve over time according to a transition probability matrix $P = (p_{ss'})$, $\beta_s > 0$ is the discount factor in state s, $p \in [0,1)$ is the birth/death probability (the infinitely-lived case corresponds to $p = 0$), $\gamma > 0$ is the relative risk aversion, c is consumption, w is the beginning of period wealth including current labor income, y_s is income in state s, R is the gross interest rate, and w is the minimum wealth (borrowing limit). The Pareto extrapolation algorithm can be described as follows.

1. Asymptotic analysis
    (a) Compute solution to the "asymptotic" policy/value functions semi-analytically
    (b) Compute the theoretical Pareto exponent ζ
2. Dynamic programming
    (a) Construct a grid
    (b) Initialize value/policy functions using a guess
    (c) Update the value/policy functions using the Bellman/Euler equations
3. Stationary distribution
    (a) Construct a (potentially different) grid
    (b) Construct the joint transition probability matrix for exogenous state and wealth using non-stochastic simulation

(c) To account for transitions in and out of the grid, approximate (1) the behavior of agents outside the grid using the asymptotic policy functions and (2) the wealth distribution outside the grid using the Pareto exponent ζ

4. Aggregation

   (a) Aggregate individual behavior of agents "on the grid"

   (b) To account for the contribution of agents outside the grid, approximate (1) the behavior of agents using the asymptotic policy functions and (2) the wealth distribution using a Pareto distribution with exponent ζ

Below, we explain each step in more detail and pay particular attention to the "new" steps, namely (1), (3c), and (4b).

2.3.1 Asymptotic analysis

The first step of the algorithm consists of characterizing the "asymptotic" properties of the model. In a nutshell, we use two (related) features of models with homothetic preferences. First, the control variables (consumption, investment, etc.) are approximately linear in wealth for wealthy agents. Second, the endogenously-determined wealth distribution has a Pareto upper tail.

Given that labor income enters additively in the budget constraint, whereas capital income is proportional to wealth, the former becomes negligible as the wealth of an agent tends to infinity. To characterize the behavior wealthy agents, we can consider a simplified problem where labor income is set to zero. Assuming that agents have homothetic preferences (e.g., additive CRRA, Epstein-Zin, etc.), which is almost always the case in applications, this simplified problem becomes a homogeneous problem in the sense that all control variables scale with wealth. We refer to this problem as the "asymptotic" problem. Such problems can be solved semi-analytically even in a Markovian (non-i.i.d.) environment as shown by Toda (2014, Theorem 5), and the decision rules become linear in wealth. (Appendix B.1 formally defines the asymptotic problem

and discusses the asymptotic linearity of policy functions in an abstract dynamic programming setting.)

For concreteness, consider the KS model. In this case, income y_s and the borrowing limit w are negligible asymptotically, so we replace the budget constraint (2.2b) and the borrowing constraint (2.2c) by

$$w' = R(w - c), \tag{2.3a}$$

$$w' \geq 0, \tag{2.3b}$$

respectively. Note that the problem is now homogeneous because the utility function is homothetic: an agent twice as rich will consume twice as much, state-by-state. We can maximize a homothetic function subject to homogeneous constraints of the form (2.3) semi-analytically quite efficiently, as explained in Toda (2014) in detail. In the case of the KS model, the asymptotic consumption rule can be compute as follows. First, conjecture a solution of the form

$$c_s(w) = \bar{c}_s w, \tag{2.4}$$

where $\{\bar{c}_s\}_{s=1}^S$ can be interpreted as the asymptotic marginal propensity to consume (MPC) out of wealth in each patience state. Then, substitute the conjectured consumption rule into the Euler equation to obtain

$$\bar{c}_s^{-\gamma} = (1-p)R^{1-\gamma}\beta_s \sum_{s'=1}^{S} p_{ss'} \left[(1-\bar{c}_s)\bar{c}_{s'}\right]^{-\gamma}. \tag{2.5}$$

The asymptotic consumption rules in the KS model can be solved semi-analytically as the solution to the asymptotic Euler equation (2.5), which admits a (necessarily unique) solution if and only if

$$(1-p)R^{1-\gamma}\rho(BP) < 1, \tag{2.6}$$

where $B = (\beta_1, \ldots, \beta_S)$ is the diagonal matrix of discount factors and $\rho(A)$ denotes the spectral

radius (largest absolute value of all eigenvalues) of the matrix A.[6]

Equipped with the solution to the asymptotic problem, we can now compute the Pareto exponent of the wealth distribution. Substituting the asymptotic consumption rule (2.4) into the budget constraint (2.3a), notice that the law of motion for wealth in the KS model is asymptotically linear:

$$w' = R(1-\bar{c}_s)w. \tag{2.7}$$

The key insight here is that, since the patience state s evolves randomly over time, the wealth accumulation process of wealthy agents is a "random growth model".

More generally, suppose that the law of motion of the asymptotic problem is

$$w_{t+1} = G_{t+1}w_t,$$

where $G_{t+1} > 0$ is the gross growth rate of wealth between time t and $t+1$ and w_t is wealth. Thus, in the asymptotic problem, the law of motion for wealth necessarily satisfies Gibrat (1931)'s law of proportional growth. Assuming that agents enter/exit the economy at constant probability $p > 0$, Beare and Toda (2017b) show that under mild conditions the stationary wealth distribution has a Pareto upper tail and characterize the Pareto exponent ζ, as follows. For $z \in$, let

$$M_{ss'}(z) = E(e^{z\log G_{t+1}}|s_t = s, s_{t+1} = s') \tag{2.8}$$

be the moment generating function of the log growth rate $\log G_{t+1}$ conditional on transitioning from state s to s', and

$$M(z) = (M_{ss'}(z)) \tag{2.9}$$

[6]See Appendix B.2.1 for a proof. If needed, the asymptotic value functions can be obtained by conjecturing that $v_s(w) = \bar{v}_s \frac{w^{1-\gamma}}{1-\gamma}$ and using the Bellman equation (2.2a) combined with the asymptotic MPCs $\{\bar{c}_s\}_{s=1}^S$:

$$\bar{v}_s = \bar{c}_s^{1-\gamma} + (1-p)R^{1-\gamma}\beta_s \sum_{s'} p_{ss'}\bar{v}_{s'}(1-\bar{c}_s)^{1-\gamma}.$$

be the $S \times S$ matrix of conditional moment generating functions (2.8). Then under mild conditions Beare and Toda (2017b) show that the equation

$$(1-p)\rho(P \odot M(z)) = 1 \tag{2.10}$$

(here $P \odot M(z)$ denotes the Hadamard (entry-wise) product of P and $M(z)$) has a unique positive solution $z = \zeta > 0$, and that the stationary wealth distribution has a Pareto upper tail with exponent ζ. The following proposition gives a simple test for the solvability of (2.10).

Proposition 9. *If $G_{t+1} \leq 1$ always, then* (2.10) *does not have a solution $z > 0$. If $M(z)$ is finite for all $z > 0$, P is irreducible, and $p_{ss} \Pr(G_{t+1} > 1 | s_t = s_{t+1} = s) > 0$ for some s, then* (2.10) *has a unique solution $z = \zeta > 0$.*

For some parametrization, the model might generate a bounded wealth distribution. Intuitively, $G_{t+1} \leq 1$ means that wealth shrinks (or stay the same), so there is no random *growth*. In this case the wealth distribution does not have a Pareto tail, and the Pareto extrapolation algorithm reduces to the conventional one. If P is irreducible, every state is visited eventually. The condition $p_{ss} \Pr(G_{t+1} > 1 | s_t = s_{t+1} = s) > 0$ says that in state s, with positive probability the agent's wealth grows *and* the agent can remain in that state. Because there is random growth, the wealth distribution has a Pareto tail.

Toda (2019) argues that if agents are infinitely lived but there exists a stationary distribution due to other mechanisms than random entry/exit (e.g., borrowing constraint), then we can just set $p = 0$ in (2.10) to compute the theoretical Pareto exponent.

We can summarize this step as follows.

1. Define the asymptotic problem (e.g., no labor income, no borrowing constraint).
2. Solve the asymptotic problem and collect the asymptotic policy functions.
3. Derive the asymptotic law of motion for wealth $w' = G_{ss'} w$.

4. Define the matrix (2.9) of conditional moment generating functions.

5. Compute the theoretical Pareto exponent $\zeta > 0$ as the solution to (2.10).

2.3.2 Dynamic programming

Dynamic programming methods such as value function iteration (Blackwell 1965) and policy function iteration (Coleman 1990) consists of numerically solving the individual optimization problem over a finite grid for wealth, which we define $\mathcal{W}_N = \{w_n\}_{n=1}^N$. Many different algorithms exist, but they all share the same structure: the researcher starts with a guess for the policy/value functions and uses optimality conditions such as the Euler/Bellman equations to update those guesses until convergence. Given that dynamic programming methods are well known, we do not provide a step-by-step procedure. From now on, we denote by x_{sn} the numerical approximation of an arbitrary policy function $x_s(w)$ evaluated at $w = w_n$.

2.3.3 Stationary distribution

Equipped with the policy functions, their asymptotic counterpart, and the theoretical Pareto exponent $\zeta > 0$, the next step is to compute the stationary distribution of wealth. For that, we need to construct the joint transition probability matrix for all state variables. Let $w' = g_{ss'}(w)$ be the law of motion for wealth (conditional on transitioning from state s to s') implied by the policy functions obtained in the dynamic programming step and $\mathcal{W}_N = \{w_n\}_{n=1}^N$ be the grid for wealth. Let

$$I_n = [w_n, w_{n+1}), \quad n = 1, \ldots, N-1$$

be the half-open interval with endpoints w_n and w_{n+1} and $I_N = [w_N, \infty)$. For $n = 1, \ldots, N$, we construct the transition probability as follows. For simplicity, we focus on the case when there is no death ($p = 0$). The case $p > 0$ is similar and it is a matter of taking the weighted average of transition probabilities conditional on survival and death, weighting by $1 - p$ and p.

For the case where $n < N$, take the lower grid point of I_n, which is w_n. If $g_{ss'}(w_n) \in I_k$ for some $k < N$, then we can take $\theta \in [0,1)$ such that

$$g_{ss'}(w_n) = (1-\theta)w_k + \theta w_{k+1} \iff \theta = \theta_{nk} = \frac{g_{ss'}(w_n) - w_k}{w_{k+1} - w_k}. \tag{2.11}$$

We can then assign probabilities $1-\theta, \theta$ to the grid points w_k, w_{k+1} (, states k and $k+1$), respectively. If $g_{ss'}(w_n) < w_1$ or $g_{ss'}(w_n) \geq w_N$, then just assign probability 1 to state 1 or N. (Assigning probabilities to neighboring grid points to match the law of motion this way is essentially the same as what Young (2010) calls "non-stochastic simulation".)

More formally, let $Q = (q_{sn,s'n'})$ be the $SN \times SN$ joint transition probability matrix of exogenous state s and wealth $\{w_n\}_{n=1}^N$. Then for $n < N$ we define

$$q_{sn,s'n'} = p_{ss'} \times \begin{cases} 1, & \text{if } g_{ss'}(w_n) < w_1 \text{ and } n' = 1 \\ 1-\theta_{nk}, & \text{if } g_{ss'}(w_n) \in I_k \text{ and } n' = k \\ \theta_{nk}, & \text{if } g_{ss'}(w_n) \in I_k \text{ and } n' = k+1 \\ 1, & \text{if } g_{ss'}(w_n) \geq w_N \text{ and } n' = N \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

where θ_{nk} is defined by (2.11).

For the case where $n = N$, suppose for the moment that there is an untruncated grid $\mathcal{W}_\infty = \{w_n\}_{n=1}^\infty$, and for $n \geq N$ we know the probability of $w = w_n$ conditional on $w \in I_N \cap \mathcal{W}_\infty$. Let this probability be denoted by r_n. By definition, we have $\sum_{n=N}^\infty r_n = 1$. Now for each $n \geq N$, we can do precisely as in the previous case, and add probabilities $(1-\theta_{nk})r_n$ and $\theta_{nk}r_n$ (where θ_{nk} is defined by (2.11)) to the grid points w_k, w_{k+1} whenever $w' = g_{ss'}(w_n) \in I_k$ for $k < N$. If $g_{ss'}(w_n) < w_1$ or $g_{ss'}(w_n) \geq w_N$, then just add probability r_n to the transition to state 1 or N. The nice thing is that for large enough n, the next period's state $w' = g_{ss'}(w_n)$ is likely large (contained in I_N), so we only need to compute θ_{nk} for finitely many n (say $n = N, \ldots, N'$). For the probability

r_n, because the theoretical density is Pareto with exponent ζ, we can simply set $r_n \propto n^{-\zeta-1}$ if the grid spacing $w_{n+1} - w_n$ is constant for $n \geq N$.

More formally, we do as follows. First, choose the grid spacing $h > 0$ of the hypothetical grid points $\{w_n\}_{n=N+1}^{\infty}$. Define the untruncated grid $\mathcal{W}_\infty = \{w_n\}_{n=1}^{\infty}$ by $w_n = w_N + (n-N)h$ for $n > N$. Compute the number of relevant grid points $N' \geq N$ such that $g_{ss'}(w_{N'}) > w_N$ for all s, s':

$$N' = \min\{n \geq N | \forall s, s', g_{ss'}(w_N + (n-N)h) > w_N\}. \tag{2.13}$$

To evaluate the law of motion outside the grid, we can simply linearly extrapolate the law of motion for $w \geq w_N$ as

$$g_{ss'}(w) = g_{ss'}(w_N) + G_{ss'}(w - w_N), \tag{2.14}$$

where $G_{ss'} > 0$ is the theoretical slope obtained in the asymptotic analysis step. Combining (2.13) and (2.14), after some algebra we obtain

$$N' = N + \max_{s,s'} \max\{\frac{w_N - g_{ss'}(w_N)}{G_{ss'}h}, 0\}, \tag{2.15}$$

where x denotes the smallest integer exceeding x.

To compute the conditional probability r_n, because theoretically the stationary distribution has a Pareto upper tail with exponent $\zeta > 1$,[7] using the density (conditional on $w \geq w_N$) $f(x) = \zeta w_N^{\zeta} x^{-\zeta-1}$, we set

$$r_n \approx \zeta w_N^{\zeta} (w_N + (n-N)h)^{-\zeta-1} h = \zeta \frac{h}{w_N} \left(1 + (n-N)\frac{h}{w_N}\right)^{-\zeta-1}$$

for $n \geq N$. Since for $n \geq N'$ the next state will always be N ($w' = g_{ss'}(w_n) \in I_N$), there is no need

[7] If $\zeta \leq 1$, then the mean is infinite, which is impossible in equilibrium.

to compute r_n individually. Using the theoretical Pareto density, we obtain

$$\sum_{n=N'}^{\infty} r_n \approx \int_{w_{N'}}^{\infty} \zeta w_N^{\zeta} x^{-\zeta-1} x = (w_{N'}/w_N)^{-\zeta} = \left(1 + (N'-N)\frac{h}{w_N}\right)^{-\zeta}.$$

Imposing the condition $\sum_{n=N}^{\infty} r_n = 1$, we obtain

$$\begin{cases} r_n = C\zeta \frac{h}{w_N} \left(1 + (n-N)\frac{h}{w_N}\right)^{-\zeta-1}, & (N \le n < N') \\ \sum_{n=N'}^{\infty} r_n = C\left(1 + (N'-N)\frac{h}{w_N}\right)^{-\zeta}, \end{cases} \tag{2.16}$$

where the constant of proportionality C is given by

$$\frac{1}{C} = \left(1 + (N'-N)\frac{h}{w_N}\right)^{-\zeta} + \sum_{n=N}^{N'-1} \zeta \frac{h}{w_N} \left(1 + (n-N)\frac{h}{w_N}\right)^{-\zeta-1}. \tag{2.17}$$

Now for each s and extra grid point $n = N, \ldots, N'$, define the transition probability $\widetilde{q}_{sn,s'n'}$ exactly as in (2.12). The remaining elements of the joint transition probability matrix $Q = (q_{sn,s'n'})$ can be computed as

$$q_{sN,s'n'} = \sum_{n=N}^{N'} r_n \widetilde{q}_{sn,s'n'}, \tag{2.18}$$

where $r_{N'} := 1 - \sum_{n=1}^{N'-1} r_n = C\left(1 + (N'-N)\frac{h}{w_N}\right)^{-\zeta}$.

Equipped with the $SN \times SN$ joint transition probability matrix Q, we can compute the stationary distribution $\pi = (\pi_{sn})$, where π_{sn} is the probability of being in state (s,n), as the (unique) eigenvector of Q' corresponding to the eigenvalue 1.

We can summarize this step as follows. Let $\mathcal{W}_N = \{w_n\}_{n=1}^{N}$ be the grid, $\zeta > 1$ be the theoretical Pareto exponent from asymptotic analysis, and $\{g_{ss'}(w)\}_{s,s'=1}^{S}$ be the law of motion from dynamic programming.

1. Choose a spacing parameter $h > 0$ for hypothetical grid points.

2. Construct the $SN \times SN$ joint transition probability matrix Q of exogenous state and wealth.

For $n < N$, use (2.12). For $n = N$, use (2.13)–(2.18).

3. Compute the stationary distribution $\pi = (\pi_{sn})$ as the (unique) eigenvector of Q' corresponding to the eigenvalue 1.

A few remarks are in order. First, the algorithm for constructing Q has essentially zero additional computational cost, despite its complicated appearance. The reason is that extrapolation from the Pareto distribution is used *only* at the largest grid point w_N. Thus, although we are computing transition probabilities from SN points, which the conventional solution algorithm needs to compute anyway, the Pareto extrapolation algorithm requires only $S \times 1 = S$ additional operations, which is negligible. In our numerical implementation in Section 2.4, we find that the computing time of this step is trivial, and therefore we do not report it.

Second, the $SN \times SN$ transition probability matrix Q is sparse. To see this, let us evaluate the number of nonzero elements of Q. For each s, s' and $n < N$, there are at most two states the next wealth can take. For $n = N$, in principle the next wealth state can be anything. Therefore the number of nonzero elements of Q is at most

$$2S^2(N-1) + S^2N = S^2(3N-2).$$

Thus the fraction of nonzero elements of Q is bounded above by

$$\frac{S^2(3N-2)}{(SN)^2} = \frac{3N-2}{N^2} \to 0$$

as $N \to \infty$, so Q is sparse.[8] Therefore, computing the sationary distribution π is not an issue, despite the fact that Q is in practice a very large matrix.

[8] Achdou et al. (2017) mention that "[c]ontinuous time imparts a number of computational advantages relative to discrete time [..., which] relate to [...] the fact that continuous-time problems with discretized state space are, by construction, very sparse." While it is true that continuous-time problems have some advantages over discrete-time problems (e.g., partial differential equations versus nonlinear difference equations), discrete-time problems also do possess sparsity if appropriately solved.

Third, although we have implicitly assumed that N' in (2.15) is larger than N, for particular models it may be $N' \leq N$, which is true if and only if $g_{ss'}(w_N) \leq w_N$ for all s. In that case we do not need to consider any extrapolation since the true distribution is not fat-tailed, and the algorithm becomes identical to the truncation method.

Finally, constructing the transition probability matrix Q requires the spacing parameter $h > 0$. Since the Pareto extrapolation algorithm uses a hypothetical evenly-spaced grid (with grid spacing h) beyond the largest grid point w_N, the most natural choice for h is $w_N - w_{N-1}$, the distance between the two largest actual grid points. Conducting numerical experiments similar to those in Section 2.4 and Appendix B.3, we have found that this choice is numerically optimal.

2.3.4 Aggregation

When computing the equilibrium of a heterogeneous-agent model, we need to impose market clearing conditions in some way or another. To do so, we need to aggregate individual behavior. Let us first focus on computing the aggregate capital supply in the KS model. Given that the capital supply of an agent in state s with wealth w is $w - c_s(w)$, the aggregate capital supply is $K^s = \int (w - c_s(w))\Gamma(w,s)$, where $\Gamma(w,s)$ is the theoretical joint distribution of wealth and exogenous state. The truncation method approximates K^s with

$$K^s_{\text{trunc}} = \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn}(w_n - c_{sn}), \tag{2.19}$$

where c_{sn} and π_{sn} are consumption and unconditional (stationary) probability at state (s,n).

The only caveat is that for the wealth state N, the probability is not concentrated on the grid point w_N but in principle distributed over the entire interval $w \in [w_N, \infty)$. We can easily overcome this problem by (i) extrapolating the policy functions outside the grid using the asymptotic policy functions, and (ii) extrapolating the wealth distribution outside the grid using the theoretical Pareto exponent ζ.

First, notice that individual capital supply is asymptotically equivalent to $(1-\bar{c}_s)w$. We can therefore approximate the capital supply of agents with $w \geq w_N$ by

$$w - c_s(w) \approx w_N - c_{sN} + (1-\bar{c}_s)(w - w_N).$$

Second, notice that the density of wealth conditional on $w \geq w_N$ is approximately a Pareto distribution with exponent ζ and minimum size w_N, which has density $f(x) = \zeta w_N^{\zeta} x^{-\zeta-1}$. Therefore

$$E[w|w \geq w_N] \approx \int_{w_N}^{\infty} \zeta w_N^{\zeta} x^{-\zeta} x = \frac{\zeta}{\zeta - 1} w_N.$$

Combining both observations, we can approximate the capital supply of an agent in state (s,N) by

$$E[w - c_s(w)|w \geq w_N] \approx w_N - c_{sN} + \frac{1}{\zeta - 1}(1-\bar{c}_s)w_N.$$

Therefore, using Pareto extrapolation, the correct approximation of the aggregate capital supply K^s is

$$K^s_{\text{PE}} = \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn}(w_n - c_{sn}) + \frac{1}{\zeta - 1} \sum_{s=1}^{S} \pi_{sN}(1-\bar{c}_s)w_N. \tag{2.20}$$

Comparing (2.19) to (2.20), we can see that the truncation method introduces an error because the correction term $\frac{1}{\zeta-1}\sum_{s=1}^{S} \pi_{sN}(1-\bar{c}_s)w_N$ is absent. If ζ is close to 1 (Zipf's law), then failing to account for this term will introduce significant error. The formula (2.20) highlights the fact that choosing a large grid point w_N (so that the fraction of agents at the largest grid point $\sum_{s=1}^{S} \pi_{sN}$ is small) does not necessarily solve the problem. The reason is that w_N and $\sum_{s=1}^{S} \pi_{sN}$ enter multiplicatively in the correction term. For example, if $\sum_{s=1}^{S} \pi_{sN} = 10^{-6}$ and $w_N = 10^6$, then the correction term is $\frac{1}{\zeta-1}\sum_{s=1}^{S}(1-\bar{c}_s)$, which can be substantial as a fraction of K^s.

The insight carries well beyond the KS model. In general, we can approximate the integral of any policy function $x_s(w)$ that is asymptotically equivalent to $\bar{x}_s w$ against the stationary

distribution using

$$X \approx \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} x_{sn} + \frac{1}{\zeta - 1} \sum_{s=1}^{S} \pi_{sN} \bar{x}_s w_N. \tag{2.21}$$

A special case is the computation of aggregate wealth, which is simply

$$E[w] \approx \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} w_n + \frac{1}{\zeta - 1} \sum_{s=1}^{S} \pi_{sN} w_N. \tag{2.22}$$

Notice that computing the correction term is not computationally intensive. The reason is that we provide a closed-form expression for the correction term whose inputs (asymptotic policy functions and theoretical Pareto exponent) were already computed in the asymptotic analysis step.

2.3.5 Additional considerations

Dynamic programming is by far the most computationally intensive step when solving a heterogeneous-agent model. Compared to the asymptotic problem, which reduces to solving for an $S \times 1$ object, dynamic programming needs to solve for an $S \times N$ object. However, we can use the solution to the asymptotic problem to construct a "good" initial guess, which helps speed up the convergence. For example, when solving the KS model using policy function iteration, we suggest using the initial guess

$$c_s^{(0)}(w) = \bar{c}_s(w - c), \tag{2.23}$$

where $\bar{c}_s$ is the asymptotic MPC and $c < w_1$ is an arbitrary number that ensures that consumption is positive. Intuitively, the reason why this guess speeds up the convergence is that the distance between the true solution $c_s(w)$ and the guess $c_s^{(0)}(w)$ is already small, especially at the upper end of the grid.

The description of our algorithm above has implicitly assumed that the researcher has already chosen the grid $\mathcal{W} = \{w_n\}_{n=1}^N$ and, in particular, the truncation point w_N. We discuss the choice of the grid in Section 2.4 and Appendix B.3. Here we provide a practical guidance on how to choose the truncation point w_N.

We suggest choosing a truncation point that implies a small difference between the marginal propensity to consume (MPC) at the largest grid point, $(c_{sN} - c_{s,N-1})/(w_N - w_{N-1})$, and the asymptotic MPC $\bar{c}_s$ determined by solving (2.5). Therefore the researcher should choose a truncation point w_N that implies a small MPC relative error

$$\frac{1}{\bar{c}_s}\frac{c_{sN} - c_{s,N-1}}{w_N - w_{N-1}} - 1, \tag{2.24}$$

where c_{sn} denotes the policy function in state s at the grid point w_n and w_N is the largest grid point (truncation point). The idea is that, if the MPC at the largest grid point is close to its asymptotic value, then the consumption function should already be approximately linear, making all of the above approximations accurate.

So far, we have considered evaluating the market clearing condition for a guess of equilibrium prices. To solve for equilibrium prices, we can apply the Pareto extrapolation method for successive guesses of equilibrium prices and update the guesses using the excess supply computed in the aggregation step. Here the asymptotic analysis step can be used to narrow down the set of prices consistent with an equilibrium.

Notice that we can rule out any prices such that (i) there exists no solution to the asymptotic individual problem or (ii) the theoretical Pareto exponent ζ is below or equal one. When $\zeta \leq 1$, aggregate wealth is infinite, which is inconsistent with market clearing. In the KS model, narrowing down the set of equilibrium interest rates R consistent with an equilibrium amounts to evaluating (2.6) and (2.10) for a range of values of R and computing a bound $(R, \bar{R})$. Since the asymptotic analysis step is not computationally intensive, we can make substantial efficiency gains by avoiding the dynamic programming step as much as possible.

Suppose we want to compute the expectation of the power function w^ν for some power ν. For example, $\nu = 1$ corresponds to aggregate wealth, $\nu = 2$ the variance of wealth, and $\nu = 1 - \gamma$ with $\gamma > 0$ appears in calculating the welfare for CRRA preferences with relative risk aversion $\gamma > 0$. Assuming that the power ν is below the theoretical Pareto upper tail with exponent ζ, the

conditional expectation of the upper tail is

$$E[w^{\nu}|w \geq w_N] = \int_{w_N}^{\infty} \zeta w_N^{\zeta} x^{\nu-\zeta-1} x = \frac{\zeta}{\zeta-\nu} w_N^{\nu}.$$

Therefore the analog of (2.22) is

$$E[w^{\nu}] \approx \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} w_n^{\nu} + \frac{\nu}{\zeta-\nu} \sum_{s=1}^{S} \pi_{sN} w_N^{\nu}. \tag{2.25}$$

Similarly, noting that

$$E[w^{\nu} \log w|w \geq w_N] = E[\frac{d}{d\nu} w^{\nu}|w \geq w_N] = \frac{\zeta}{(\zeta-\nu)^2} w_N^{\nu} + \frac{\zeta}{\zeta-\nu} w_N^{\nu} \log w,$$

setting $\nu = 0$ we obtain

$$E[\log w] \approx \sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} \log w_n + \frac{1}{\zeta} \sum_{s=1}^{S} \pi_{sN}. \tag{2.26}$$

In the KS model, the value function is asymptotically equivalent to $\bar{v}_s \frac{w^{1-\gamma}}{1-\gamma}$ (see Footnote 6). To approximate the welfare function $\mathcal{W}$, which is defined as the integral of the value function against the stationary distribution, we can apply (2.25) to $\nu = 1-\gamma$ and dividing by $1-\gamma$. Hence the welfare in consumption equivalent is

$$\mathcal{W} \approx \left(\sum_{s=1}^{S} \sum_{n=1}^{N} \pi_{sn} v_{sn} + \frac{1}{\zeta-1+\gamma} \sum_{s=1}^{S} \pi_{sN} \bar{v}_s w_N^{1-\gamma} \right)^{\frac{1}{1-\gamma}}. \tag{2.27}$$

Oftentimes, the object of interest is the wealth distribution itself. One common way to summarize the concentration of wealth (other than to report the Pareto exponent) is to compute "top wealth shares" (, the share of wealth owned by the top fraction $p \in (0,1)$ agents). We compute the top wealth shares as follows. For each grid point, we can compute the aggregate wealth held by agents at least as rich as that grid point. Dividing that number by aggregate wealth gives the

top wealth share at that grid point. By interpolating between points, we can define the top wealth shares inside the grid. To compute the top wealth shares outside the grid, we suggest using the theoretical Pareto exponent ζ to extrapolate the wealth share beyond the largest grid point. More precisely, let $\pi_N = \sum_{s=1}^{S} \pi_{sN}$ be the probability mass on the largest grid point w_N. The density for $x \geq w_N$ is then $f(x) = \pi_N \zeta w_N^{\zeta} x^{-\zeta-1}$. Using this, the tail probability is

$$\Pr(X \geq x) = \int_x^{\infty} \pi_N \zeta w_N^{\zeta} x^{-\zeta-1} x = \pi_N w_N^{\zeta} x^{-\zeta}. \tag{2.28}$$

On the other hand, the total wealth held by wealthy agents is

$$E[X|X \geq x] = \int_x^{\infty} \pi_N \zeta w_N^{\zeta} x^{-\zeta} x = \frac{\zeta}{\zeta - 1} \pi_N w_N^{\zeta} x^{-\zeta+1}. \tag{2.29}$$

Therefore letting W be the aggregate wealth, setting $p = \Pr(X \geq x)$, and eliminating x, the wealth share $s(p)$ of the wealthiest fraction $p \in (0,1)$ of agents is

$$s(p) = \frac{\zeta}{\zeta - 1} \pi_N^{1/\zeta} \frac{w_N}{W} p^{1-1/\zeta}. \tag{2.30}$$

2.4 Evaluating solution accuracy

As in any new numerical method, the first order of business is to evaluate the solution accuracy. In this regard, Den Haan, Judd, and Juillard (2010) "find it troublesome that [...] the accuracy of numerical solutions obtains so little attention by so many authors these days." One reason why accuracy gets little attention may be due to the lack of benchmark closed-form solutions for heterogeneous-agent models.[9] For this purpose, we present a simple (minimal) heterogeneous-agent model with idiosyncratic investment risk that admits a semi-analytical solution, which we use as a benchmark for evaluating numerical solutions.

[9] In the context of representative-agent asset pricing models, several authors such as Collard and Juillard (2001), Schmitt-Grohé and Uribe (2004), and Farmer and Toda (2017) use the closed-form solution of Burnside (1998) to evaluate the solution accuracy.

2.4.1 Model

We consider a standard Aiyagari (1994) model, except that the model features no idiosyncratic labor income risk (to make the model analytically tractable) but only investment risk (to generate a fat-tailed wealth distribution). The production side is completely standard: there is a representative firm with Cobb-Douglas production function $F(K,L) = AK^{\alpha}L^{1-\alpha}$, where $A > 0$ is productivity and $\alpha \in (0,1)$ is the capital share. Capital depreciates at rate δ each period. There are two types of agents, capitalists and workers, of whom there is a mass 1 continuum each. Workers are identical, supply one unit of labor inelastically, and consume the entire wage (hand-to-mouth).[10]

To obtain stationarity, we assume that capitalists are born and go bankrupt with probability p each period (Yaari (1965)–Blanchard (1985) perpetual youth model). Newborn agents are exogenously endowed with initial wealth $w_0 > 0$, and capital is destroyed after bankruptcy. Capitalists have constant relative risk aversion (CRRA) utility

$$E_0 \sum_{t=0}^{\infty} [\beta(1-p)]^t \frac{c_t^{1-\gamma}}{1-\gamma} \tag{2.31}$$

and supply capital to the firm. Importantly, the gross return on capital is *not* risk-free as

$$R_f = F_K(K,1) + 1 - \delta = A\alpha K^{\alpha-1} + 1 - \delta, \tag{2.32}$$

but rather $z_s R_f$, where $s = 1,\ldots,S$ denotes the exogenous Markov state and $z_s > 0$ is the gross return on capital *relative* to the risk-free rate (essentially the excess return). Let $P = (p_{ss'})$ be the transition probability matrix, which we assume to be irreducible. We assume that $E[z_s] = 1$, so capital income is just a zero-sum redistribution of aggregate capital income across capitalists. An interpretation is that capitalists earn persistent heterogeneous returns (Fagereng, Guiso, Malacrino,

[10]The hand-to-mouth assumption is only for simplicity. Although we can also assume that workers behave optimally, it is inessential for our purpose of discussing numerical algorithms and evaluating the solution accuracy.

and Pistaferri 2016b; Cao and Luo 2017) because some are more skillful in using capital (or just lucky) than others. The initial state of a newborn capitalist is drawn from the stationary distribution $\pi = (\pi_1, \ldots, \pi_S)'$ of the transition probability matrix P.

The timing is as follows. A capitalist enters period t with some resource (units of consumption good) w_t. He decides how much to consume c_t, and the remaining amount $k_{t+1} := w_t - c_t$ is installed as capital. At the beginning of period $t+1$, production takes place by pooling all capital, and the capitalist receives the proceed $w_{t+1} = z_{s_t} R_f k_{t+1}$, where R_f is the gross risk-free rate in (2.32) and z_{s_t} is the predetermined gross excess return.[11] Thus the budget constraint of a capitalist is

$$w' = z_s R_f (w - c). \tag{2.33}$$

A stationary equilibrium consists of aggregate capital K, gross risk-free rate R_f, optimal consumption rule $\{c_s(w)\}_{s=1}^S$, and a stationary distribution $\Gamma(w,s)$ such that (i) given R_f, the optimal consumption rule maximizes the utility (2.31) subject to the budget constraint (2.33), (ii) firms maximize profits, so (2.32) holds, (iii) the capital market clears, so

$$K = \int (w - c_s(w)) \Gamma(w,s), \tag{2.34}$$

and (iv) $\Gamma(w,s)$ is the stationary distribution of the law of motion

$$(w,s) \mapsto \begin{cases} (z_s R_f (w - c_s(w)), s'), & \text{with probability } (1-p) p_{ss'}, \\ (w_0, s'), & \text{with probability } p \pi_{s'}. \end{cases}$$

By exploiting homotheticity, we can solve the model semi-analytically as discussed in Appendix B.2.2. We also prove that a stationary equilibrium exists and the wealth distribution has a Pareto upper tail.

We use a numerical example to evaluate the solution accuracy. We consider the parameter

[11] We can also allow for the possibility that the gross excess returns are risky by using $z_{ss'}$ instead of z_s.

values in Table 2.2. By solving the equilibrium conditions discussed in Appendix B.2.2, we obtain the gross risk-free rate $R_f = 1.0972$, aggregate capital $K = 3.4231$, and Pareto exponent $\zeta = 1.2826$.

Table 2.2: Parameter values of the Aiyagari model.

Parameter	Symbol	Value
Discount factor	β	0.96
Relative risk aversion	γ	2
Bankruptcy probability	p	0.025
Gross excess return	z	(0.95,1.05)
Transition probability matrix	P	$\begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$
Productivity	A	1
Capital share	α	0.38
Capital depreciation rate	δ	0.08
Initial wealth	w_0	1

For the numerical solution, we consider both the conventional truncation method as well as the proposed Pareto extrapolation method with various wealth grid, truncation point, and number of grid points. For the Pareto extrapolation spacing parameter h, we always take $h = w_N - w_{N-1}$, the distance between the two largest grid points.

2.4.2 Solution accuracy in partial equilibrium

We first evaluate the solution accuracy in partial equilibrium. To ensure that all the differences of the numerical solutions from the analytical one are entirely due to the construction of the transition probability matrix, instead of solving for the equilibrium numerically for each method, we use the equilibrium risk-free rate and consumption policies from the semi-analytical solution to compute the stationary distribution on the wealth grid, and then compute the implied aggregate capital using (2.1) and (2.22) for the Truncation and Pareto extrapolation methods, respectively. For this exercise, our primary interest is the relative error $\widehat{K}/K - 1$, where K and $\widehat{K}$ are the aggregate capital from the semi-analytical and numerical solutions, respectively.

To implement our algorithm, we first need to specify the wealth grid $\{w_n\}_{n=1}^N$. Two natural candidates are the evenly- and exponentially-spaced grids, which we discuss in detail in Appendices B.3.1 and B.3.2, respectively. These two grids have both advantages and disadvantages. An ideal grid is one such that

1. the grid spacing $w_n - w_{n-1}$ is sufficiently small in the bulk of the wealth distribution so that we can approximate the law of motion accurately, and

2. the largest point w_N is sufficiently large so that the grid covers the (potentially) nonlinear part of the policy functions.

The evenly-spaced grid achieves the first objective but fails the second, while the opposite is true for the exponentially-spaced grid.

As a compromise, we suggest using the hybrid (affine-exponential) grid: construct the exponentially-spaced grid as discussed in Appendix B.3.2, but replace the bottom by an evenly-spaced grid. This way, we can choose a relatively large truncation point w_N, while keeping the grid spacing $w_n - w_{n-1}$ small for at least the bottom points, which contain the bulk of the wealth distribution. In particular, we construct the grid as follows. First, we compute the aggregate capital in a corresponding representative-agent model, which is

$$K_{\mathrm{RA}} = ((1/(\beta(1-p)) - 1 + \delta)/(A\alpha))^{\frac{1}{\alpha-1}} = 4.5577. \tag{2.35}$$

K_{RA} serves as the transition point between the evenly- and exponentially-spaced grids. Second, we construct an N-point exponential grid on $(0, \bar{w}]$ such that the median grid point corresponds to K_{RA}, and we replace the points on $(0, K_{\mathrm{RA}}]$ by an evenly-spaced grid. Table 2.3 shows the relative error $\widehat{K}/K - 1$ in the aggregate capital using this affine-exponential grid for various truncation point $\bar{w}$ and number of points N, both for the truncation and Pareto extrapolation methods.

We can make a few observations from Table 2.3. First, the conventional truncation method is extremely poor at calculating the aggregate capital with a moderate truncation point $\bar{w}$: the

Table 2.3: Relative error (%) in aggregate capital for the truncation and Pareto extrapolation methods with the affine-exponential grid.

Method:	Truncation			Pareto extrapolation		
$\bar{w}/K_{\text{RA}}$	$N=25$	50	100	25	50	100
10^1	-31.240	-27.620	-26.250	-1.110	0.292	0.422
10^2	-21.500	-16.860	-14.480	-2.172	-0.642	0.128
10^3	-16.510	-11.210	-8.590	-2.303	-0.827	-0.141
10^4	-13.610	-7.950	-5.360	-2.234	-0.804	-0.205
10^5	-11.750	-5.930	-3.490	-2.125	-0.727	-0.200
10^6	-10.530	-4.610	-2.360	-2.029	-0.643	-0.174

Note: N: number of grid points; $\bar{w}$: wealth truncation point.

relative error is about 26% with $\bar{w} = 10K_{\text{RA}} = 45.6$ and $N = 100$, which is similar to the case with an evenly-spaced grid with $\bar{w} = 40$ and $N = 100$ in Table B.1 (27%). On the other hand, the Pareto extrapolation method is astonishingly more accurate, with relative errors ranging from 0.13% to 2.3% depending on the specification. Second, for the truncation method, choosing a larger truncation point $\bar{w}$ improves the accuracy because it misses less of the upper tail. However, even with a huge truncation point such as $\bar{w}/K_{\text{RA}} = 10^6$, the errors exceed the largest errors with Pareto extrapolation. Finally, the accuracy of the Pareto extrapolation method is almost independent of $\bar{w}$. This is probably because the upper tail is well-approximated by a Pareto distribution and our method corrects for the truncation error using the theoretical Pareto exponent, so the choice of $\bar{w}$ is not so important for accuracy.

2.4.3 Solution accuracy in general equilibrium

So far we have evaluated the accuracy of each solution method by comparing the implied aggregate capital to that from the semi-analytical solution. However, aggregate capital is usually not a quantity of interest. Therefore we now evaluate the solution accuracy by solving for the entire equilibrium.

We consider both the truncation and Pareto extrapolation methods with the affine-

exponential grid, where the number of grid points is $N = 100$, the median grid point is K_{RA}, and the largest grid point $\bar{w} = w_N$ is such that $\bar{w}/K_{\text{RA}} = 10^1, 10^2, 10^3, 10^4, 10^5, 10^6$ as before. To ensure that all the differences of the numerical solutions from the analytical one are entirely due to the construction of the transition probability matrix, for each guess of the equilibrium risk-free rate, we use the semi-analytical solution to the optimal consumption-savings problem to compute the law of motion for wealth. Table 2.4 shows the relative errors in equilibrium quantities (gross risk-free rate R_f, aggregate capital K, and Pareto exponent ζ).

Table 2.4: Relative errors (%) in equilibrium quantities.

Method:	Truncation			Pareto extrapolation		
$\bar{w}/K_{\text{RA}}$	R_f	K	ζ	R_f	K	ζ
10^1	0.761	-7.159	-13.317	-0.005	0.051	0.100
10^2	0.287	-2.801	-5.399	-0.002	0.016	0.031
10^3	0.142	-1.402	-2.732	0.002	-0.017	-0.034
10^4	0.080	-0.793	-1.553	0.003	-0.025	-0.049
10^5	0.049	-0.486	-0.955	0.002	-0.024	-0.048
10^6	0.032	-0.316	-0.622	0.002	-0.021	-0.042

Note: $\bar{w}$: wealth truncation point. The number of grid points is $N = 100$. The actual values are: $R_f = 1.0972$, $K = 3.4231$, $\zeta = 1.2826$.

As opposed to partial equilibrium, in general equilibrium the truncation method with a moderately large truncation point performs reasonably well. For example, the relative errors are 0.1–3% when $\bar{w}/K_{\text{RA}} = 10^3$. However, the relative errors for the Pareto extrapolation method are 0.002–0.03% with the same $\bar{w}$, so it is 100 times more accurate. The reason why the truncation method performs better in general equilibrium than in partial equilibrium is probably due to the general equilibrium effect: because the truncation method underestimates the aggregate capital (given the risk-free rate, as seen in Table 2.3), the equilibrium risk-free rate must rise to clear the market. In fact, in Table 2.4, R_f is upwardly biased for the truncation method. Since wealth (and hence aggregate capital) rises with a higher risk-free rate, the downward bias in aggregate capital with the truncation method gets mitigated in general equilibrium.

Given that the truncation method with a moderately large truncation point $\bar{w}$ is reasonably accurate, one may wonder what is the point of improving it further. There are several reasons to prefer the Pareto extrapolation method. First, we can obtain 10–100 times more accurate results at no additional computational cost, so there is just no reason not to use it. Second, and more importantly, the performance of the Pareto extrapolation method is robust across the choice of the truncation point $\bar{w}$, whereas it is sensitive for the truncation method. Thus, with the Pareto extrapolation method, the researcher need not worry about the choice of the truncation point, while extra care is needed with the truncation method. Finally, the interest rate, aggregate capital, and Pareto exponent may not be the only quantities of interest. One may be interested in other quantities, such as the top 1% wealth share.

To address the last point, we compute top wealth shares using the Pareto extrapolation method as described in Section 2.3.5. For the truncation method, since it is not obvious how to extrapolate the top wealth share beyond the largest grid point, we simply interpolate by a cubic spline using the point $(0,0)$ (by definition, the top 0% wealth share is 0) and all the grid points. Table 2.5 shows some representative top wealth shares.[12] Because top wealth shares need to be computed only once after solving for the equilibrium, for both truncation and Pareto extrapolation methods, we use a grid that is 10 times finer than the one used for solving the equilibrium.

According to Table 2.5, the truncation method vastly underestimates top wealth shares when the truncation point $\bar{w}$ is small, as expected. On the other hand, the Pareto extrapolation method gives numbers that are accurate up to two or three significant digits.

[12]Technically, for the semi-analytical solution we cannot compute the exact top wealth shares because the functional form of the wealth distribution is unknown (we only know the tail behavior characterized by the Pareto exponent ζ). For this case, to compute the stationary distribution, we use the Pareto extrapolation method with a highly accurate 2,000-point affine-exponential grid with truncation $\bar{w} = 10^6 \times K_{\mathrm{RA}}$, which we take as the truth.

Table 2.5: Top wealth shares (%) in equilibrium.

Method:	Truncation				Pareto extrapolation			
$\bar{w}/K_{\text{RA}}$	Top 0.01%	0.1%	1%	10%	Top 0.01%	0.1%	1%	10%
10^1	0.11	1.16	12.76	50.87	13.11	21.81	36.27	60.31
10^2	1.20	10.08	29.07	57.55	13.20	21.92	36.37	60.39
10^3	7.08	17.43	33.76	59.54	13.27	21.99	36.46	60.46
10^4	10.65	20.16	35.51	60.28	13.29	22.01	36.48	60.47
10^5	12.17	21.32	36.23	60.55	13.30	22.03	36.49	60.47
10^6	12.83	21.80	36.50	60.63	13.28	22.00	36.47	60.46
Analytical	13.21	21.92	36.39	60.40	13.21	21.92	36.39	60.40

Note: $\bar{w}$: wealth truncation point; Top x%: wealth share (%) of the wealthiest x%. The number of grid points is $N = 100$ for solving the equilibrium and $10N = 1{,}000$ for computing top wealth shares. "Truncation" and "Pareto extrapolation" refer to the truncation and Pareto extrapolation methods for solving the equilibrium, and "Analytical" shows results from the semi-analytical solution. Top wealth shares for "Analytical" are computed using the grid in Footnote 12.

2.4.4 Constructing the grid for wealth

What do we learn from these exercises? The good news is that the conventional truncation method is able to solve models reasonably accurately, provided that we use an exponentially-spaced grid with a large enough number of points and a large enough truncation point—say a million times the typical scale. The bad news is that we do not a priori know how large is large enough. There is a superior alternative: Pareto extrapolation is far more accurate, it has no additional computational cost, and its performance is robust across the grid specification.

Based on the above observations as well as the results in Appendix B.3, we recommend the following strategy for solving heterogeneous-agent models with fat-tailed wealth distribution. [Constructing the grid for wealth]

1. Before solving the model, find out a typical scale for the state variable (wealth), perhaps by solving a representative-agent model without any shock.

2. Solve the heterogeneous-agent model using the Pareto extrapolation method with the hybrid affine-exponential grid. More concretely,

(a) Construct the exponentially-spaced grid with a truncation point about 1,000 times the typical scale for the state variable.

(b) Replace the bottom half grid points by an evenly-spaced grid.

3. After solving the model, if necessary, recompute the wealth distribution on a finer grid from the already computed equilibrium law of motion.

Note that there are many other possibilities. In the value function iteration step, since we do not need the wealth distribution, we can just use an exponentially-spaced grid with relatively few points to increase the speed. When computing the market clearing condition, we can interpolate the policy functions on a finer grid and then use Pareto extrapolation method for accuracy.

2.5 Merton-Bewley-Aiyagari model

Having now established that the solution method we propose is accurate, we apply it to study wealth inequality in an incomplete market general equilibrium model in the spirit of 1994. Agents face an income fluctuation problem as in 1977, 1983 and those who choose to invest face uninsurable investment risk that leads to an investor's problem similar to 1969 and 1969, although in a Markovian setting as studied in 2006 and 2014. The model (which we refer to as the Merton-Bewley-Aiyagari, or MBA model) generates a fat-tailed wealth distribution, where the Pareto exponent is shaped by rich general equilibrium effects. We provide a step-by-step approach to solving the model and analyzing its quantitative implications.

2.5.1 Model

Time is discrete and denoted by $t = 0, 1, \ldots$.

The economy is populated by a unit measure of infinitely-lived agents with Epstein-Zin

preferences

$$U_t = \left((1-\beta) c_t^{1-1/\varepsilon} + \beta E_t [U_{t+1}^{1-\gamma}]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}, \tag{2.36}$$

where $c_t > 0$ is consumption, $U_t > 0$ is continuation utility, $\beta \in (0,1)$ is the discount factor, $\gamma > 0$ is the coefficient of relative risk aversion, and $\varepsilon > 0$ is the elasticity of intertemporal substitution.[13] Agents differ in productivity/ability states denoted by $s \in S = \{1, \ldots, S\}$, which evolve over time according to a Markov chain with irreducible transition probability matrix $P = (p_{ss'})$. The idiosyncratic productivity states are independent and identically distributed across agents and we assume the law of large numbers for the continuum as in 2006. Therefore if $\pi = (\pi_1, \ldots, \pi_S)'$ denotes the (unique) stationary distribution of the transition probability matrix P, at any point in time exactly fraction $\pi_s > 0$ of agents are in state s.

A type s agent has labor productivity $h_s \geq 0$ and earns pre-tax labor income ωh_s, where $\omega > 0$ is the "piece-rate" wage determined in equilibrium. A type s agent has investment ability $z_s > 0$ and earns excess returns in the financial market, as described below. Without loss of generality we assume that $z_1 \leq \cdots \leq z_S$, so the Markov state s indexes the agent's investment ability. There are two types of assets, risk-free and risky. Let R_f be the gross risk-free rate determined in equilibrium. We assume that the ex post pre-tax gross return on risky investment for an investor in state s is

$$R_{sj} = z_{sj} R_f, \tag{2.37}$$

where z_{sj} is defined as the sum of the investment ability z_s and a zero-mean i.i.d. random variable ε_j that can take J possible values $\varepsilon_1 < \cdots < \varepsilon_J$. Let $p_j > 0$ be the probability of state j. Thus high-skilled investors earn higher returns on average (z_s), but there is some element of luck (ε_j). We can interpret this as lack of diversification. We assume that $z_1 + \varepsilon_1 > 0$, so agents have limited liability even in the worst possible state. To prevent arbitrage, we also assume that $z_S + \varepsilon_1 < 1$, so

[13] By considering the limit $\gamma \to 1$, we interpret $[U^{1-\gamma}]^{\frac{1}{1-\gamma}}$ as $\mathbb{E}([\log U])$ if $\gamma = 1$. Similarly, we interpret $((1-\beta)c^{1-1/\varepsilon} + \beta v^{1-1/\varepsilon})^{\frac{1}{1-1/\varepsilon}}$ as $c^{1-\beta} v^{\beta}$ if $\varepsilon = 1$.

even the most skilled investor underperforms the risk-free asset with positive probability. From the above assumptions, note that (i) z_{sj} is increasing in both s and j, (ii) $z_{sj} > 0$ for all s, j, and (iii) $z_{S1} < 1$.

Technology is represented by a representative firm with a constant-returns-to-scale production function $F(K,L)$. Capital depreciates at rate $\delta \in [0,1]$. Therefore the firm's problem is

$$\max_{K,L\geq 0} \left[-K + \frac{1}{R_f}(F(K,L) - \omega L + (1-\delta)K) \right]. \tag{2.38}$$

That is, the firm buys capital K at the end of time t, hires labor to produce, and pays the profit and depreciated capital to the shareholders (who discount using the risk-free rate since there is no aggregate risk).

Letting w be the financial wealth at the beginning of the period, the budget constraint of an agent is

$$w' = (1-\tau_w)\Big((1+(1-\tau_k)(R_f-1))(w+(1-\tau_h)\omega h_s - I - c) + (1+(1-\tau_k)(R_{sj}-1))I\Big) \geq w,$$

where w is an exogenous minimum wealth constraint and $I \geq 0$ is the investment in the risky asset. We assume that there are proportional taxes on labor income (at rate τ_h), capital income (τ_k), and wealth (τ_w). Notice that the capital income tax applies to the net return on the risk-free and risky assets while the wealth tax applies to the beginning of the period wealth. We assume that the tax proceeds are wasted. To simplify the notation, we denote by $\widetilde{R}_f, \widetilde{R}_{sj}$ the after-tax gross returns on the risk-free and risky assets, respectively:

$$\widetilde{R}_f = (1-\tau_w)(1+(1-\tau_k)(R_f-1)), \tag{2.39a}$$

$$\widetilde{R}_{sj} = (1-\tau_w)(1+(1-\tau_k)(R_{sj}-1)). \tag{2.39b}$$

The budget constraint thus simplifies to

$$w' = \widetilde{R}_f(w + (1-\tau_h)\omega h_s - I - c) + \widetilde{R}_{sj} I \geq w. \tag{2.40}$$

Our equilibrium concept is the stationary equilibrium defined as follows.

Definition 1 (Stationary equilibrium). A stationary equilibrium consists of a gross risk-free rate R_f, a piece-rate wage ω, aggregate capital K, aggregate labor L, optimal decision rules $\{c_s(w), I_s(w)\}_{s=1}^S$, value functions $\{v_s(w)\}_{s=1}^S$, and a stationary distribution $\Gamma(w,s)$ such that

1. given R_f and ω, aggregate capital K and aggregate labor L solves the profit maximization problem (2.38),

2. given R_f and ω, for each s the optimal decision rule $(c_s(w), I_s(w))$ maximizes the recursive utility (2.36) subject to the budget and borrowing constraint (2.40), ,

$$v_s(w) = \max_{c,I \geq 0} \left((1-\beta)c^{1-1/\varepsilon} + \beta E[v_{s'}(w')^{1-\gamma}|s]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}, \tag{2.41}$$

3. the capital market clears, so

$$K = \int (w + (1-\tau_h)\omega h_s - I_s(w) - c_s(w))\Gamma(w,s) + \int z_s I_s(w)\Gamma(w,s), \tag{2.42}$$

4. the labor market clears, so

$$L = \sum_{s=1}^{S} \pi_s h_s, \tag{2.43}$$

5. $\Gamma(w,s)$ is the stationary distribution of the law of motion for $(w,s) \in [w,\infty) \times S$ defined by

$$(w,s) \mapsto \left(\widetilde{R}_f(w + (1-\tau_h)\omega h_s - I - c) + \widetilde{R}_{sj} I, s' \right) \tag{2.44}$$

with probability $p_{ss'} p_j$, where $\widetilde{R}_f$ and $\widetilde{R}_{sj}$ are as in (2.39).

Note that an investor who invests I units of wealth supplies $z_s I$ units of capital to the

firm (see the right-most term in (2.42)), so investors not only supply funds but also "expertise". This assumption (together with $E[\varepsilon_j] = 0$) makes the (pre-tax) gross return to investment $R_{sj} = (z_s + \varepsilon_j)R_f$ consistent with the aggregate resource constraint.

2.5.2 Asymptotic analysis

As discussed in Section 2.3, the solution to the asymptotic problem plays an important role in the Pareto extrapolation algorithm (especially for computing the theoretical Pareto exponent). In this section, we discuss the properties of the asymptotic problem of the MBA model and derive parametric restrictions in general equilibrium.

Asymptotic problem

We first derive the asymptotic problem and convert it in a more convenient form. Since there is no labor income in the asymptotic problem, the budget constraint (2.40) becomes

$$w' = \widetilde{R}_f(w - I - c) + \widetilde{R}_{sj} I \geq 0, \tag{2.45}$$

where $\widetilde{R}_f, \widetilde{R}_{sj}$ are as in (2.39). Letting $\theta = \frac{I}{w-c} \geq 0$, (2.45) becomes

$$w' = \left(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta\right)(w-c) \geq 0. \tag{2.46}$$

The following proposition characterizes the solution to the asymptotic problem.

Proposition 10 (Asymptotic problem)**.** *Suppose $\gamma \neq 1$. For $s = 1, \ldots, S$, define*

$$\rho_s = \max_{0 \leq \theta \leq \bar{\theta}_s} E[\left(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta\right)^{1-\gamma} | s]^{\frac{1}{1-\gamma}}, \tag{2.47}$$

where the upper bound is $\bar{\theta}_s := \frac{\widetilde{R}_f}{\widetilde{R}_f - \widetilde{R}_{s1}} > 0$. Then ρ_s is well-defined and there exists a unique

maximizer θ_s^*. *Letting* $D = \left(\rho_1^{1-\gamma}, \ldots, \rho_S^{1-\gamma}\right)$, *the asymptotic problem has a solution if and only if*

$$\beta\rho(DP)^{\frac{1-1/\varepsilon}{1-\gamma}} < 1, \tag{2.48}$$

where $\rho(DP)$ *is the spectral radius of DP. Under this condition, the value function of the asymptotic problem is given by* $v_s(w) = b_s w$, *where* $b = (b_1, \ldots, b_S) \gg 0$ *is the unique positive solution to the system of nonlinear equations*

$$b_s = \begin{cases} \left((1-\beta)^\varepsilon + \beta^\varepsilon \left(\rho_s E[b_{s'}^{1-\gamma}|s]^{\frac{1}{1-\gamma}}\right)^{\varepsilon-1}\right)^{\frac{1}{\varepsilon-1}} & (\varepsilon \neq 1) \\ (1-\beta)^{1-\beta}\beta^\beta \left(\rho_s E[b_{s'}^{1-\gamma}|s]^{\frac{1}{1-\gamma}}\right)^\beta & (\varepsilon = 1) \end{cases} \tag{2.49}$$

for $s = 1, \ldots, S$. *The optimal consumption-investment rules of the asymptotic problem are*

$$c_s(w) = \bar{c}_s w := (1-\beta)^\varepsilon b_s^{1-\varepsilon} w, \tag{2.50a}$$

$$I_s(w) = \bar{I}_s w := \theta_s^*(1-(1-\beta)^\varepsilon b_s^{1-\varepsilon})w. \tag{2.50b}$$

Parametric restrictions in general equilibrium

Since the diagonal matrix $D = \left(\rho_1^{1-\gamma}, \ldots, \rho_S^{1-\gamma}\right)$ depends on the equilibrium interest rate R_f, the spectral condition (2.48) puts a restriction on R_f. The following lemma puts further restrictions based on equilibrium considerations.

Lemma 3 (Finite aggregate wealth)**.** *Let everything be as in Proposition 10 and suppose that* (2.48) *holds. Define*

$$G_s := (1-(1-\beta)^\varepsilon b_s^{1-\varepsilon})(\widetilde{R}_f(1-\theta_s^*) + E[\widetilde{R}_{sj}|s]\theta_s^*) \tag{2.51}$$

and $G = (G_1, \ldots, G_S)'$. *Then in equilibrium it must be*

$$\rho(PG) < 1. \tag{2.52}$$

The intuition for Lemma 3 is as follows. Using the budget constraint and the optimal consumption and portfolio rules for asymptotic agents established in Proposition 10, the expected growth rate of wealth in state s becomes G_s in (2.51). The spectral condition (2.52) ensures that the wealth of rich agents does not grow on average and makes the aggregate wealth finite, which must be the case in stationary equilibrium.

Under the assumptions of Lemma 3, the following lemma provides an explicit algorithm for computing the Pareto exponent ζ.

[Pareto exponent] Let everything be as in Lemma 3 and

$$G_{sj} = (1-(1-\beta)^{\varepsilon} b_s^{1-\varepsilon})(\widetilde{R}_f(1-\theta_s^*) + \widetilde{R}_{sj}\theta_s^*) > 0 \tag{2.53}$$

be the ex post gross growth rate of wealth for asymptotic agents in state (s, j). Define the conditional moment generating function of log growth rate by

$$M_s(z) = E[G_{sj}^z | s] = \sum_{j=1}^{J} p_j G_{sj}^z \tag{2.54}$$

and the diagonal matrix $D(z) = (M_1(z), \ldots, M_S(z))$. Suppose that $p_{ss} > 0$ for all s and $G_{sJ} > 1$ for some s. Then there exists a unique solution $z = \zeta > 1$ to

$$\rho(PD(z)) = 1. \tag{2.55}$$

The Pareto exponent of the wealth distribution is $\zeta > 1$. If $G_{sJ} \leq 1$ for all s, then the wealth distribution does not have a Pareto tail.

The reason why the Pareto exponent formulas (2.10) and (2.55) are slightly different is because the conditional moment generating function (2.54) depends only on the current state s. If it depended on two consecutive states, we should use (2.10). It is straightforward to see that (2.10) reduces to (2.55) when dependence is only on the current state.

Finally, it is trivial to show that the equilibrium risk-free rate must satisfy

$$R_f > 1 - \delta, \tag{2.56}$$

otherwise the demand for capital would be infinite.

2.5.3 Calibration

A time period represents a year and we calibrate the model to the U.S. economy.

We assume a Cobb-Douglas production function $F(K,L) = K^{\alpha}L^{1-\alpha}$, where $\alpha \in (0,1)$ represents the capital share. We set the technological and preference parameters (β, γ, α) to standard values (see Table 2.6). For the elasticity of intertemporal substitution, most macro papers assume that it is less than 1, while most finance papers assume that it is greater than 1. To be neutral, we set $\varepsilon = 1$, which is also supported by studies using disaggregated data to estimate the elasticity (Mankiw and Zeldes 1991; Attanasio and Weber 1993; Beaudry and Wincoop 1996; Vissing-Jørgensen 2002). We set the labor income, capital income, and wealth tax rate respectively to $\tau_h = 0.224$, $\tau_k = 0.25$, $\tau_w = 0$ as in 2018, who use estimates from 2007. Finally, we set the borrowing limit to $1/4$ of average annual labor income as in 2018. Our calibration implies that average value of h_s in a stationary equilibrium is 1, so $w = -\omega/4$.

Table 2.6: Parameter values.

Parameter	Symbol	Value
Discount factor	β	0.96
Relative risk aversion	γ	2
Elasticity of intertemporal substitution	ε	1
Capital share	α	0.38
Labor income tax	τ_h	0.224
Capital income tax	τ_k	0.25
Wealth tax	τ_w	0
Borrowing limit	w	$-\omega/4$

We assume that labor productivity $\{h_{s_t}\}_{t=0}^{\infty}$ and investment ability $\{z_{s_t}\}_{t=0}^{\infty}$ depend con-

temporaneously on two Markov state variables s^{π} and s^{τ}, with associated transition probability matrices P^{π} and P^{τ}. We interpret the first state s^{π} as a "permanent component", which affects both labor productivity and investment ability and takes three states: low, high, and high-entrepreneur. The second state s^{τ}, which we call the "transitory component" affects only labor productivity and takes three values: low, average, and high. The index $s = s^{\pi} \times s^{\tau}$ can thus take 9 states.

Labor productivity in state s is the product of a permanent component and a transitory component:

$$h_s = h_s^{\pi} h_s^{\tau}.$$

The permanent component h_s^{π} takes two values: 0.3980 (low) and 1.6020 (high and high-entrepreneur). The high state workers (high and high-entrepreneur) thus earn a wage rate 4.03 times higher than low state workers, in line with the ratio of the mean annual income of the top half to the bottom half of full-time workers in the U.S.[14] We interpret agents as dynasties with perfect altruism, and assume that the permanent component of labor productivity is very persistent and changes on average every 40 years. Moreover, we choose transition probabilities for the permanent component as to imply that, in a stationary equilibrium, 50% of the agents are in the low state and 3.7% are in the high-entrepreneur state. We choose the value of 3.7% to match the fraction of households that invest at least half of their net worth in a business (see Footnote 15). The transition probability matrix P^{π} is thus given by

$$P^{\pi} = \begin{bmatrix} 0.9875 & 0.0116 & 0.0009 \\ 0.0125 & 0.9866 & 0.0009 \\ 0.0125 & 0.0116 & 0.9759 \end{bmatrix}. \tag{2.57}$$

We model the process for the transitory component of labor productivity h^{τ} as an AR(1) in

[14] We use data from the American Community Survey and restrict the sample to employed individuals aged between 20 and 60. For every year, we truncate the sample at the 5th and 95th percentile and then compute the ratio of the mean annual income in the bottom and top half of the sample. We then average the ratio over the 2000-2016 period and obtain an average ratio of 4.03.

logarithm

$$\log h_{s_t}^{\tau} = \rho \log h_{s_{t-1}}^{\tau} + \sigma \eta_t, \tag{2.58}$$

where $\eta_t \sim N(0,1)$. 2009a obtains values of $\rho = 0.821$ and $\sigma = 0.170$ by estimating a model that allows for heterogeneous lifetime earnings profiles. We discretize the process (2.58) over a three-point grid using the method proposed by 2017, while imposing that the unconditional mean of h_s^{τ} is equal to one. We obtain values for h_s^{τ} of 0.6584 (low state), 0.9150 (average state), and 1.5115 (high state). The resulting transition probability matrix P^{τ} is given by

$$P^{\tau} = \begin{bmatrix} 0.8290 & 0.1630 & 0.0080 \\ 0.0815 & 0.8370 & 0.0815 \\ 0.0080 & 0.1630 & 0.8290 \end{bmatrix}. \tag{2.59}$$

Since h_s^{π} and h_s^{τ} are independent and both have an unconditional mean of one, the average labor productivity $h_s = h_s^{\pi} h_s^{\tau}$ is one in equilibrium.

We assume that investment ability z is fully determined by the permanent component s^{π}. For the low and high states, we set $z = 1$, which implies that those agents do not earn an excess return on their investments (, they earn the risk-free rate R_f). Since investment returns are risky, those agents will never invest. For the high-entrepreneur state, we set $z = 1.028$, which implies an annual excess return of 2.8%. This is roughly the difference between the average return on financial assets of households at the 90th and 10th percentiles of the financial wealth distribution in Norway as reported in Fagereng, Guiso, Malacrino, and Pistaferri (2016a, Figure 2).

To calibrate the distribution of idiosyncratic investment return shocks ε, we use microdata from the Survey of Consumer Finances and construct a measure of the rate of return on business investment (business income over the market value of the business) for each household.[15] Using

[15] We use data from the Survey of Consumer Finances for the years 2001, 2004, 2007, 2010, 2013, and 2016 and keep only households who have at least 50% of their net worth invested in businesses. Business income `busseincfarm` is defined as "Income from business, sole proprietorship, and farm" while the market value of the businesses `bus` is defined as "Total value of business(es) in which the household has either an active or nonactive interest". We winsorize the rate of return at the 5% level.

the nonparametric discretization method proposed by 2018 on the de-meaned data, we obtain a discrete distribution for ε, which takes values $(-0.0836, 0.0761, 0.3795)$ with probability $(0.6345, 0.2822, 0.0833)$.

Putting all the pieces together, we have $S = 9$ exogenous individual states and $J = 3$ idiosyncratic investment return shock states. The transition probability matrix for the exogenous individual states is given by $P = P^{\pi} \otimes P^{\tau}$, where $\otimes$ is the Kronecker product and P^{π}, P^{τ} are defined in (2.57) and (2.59), respectively. Table 2.7 summarizes the dependence of labor productivity and investment ability on the exogenous individual state s.

Table 2.7: Exogenous individual states.

State (s)	Component		Productivity/ability	
	Permanent	Transitory	Labor (h_s)	Investment (z_s)
1	low	low	0.2620	1.00
2	low	average	0.3642	1.00
3	low	high	0.6016	1.00
4	high	low	1.0547	1.00
5	high	average	1.4659	1.00
6	high	high	2.4215	1.00
7	high-entrepreneur	low	1.0547	1.028
8	high-entrepreneur	average	1.4659	1.028
9	high-entrepreneur	high	2.4215	1.028

2.5.4 Solving the model

To solve the calibrated model, we use the Pareto extrapolation algorithm. We construct a 100-point affine-exponential grid (as described in Section 2.4.3) to approximate the optimal decision rules $c_s(w)$ and $I_s(w)$ as well as the wealth distribution $\Gamma(w,s)$ and use an error tolerance of 10^{-6} in the dynamic programming step. Once the model is solved, we recompute the wealth distribution over a finer grid with 1,000 points to compute top wealth shares. In both cases, we

use a truncation point of $\bar{w} = 10^3 \times K_{\text{RA}}$, where

$$K_{\text{RA}} = ((1/\beta - 1 + \delta)/(A\alpha))^{\frac{1}{\alpha-1}} = 6.2771 \tag{2.60}$$

is the capital stock in the corresponding representative-agent model. Below, we adapt the Pareto extrapolation algorithm in the context of the MBA model.

1. Given a guess of R_f that satisfies the finite capital demand restriction (2.56), solve the problem of the asymptotic agents (2.47), (2.49) and compute the Pareto exponent (2.55) (**asymptotic analysis step**). If a solution to the asymptotic problem exists (which can be checked using the spectral condition (2.48)), verify that aggregate capital supply is finite using (2.52). If there is no solution to the asymptotic problem or if capital supply is infinite, update R_f.

2. Given R_f, compute the capital demand K^d and wage ω implied by profit maximization. Recall that aggregate labor supply is determined only by exogenous parameters and is normalized to $L = 1$. Thus

$$K^d = \left(\frac{\alpha}{R_f - 1 + \delta}\right)^{\frac{1}{1-\alpha}}, \quad \omega = (1-\alpha)\left(\frac{\alpha}{R_f - 1 + \delta}\right)^{\frac{\alpha}{1-\alpha}}.$$

3. Given R_f and ω, solve the individual optimization problem using dynamic programming (**dynamic programming step**, see Appendix B.4 for details), compute the stationary distribution from the law of motion (2.44) (**stationary distribution step**), and compute the aggregate capital supply K^s (2.42) (**aggregation step**).

4. If excess demand $K^d - K^s$ is within error tolerance, stop. Otherwise, update R_f.

Using a machine equipped with an `Intel Xeon E3-1245 3.5GHz` processor and 16GB of memory, the computing time for a guess of R_f is roughly 840 seconds and it takes 9 iterations to find the equilibrium R_f using Matlab's `fzero` function with an error tolerance of 10^{-6}.

2.5.5 Quantitative results

We now discuss the quantitative implications of the calibrated model. Figure 2.1 shows the aggregate capital supply and demand curves for a range of values of R_f. The demand curve is determined by profit maximization of the representative firm, while the supply curve is obtained by aggregating net capital supply from households, either supplied directly to the firm or intermediated by investors (see (2.42)). The intersection of the two curves pins down the risk-free rate that clears the capital market. We obtain a value of 1.0245 (Table 2.8), or 2.45%.

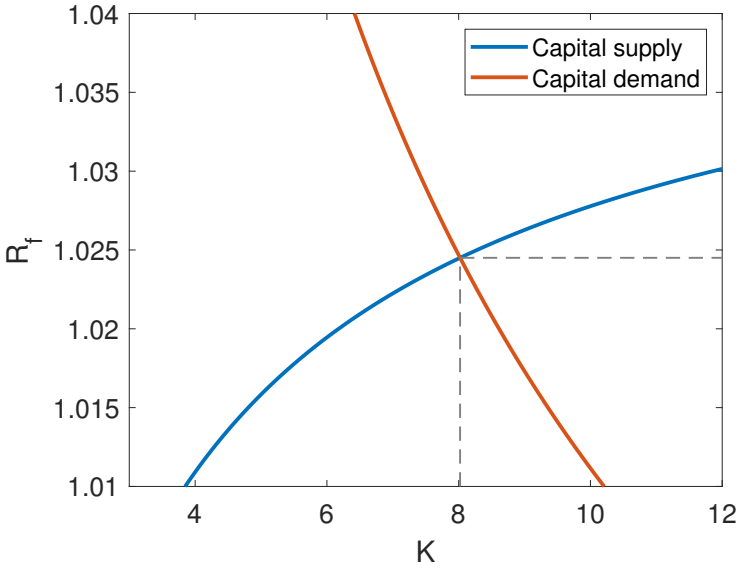

Figure 2.1: Capital demand and supply.

The equilibrium interest rate in turn determines the Pareto exponent ζ of the wealth distribution (Figure 2.2). Higher interest rates R_f are associated with lower Pareto exponents ζ (higher inequality). Intuitively, the "pace" at which rich agents get richer is increasing in the interest rate, so there is more concentration of wealth at the top of the distribution in economies with higher interest rates. We obtain an equilibrium value of $\zeta = 1.69$ (Table 2.8), which is close to but higher than what is estimated for the U.S. (1.52 according to Table 8 of Vermeulen 2018).

While the wealth distribution exhibits a Pareto upper tail, the Pareto exponent ζ does not fully summarize the wealth distribution. For example, the borrowing constraint is an important

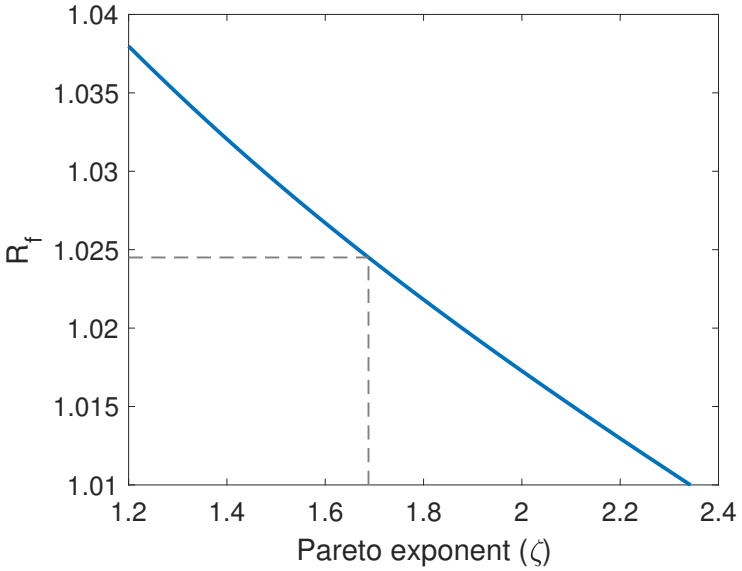

Figure 2.2: Determination of Pareto exponent.

Table 2.8: Equilibrium objects.

Object	Symbol	Value
Capital	K	8.02
Labor	L	1.00
Risk-free rate	$R_f - 1$	2.45%
Wage rate	ω	1.37
Pareto exponent	ζ	1.69

determinant of the wealth share of the poorest 50% of agents, yet ζ depends only on the behavior of rich agents, who are not affected by the borrowing constraint.[16] Table 2.9 presents the wealth shares in the model and in the data (Survey of Consumer Finances).[17] The model qualitatively replicates two important features of the data: the poorest 50% of households hold little wealth (1.85% in the model, 1.79% in the data) while the top 1% accounts for a large share (39.91% in the model, 34.95% in the data).

The last column of Table 2.9 presents the wealth shares associated with a "pure-Pareto"

[16]To be precise, the Pareto exponent ζ can be computed using only the solution to the asymptotic problem (see Section 2.5.2, especially Lemma 2.5.2), which does not depend on the borrowing constraint.

[17]We compute the wealth shares in the data using the Survey of Consumer Finances and average over the survey years 2001, 2004, 2007, 2010, 2013, and 2016.

Table 2.9: Wealth shares (%).

Groups	Model	Data	Pure Pareto
$[0,50)$	1.85	1.79	24.61
$[50,90)$	31.56	25.09	36.27
$[90,99)$	26.68	38.17	23.82
$[99,100]$	39.91	34.95	15.31

distribution with the same exponent ζ as in the model.[18] The Pareto distribution generates a bottom 50% wealth share more than ten times as high as in the model (24.61% versus 1.85%). One drawback of analytical models that imply a "pure-Pareto" wealth distribution such as 2017 is that they do not allow for negative wealth levels, and therefore cannot match the low wealth share of the bottom 50% of agents. In contrast, our model naturally matches both the upper and lower tails of the wealth distribution.

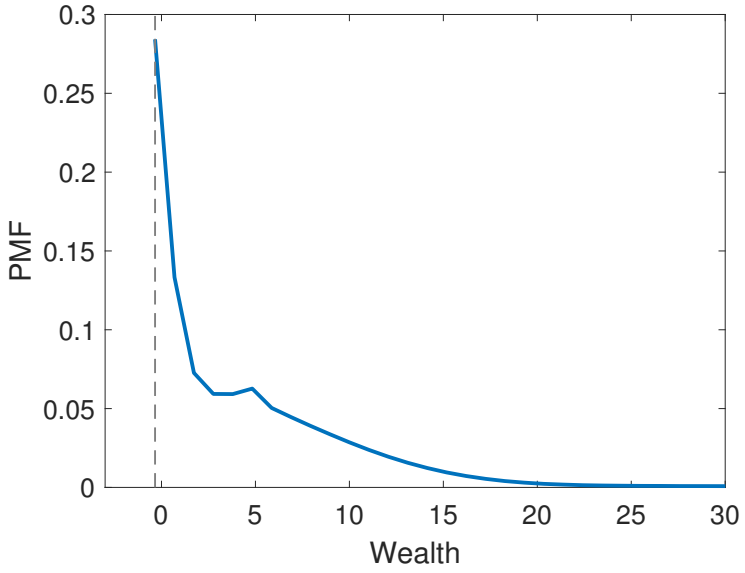

Figure 2.3: Probability mass function.

Figure 2.3 presents the probability mass function (PMF) for wealth levels $[w,30]$. Visually, it would appear that truncating the distribution at $\bar{w} = 30$ is a reasonable choice.[19] While it is true

[18] The cumulative distribution function (CDF) of a Pareto distribution with exponent ζ and minimum size $x > 0$ is $F(x) = 1 - (x/x)^{-\zeta}$ over $[x,\infty)$. Its top wealth shares are independent of x.

[19] The "typical scale" computed as in (2.60) is $K_{\text{RA}} = 6.28$, so a truncation point of 30 represents roughly 5 times the typical scale.

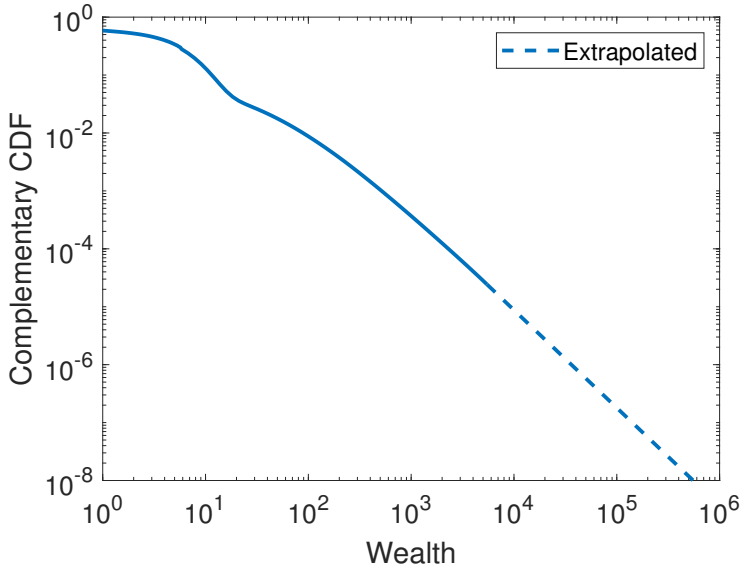

Figure 2.4: Complementary CDF.

that 97.2% of households in the model have wealth levels in the range $[w, 30]$, the remaining 2.8% of households account for 52% of aggregate wealth. Is our choice of $\bar{w} = 10^3 \times K_{\text{RA}}$ appropriate? Table 2.10 reports the MPC relative error (in percentage points) defined by (2.24) for a range of truncation points. With $\bar{w}/K_{\text{RA}} = 10^3$, the MPC relative error is at most 0.0223%, suggesting that it is an appropriate choice.

Table 2.10: Relative error (%) in marginal propensities to consume.

State (s)	Truncation point ($\bar{w}/K_{\text{RA}}$)					
	10^1	10^2	10^3	10^4	10^5	10^6
1	6.2566	0.6470	0.0212	0.0004	0	0
2	6.2487	0.6490	0.0213	0.0004	0	0
3	6.2173	0.6511	0.0213	0.0004	0	0
4	10.1042	0.7505	0.0222	0.0004	0	0
5	9.9959	0.7515	0.0222	0.0004	0	0
6	9.7282	0.7505	0.0223	0.0004	0	0
7	5.1534	0.4597	0.0166	0.0004	0	0
8	4.9042	0.4598	0.0167	0.0004	0	0
9	4.5197	0.4583	0.0168	0.0004	0	0

Note: the individual states $s = 1, \ldots, 9$ are defined in Table 2.7.

Figure 2.4 shows the complementary CDF of the wealth distribution in a log-log scale for the range $[10^0, 10^6]$. As predicted by theory, the upper tail of the distribution appears to converge to the theoretical Pareto slope.

2.5.6 Taxing wealth? A bad idea

We now use the model to quantify the welfare effects of introducing a 1% wealth tax. To do so, we solve the model again with $\tau_w = 0.01$. We assume that any increase (decrease) in tax revenue, which we denote ΔT, is rebated to the household in the form of a (potentially negative) consumption subsidy τ_c. Noting that a typical agent's net capital income is

$$(R_f - 1)(w + (1-\tau_h)\omega h_s - c_s(w)) + R_f(z_{sj} - 1)I_s(w)$$

and $z_{sj}|s = z_s + \varepsilon_j|s = z_s$, the total tax revenue is

$$T = \int \Bigg[\underbrace{\tau_h \omega h_s}_{\text{Labor income tax}} + \underbrace{\tau_k\big((R_f-1)(w+(1-\tau_h)\omega h_s - c_s(w)) + R_f(z_s-1)I_s(w)\big)}_{\text{Capital income tax}}$$
$$+ \underbrace{\tau_w\big((1+(1-\tau_k)(R_f-1))(w+(1-\tau_h)\omega h_s - c_s(w)) + (1-\tau_k)R_f(z_s-1)I_s(w)\big)}_{\text{Wealth tax}} \Bigg] \Gamma(w,s).$$

Due to homothetic preferences, we only need to solve the model without consumption subsidy and then compute aggregate consumption $C := \int c_s(w)\Gamma(w,s)$ as well as the change in tax revenue ΔT. We then compute the consumption tax as

$$\tau_c = -\frac{\Delta T}{C},$$

and the welfare function as

$$\mathcal{W} = (1-\tau_c)\left(\int v_s(w)^{1-\gamma}\Gamma(w,s)\right)^{\frac{1}{1-\gamma}}, \tag{2.61}$$

which is the certainty equivalent of the value function in the stationary equilibrium. Note that is in units of consumption because the Epstein-Zin utility (2.36) is also in units of consumption.[20] Intuitively, the welfare measure (2.61) is the risk-adjusted utility (in units of consumption) of an agent who is randomly thrown into the stationary equilibrium.

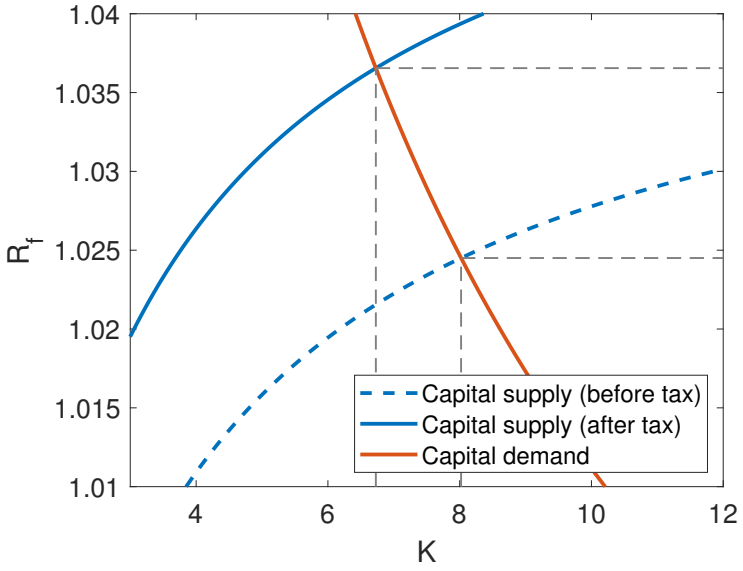

Figure 2.5: Capital demand and supply.

The wealth tax reduces the after tax return on investment, which induces a shift to the left of the capital supply curve (Figure 2.5). Capital demand remains unchanged, so the equilibrium interest rate increases in order to clear the market. The Pareto exponent curve shifts to the right (Figure 2.6), meaning that for a given interest rate, the equilibrium Pareto exponent is larger. Yet, the increase of the interest rate (+1.2 p.p.) is so large that the equilibrium Pareto exponent increases only slightly (from 1.69 to 1.76). Loosely speaking, general equilibrium effects undo most of the effect of the wealth tax on inequality. Overall, we find that the introduction of a wealth tax is a lose-lose policy: wealth inequality remains mostly unchanged, while tax revenue, output, and welfare all decrease. Table 2.11 shows the percentage change in aggregate variables in response to the introduction of the wealth tax. First, notice that capital decreases by 16.13%.

[20] To compute (2.61) numerically, we use the correction term from (2.25) with $\nu = 1-\gamma$ to extrapolate the term $v_s(w)^{1-\gamma}$ off the grid.

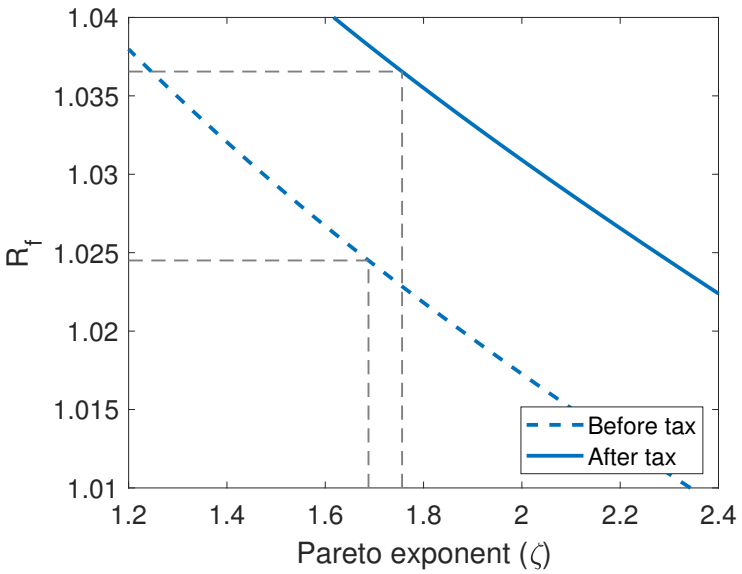

Figure 2.6: Determination of Pareto exponent.

The decrease in the capital stock, combined with the fact that aggregate labor supply is inelastic, leads to a decline in output and wages (-6.47%).

Table 2.11: Equilibrium objects.

Object	Symbol	No Wealth Tax	1% Wealth Tax	Change (%)
Output	Y	2.21	2.06	-6.47
Capital	K	8.02	6.73	-16.13
Labor	L	1.00	1.00	0.00
Risk-free rate	$R_f - 1$	2.45%	3.65%	1.20
Wage rate	ω	1.37	1.28	-6.47
Pareto exponent	ζ	1.69	1.76	4.05
Welfare		0.75	0.69	-7.65

The decline in output alone need not imply a decrease in welfare. For instance, it could be compensated by a reduction in consumption inequality driven by lower wealth concentration. We find that this is not the case. Welfare decreases (in consumption equivalent) by 7.65%. Moreover, wealth inequality remains mostly unchanged. Table 2.12 shows the change in wealth shares. While it is true that the top 1% wealth share declines modestly (from 39.91% to 39.19%), the bottom 50% also declines (from 1.85% to 1.54%).

Table 2.12: Wealth shares (%).

Groups	No Wealth Tax	1% Wealth Tax	Change
$[0,50)$	1.85	1.54	-0.30
$[50,90)$	31.56	32.05	0.49
$[90,99)$	26.68	27.21	0.53
$[99,100]$	39.91	39.19	-0.72

A surprising result is that the decline in output and labor income is so large that total tax revenue actually *declines*. Table 2.13 shows the change in tax revenue by components (labor income tax, capital income tax, and wealth tax). The introduction of the wealth tax provides a new tax revenue of 0.020 (5.51% of initial total tax revenue). Yet, the decline in labor income tax (-6.47%) and capital income tax (-4.82%) dominate, leading to a 0.72% decline in total tax revenue. As a result, the shortfall in tax revenue is compensated by a consumption tax of 0.21% ($\tau_c = 0.0021$), which further contributes to decreasing welfare.

Table 2.13: Tax revenue.

Tax	No Wealth Tax	1% Wealth Tax	Change (%)
Labor income tax	0.306	0.287	-6.47
Capital income tax	0.054	0.051	-4.82
Wealth tax	0	0.020	-
Total	0.360	0.358	-0.72

Figure 2.7 shows the welfare change conditional on wealth and the three persistent types (low, high, high-entrepreneur).[21] Notice that the welfare change is negative at all wealth levels but is larger (in absolute value) for poor households. Given that households with low wealth levels rely mostly on labor income, they are most affected by the decline in wages that operate through general equilibrium effects, and less by the wealth tax. In fact, agents with zero wealth see a welfare decrease of roughly 7% (see table 2.14) which is close to the wage decline of 6.47%. On the other hand the welfare cost declines monotonically with wealth and asymptotes to roughly

[21] We average the value function for each type over the three transitory states using the certainty equivalent.

3%. While the rich bear most of the wealth tax, it is the poor who see the largest welfare losses.[22]

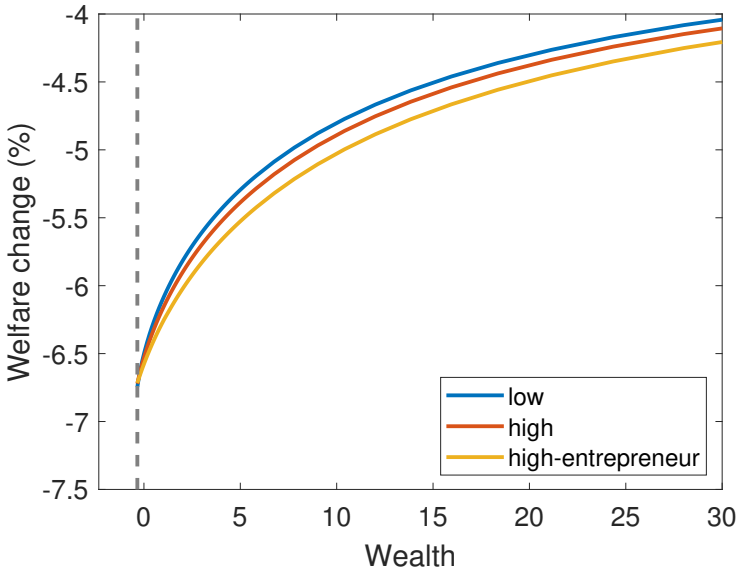

Figure 2.7: Welfare change along the wealth distribution.

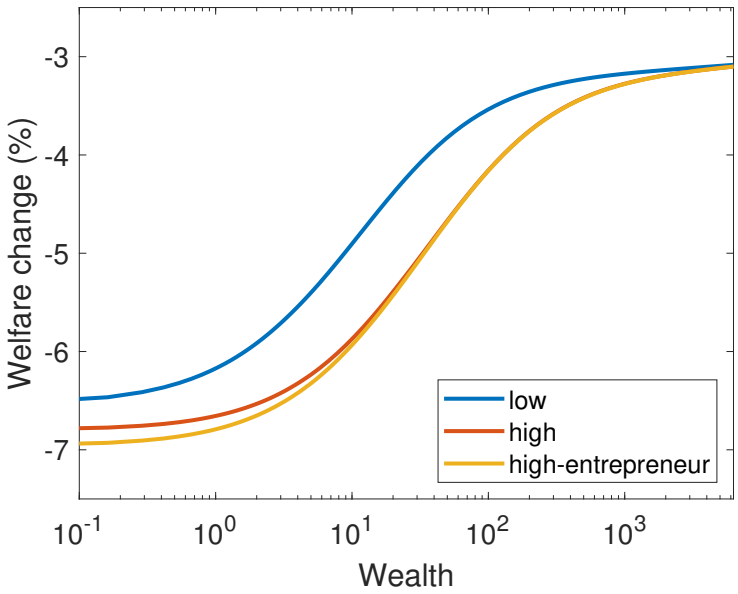

Figure 2.8: Welfare change along the wealth distribution.

The analysis thus far predicts a quantitatively large response of output and welfare to the introduction of a wealth tax. Since we focus on the stationary equilibrium of the model,

[22]The reason why the aggregate welfare loss of 7.65% is larger than those in Figure 2.7 is because the overall wealth distribution shifts to the left, as can be seen in the declining bottom 50% wealth share in Table 2.12.

Table 2.14: Decomposition of welfare change.

Permanent type	Welfare change (%)	
	$w = 0$	$w \to \infty$
Low	-6.72	-3.09
High	-6.86	-3.09
High-entrepreneur	-7.03	-3.09

it is not clear how fast the economy would converge to its new steady state. One way to approximate the speed of convergence is to compute the second largest eigenvalue of the transition probability matrix for wealth Q in the benchmark economy (Rosenthal 1995, Section 4). We find a value of 0.9799, which implies a relatively slow rate of convergence. (The half-life is $\log 0.5 / \log 0.9799 = 34.1$ years.) Therefore, the adverse effects of the wealth tax, which operate partially through general equilibrium effects, might take many decades to materialize. The benefits, however, are immediate (higher tax revenue).

2.6 Concluding remarks

This paper proposes a simple, systematic approach—Pareto extrapolation—to analyze and solve heterogeneous-agent models that endogenously generate fat-tailed wealth distributions. The core insight that we take advantage of is due to Pareto, who noticed that wealth microdata displayed a striking empirical regularity.

> Nous sommes tout de suite frappé du fait que les points ainsi déterminés, ont une tendance très marqué à se disposer en ligne droite.
>
> (We are instantly struck by the fact that the points determined this way have a very marked tendency to be disposed in straight line.)
>
> —Pareto (1897, pp. 304–305)

We put Pareto's insight to work to tackle *models* of wealth inequality. Our approach makes the solution algorithm more transparent, efficient, and accurate with zero additional computational

cost. We leave for future work the implementation of the Pareto extrapolation algorithm along deterministic transition paths as well as in the context of heterogeneous-agent models with aggregate shocks.

2.7 Acknowledgements

Chapter 2, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien; Toda, Alexis Akira. “Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality”. The dissertation author was a primary author of this paper.

Chapter 3

Earnings Dynamics and Inequality Within the Firm

The topic of pay inequality within the firm has attracted a lot attention amongst academics and policymakers.[1] Empirical evidence points to large within-firm inequality (Mueller, Ouimet, and Simintzi 2017b, Song et al. 2018, Bayer and Kuhn 2018) especially between executives and the rest (Frydman and Saks 2010, Edmans, Gabaix, and Jenter 2017). In a context of rising income inequality, is there a case to be made for limiting pay ratios between employees within the same firm? What is the socially optimal level of inequality? To shed light on these questions, this paper develops a model of earnings dynamics and inequality within the firm motivated by novel evidence from personnel data.

Using a proprietary dataset containing personnel data for a sample of Canadian firms over the 2013-2018 period, I provide new evidence on inequality and mobility within the firm. For each firm and year, the dataset contains worker-level information on job title, base pay, and performance bonus. While job titles differ across firms and industries, the dataset contains

[1] For example, on April 10, 2019, the CEOs of the largest seven financial institutions in the U.S. were invited to testify before the House Financial Services Committee and part of the hearing focused on the pay ratio between CEOs and the average employee. The transcript can be found at https://www.npr.org/templates/transcript/transcript.php?storyId=711732570.

14 harmonized "hierarchy levels" that summarize the ranking of job titles within the firm. To motivate my modelling choices, I document three sets of facts. First, most of inequality within the firm is between hierarchy levels rather than within. For example, a regression of log earnings on 14 hierarchy levels yields an R^2 of 0.78. Second, upward mobility within the firm is on average low. In a given year, less than one out of eight employees are promoted to a higher rank. Third, there is a lot of heterogeneity in earnings trajectories. Over a five year period, the median employee experiences a 20% earnings growth, while those at the 95th percentile see their earnings increase by 55%.

The model developed in this paper combines insights from two distinct literatures. First, I model firms as production hierarchies. Workers are assigned to different hierarchy levels and interact only with their direct superiors and subordinates. This view of the production process finds its roots in early theoretical work (see Lucas 1978 and Rosen 1982) and more recently in the "knowledge hierarchy" literature (Garicano 2000, Garicano and Rossi-Hansberg 2006, Caicedo, Lucas Jr, and Rossi-Hansberg 2019). Second, I build on the idea that wage inequality within the firm generates an incentive for workers to exert effort (in order to increase their likelihood of being promoted). This idea can be traced to Lazear and Rosen 1981, who show that a "rank order tournament" compensating scheme—i.e., where two workers compete for a single prize—leads to an efficient level of effort.

The mechanics of the model are as follows. Firms offer labor contracts specifying promotion rates as well as wages for each hierarchy level. Heterogeneous workers join the workforce at the bottom of the job hierarchy and at each period they choose how much effort to exert. Firms observe effort with noise and promote workers who they perceive to be the most productive. As a result, workers face a stochastic process for earnings that is characterized by infrequent but large and persistent jumps in earnings. The model provides a positive theory of these facts and sheds light on the determinants of inequality within the firm.

This paper relates to the theoretical literature on "production hierarchies" (Lucas 1978,

Rosen 1982, Garicano 2000, and Garicano and Rossi-Hansberg 2006). The labor contract that firms offer is a "rank order tournament" scheme as in Lazear and Rosen 1981. Compared to their model, I consider a dynamic environment instead of a one-shot game. Compared to the canonical "job-ladder" model (Burdett and Mortensen 1998), the "job ladder" in my model is within firms rather than between (i.e., workers climb the job ladder by being promoted to a higher ranked job rather than by switching employer).

The paper is also related to the empirical literature that documents an important role of firms in shaping trends in inequality (Song et al. 2018, Barth et al. 2014, Mueller, Ouimet, and Simintzi 2017a). Existing evidence from personnel data on earnings and promotion dynamics includes Gibbons and Waldman 1999 and Kahn and Lange 2014. Existing evidence from administrative data on the relation between hierarchy level and wages include Caliendo, Monte, and Rossi-Hansberg 2015 and Bayer and Kuhn 2018. Existing empirical evidence on heterogeneity in earnings trajectories include Guvenen 2009b and Guvenen et al. 2015.

3.1 Empirical facts

3.1.1 Data

I use a proprietary dataset compiled by Normandin Beaudry, a Canadian a consulting firm that provides actuarial services to pension and savings plans, asset management consulting, group benefits, compensation and other consulting services. The dataset contains de-identified firm-level data covering a sample of Canadian firms over the 2013-2018 period. For each firm and year, I observe detailed compensation and job characteristics information for all of their employees. Worker-level records contains information on job title, base pay, performance bonus, and hours worked. While job titles differ across firms and industries, the dataset contains 14 harmonized "hierarchy levels" that describe the organizational structure of job titles within the firm (see Table 3.1). Appendix C.1 contains a detailed description of the contribution of each

hierarchy level to the production process of a typical firm.

Table 3.1: Organizational hierarchy and job titles.

Generic hierarchy level		Detailed hierarchy level		# Observations
Code	Title	Code	Title	
1	Management	110	President	93
		111	Vice-President	13,13
		112	Director	9,602
		113	Manager	14,026
		114	Supervisor	7,184
2	Professional	211	Professional Expert	8,341
		212	Professional III	35,463
		213	Professional II	25,769
		214	Professional I	11,646
3	Technical	311	Technical III	14,518
		312	Technical II	21,124
		313	Technical I	14,765
4	Support	411	Support II	14,948
		412	Support I	15,042

Note: more information on hierarchy levels are available in Appendix C.1.

Restricting the sample to private sector firms with at least 20 employees I obtain an employer-employee matched panel dataset containing 205 unique firms and 97,345 unique employees (see Table 3.3). Longitudinal coverage is uneven and ranges from one to five years (see Table 3.1). Compared to administrative employer-employee matched dataset, I can not observe employees who switch employers, the reason being that the employee identifiers are firm-specific. The main measure of earnings that I will consider is "annual base salary". Note that when an employee works only for part of the year, the base salary in the dataset is "annualized".

Table 3.2: Number of observations by year.

Year	2013	2014	2015	2016	2017	2018	Total	Unique IDs
Employees	24,962	26,120	30,957	34,501	34,423	42,871	193,834	97,345
Firms	77	73	73	85	95	122	525	205

Table 3.3: Number of longitudinal links.

Horizon	1 year	2 years	3 years	4 years	5 years
Employees	18,545	10,921	6,587	5,218	4,995
Firms	39	17	16	14	32

To complement the analysis, I use microdata from the Canadian Labour Force Survey (henceforth LFS) for the 2013-2018 period. The LFS is a large monthly survey that is representative of the Canadian population. For comparability, I restrict the sample to private sector full-time workers at firms with at least 20 employees. I compute annual earnings by multiplying usual hours worked by hourly earnings.

3.1.2 Variance decomposition

Figure 3.1 plots the distribution of earnings for the year 2018. The top 1% of employees earn roughly 200 K$ (Canadian dollars) while the bottom 1% earn roughly 30K$. Notice that for the top percentiles, the relationship between tail probability and earnings appears to be linear on a logarithmic scale, which is consistent with the distribution of earnings having a Pareto upper tail.

I now summarize the distribution of log earnings. To make observations comparable across years, I first remove year fixed effects. The first row of Table 3.4 contains two measures of inequality: the variance and the Pareto exponent. The variance (0.16) measures the dispersion of log earnings around the mean. The Pareto exponent measures the shape of the upper tail of the earnings distribution. A low Pareto exponent implies a high level of "top earnings inequality" (i.e., inequality between high earners). I estimate a Pareto exponent of 3.8 using the maximum likelihood estimator on the top 25% of observations.[2] In Appendix C.2, I compute the same summary statistics using survey data and obtain very similar results, suggesting that the personnel data is representative of the Canadian population in terms of earnings inequality.

[2]Let $\{y_i\}_{i=1}^N$ denote the top 25% of earnings observations. The maximum likelihood estimator is $\left(\frac{1}{N}\sum_{i=1}^N \log y_i - \log y_{P70}\right)^{-1}$, where $y_{P75} \equiv \min\{y_i\}_{i=1}^N$ is the 75th percentile of y.

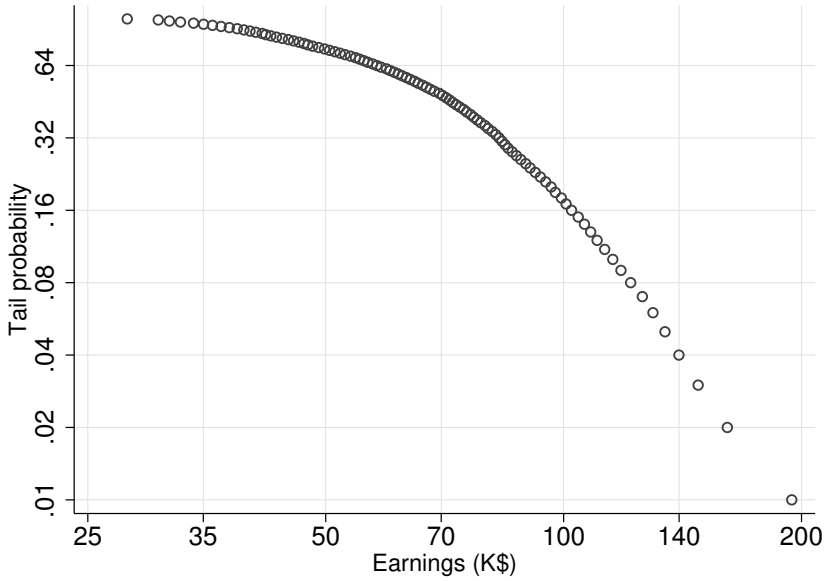

Figure 3.1: Percentiles of the distribution of earnings against tail probability (2018).

Table 3.4: Variance decomposition of log earnings.

Fixed effects	K	R^2	Properties of residuals	
			Variance	Pareto exp.
None	-	-	0.16	3.8
Firm	205	0.31	0.11	4.1
Hierarchy level	14	0.78	0.04	9.3
Firm and hierarchy level	219	0.85	0.03	10.2

Notes: The sample size is $N = 193,834$. "K" is the number of fixed effects; "Pareto exponent" is the maximum-likelihood estimator of the Pareto exponent using the largest 25% observations.

Firm fixed effects explain 31% of the variance of log earnings, meaning that 69% of the variance of log earnings is within firm (see second row of Table 3.4). Song et al. 2018 obtain a similar result using the universe of U.S. tax returns. For the year 2013, they find that 59% of the variance of log earnings is within firm. Notice that while firm fixed effects explain a significant fraction of the variance, the Pareto exponent estimated using residual earnings remains mostly unchanged (4.1 versus 3.8). This result highlights the fact that top earnings inequality is not driven by differences in average earnings between firms, but rather by differences in earnings

amongst workers within the same firm.

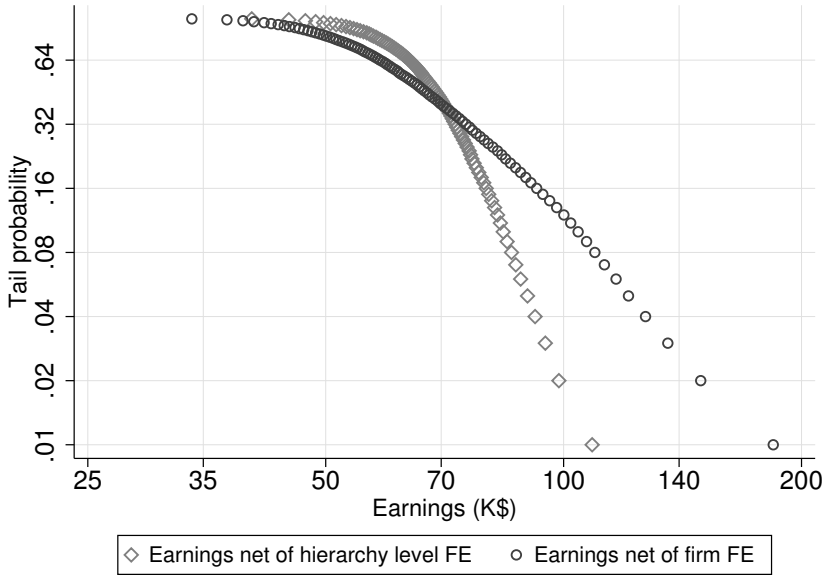

Figure 3.2: Distribution of earnings net of firm and hierarchy level fixed effects (2018)

Notes: The distribution of earnings net of firm (hierarchy level) fixed effects is computed by (i) regressing log earnings on firm (hierarchy level) fixed effects, (ii) exponentiating the sum of the residuals and the unconditional mean of log earnings.

Using the 14 harmonized job categories defined in Table 3.1 (henceforth "hierarchy levels"), I estimate a fixed effects regression. Hierarchy level fixed effects explain 78% of the variance of log earnings (see third row of Table 3.4). This is a striking result. For comparison, I estimate a "Mincer regression" using the LFS data (i.e., log earnings are regressed on education and experience) and find that it explains merely 16% of the variation (see Appendix C.2). Adding firm fixed effects in addition to hierarchy level fixed effects increase the R^2 by only 0.07, from 0.78 to 0.85. Notice that the distribution of log earnings net of hierarchy level fixed effect has Pareto exponent of 9.3, meaning that there is much less top earnings inequality within hierarchy levels. Figure 3.2 plots the distribution of earnings net of firm and hierarchy level fixed effects. Visually, the distribution of earnings net of hierarchy level fixed effects is much more compressed than the distribution of earnings net of firm fixed effects, especially the upper tail. The fact that

hierarchy level fixed effects explain so much of the variation of log earnings thus appears to be that they can account for the thick upper tail of the earnings distribution.

3.1.3 Mobility within the earnings distribution

How much mobility is there within earnings distribution? If there is a lot of within-firm mobility, then static measures of inequality will overstate the amount of "lifetime earnings inequality". For example, if employees join a firm at the bottom of the hierarchy and are then deterministically promoted to higher ranks as their tenure increases, then inequality within the firm can be quite benign. In contrast, if there is no mobility across hierarchy levels, then static measures of inequality will capture persistent differences in earnings across employees.

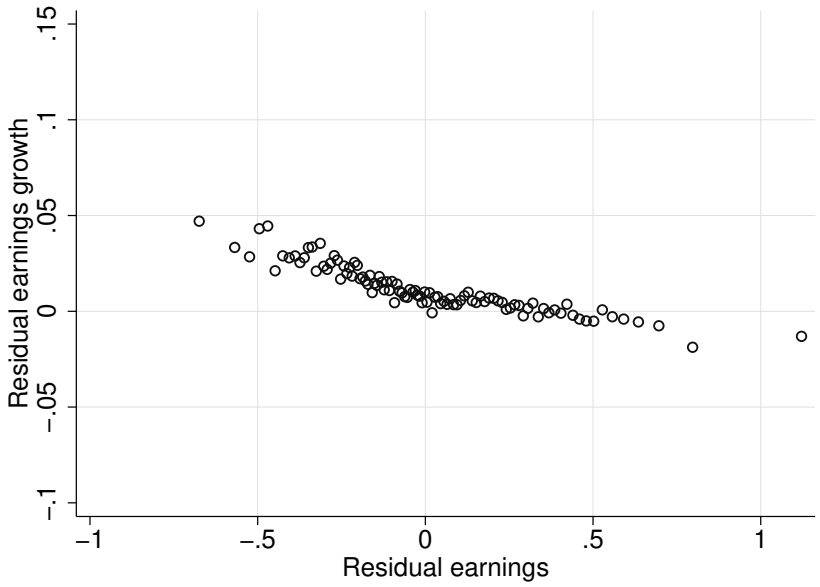

Figure 3.3: Growth of residual earnings by percentile.

Notes: residual earnings are obtained as the residuals of a regression of (log) earnings on firm and year fixed effects.

Figure 3.3 contains a "binned scatterplot"[3] of residual earnings growth (i.e., forward

[3] A binned scatter plot is a scatter plot where the variable on the x axis is binned into quantiles. Each point therefore represent the average value of (x,y) conditional on x being in a particular quantile.

difference) against level. Residual earnings are defined as the residual in the regression of log earnings on firm and year fixed effects (as in the second row of Table 3.4). The residual earning of an employee is therefore the "log difference" between her earnings and the average earning of her colleagues. Notice that there is a negative relationship between residual earnings growth and level. This implies that employees with below average earnings will tend to "move up" in the firm earnings distribution faster than those with above average earnings. To quantify the relationship, I estimate the following regression

$$y^{\text{res}}_{i,t+1} - y^{\text{res}}_{i,t} = \mu + \rho y^{\text{res}}_{i,t} + u_{i,t}, \tag{3.1}$$

where $y^{\text{res}}_{i,t}$ is the residual earning of employee i in year t. I report the results in Table 3.5.

Table 3.5: Convergence of (within-firm) residual earnings.

	Symbol	Point estimate	Standard error
Constant	μ	0.012	0.0002
AR(1) coefficient	ρ	-0.035	0.0011
Half-life	$\log(1/2)/\log(1+\rho)$	19.49	0.65

Notes: Standard error are clustered at the employee-level and the Delta method is used for the half-life.

The coefficient $\widehat{\mu}$ is positive, which means that the average worker (i.e., $y^{res} = 0$) tends to move up in the earnings distribution from one year to the next. The coefficient $\widehat{\rho}$ is negative, which means that the average change in residual earnings is higher for lower paid employees. Equipped with the estimate for ρ, I can compute the "half-life" of earnings differentials amongst continuing workers. The thought experiment is the following. Imagine sampling two workers within a firm and trying to predict how many years it will take until their earnings differential shrinks by half. Using Equation 3.1 and the estimator of ρ, the answer is that it will take $\log(1/2)/\log(1-\widehat{\rho}) \approx$

19.5 years.[4] Given that a typical worker spends 30 to 40 years in the workforce, a half-life of earnings differentials of almost 20 years indicates a very high persistence of earnings differentials.

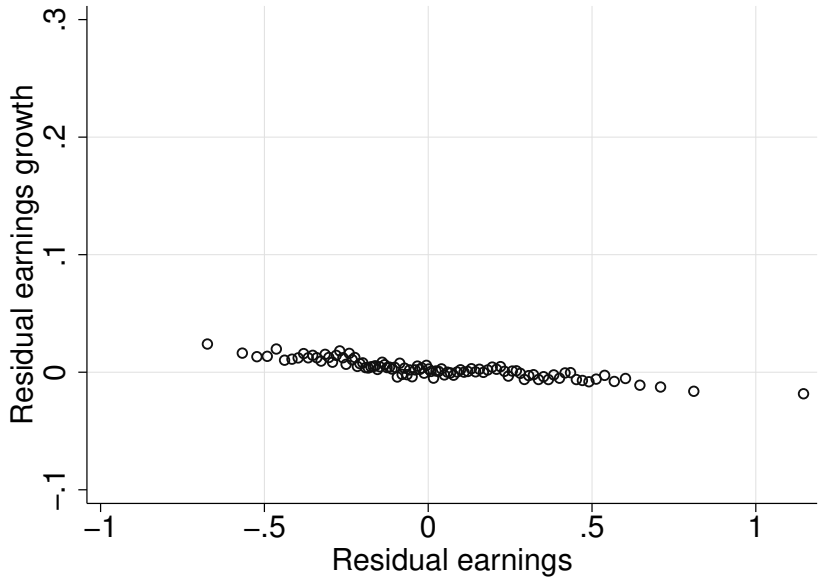

Figure 3.4: Growth of residual earnings by percentile and promotion status (no promotion, $p = 0.89$).

Notes: residual earnings are obtained as the residuals of a regression of (log) earnings on firm and year fixed effects.

In the previous section, I have established that earnings within the firm are mostly determined by the job hierarchy. Therefore, we should expect earnings mobility to occur mostly through promotions. In the sample, the annual promotion rate is $p = 0.11$.[5] Figures 3.4 and 3.5 contain a binned scatterplot of residual earnings growth by initial level and promotion status. Notice that promotions are associated with large increases in residual earnings, especially for workers with low initial residual earnings. In contrast, residual earnings growth tends to be very low and predictable for employees in years where they are are not promoted. Figures 3.6 and 3.7

[4]From Equation 3.1, we have that $\mathbb{E}_{i,j,t}(y^{\text{res}}_{i,t+h} - y^{\text{res}}_{j,t+h}) = (1+\rho)^h(y^{\text{res}}_{i,t} - y^{\text{res}}_{j,t})$. Therefore

$$\frac{\mathbb{E}_{i,j,t}(y^{\text{res}}_{i,t+h} - y^{\text{res}}_{j,t+h})}{(y^{\text{res}}_{i,t} - y^{\text{res}}_{j,t})} = \frac{1}{2} \iff h = \frac{\log(1/2)}{\log(1+\rho)}.$$

[5]A promotion is defined as a move to a higher rank in the job hierarchy, where jobs are ranked as in Table 3.1.

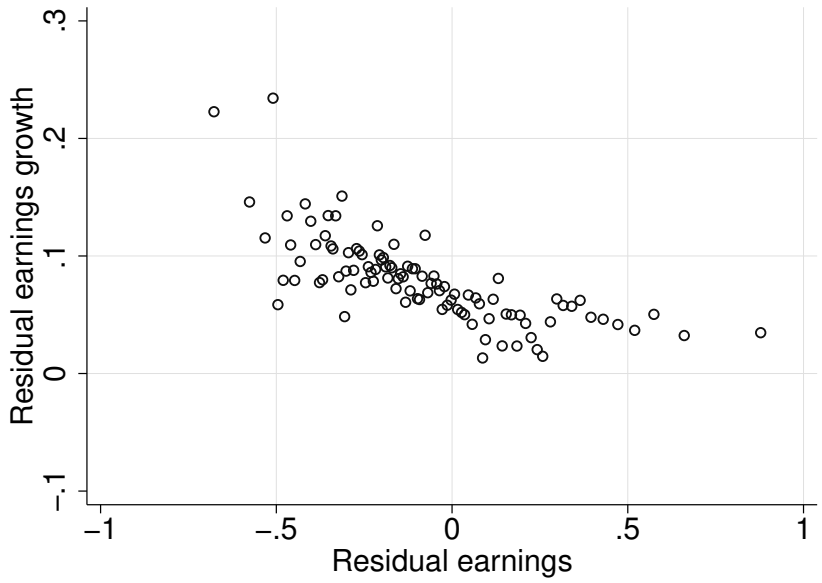

Figure 3.5: Growth of residual earnings by percentile and promotion status (promotion, $p = 0.11$).

Notes: residual earnings are obtained as the residuals of a regression of (log) earnings on firm and year fixed effects.

presents the distribution of nominal earnings growth by promotion status. For employees who are not promoted, more than half experience a nominal earnings growth of 2 or 3 percentage points. In contrast, almost half of promoted employees obtain an increase of at least 10 percentage points. Absent promotions, earnings growth tends to be barely more than inflation.[6] Upward mobility within the firm thus mostly operates through promotions.

3.1.4 Heterogeneity in earnings trajectories

While it is true that *on average*, there is little upward mobility within the firm, a small fraction of employees experience fast earnings growth. Figure 3.8 plots percentiles of the distribution of cumulative earnings growth over a five year horizon. The median of the one-year-ahead earnings growth distribution is 3.2%, yet the 95th percentile is 14.9%. Over a five

[6]Over the 2013-2018 period, annual CPI inflation averaged 1.7% in Canada.

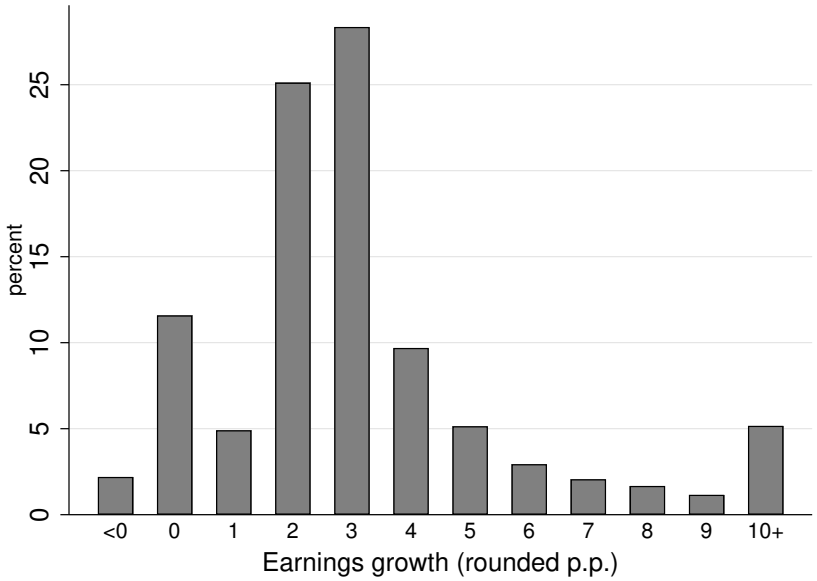

Figure 3.6: Distribution of earnings change by promotion status (no promotion, $p = 0.89$).

Notes: residual earnings are obtained as the residuals of a regression of (log) earnings on firm and year fixed effects.

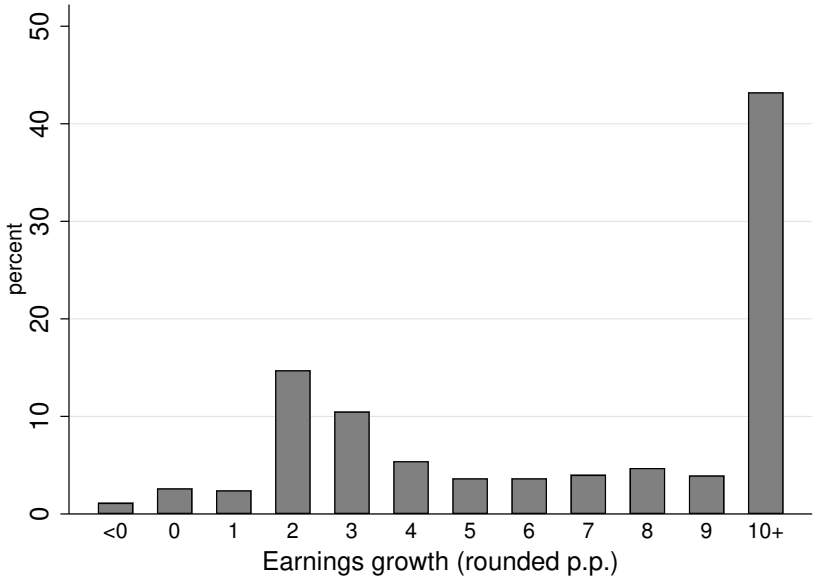

Figure 3.7: Distribution of earnings change by promotion status (promotion, $p = 0.11$).

Notes: residual earnings are obtained as the residuals of a regression of (log) earnings on firm and year fixed effects.

year horizon, the distribution of cumulative earnings growth is even more dispersed: the median earnings growth is 19.5%, yet employees at the 95th percentile see their earnings increase by 54.5%.

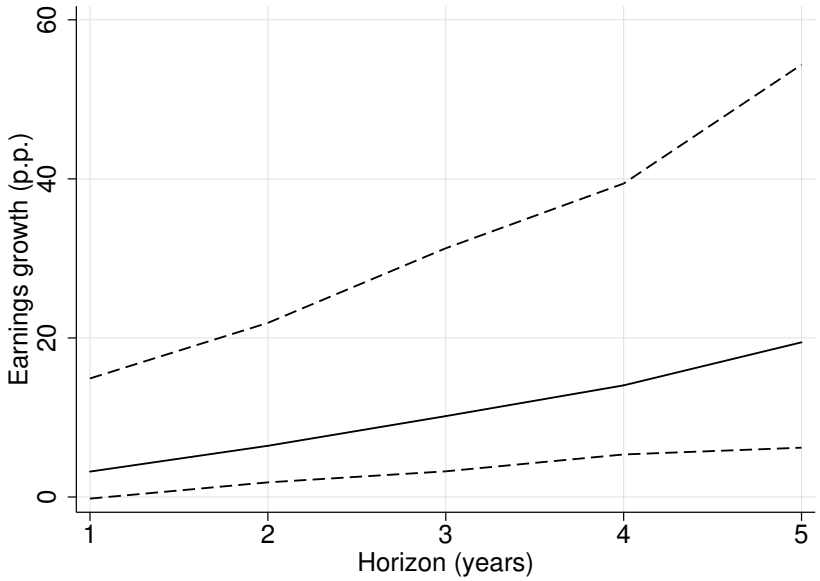

Figure 3.8: Distribution of earnings growth.

Notes: solid line represents the median; dashed lines represent the 5th and 95th percentiles.

Is the heterogeneity in cumulative earnings growth rates driven by the fact that employees at the bottom of the hierarchy experience faster earnings growth? Perhaps surprisingly, I find that this is not the case. Figure 3.9 through 3.12 plot percentiles of the earnings growth distribution by initial hierarchy level ("Management", "Professional", "Technician", "Support") as defined in Table 3.1. Within each group, the same pattern emerges: while the median employee experiences little earnings growth, some "superstar" employees experience fast earnings growth.

The fact that the distribution of earnings growth "fans out" with the horizon suggests that earnings shocks are very persistent. To estimate the dynamic earnings response to a promotion, consider the following regression model

$$\log w_{i,t+h} = \alpha_f^h + \gamma_j^h + \mu_t^h + \beta^h P_{i,t} + \rho^h \log w_{i,t} + u_{i,t}^h, \tag{3.2}$$

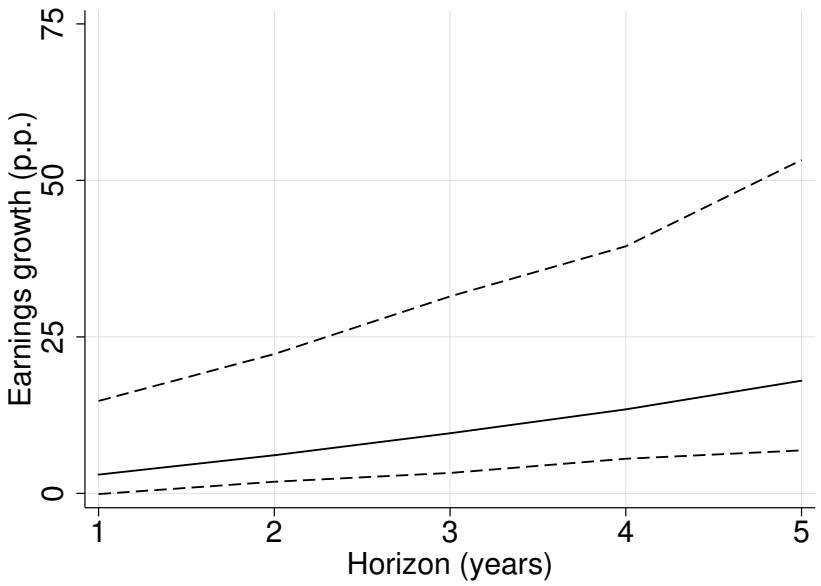

Figure 3.9: Distribution of earnings growth (manager).

Notes: the solid line represents the median; the dashed lines represent the 5th and 95th percentiles;

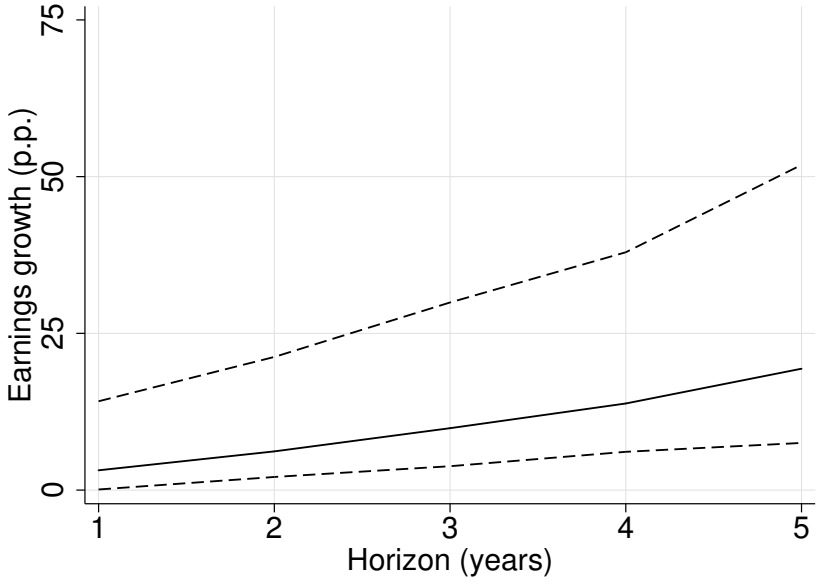

Figure 3.10: Distribution of earnings growth (professional).

Notes: the solid line represents the median; the dashed lines represent the 5th and 95th percentiles;

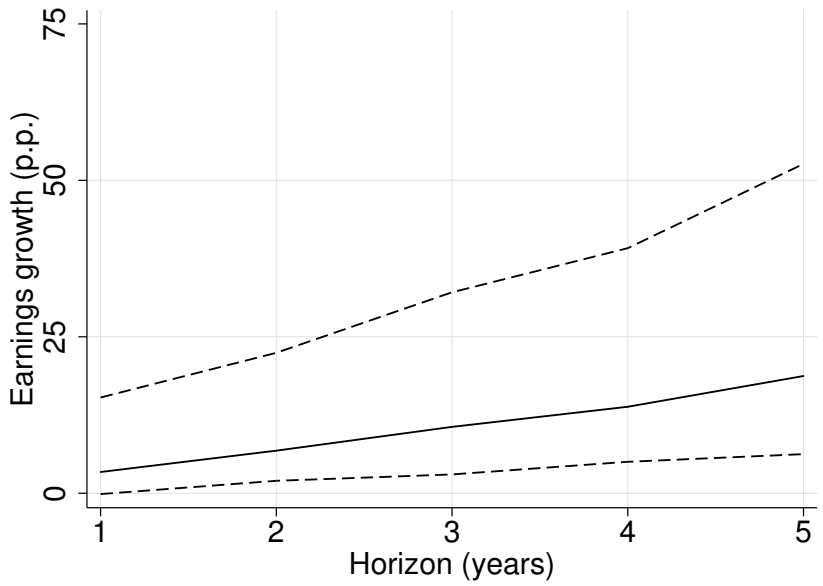

Figure 3.11: Distribution of earnings growth (technician).

Notes: the solid line represents the median; the dashed lines represent the 5th and 95th percentiles;

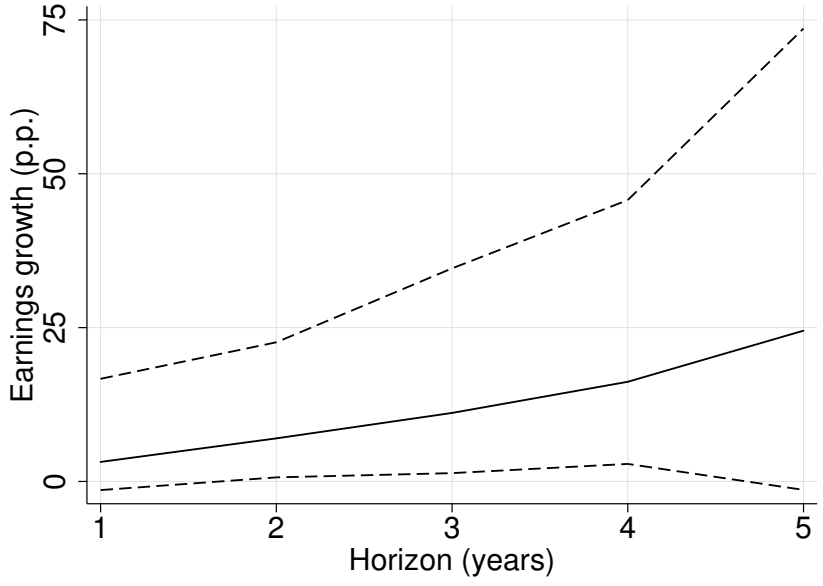

Figure 3.12: Distribution of earnings growth (support).

Notes: the solid line represents the median; the dashed lines represent the 5th and 95th percentiles;

where α_f^h, γ_j^h, and μ_t^h are respectively firm, hierarchy level and year fixed effects. The variable $w_{i,t}$ denotes earnings and $P_{i,t}$ is a promotion dummy for employee i in year t. The estimated h-year ahead response of earnings (in log points) is therfore $\widehat{\beta}^h$. Figure 3.13 plots the dynamic response with 95% confidence intervals for $h = -1, 0, \ldots, 5$. The standard errors are clustered at the employee level. A few remarks are in order. First, the effect of a promotion on earnings is large on impact. Compared to a "control group" of employees within the same firm, in the same year, and at the same hierarchy level, the earnings of promoted employees is roughly 8% higher on impact. Second, the effect of a promotion on earnings appears to be permanent. Five years after the promotion, promoted employees still earn roughly 8% more than their peers.

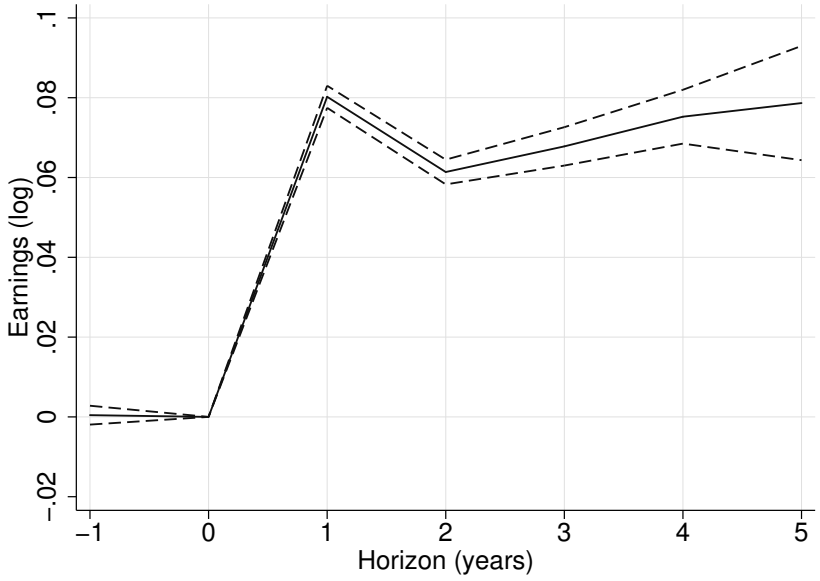

Figure 3.13: Dynamic effect of a promotion on earnings.

3.2 Model

3.2.1 Environment

The economy is populated by a unit measure of workers who differ in productivity $z \in \{z_1, \ldots, z_S\}$, where $\bar{\phi}_s \equiv \mathbb{P}(z = z_s)$. Without loss of generality, assume that $0 < z_1 < \cdots < z_S$. Workers derive utility from a bundle of wage and effort (w, a) according to a twice continuously differentiable utility function $u(w, a)$. Time is continuous and the output of a worker of productivity z exerting effort $a \geq 0$ is given by

$$y = za, \tag{3.3}$$

Workers are assigned to different jobs labeled $k \geq 0$ where a higher k denotes a job that is higher in the hierarchy. The production function is

$$Y = F\big(\{y_{s,k}, N_{s,k}\}\big), \tag{3.4}$$

where $N_{s,k}$ and denotes the measure of workers of type s in job k and $y_{s,k}$ denotes the output of workers of type s in job k. I assume that F exhibits constant returns to scale in employment

$$\forall\, \lambda > 0, \quad F\big(\{y_{s,k}, \lambda N_{s,k}\}\big) = \lambda F\big(\{y_{s,k}, N_{s,k}\}\big). \tag{3.5}$$

Firms offer labor contracts that determine job-specific wages $\{w_k\}$ and promotion rates $\{q_k\}$. The contract is implemented in the following manner. At every instant, the firm randomly selects a measure $q_k N_k$ of worker pairs in job k and, for each pair, promotes the worker with the highest "observed" output to job $k+1$. I assume that firms observe the true level of output y with a multiplicative error $\varepsilon \sim G$. The rate at which a worker with output y obtains a promotion is therefore

$$Q_k(y) \equiv 2 \times q_k \times \mathbb{P}(\varepsilon y > \varepsilon' y'), \tag{3.6}$$

where $(\varepsilon, \varepsilon')$ are independent draws from G while y' is drawn from the endogenously-determined distribution of worker output in job k. Promoting employees leads to training costs governed by a cost function $C(\{q_k, N_k\})$. I assume that C exhibits constant returns to scale in employment,

$$\forall\, \lambda > 0, \quad C\big(\{q_k, \lambda N_k\}\big) = \lambda C\big(\{q_k, N_k\}\big). \tag{3.7}$$

Workers exogenously retire at rate $\psi \geq 0$ and I normalize the continuation value of being retired to zero. Retired workers are replaced by labor market entrants who start at $k = 0$. Denote by $v_{s,k}$ the continuation value for a worker of type s in job k. The Bellman equation is given by

$$(r + \psi) v_{s,k} = \max_{a \geq 0} \Big\{ u(w_k, a) + Q_k(z_s a)(v_{s,k+1} - v_{s,k}) \Big\}, \tag{3.8}$$

Interior solutions $a_{s,k}$ for effort satisfy the following first-order condition

$$u_a\big(w_k, a\big) + Q_k'\big(z_s a\big) z_s (v_{s,k+1} - v_{s,k}) = 0. \tag{3.9}$$

From now on, denote the endogenously-determined promotion rate of a worker of type s in job k by

$$q_{s,k} \equiv Q_k\big(z_s a_{s,k}\big). \tag{3.10}$$

The law of motion for the measure of type s in job k, which I denote by $N_{s,k}$, is given by

$$\dot{N}_{s,0} = \psi \bar{\phi}_s N - (\psi + q_{s,0}) N_{s,0}, \tag{3.11}$$

$$\dot{N}_{s,k} = q_{s,k-1} N_{s,k-1} - (\psi + q_{s,k}) N_{s,k} \quad \forall k \geq 1, \tag{3.12}$$

where $N \in [0, 1]$ denotes the measure of labor market entrants that the firm hires (and incidentally the long-run employment of the firm). Similarly, the law of motion for the measure of workers in

job k, which I denote by N_k is

$$\dot{N}_0 = \psi N - (\psi + q_0 N_0), \tag{3.13}$$

$$\dot{N}_k = q_{k-1} N_{k-1} - (\psi + q_k) N_k \quad \forall k \geq 1. \tag{3.14}$$

Both sets of equations merely account for the inflow and outflow of workers due to promotions and retirements.

Lemma 4. *The stationary distributions are given by*

$$N_{s,k} = \left(\prod_{n=1}^{k} \frac{q_{s,k-1}}{\psi + q_{s,k}} \right) \frac{\psi}{\psi + q_{0,s}} \bar{\phi}_s N \tag{3.15}$$

$$N_k = \left(\prod_{n=1}^{k} \frac{q_{k-1}}{\psi + q_k} \right) \frac{\psi}{\psi + q_0} N \tag{3.16}$$

Proof. Equations (3.15) and (3.16) are obtained by (i) solving for the initial condition initial conditions in equations (3.11) and (3.13), (ii) iterating forward using equations (3.12) and (3.14). □

I assume that firms maximize long-run flow profits. The firm problem is thus given by

$$\max_{\{w_k, q_k\}_{k \geq 0}} F\left(\{y_{s,k}, N_{s,k}\}\right) - \sum_{k \geq 0} \sum_{s=1}^{S} w_k p_{s,k} - C(\{q_k, N_k\}) \tag{3.17}$$

$$\text{s.t.} \quad \sum_{s=1}^{S} \bar{\phi}_s v_{s,0} \geq \mathcal{V}. \tag{3.18}$$

I denote by $\mathcal{V}$ the minimal value that a firm must offer to its incoming workers in order to be able to attract a positive measure of workers. Intuitively, to be able to attract workers, the firm must offer a labor contract that is at least as good what the "most competitive" firm is offering. I assume that workers choose which firm to join before observing their type s, so the expected annuity value of joining a firm is $\sum_{s=1}^{S} \bar{\phi}_s v_{s,0}$.

3.2.2 Equilibrium

Assumptions (3.5) and (3.7) imply that the firm problem is homogeneous in employment N. This in turn implies that the equilibrium distribution of employment across firm is indeterminate. I therefore consider equilibria where a single firm employs all workers (henceforth the "representative firm"). An equilibrium consists of a labor contract $\{w_k, q_k\}$, worker effort levels $\{a_{s,k}\}$, a promotion rate function $\{q_{s,k}\}$, employment shares $\{N_{s,k}\}$ and $\{N_k\}$, and a value $\mathcal{V}$ such that:

1. the effort function $\{a_{sk}\}$ solves the worker problem;
2. the labor contract $\{w_k, p_k\}$ solves the firm problem;
3. the promotion rates $\{q_{sk}\}$ satisfy (3.10);
4. the employment shares $\{N_{s,k}\}$ and $\{N_k\}$ are given by (3.15) and (3.16);
5. due to competition, the value $\mathcal{V}$ implies zero profits.

Notice that $\{N_{s,k}\}$ and $\{N_k\}$ have the interpretation of employment *shares* since $N = 1$.

3.2.3 Numerical example

Suppose that workers are ex-ante identical with $z = 1$, which implies that $y = a$. The utility function is

$$u(w,a) = \log w - \beta a, \tag{3.19}$$

and the distribution (CDF) of the output measurement error is Pareto with exponent $\frac{1}{\sigma} > 1$

$$G(\varepsilon) = 1 - \varepsilon^{-\frac{1}{\sigma}} \quad \forall\, \varepsilon \geq 1. \tag{3.20}$$

Using the definition of the promotion rate (3.6), I obtain

$$Q_k(a) = q_k \times a^{\frac{1}{\sigma}} \times \mathbb{E}[(a')^{-\frac{1}{\sigma}}], \tag{3.21}$$

which implies that the promotion rate is *isoelastic* in effort. Finally, the production function is given by

$$F\left(\{a_k, N_k\}_{k\geq 1}\right) = \sum_{k\geq 1} (a_k N_k)^{1-\theta} (a_{k-1} N_{k-1})^{\theta}. \tag{3.22}$$

The parameter $\theta \in (0,1)$ governs the diminishing return in managing and is reminiscent of the "span of control" parameter in Lucas 1978. Labor productivity in job k is given by

$$\frac{(a_k N_k)^{1-\theta} (a_{k-1} N_{k-1})^{\theta}}{N_k} = a_k \left(\frac{a_{k-1} N_{k-1}}{a_k N_k}\right)^{\theta}.$$

Therefore, labor productivity is increasing in effort a_k and in the relative output between two layers (i.e., the ratio of $a_{k-1}N_{k-1}$ and $a_k N_k$). I assume that the cost function has the form

$$C(\{q_k, N_k\}) = \kappa \sum_{k\geq 0} N_k q_k, \tag{3.23}$$

which implies that each promotion incurs a cost κ.

The worker's Bellman equation is

$$(r+\psi) v_k = \max_a \{\log w_k - \beta a + Q_k(a) v_k\} \tag{3.24}$$

and the optimal effort is

$$a_k = \frac{q_k v_k}{\sigma \beta}, \tag{3.25}$$

where $Q_k(a_k) = q_k$ in equilibrium since workers are identical. Substituting the optimal level of effort in the Bellman equation and re-arranging, the value function can be expressed as the

forward solution to

$$\left(r+\psi+\frac{\sigma-1}{\sigma}q_k\right)v_k=\log w_k+\frac{\sigma-1}{\sigma}q_k v_{k+1}, \tag{3.26}$$

which highlights the fact that the effort of a worker in layer k depends on the full sequence of wages $\{w_k, w_{k+1}, \dots\}$ and promotion rates $\{q_k, q_{k+1}, \dots\}$.

The firm problem is thus given by

$$\max_{\{w_k,N_k\}_{k\geq 0}} \sum_{k\geq 1}(a_k N_k)^{1-\theta}(a_{k-1}N_{k-1})^{\theta}-\sum_{k\geq 0}w_k N_k-\kappa\sum_{k\geq 0}N_k q_k \tag{3.27}$$

$$\text{s.t.} \quad v_0\geq \mathcal{V}. \tag{3.28}$$

I restrict the contract space to

$$\mathcal{C}=\left\{\{w_k,q_k\}_{k\geq 0}\;\middle|\;\forall k\geq 0,\ w_k=w_0e^{\pi k},\ \pi\geq 0,\ q_k=q\right\}, \tag{3.29}$$

which considerably reduces the complexity of the problem.

Lemma 5. *Whenever $\{w_k,q_k\}\in\mathcal{C}$ and $\frac{q}{q+\psi}e^{\pi}<1$, the following equations hold in equilibrium*

$$N_k=\frac{\psi}{\psi+q}\left(\frac{q}{\psi+q}\right)^k, \tag{3.30}$$

$$\sum_{k\geq 0}w_k N_k=w_0\frac{\psi}{\psi-q(e^{\pi}-1)}, \tag{3.31}$$

$$a=\frac{q\pi}{(r+\psi)\sigma\beta}, \tag{3.32}$$

$$\sum_{k\geq 1}(a_k N_k)^{1-\theta}(a_{k-1}N_{k-1})^{\theta}=a\left(\frac{q}{\psi+q}\right)^{1-\theta}. \tag{3.33}$$

Moreover, the constraint can be expressed as

$$w_0\geq e^{(r+\psi)\mathcal{V}-\frac{\sigma-1}{\sigma}\frac{q\pi}{r+\psi}}. \tag{3.34}$$

Proof. See section C.3.1 of the Appendix. □

Using results from Lemma 5, the firm problem can be written as

$$\max_{w_0,\pi,q} \frac{q\pi}{(r+\psi)\sigma\beta} \times \left(\frac{q}{\psi+q}\right)^{1-\theta} - w_0 \times \frac{\psi}{\psi - q(e^{\pi}-1)} - \kappa q \tag{3.35}$$

$$\text{s.t.} \quad w_0 \geq e^{(r+\psi)\mathcal{V} - \frac{\sigma-1}{\sigma}\frac{q\pi}{r+\psi}}. \tag{3.36}$$

To compute the equilibrium, I first solve the solution to the firm problem (w_0, π, q) for a guess of $\mathcal{V}$ and then update $\mathcal{V}$ until equilibrium profits are zero.

As a first step, I set the retirement ψ to 0.025, which implies and average worklife of 40 years. I also set the discount rate to $r = 0.02$. As a proof of concept, I set the parameters (σ, κ) to $(1/3, 5/4)$. The remaining parameters β (cost of effort) and θ (span of control) are chosen as to imply a promotion rate of 11% and an earnings growth conditional on promotion of 8%.

Table 3.6: Calibrated parameters.

Parameter	Symbol	Value	Target
Retirement rate	ψ	0.025	average work-life of 40 years
Rate of time preference	r	0.02	real interest rate of 2%

Figure 3.14 plots the equilibrium wage schedule $\{w_k\}$ for $k = 0, \ldots, 10$. The wage differential between successive layers is 8%, which is direct result of the calibration strategy. Figure 3.15 plots the equilibrium employment shares $\{N_k\}$ for $k = 0, \ldots, 10$. Each hierarchy level employs 81% as many workers as the previous one.

Figure 3.16 plots the wage distribution for the first 20 hierarchy levels. Notice that the distribution appears to be fat-tailed. In fact, the wage distribution is a (discrete) Pareto distribution. This is due to the combination of (i) wages are an exponential function of the hierarchy and (2) the distribution of employment across hierarchy level is a geometric distribution. The equilibrium Pareto exponent is $\log(1+\psi/q)/\pi \approx 1.9$.

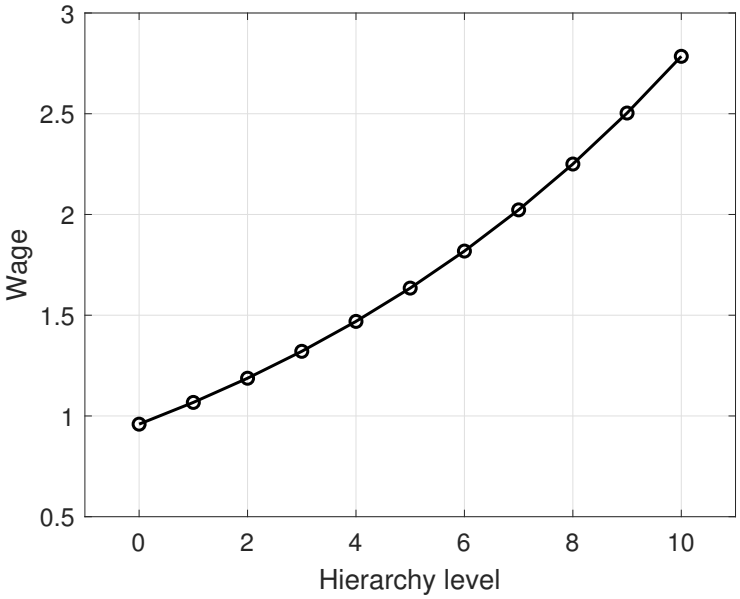

Figure 3.14: Wage schedule $\{w_k\}$ (model).

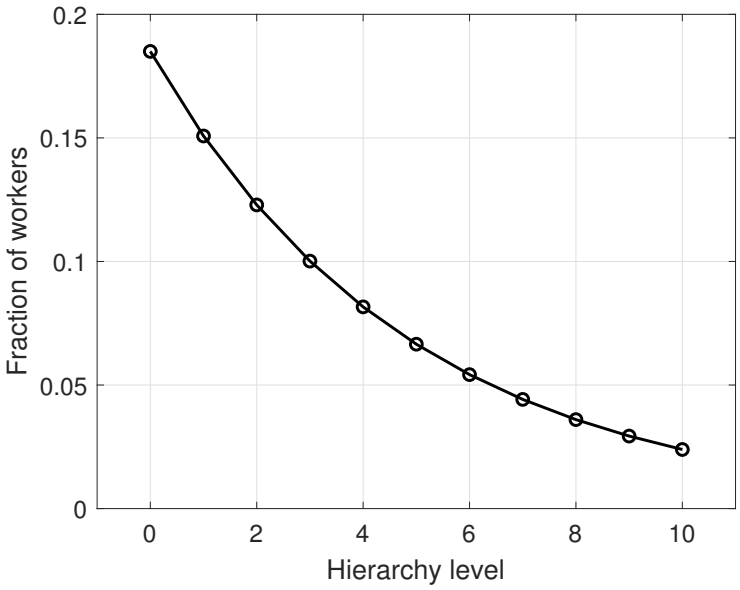

Figure 3.15: Employment shares $\{N_k\}$ (model).

3.3 Concluding remarks

In this paper, I propose a model of earnings dynamics and inequality within the firm motivated by evidence from personnel data. In the model, firms compete for workers by offering labor contracts specifying job-specific wages and promotion rates. On the one hand, large wage

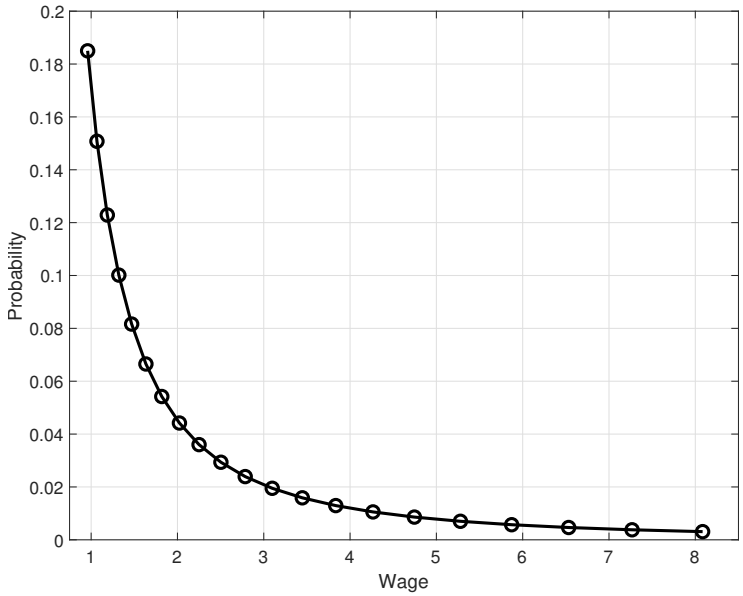

Figure 3.16: Wage distribution (model).

differentials between successive hierarchy levels induce workers to exert effort in order to increase their probability of being promoted. On the other hand, wage inequality decreases welfare due to diminishing marginal returns to consumption. The model clarifies how competition between firms balances these two margins and shapes the equilibrium distribution of wages. Future work should study optimal nonlinear labor income taxation in such a framework.

3.4 Acknowledgements

Chapter 3, in full, is currently being prepared for submission for publication of the material. Gouin-Bonenfant, Émilien. “Earnings Dynamics and Inequality Within the Firm”. The dissertation author was the sole investigator and author of this material.

Bibliography

Acemoglu, Daron, and Dan Cao. 2015. "Innovation by Entrants and Incumbents". *Journal of Economic Theory* 157 (): 255–294. doi:10.1016/j.jet.2015.01.001.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2017. *Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach*. NBER Working Paper 23732. http://www.nber.org/papers/w23732.

Achdou, Yves, Francisco Buera, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2014. "PDE models in macroeconomics". *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences*.

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving". *Quarterly Journal of Economics* 109, no. 3 (): 659–684. doi:10.2307/2118417.

Algan, Yann, Olivier Allais, and Wouter J. Den Haan. 2008. "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions". *Journal of Economic Dynamics and Control* 32, no. 3 (): 875–908. doi:10.1016/j.jedc.2007.03.007.

Algan, Yann, Olivier Allais, Wouter J. Den Haan, and Pontus Rendahl. 2014. "Solving and Simulating Models with Heterogeneous Agents and Aggregate Uncertainty". Chap. 6 in *Handbook of Computational Economics*, ed. by Karl Schmedders and Kenneth L. Judd, 3:277–324. Elsevier. doi:10.1016/B978-0-444-52980-0.00006-2.

Andrews, Dan, Chiara Criscuolo, Peter N Gal, et al. 2016. *The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy*. Tech. rep. OECD Publishing.

Aoki, Shuhei, and Makoto Nirei. 2017. "Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the United States". *American Economic Journal: Macroeconomics* 9, no. 3 (): 36–71. doi:10.1257/mac.20150051.

Arkolakis, Costas. 2016. "A Unified Theory of Firm Selection and Growth". *Quarterly Journal of Economics* 131, no. 1 (): 89–155. doi:10.1093/qje/qjv039.

Ashenfelter, Orley C, Henry Farber, and Michael R Ransom. 2010. "Labor market monopsony". *Journal of Labor Economics* 28 (2): 203–210.

Attanasio, Orazio P., and Guglielmo Weber. 1993. “Consumption Growth, the Interest Rate and Aggregation”. *Review of Economic Studies* 60 (3): 631–649. doi:10.2307/2298128.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, John Van Reenen, et al. 2017. *The fall of the Labor share and the rise of superstar firms*. Tech. rep. National Bureau of Economic Research.

Axtell, Robert L. 2001a. “Zipf distribution of US firm sizes”. *Science* 293 (5536): 1818–1820.

Axtell, Robert L. 2001b. “Zipf Distribution of U.S. Firm Sizes”. *Science* 293 (5536): 1818–1820. doi:10.1126/science.1062081.

Baily, Martin, Charles Hulten, and David Campbell. 1992. “Productivity Dynamics in Manufacturing Plants”. *Brookings Papers on Economic Activity (1992)*: 187–267.

Barkai, Simcha. 2016. “Declining labor and capital shares”. *Stigler Center for the Study of the Economy and the State New Working Paper Series* 2.

Barth, Erling, Alex Bryson, James C Davis, and Richard Freeman. 2014. *It’s where you work: Increases in earnings dispersion across establishments and individuals in the US*. Tech. rep. National Bureau of Economic Research.

— . 2016. “Its where you work: Increases in the dispersion of earnings across establishments and individuals in the United States”. *Journal of Labor Economics* 34 (S2): S67–S97.

Bayer, Christian, and Moritz Kuhn. 2018. “Which Ladder to Climb? Wages of workers by job, plant, and education”.

Beare, Brendan K, and Alexis Akira Toda. 2017a. “Geometrically stopped Markovian random growth processes and Pareto tails”. *arXiv preprint arXiv:1712.01431*.

Beare, Brendan K., and Alexis Akira Toda. 2017b. “Geometrically Stopped Markovian Random Growth Processes and Pareto Tails”. https://arxiv.org/abs/1712.01431.

Beaudry, Paul, and Eric van Wincoop. 1996. “The Intertemporal Elasticity of Substitution: An Exploration using a US Panel of State Data”. *Economica* 63 (251): 495–512. doi:10.2307/2555019.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. “The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents”. *Econometrica* 79, no. 1 (): 123–157. doi:10.3982/ECTA8416.

— . 2016. “The Distribution of Wealth in the Blanchard-Yaari Model”. *Macroeconomic Dynamics* 20 (): 466–481. doi:10.1017/S1365100514000066.

— . 2015. “The Wealth Distribution in Bewley Economies with Capital Income Risk”. *Journal of Economic Theory* 159, no. A (): 489–515. doi:10.1016/j.jet.2015.07.013.

Berlingieri, Giuseppe, Patrick Blanchenay, and Chiara Criscuolo. 2017. *The Great Divergence (s) Giuseppe Berlingieri*. Tech. rep.

Bewley, Truman F. 1983. "A Difficulty with the Optimum Quantity of Money". *Econometrica* 51, no. 5 (): 1485–1504. doi:10.2307/1912286.

— . 1977. "The Permanent Income Hypothesis: A Theoretical Formulation". *Journal of Economic Theory* 16, no. 2 (): 252–292. doi:10.1016/0022-0531(77)90009-6.

Blackwell, David. 1965. "Discounted Dynamic Programming". *Annals of Mathematical Statistics* 36 (1): 226–235. doi:10.1214/aoms/1177700285.

Blanchard, Olivier J. 1985. "Debt, Deficits, and Finite Horizons". *Journal of Political Economy* 93, no. 2 (): 223–247. doi:10.1086/261297.

Blanchard, Olivier J, William D Nordhaus, and Edmund S Phelps. 1997. "The medium run". *Brookings Papers on Economic Activity* 1997 (2): 89–158.

Bontemps, Christian, Jean-Marc Robin, and Gerard J Van den Berg. 2000. "Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation". *International Economic Review* 41 (2): 305–358.

Borovička, Jaroslav, and John Stachurski. 2019. "Necessary and Sufficient Conditions for Existence and Uniqueness of Recursive Utilities". https://arxiv.org/abs/1710.06526.

Brown, Charles, and James Medoff. 1989. "The employer size-wage effect". *Journal of Political Economy* 97 (5): 1027–1059.

Burdett, Kenneth, and Dale T Mortensen. 1998. "Wage differentials, employer size, and unemployment". *International Economic Review*: 257–273.

Burnside, Craig. 1998. "Solving Asset Pricing Models with Gaussian Shocks". *Journal of Economic Dynamics and Control* 22 (3): 329–340. doi:10.1016/S0165-1889(97)00075-4.

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth". *Journal of Political Economy* 114, no. 5 (): 835–870. doi:10.1086/508032.

Caicedo, Santiago, Robert E Lucas Jr, and Esteban Rossi-Hansberg. 2019. "Learning, career paths, and the distribution of wages". *American Economic Journal: Macroeconomics* 11 (1): 49–88.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The anatomy of French production hierarchies". *Journal of Political Economy* 123 (4): 809–852.

Cao, Dan, and Wenlan Luo. 2017. "Persistent Heterogeneous Returns and Top End Wealth Inequality". *Review of Economic Dynamics* 26 (): 301–326. doi:10.1016/j.red.2017.10.001.

Carroll, Christopher D., Kiichi Tokuoka, and Weifeng Wu. 2012. “The Method of Moderation”.

Carroll, Christopher D., Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. 2017. “The Distribution of Wealth and the Marginal Propensity to Consume”. *Quantitative Economics* 8, no. 3 (): 977–1020. doi:10.3982/QE694.

Coleman, Wilbur John, II. 1990. “Solving the Stochastic Growth Model by Policy-Function Iteration”. *Journal of Business and Economic Statistics* 8, no. 1 (): 27–29. doi:10.1080/07350015.1990.10509769.

Coles, Melvyn G, and Dale T Mortensen. 2016. “Equilibrium Labor Turnover, Firm Growth, and Unemployment”. *Econometrica* 84 (1): 347–363.

Collard, Fabrice, and Michel Juillard. 2001. “Accuracy of Stochastic Perturbation Methods: The Case of Asset Pricing Models”. *Journal of Economic Dynamics and Control* 25 (6-7): 979–999. doi:10.1016/S0165-1889(00)00064-6.

Dao, Mai, Mitali Das, Zsoka Koczan, and Weicheng Lian. 2017. *Why is Labor Receiving a Smaller Share of Global Income? Theory and Empirical Evidence*. Tech. rep. IMF Working Paper.

De Loecker, Jan, and Jan Eeckhout. 2017. *The rise of market power and the macroeconomic implications*. Tech. rep. National Bureau of Economic Research.

Decker, Ryan A, John C Haltiwanger, Ron S Jarmin, and Javier Miranda. 2018. *Changing business dynamism and productivity: Shocks vs. responsiveness*. Tech. rep. National Bureau of Economic Research.

Den Haan, Wouter J. 2010a. “Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents”. *Journal of Economic Dynamics and Control* 34, no. 1 (): 79–99. doi:10.1016/j.jedc.2008.12.009.

— . 2010b. “Comparison of Solutions to the Incomplete Markets Model with Aggregate Uncertainty”. *Journal of Economic Dynamics and Control* 34, no. 1 (): 4–27. doi:10.1016/j.jedc.2008.12.010.

Den Haan, Wouter J., Kenneth L. Judd, and Michel Juillard. 2010. “Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty”. *Journal of Economic Dynamics and Control* 34, no. 1 (): 1–3. doi:10.1016/j.jedc.2009.07.001.

Dixon, Jay, and Anne-Marie Rollin. 2012. “Firm dynamics: Employment growth rates of small versus large firms in Canada”. *Working paper*.

Durrett, Richard. 2010. *Probability: Theory and Examples*. Fourth. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press.

Edmans, Alex, Xavier Gabaix, and Dirk Jenter. 2017. "Executive compensation: A survey of theory and evidence". In *The Handbook of the Economics of Corporate Governance*, 1:383–539. Elsevier.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, markups, and the gains from international trade". *The American Economic Review* 105 (10): 3183–3221.

Eggertsson, Gauti B, Jacob A Robbins, and Ella Getz Wold. 2018. *Kaldor and Pikettys Facts: The Rise of Monopoly Power in the United States*. Tech. rep. National Bureau of Economic Research.

Elsby, Michael WL, Bart Hobijn, and Ayşegül Şahin. 2013. "The decline of the US labor share". *Brookings Papers on Economic Activity* 2013 (2): 1–63.

Elsby, Michael WL, and Ryan Michaels. 2013. "Marginal jobs, heterogeneous firms, and unemployment flows". *American Economic Journal: Macroeconomics* 5 (1): 1–48.

Engbom, Niklas, and Christian Moser. 2017. "Earnings inequality and the minimum wage: Evidence from Brazil". *Working paper*.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2016a. *Heterogeneity and Persistence in Returns to Wealth*. NBER Working Paper 22822. http://www.nber.org/papers/w22822.

— . 2016b. "Heterogeneity in Returns to Wealth and the Measurement of Wealth Inequality". *American Economic Review: Papers and Proceedings* 106, no. 5 (): 651–655. doi:10.1257/aer.p20161022.

Farmer, Leland E., and Alexis Akira Toda. 2017. "Discretizing Nonlinear, Non-Gaussian Markov Processes with Exact Conditional Moments". *Quantitative Economics* 8, no. 2 (): 651–683. doi:10.3982/QE737.

Frydman, Carola, and Raven E Saks. 2010. "Executive compensation: A new view from a long-term perspective, 1936–2005". *The Review of Financial Studies* 23 (5): 2099–2138.

Gabaix, Xavier. 2009. "Power Laws in Economics and Finance". *Annual Review of Economics* 1:255–293. doi:10.1146/annurev.economics.050708.142940.

— . 2011. "The Granular Origins of Aggregate Fluctuations". *Econometrica* 79, no. 3 (): 733–772. doi:10.3982/ECTA8769.

— . 1999. "Zipf's law for cities: an explanation". *The Quarterly journal of economics* 114 (3): 739–767.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality". *Econometrica* 84, no. 6 (): 2071–2111. doi:10.3982/ECTA13569.

Garicano, Luis. 2000. “Hierarchies and the Organization of Knowledge in Production”. *Journal of political economy* 108 (5): 874–904.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. “Organization and inequality in a knowledge economy”. *The Quarterly Journal of Economics* 121 (4): 1383–1435.

Gavazza, Alessandro, Simon Mongey, and Giovanni L Violante. 2018. “Aggregate recruiting intensity”. *American Economic Review* 108 (8): 2088–2127.

Gibbons, Robert, and Michael Waldman. 1999. “A theory of wage and promotion dynamics inside firms”. *The Quarterly Journal of Economics* 114 (4): 1321–1358.

Gibrat, Robert. 1931. *Les Inégalités Économiques*. Paris: Librairie du Recueil Sirey.

Gouin-Bonenfant, Emilien, and Alexis Akira Toda. 2018. “Pareto Extrapolation: Bridging Theoretical and Quantitative Models of Wealth Inequality”. *Working paper*.

Gourieroux, Christian, Alain Monfort, and Eric Renault. 1993. “Indirect inference”. *Journal of Applied Econometrics* 8 (S1).

Grullon, Gustavo, Yelena Larkin, and Roni Michaely. 2017. “Are US Industries Becoming More Concentrated?” *Available at SSRN*.

Guvenen, Fatih. 2009a. “An Empirical Investigation of Labor Income Processes”. *Review of Economic Dynamics* 12, no. 1 (): 58–79. doi:10.1016/j.red.2008.06.004.

— . 2009b. “An empirical investigation of labor income processes”. *Review of Economic dynamics* 12 (1): 58–79.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo-Diaz, and Daphne Chen. 2018. “Use It or Lose It: Efficiency Gains from Wealth Taxation”. https://fguvenendotcom.files.wordpress.com/2018/11/gkkoc-2018_fg_v107.pdf.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2015. *What do data on millions of US workers reveal about life-cycle earnings risk?* Tech. rep. National Bureau of Economic Research.

Hagedorn, Marcus, and Iourii Manovskii. 2008. “The cyclical behavior of equilibrium unemployment and vacancies revisited”. *American Economic Review* 98 (4): 1692–1706.

Hartman-Glaser, Barney, Hanno Lustig, and Mindy X Zhang. 2018. “Capital Share Dynamics When Firms Insure Workers”. *Journal of Finance (forthcoming)*.

Heise, Sebastian, and Tommaso Porzio. 2018. “Why Do Spatial Wage Gaps Persist?” *Working paper*.

Hopenhayn, Hugo, and Richard Rogerson. 1993. “Job turnover and policy evaluation: A general equilibrium analysis”. *Journal of political Economy* 101 (5): 915–938.

Huggett, Mark. 1993. “The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies”. *Journal of Economic Dynamics and Control* 17 (5-6): 953–969. doi:10.1016/0165-1889(93)90024-M.

Jones, Charles I., and Jihee Kim. 2018. “A Schumpeterian Model of Top Income Inequality”. *Journal of Political Economy* 126, no. 5 (): 1785–1826. doi:10.1086/699190.

Jorgenson, Dale W. 2012. “The world KLEMS initiative”. *International Productivity Monitor*, no. 24: 5.

Kaas, Leo, and Philipp Kircher. 2015. “Efficient firm dynamics in a frictional labor market”. *The American Economic Review* 105 (10): 3030–3060.

Kahn, Lisa B, and Fabian Lange. 2014. “Employer learning, productivity, and the earnings distribution: Evidence from performance measures”. *The Review of Economic Studies* 81 (4): 1575–1613.

Kaldor, Nicholas. 1961. “Capital accumulation and economic growth”. In *The theory of capital*, 177–222. Springer.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. “Monetary Policy According to HANK”. *American Economic Review* 108, no. 3 (): 697–743. doi:10.1257/aer.20160042.

Karabarbounis, Loukas, and Brent Neiman. 2014. “The global decline of the labor share”. *The Quarterly Journal of Economics* 129 (1): 61–103.

Kasa, Kenneth, and Xiaowen Lei. 2018. “Risk, Uncertainty, and the Dynamics of Inequality”. *Journal of Monetary Economics* 94 (): 60–78. doi:10.1016/j.jmoneco.2017.11.008.

Kehrig, Matthias. 2015. “The Cyclical Nature of the Productivity Distribution”. *Working paper*.

Kehrig, Matthias, and Nicolas Vincent. 2017. “Growing Productivity without Growing Wages: The Micro-Level Anatomy of the Aggregate Labor Share Decline”. *Working paper*.

Kesten, Harry. 1973. “Random Difference Equations and Renewal Theory for Products of Random Matrices”. *Acta Mathematica* 131 (1): 207–248. doi:10.1007/BF02392040.

Klass, Oren S., Ofer Biham, Moshe Levy, Ofer Malcai, and Sorin Solomon. 2006. “The Forbes 400 and the Pareto Wealth Distribution”. *Economics Letters* 90, no. 2 (): 290–295. doi:10.1016/j.econlet.2005.08.020.

Koh, Dongya, Raül Santaeulàlia-Llopis, and Yu Zheng. 2016. “Labor share decline and intellectual property products capital”. *Working paper*.

Krebs, Tom. 2006. "Recursive Equilibrium in Endogenous Growth Models with Incomplete Markets". *Economic Theory* 29 (3): 505–523. doi:10.1016/S0165-1889(03)00062-9.

Krueger, Dirk, Kurt Mitman, and Fabrizio Perri. 2016. "Macroeconomics and Household Heterogeneity". Chap. 11 in *Handbook of Macroeconomics*, ed. by John B. Taylor and Harald Uhlig, 2:843–921. Elsevier. doi:10.1016/bs.hesmac.2016.04.003.

Krusell, Per, and Anthony A. Smith Jr. 1998. "Income and Wealth Heterogeneity in the Macroeconomy". *Journal of Political Economy* 106, no. 5 (): 867–896. doi:10.1086/250034.

— . 2006. "Quantitative Macroeconomic Models with Heterogeneous Agents". Chap. 8 in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by Richard Blundell, Whitney K. Newey, and Torsten Persson, 1:298–340. Econometric Society Monograph 41. New York: Cambridge University Press.

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts". *Journal of political Economy* 89 (5): 841–864.

Lucas, Robert E. 1978. "On the size distribution of business firms". *The Bell Journal of Economics*: 508–523.

Ma, Qingyin, and John Stachurski. 2018. "Dynamic Programming Deconstructed". http://johnstachurski.net/_downloads/dpd6.pdf.

Mankiw, N. Gregory, and Stephen P. Zeldes. 1991. "The Consumption of Stockholders and Nonstockholders". *Journal of Financial Economics* 29 (1): 97–112. doi:10.1016/0304-405X(91)90015-C.

Manning, Alan. 2011. "Imperfect competition in the labor market". In *Handbook of labor economics*, 4:973–1041. Elsevier.

McDaniel, Cara. 2007. "Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD 1950-2003".

McKay, Alisdair. 2017. "Time-Varying Idiosyncratic Risk and Aggregate Consumption Dynamics". *Journal of Monetary Economics* 88 (): 1–14. doi:10.1016/j.jmoneco.2017.05.002.

Meghir, Costas, Renata Narita, and Jean-Marc Robin. 2015. "Wages and informality in developing countries". *American Economic Review* 105 (4): 1509–46.

Merton, Robert C. 1969. "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case". *Review of Economics and Statistics* 51 (3): 247–257. doi:10.2307/1926560.

Moll, Benjamin. 2014. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review* 104, no. 10 (): 3186–3221. doi:10.1257/aer.104.10.3186.

Mueller, Holger M, Paige P Ouimet, and Elena Simintzi. 2017a. "Wage inequality and firm growth". *American Economic Review* 107 (5): 379–83.

— . 2017b. "Within-firm pay inequality". *The Review of Financial Studies* 30 (10): 3605–3635.

Nekarda, Christopher J, and Valerie A Ramey. 2013. *The cyclical behavior of the price-cost markup*. Tech. rep. National Bureau of Economic Research.

Nirei, Makoto, and Shuhei Aoki. 2016. "Pareto Distribution of Income in Neoclassical Growth Models". *Review of Economic Dynamics* 20 (): 25–42. doi:10.1016/j.red.2015.11.002.

Nirei, Makoto, and Wataru Souma. 2007. "A Two Factor Model of Income Distribution Dynamics". *Review of Income and Wealth* 53, no. 3 (): 440–459. doi:10.1111/j.1475-4991.2007.00242.x.

Olley, G Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica* 64 (6): 1263–1297.

O'Mahony, Mary, and Marcel P Timmer. 2009. "Output, input and productivity measures at the industry level: the EU KLEMS database". *The Economic Journal* 119 (538): F374–F403.

Pareto, Vilfredo. 1897. *Cours d'Économie Politique*. Vol. 2. Lausanne: F. Rouge.

— . 1896. *La Courbe de la Répartition de la Richesse*. Lausanne: Imprimerie Ch. Viret-Genton.

— . 1895. "La Legge della Demanda". *Giornale degli Economisti* 10 (): 59–68.

Quadrini, Vincenzo. 2000. "Entrepreneurship, Saving, and Social Mobility". *Review of Economic Dynamics* 3, no. 1 (): 1–40. doi:10.1006/redy.1999.0077.

Reed, William J. 2001. "The Pareto, Zipf and Other Power Laws". *Economics Letters* 74, no. 1 (): 15–19. doi:10.1016/S0165-1765(01)00524-9.

Reiter, Michael. 2009. "Solving Heterogeneous-Agent Models by Projection and Perturbation". *Journal of Economic Dynamics and Control* 33 (3): 649–665. doi:10.1016/j.jedc.2008.08.010.

— . 2010. "Solving the Incomplete Markets Model with Aggregate Uncertainty by Backward Induction". *Journal of Economic Dynamics and Control* 34, no. 1 (): 28–35. doi:10.1016/j.jedc.2008.11.009.

Rosen, Sherwin. 1982. "Authority, control, and the distribution of earnings". *The Bell Journal of Economics*: 311–323.

Rosenthal, Jeffrey S. 1995. “Convergence Rates for Markov Chains”. *SIAM Review* 37, no. 3 (): 387–405. doi:10.1137/1037083.

Samuelson, Paul A. 1969. “Lifetime Portfolio Selection by Dynamic Stochastic Programming”. *Review of Economics and Statistics* 51, no. 3 (): 239–246. doi:10.2307/1926559.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2004. “Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function”. *Journal of Economic Dynamics and Control* 28 (4): 755–775. doi:10.1016/S0165-1889(03)00043-5.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2018. “Firming up inequality”. *The Quarterly Journal of Economics* 134 (1): 1–50.

Stachurski, John, and Alexis Akira Toda. 2019. “An Impossibility Theorem for Wealth in Heterogeneous-agent Models with Limited Heterogeneity”. *Journal of Economic Theory* 182 (): 1–24. doi:10.1016/j.jet.2019.04.001.

Sun, Yeneng. 2006. “The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks”. *Journal of Economic Theory* 126, no. 1 (): 31–69. doi:10.1016/j.jet.2004.10.005.

Syverson, Chad. 2011. “What determines productivity?” *Journal of Economic literature* 49 (2): 326–365.

Toda, Alexis Akira. 2018. “Data-based Automatic Discretization of Nonparametric Distributions”. http://arxiv.org/abs/1805.00896.

— . 2014. “Incomplete Market Dynamics and Cross-Sectional Distributions”. *Journal of Economic Theory* 154 (): 310–348. doi:10.1016/j.jet.2014.09.015.

— . 2019. “Wealth Distribution with Random Discount Factors”. *Journal of Monetary Economics*. doi:10.1016/j.jmoneco.2018.09.006.

— . 2016. “Zipf’s Law: A Microfoundation”. *Working paper*.

Toda, Alexis Akira, and Kieran Walsh. 2015. “The Double Power Law in Consumption and Implications for Testing Euler Equations”. *Journal of Political Economy* 123, no. 5 (): 1177–1200. doi:10.1086/682729.

Toda, Alexis Akira, and Kieran James Walsh. 2017. “Fat Tails and Spurious Estimation of Consumption-Based Asset Pricing Models”. *Journal of Applied Econometrics* 32 (6): 1156–1177. doi:10.1002/jae.2564.

Vermeulen, Philip. 2018. “How Fat is the Top Tail of the Wealth Distribution?” *Review of Income and Wealth* 64, no. 2 (): 357–387. doi:10.1111/roiw.12279.

Vissing-Jørgensen, Annette. 2002. “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution”. *Journal of Political Economy* 110 (4): 825–853. doi:10.1086/340782.

Winberry, Thomas. 2018. “A Method for Solving and Estimating Heterogeneous Agent Macro Models”. *Quantitative Economics* 9, no. 3 (): 1123–1151. doi:10.3982/QE740.

Yaari, Menahem E. 1965. “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer”. *Review of Economic Studies* 32, no. 2 (): 137–150. doi:10.2307/2296058.

Young, Eric R. 2010. “Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell-Smith Algorithm and Non-Stochastic Simulations”. *Journal of Economic Dynamics and Control* 34, no. 1 (): 36–41. doi:10.1016/j.jedc.2008.11.010.

Zhu, Shenghao. 2018. “A Becker-Tomes Model with Investment Risk”. Forthcoming, *Economic Theory*. doi:10.1007/s00199-018-1103-2.

Appendix A

Appendix of Chapter 1

A.1 Proofs

A.1.1 Proof of Proposition 1

First, I show that the value of unemployment $W(w)$ is strictly increasing in the current wage w. From Equation 1.6, we have that

$$rW'(w) = 1 - \chi_s W'(w) + \lambda(1-u)\int \frac{\partial}{\partial w} \max\left\{W(w') - W(w), 0\right\} d\widetilde{P}(w') - \chi_x W'(w).$$

I then bound the derivative from below

$$rW'(w) \geq 1 - \chi_s W'(w) - \lambda(1-u)W'(w) - \chi_x W'(w).$$

Re-arranging, I obtain the desired result

$$W'(w) \geq \frac{1}{r + \chi_s + \chi_x + \lambda(1-u)} > 0.$$

Second, I derive the expression for the reservation wage w_r. Setting $W(w_r) = U$ in

Equation 1.6, I obtain

$$rU = w_r + \chi_s \int \Big(W(\widehat{w}(z)) - U\Big) d\Gamma(z) + \lambda(1-u) \int \max\Big\{W(w'), U\Big\} d\widetilde{P}(w').$$

Using the the inequality $W(w) \geq U$,

$$rU = w_r + \chi_s \int \max\Big\{W(\widehat{w}(z)) - U, 0\Big\} d\Gamma(z) + \lambda(1-u) \int \max\Big\{W(w'), U\Big\} d\widetilde{P}(w').$$

and combining with Equation 1.5, I obtain

$$w_r = b + (\chi_e - \chi_s) \int \max\Big\{W(\widehat{w}(z)) - U, 0\Big\} d\Gamma(z).$$

Finally, using the definitions of χ_e and χ_s (Equation 1.3), I obtain the desired expression

$$w_r = b + (\mu - \chi)(1 - \Gamma_0(z_l)) \int \max\Big\{W(\widehat{w}(z)) - U, 0\Big\} d\Gamma(z).$$

Since $\Gamma(z)$ is the truncation of Γ_0 at z_l (Equation 1.3), then the expressions is equivalent to

$$w_r = b + (\mu - \chi) \int_{z_l}^{\infty} \max\Big\{W(\widehat{w}(z)) - U, 0\Big\} d\Gamma_0(z).$$

A.1.2 Proof of Lemma 2

First, I show that the value function is homogeneous. Dividing by N on both sides of Equation 1.12 and defining $n_s \equiv N_s/N_0$ and $k_s \equiv K_s/N_s$, we obtain

$$\frac{v(z,N)}{N} = \max_{T, \{w_s, k_s\}_{s=0}^{T}} \mathbb{E}_0 \int_0^T e^{-rs} \Big(z_s k_s^{\alpha} - w_s - R k_s\Big) n_s ds$$

$$n_0 = 1, \quad z_0 = z$$

$$w_s \geq w_r, \quad dn_s = \widetilde{g}(w_s) n_s ds, \quad dz_s = d\mathcal{J}_s(z_s' - z_s)$$

The expression for $\frac{v(z,N)}{N}$ is equal to Equation 1.12 when $N = 1$, so we obtain the desired result that $v(z,N) = v(z,1)N$. I now prove that the optimal stopping time is a threshold rule. Given that z is the only state variable, we now only need to show that $v'(z) \geq 0$. Using the homogeneity result, Equation 1.13 reduces to

$$rv(z) = \min\Big\{ \max_{w,k} \Big\{ zk^{\alpha} - w - Rk + v(z)\widetilde{g}(w) \Big\} + \chi\Big(\int v(\zeta)\Gamma_0(d\zeta) - v(z) \Big), 0 \Big\}.$$

Using the Envelope theorem, we have

$$rv'(z) \in \big\{ k(z)^{\alpha} + v'(z)g(z) - \chi v(z), 0 \big\} \implies v'(z) \in \left\{ \frac{k(z)^{\alpha}}{r + \chi - g(z)}, 0 \right\}.$$

Assumption 3 combined with Lemma 1 implies that $\chi - g(z) > 0$, so we obtain the desired result that $v'(z) \geq 0$. The fact that the exit threshold z_l coincides with the entry threshold follows directly from homogeneity and free entry: $v(z)N_0 > 0 \iff v(z) > 0 \iff z > z_l$.

A.1.3 Proof of Proposition 2

From the first-order condition (Equation 1.19), we have that

$$1 = v(z)\widetilde{g}'(w(z)).$$

Using the definition $g(z) \equiv \widetilde{g}(w(z))$, it follows from the chain rule that $g'(z) = \widetilde{g}'(w(z))w'(z)$. Plugging back in the first-order condition, I obtain

$$w'(z) = v(z)g'(z).$$

To solve for $w(z)$, I need a a boundary condition. I now use the fact that the marginal exiting firm is constrained by the worker's reservation wage w_r. Without the constraint $w \geq w_r$, firms would never exit as they would be able to achieve zero flow profit by setting $w = 0$ and $k = 0$. Using the

fact that $w(z_l) = w_r$, I obtain the desired result

$$w(z) = w_r + \int_{z_l}^{z} v(\zeta) g'(\zeta) d\zeta.$$

I now verify that the second order condition $v(z)\widetilde{g}''(w(z)) < 0$ holds for interior solutions (i.e. for all $z > z_l$). Since $z > z_l \implies v(z) > 0$, I only need to verify that $\widetilde{g}''(w(z)) \leq 0$. From the definition of $g(z)$, we have that

$$g'(z) = \widetilde{g}'(w(z))w'(z) \implies g''(z) = \widetilde{g}''(w(z))(w'(z))^2 + \widetilde{g}'(w(z))w''(z).$$

And from the first-order condition, we have that $w'(z) = g'(z)v(z) \implies w''(z) = g''(z)v(z) + g'(z)v'(z)$. Putting together,

$$g''(z) = \widetilde{g}''(w(z))(w'(z))^2 + \widetilde{g}'(w(z))\Big(g''(z)v(z) + g'(z)v'(z)\Big)$$

$$\implies g''(z)\Big(1 - \widetilde{g}'(w(z))v(z)\Big) = \widetilde{g}''(w(z))(w'(z))^2 + \widetilde{g}'(w(z))g'(z)v'(z)$$

But from the first-order condition, we have that $1 - \widetilde{g}'(w(z))v(z) = 0$, so

$$\widetilde{g}''(w(z)) = -\frac{\widetilde{g}'(w(z))g'(z)v'(z)}{(w'(z))^2} = -\frac{(g'(z))^2 v'(z)}{w'(z)}$$

The sign of $\widetilde{g}''(w(z))$ therefore depends on the sign of $w'(z)$. Since I restrict the analysis to equilibria where wages are increasing in productivity, we have that $\widetilde{g}''(w(z)) < 0$.

A.1.4 Proof of Proposition 3

First, the steady-state unemployment can be solved independently of $P(z)$. The law of motion (Equation 1.9) is

$$\dot{u} = (1-u)\chi_x - u\Big(\chi_e + \lambda(1-u)\Big)$$

Setting $\dot{u} = 0$, we have a quadratic equation of the form $au^2 + bu + c = 0$ with constants given by

$$a = \lambda, \quad b = -(\lambda + \chi_e + \chi_x), \quad c = \chi_x.$$

First, I verify that the discriminant $(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x$ is non-negative.

$$(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x \geq (\lambda + \chi_x)^2 - 4\lambda\chi_x = (\lambda - \chi_x)^2 \geq 0$$

There are therefore two candidate solutions

$$u^{(1)} = \frac{\lambda + \chi_e + \chi_x - \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x}}{2\lambda}, \quad u^{(2)} = \frac{\lambda + \chi_e + \chi_x + \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x}}{2\lambda}$$

First, I show that $u^{(2)}$ is not a valid solution since it is greater than one.

$$u^{(2)} > 1 \iff \chi - \lambda + \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x} > 0$$

Under Assumption 3, the first term is positive which concludes the proof. Then, I show that $u^{(1)}$ is a valid solution (i.e. it satisfies $0 \leq u^{(1)} \leq 1$). Using the fact that $\chi_x \geq 0$, we have that

$$u^{(1)} \geq \frac{\lambda + \chi_e + \chi_x - (\lambda + \chi_e + \chi_x)}{2\lambda} = 0$$

Now, I show that $u^{(1)} \leq 1$.

$$u^{(1)} \leq 1 \iff \lambda + \chi_e + \chi_x - \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x} \leq 2\lambda$$

$$\iff \chi_e + \chi_x - \lambda \leq \sqrt{(\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x}$$

So, if $\chi_e + \chi_x - \lambda < 0$, the proof is done. Otherwise, I square on both sides.

$$(\chi_e + \chi_x - \lambda)^2 \leq (\lambda + \chi_e + \chi_x)^2 - 4\lambda\chi_x \iff \chi_e \geq 0$$

Which concludes the proof.

I now provide a derivation of each term of *Kolmogorov Forward Equation* for the employment-weighted productivity distribution $P(z)$.

$$\dot{P}(z) = \underbrace{(1-u)\lambda P(z)(P(z)-1)}_{\text{Job-to-job flows}} + \underbrace{\frac{u}{1-u}\chi_e \Gamma(z) + u\lambda P(z)}_{\text{Employment inflows}} - \underbrace{\chi_x P(z)}_{\substack{\text{Employment} \\ \text{outflows}}} + \underbrace{\chi_s(\Gamma(z)-P(z))}_{\text{Productivity shocks}}.$$

At rate $\lambda(1-u)$, an employed worker receives a competing job offer. The distribution of productivity (CDF) of such workers, after they have made their decision whether or not to accept the competing job offer, is P^2. The reason is that the distribution of z in the population of workers is precisely the employment-weighted productivity distribution P while the distribution of job offers z' is also P, since firms meet worker proporitonally to their size. The CDF of $\max\{z,z'\}$ is therefore P^2.[1] Using the KFE formula for jump processes, the term that accounts for job-to-job flows is therefore

$$\lambda(1-u)(P^2(z)-P(z)) = \lambda(1-u)P(z)(P(z)-1).$$

At rate χ_e, an unemployed worker meets an entrepreneur and enters the workforce with productivity distributed according to $\Gamma(z)$. At rate $\lambda(1-u)$, an unemployed worker meets a firm and enters the workforce with productivity distributed according to $P(z)$. Since the ratio of unemployment to employment is $\frac{u}{1-u}$, the term that accounts for unemployment inflows is

$$\frac{u}{1-u}\Big(\chi_e\Gamma(z) + \lambda(1-u)P(z)\Big) = \frac{u}{1-u}\chi_e\Gamma(z) + u\lambda P(z)$$

At rate χ_x, a worker with productivity distributed according to $P(z)$ is sent to unemployment due to firm exit so that the term which accounts for employment outflows is $-\chi_x P(z)$. At rate χ_s, an employed worker's productivity resets to a draw from Γ, so that the term which accounts for productivity shocks is $\chi_s(\Gamma(z)-P(z))$.

[1] If X and Y are two independent random variables with CDFs F, G, then $\mathbb{P}(\max\{X,Y\} \leq z) = F(z)G(z)$.

I now compute the stationary solution of the the law of motion for $P(z)$. Setting $\dot{P}(z) = 0$, we obtain a quadratic equation in $P(z)$ of the form $aP^2(z) + bP(z) + c = 0$, with coefficients given by

$$a = (1-u)\lambda, \quad b = -\big(\lambda(1-2u) + \chi\big), \quad c(z) = \big(\frac{u}{1-u}\chi_e + \chi_s\big)\Gamma(z),$$

I will now use the fact that, in steady-state, $\chi_e u = (1-u)\chi_x - \lambda(1-u)u$ (Equation 1.9). The expression for $c(z)$ thus simplifies to $c(z) = (\chi - \lambda u)\Gamma(z)$, so the discriminant is

$$\Delta(z) \equiv \big(\lambda(1-u) - \chi - \lambda u\big)^2 - 4\lambda(1-u)(\chi - \lambda u)\Gamma(z).$$

I will repeatedly use the fact that $\sqrt{\Delta(z)} \geq |\chi - \lambda|$. Under Assumption 3, the discriminant is strictly positive, which implies that there are two real roots.

$$P^{(1)}(z) = \frac{\lambda(1-2u) + \chi - \sqrt{\Delta(z)}}{2(1-u)\lambda}, \quad P^{(2)}(z) = \frac{\lambda(1-2u) + \chi + \sqrt{\Delta(z)}}{2(1-u)\lambda}.$$

First, I show that $P^{(2)}(z) > 1$ for all $z > z_l$, which implies that $P^{(2)}$ it is not a valid CDF. Using the fact that $\sqrt{\Delta(z)} \geq \chi - \lambda$, we have that

$$P^{(2)}(z) \geq \frac{\lambda(1-2u) + \chi - \chi + \lambda}{\lambda(1-u)} = 1$$

To conclude, I show that $P^{(1)}$ is a valid CDF, meaning that (1) $P^{(1)\prime}(z) \geq 0$ and (2) $P^{(1)}(z_l) = 0$, and (3) $\lim_{z\to\infty} P^{(1)}(z) = 1$. First,

$$P^{(1)\prime}(z) = \frac{\chi - \lambda u}{\sqrt{\Delta(z)}}\Gamma'(z) \geq 0.$$

Second,

$$P^{(1)}(z_l) = \frac{\lambda(1-2u) + \chi - \Big(\lambda(1-2u) + \chi\Big)}{2(1-u)\lambda} = 0.$$

Third,

$$\lim_{z\to\infty} P^{(1)}(z) = \frac{\lambda(1-2u)+\chi-(\chi-\lambda)}{2\lambda(1-u)} = 1$$

Going from the second to last equation to the last one involves basic algebra as well.

A.1.5 Proof of Propositions 4 and 5

First, I will prove 3 lemmas.

Lemma 6. *There exists constants $\bar{v}_1, \bar{v}_2, \bar{g}' < \infty$ such that, for all $z \geq z_l$*

$$v(z) \leq \bar{v}_1 + \bar{v}_2 z^{\frac{1}{1-\alpha}}$$

$$g'(z) \leq \bar{g}' \Gamma_0(z)$$

$$w(z) \leq \bar{w}$$

Proof. From the Equation 1.14 and optimizing over k, we have

$$rv(z) = \max_{w \geq w_r} \left\{ (1-\alpha)(\alpha/R)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}} - w - v(z)\widetilde{g}(w) \right\} + \chi\Big(\int v(z')\Gamma_0(dz') - v(z) \Big)$$

Since $w \geq w_r$ and $\widetilde{g}(w) \leq \lambda$, we have

$$rv(z) \leq (1-\alpha)(\alpha/R)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}} - w_r - v(z)\lambda + \chi \int v(z')\Gamma_0(dz') - \chi v(z)$$

$$\implies v(z) \leq \underbrace{\frac{\chi \int v(z')\Gamma_0(dz') - w_r}{r+\chi-\lambda}}_{\equiv \bar{v}_1} + \underbrace{\frac{(1-\alpha)(\alpha/R)^{\frac{\alpha}{1-\alpha}}}{r+\chi-\lambda}}_{\equiv \bar{v}_1} z^{\frac{1}{1-\alpha}}$$

From Equation 1.25, we have

$$g'(z) = 2\lambda(1-u)P'(z).$$

From Equation 1.24, we have

$$P'(z) = \Delta(z)^{-\frac{1}{2}}(\chi - \lambda u)(1-\Gamma(z_l))\Gamma_0(z),$$

where $\Delta(z) = \big(\lambda(1-u) - \chi - \lambda u\big)^2 - 4\lambda(1-u)(\chi - \lambda u)\Gamma(z)$. Putting together and using the fact that $\Delta(z) \geq \chi - \lambda > 0$

$$g'(z) \leq \underbrace{\frac{2\lambda(1-u)(\chi-\lambda u)(1-\Gamma(z_l))}{\chi-\lambda}}_{\equiv \bar{g}'} \Gamma_0(z)$$

□

Lemma 7. *There exists a constant $\bar{w} < \infty$ such that, for all $z \geq z_l$*

$$w(z) \leq \bar{w}$$

Proof. Since $w'(z) > 0$, then

$$w(z) \leq \lim_{z\to\infty} w(z) = w_r + \int_0^\infty g'(x)v(x)dx$$

Using the results from Lemma 6,

$$w(z) \leq \underbrace{w_r + \kappa \int_0^\infty z^{\frac{1}{1-\alpha}} \Gamma_0'(z)dz}_{\equiv \bar{w}} < \infty.$$

The last inequality comes from Assumption 1. □

Lemma 8. *The expression $w'(z)z$ converges to zero as $z \to \infty$*

Proof. From Assumption 1, we have that that $\int z^{\frac{1}{1-\alpha}} \Gamma_0'(z) < \infty$, which implies that $z^{1+\frac{1}{1-\alpha}} \Gamma_0'(z) \to 0$ Using the results from Lemma 6,

$$w'(z)z = g'(z)v(z)z \leq \bar{v}_0 \bar{g}' z \Gamma_0'(z) + \bar{v}_1 \bar{g}' z^{1+\frac{1}{1-\alpha}} \Gamma_0'(z) \to 0$$

□

I now prove Proposition 4 that $\frac{dw}{dLP}(z) \to 0$.

$$\frac{dw}{dLP}(z) = \frac{w'(z)}{LP'(z)} = \frac{w'(z)z}{(1-\alpha)(\alpha/R)^{\frac{\alpha}{1-\alpha}} z^{\frac{1}{1-\alpha}}} \to 0$$

The last step follows from $w'(z)z \to 0$ (Lemma 8) and $z^{\frac{1}{1-\alpha}} \to \infty$.

I now prove Proposition 5. I will show that there exists a z' such that, for every $z > z'$, $\frac{dLS}{dz}(z) < 0$, which is equivalent as showing that $\frac{d\log LS}{dz}(z) < 0$. Notice that

$$\log LS(z) = \log w(z) - \log LP(z) \implies \frac{d\log LS}{dz}(z) = \frac{w'(z)}{w(z)} - (1-\alpha)\frac{1}{z}$$

So, we have that

$$\frac{d\log LS}{dz}(z) < 0 \iff \frac{w'(z)z}{w(z)} < 1-\alpha$$

The right-hand-side is positive for all $\alpha < 1$, while the left hand side converges to zero, since $w'(z)z \to$ and $w(z) \to \bar{w}$. By continuity, there exists a z' such that, for every $z > z'$, $\frac{w'(z)z}{w(z)} < 1-\alpha$.

A.1.6 Proof of Proposition 6

First, the average firm size is given by the measure of employed workers $1-u$ over the measure of firm F. The law of motion for the measure of firms is given by

$$\dot{F} = \chi_e u - \chi_x F,$$

which implies that, in a stationary equilibirum, $F = u\frac{\chi_e}{\chi_x}$ and the average firm size is thus given by

$$\mathbb{E}(N) = \frac{1-u}{u}\frac{\chi_x}{\chi_e}$$

To compute the conditional expectation $\mathbb{E}(N|z)$, I will use the employment-weighted distribution P. For a set $S \in \mathbb{R}_+$, the measure of employment at firms with $z \in S$ is given by $(1-u)\int_S dP(\zeta)$

while the measure of firms is given by $F\int d\Gamma(\zeta)$. It follows that the average firm size for firms with $z \in S$ is given by

$$\mathbb{E}(N|z \in S) = \frac{(1-u)\int_S dP(\zeta)}{F\int_S d\Gamma(\zeta)}$$

In the limit, when $S = \{z\}$, I obtain the following formula by using the expression

$$\mathbb{E}(N|z) = \frac{1-u}{F}\frac{P'(z)}{\Gamma'(z)} = \frac{1-u}{u}\frac{\chi_x}{\chi_e}\frac{\chi - \lambda u}{\sqrt{\left(\lambda(1-u)+\chi-\lambda u\right)^2 - 4\lambda(1-u)(\chi-\lambda u)\Gamma(z)}},$$

which is increasing in z.

A.1.7 Proof of Proposition 8

Lemma 9. *Let $U^*, W^*(w), w_r^*$ the worker's value functions and reservation wage after the broad-based increase in productivity $z \to \pi z$ and $b \to \pi^{\frac{1}{1-\alpha}} b$.*

$$\pi^{-\frac{1}{1-\alpha}} U^* = U$$

$$W^*(\pi^{-\frac{1}{1-\alpha}} w) = W(w)$$

Lemma 10. *Let $v^*(z)$ and $w^*(z)$ the firm's value function and reservation wage after the broad-based increase in productivity $z \to \pi z$ and $b \to \pi^{\frac{1}{1-\alpha}} b$.*

$$\pi^{-\frac{1}{1-\alpha}} v^*(z) = v(z)$$

$$\pi^{-\frac{1}{1-\alpha}} w^*(z) = w(z)$$

A.1.8 Burdett and Mortensen 1998 with capital

I now present a derivation of Burdett and Mortensen 1998, the only extension being the addition of capital as a factor of production. The laws of motion for the wage distribution $P(w,t)$[2]

[2] $P(0,t)$ is, by convention, the unemployment rate (i.e. the measure of workers working at firms which pay $w = 0$).

and firm-level employment $N(w,t)$ are respectively given by

$$\frac{\partial}{\partial t}N(w,t) = \underbrace{\lambda P(w)}_{\text{Hires}} - \underbrace{\delta N(w,t)}_{\text{Layoffs}} - \underbrace{\lambda \bar{F}(w) N(w,t)}_{\text{Quits}}$$

$$\frac{\partial}{\partial t}P(w,t) = \underbrace{\delta\big(H(w) - P(w,t)\big)}_{\text{Job destruction}} + \underbrace{\lambda\big(F(w)P(w,t) - P(w,t)\big)}_{\text{Job creation + Churn}}.$$

The stationary solution for employment is given by

$$N(w) = \frac{\lambda\delta}{\Big(\delta + \lambda\bar{F}(w)\Big)^2}.$$

Firms chose the wage policy $w \geq b$ and capital stock per worker $k \geq 0$ as to maximize long-run profits

$$\max_{w,k}\ zk^{\alpha}N(w) - RkN(w) - wN(w).$$

The optimal capital stock is $k(z) = \left(\frac{z\alpha}{R}\right)^{\frac{1}{1-\alpha}}$. Substituting it in the objective, I obtain

$$\max_{w}\ c(z)N(w) - wN(w),$$

where $c(z) \equiv (1-\alpha)zk(z)^{\alpha}$. The first-order condition given by

$$c(z)N'(w(z)) = N(w(z)) + w(z)N'(w(z)).$$

Now define $N(z) \equiv N(w(z))$ which means that $N'(z) = N'(w(z))w'(z)$, which implies that

$$c(z)N'(z) = w'(z)N(z) + w(z)N'(z) \implies \Big(w(z)N(z)\Big)' = c(z)N'(z).$$

This is an ODE. Combined with the initial condition $w(z_l) = b$—i.e. the worst operating firm is constrained by the worker's participation constraint—I obtain a unique solution.

$$w(z)N(z) - w(z_l)N(z_l) = N(z)c(z) - N(z_l)c(z_l) - \int_{z_l}^{z} c'(\zeta)N(\zeta)d\zeta$$

$$w(z) = c(z) + \big(b - c(z_l)\big)\frac{N(z_l)}{N(z)} - \int_{z_l}^{z} c'(\zeta)\frac{N(\zeta)}{N(z)}d\zeta$$

The entry threshold z_l satisfies $c(z_l) = b$. The formula simplifies to

$$w(z) = c(z) - \int_{z_l}^{z} c'(\zeta)\frac{N(\zeta)}{N(z)}d\zeta$$

Using the fact that the wage schedule is increasing in z, we have that

$$N(z) = \frac{\lambda\delta}{\big(\delta + \lambda\bar{\Gamma}(z)\big)^2}, \quad P(z) = \frac{\delta}{\delta + \lambda\bar{\Gamma}(z)}$$

where Γ is the distribution of productivity at active firms and Γ_0 is the distribution of productivity of potential entrants

$$\Gamma(z) = \frac{\Gamma_0(z) - \Gamma_0(z_l)}{1 - \Gamma_0(z_l)}$$

The threshold z_l is given by

$$z_l = \frac{b^{1-\alpha}R^{\alpha}}{(1-\alpha)^{1-\alpha}\alpha^{\alpha}}.$$

Notice that the equilibrium functions N and P only depend on (λ, δ) through their ratio $\psi \equiv \lambda/\delta$. Since the wage schedule (and therefore the labor share) depend only on N, it also only depends on ψ. Hence, I normalize $\delta = 1$. The free parameters of the model are therefore $(\alpha, R, \lambda, b, \eta)$. First, I calibrate (α, R, η) exactly as in Section 1.4.1. Then, I choose λ as to match exactly the unemployment in the full model. Finally, b is chosen such that the entry threshold in the *BM* model coincides with the entry threshold in the full model. This last step implies that the distribution of productivity amongst active firms Γ coincides in both models.

I now provide a characterization of the firm size distribution. Using the definition of $N(z)$, we have

$$\mathbb{P}(N(z) \leq n) = \mathbb{P}\left(\Gamma(z) \leq 1 + \frac{\delta}{\lambda} - \sqrt{\frac{\delta}{\lambda}} n^{-\frac{1}{2}}\right) = 1 + \frac{\delta}{\lambda} - \sqrt{\frac{\delta}{\lambda}} n^{-\frac{1}{2}}.$$

the last equation uses the fact that $\Gamma(z) \sim U[0,1]$. So that the firm size distribution obeys a power law with Pareto exponent $\zeta = 1/2$. But the distribution is bounded

$$\frac{\lambda\delta}{(\delta+\lambda)^2} \leq N(z) \leq \frac{\lambda}{\delta},$$

So the firm size distribution is

$$\mathbb{P}(N \leq n) = \begin{cases} 0 & \text{if } n \leq \frac{\lambda\delta}{(\delta+\lambda)^2}\frac{\lambda}{\delta} \\ 1 + \frac{\delta}{\lambda} - \sqrt{\frac{\delta}{\lambda}} n^{-\frac{1}{2}} & \text{if } \frac{\lambda\delta}{(\delta+\lambda)^2} \leq n \leq \frac{\lambda}{\delta} \\ 1 & \text{if } n \geq \frac{\lambda}{\delta} \end{cases}$$

I now provide a closed-form solution for the aggregate labor share in the special case where there is no capital ($\alpha = 0$) and the distribution of firm productivity is Pareto ($\Gamma(z) = 1 - (z/b)^{-\eta}$ over $z \geq b$). I require $\eta > 1$ to ensure that output is finite. The formula for the wage is

$$w(z) = z - \int_b^z N(x)dx$$

Aggregating, I obtain

$$\underbrace{\int_b^\infty w(x)N(x)d\Gamma(x)}_{wN} = \underbrace{\int_b^\infty xN(x)d\Gamma(x)}_{Y} - \underbrace{\int_b^\infty \int_b^z N(y)dyd\Gamma(x)}_{\Pi}$$

where wN is worker compensation, Y is output, and Π is firm profits. I now focus on the last term and will use the fact that $(1-\Gamma(z)) = \frac{1}{\eta} z \frac{d\Gamma(z)}{dz}$

$$\int_b^\infty \int_b^z N(y)dyd\Gamma(x) = \int_b^\infty N(y)\Big(\int_b^\infty 1\{y \leq x\}d\Gamma(x)\Big)dy = \int_b^\infty N(y)(1-\Gamma(y))dy = \frac{1}{\eta}Y$$

The labor share $LS = wN/Y$ is therefore given by

$$LS = 1 - \frac{1}{\eta}$$

When productivity dispersion is high (η is low), then the labor share is low.

A.1.9 Solution method for $\mu = \chi$ case

I now present a solution method for the case $\mu = \chi$ that I use to solve for the firm value function and threshold (v, z_l) as well as the productivity and size distribution φ.

First, I combine the *HJB* (Equation 1.14) with the expressions for the wage schedule $w(z)$ as a function of $v(\cdot)$ (Equation 1.22), labor productivity $LP(z)$ (Equation 1.21) and the employment growth function $g(z)$ (Equation 1.25). I obtain the following equations.

$$rv(z) = (1-\alpha)LP(z) - b - \int_{z_l}^{z} v(\zeta)g'(\zeta)d\zeta + v(z)g(z) + \chi \int v(\zeta)\Gamma_0(d\zeta) - \chi v(z) \quad \forall z \geq z_l$$

$$v(z) = 0 \quad \forall z < z_l$$

To simplify notation, I define the the linear operators $\mathcal{A}, \mathcal{B}, \mathcal{C}$ respectively defined by the actions

$$\mathcal{A}f(z) = \int_{z_l}^{z} f(\zeta)g'(\zeta)d\zeta, \quad \mathcal{B}f(z) = f(z)g(z), \quad \mathcal{C}f(z) = \int v(\zeta)\Gamma_0(d\zeta), \quad If(z) = f(z)$$

The HJB for $v(z)$ can rewritten as

$$rv(z) = (1-\alpha)LP(z) - b - \mathcal{A}v(z) + \mathcal{B}v(z) + \chi\mathcal{C}v(z) - \chi Iv(z) \quad \forall z \geq z_l$$

$$v(z) = 0 \quad \forall z < z_l.$$

The system can be equivalently written as

$$v(z)\Big(\mathcal{M}v(z)+q(z)\Big)=0,\quad \mathcal{M}v(z)+q(z)\geq 0,\quad v(z)\geq 0$$

where $\mathcal{M}\equiv(r+\chi)I+\mathcal{A}-\mathcal{B}-\chi\mathcal{C}$ and $q(z)\equiv b-(1-\alpha)LP(z)$.

The threshold z_l does not appear anywhere but can recovered from the solution $v(z)$ as $z_l\equiv\inf\big\{z:v(z)>0\big\}$. I now consider a discrete approximation of the system of equations over a grid $G_z=\{z_1,\ldots,z_{N_z}\}$. Let $v\equiv(v(z_1),\ldots,v(z_{N_z}))'$ and $LP\equiv(LP(z_1),\ldots,LP(z_{N_z}))'$. I approximate the operators $\mathcal{A},\mathcal{B},\mathcal{C},I$ by the finite difference method and obtain $N_z\times N_z$ matrices A,B,C,I and M. The resulting system of equations is given by

$$v^T\Big(Mv+q\Big)=0,\quad Mv+q\geq 0,\quad v\geq 0,$$

which is a plain-vanilla *Linear Complementarity Problem* (LCP) that can be solved with standard routines! I recover the threshold as $z_l\equiv\min_i\{\,i\mid v_i>0\}$

To solve for the equilibrium, I need to iterate over the LCP. The reason is that the firm growth function $g(z)$ depends on z_l. First, I start with a guess for z_l, compute g and solve the LCP. I then use the new threshold z_l' to compute g ans solve the LCP again. I stop when the difference between the productivity threshold in between iterations satisfies $|z_l'-z_l|<\varepsilon$ where $\varepsilon>0$ is a pre-set tolerance level.

To compute the joint-distribution of firm size N and productivity z, I use the Pareto Extrapolation method developed in Gouin-Bonenfant and Toda 2018. Since the method applies to discrete-time model, I "time discretize" the model over short time intervals of one week. The first step of the Pareto extrapolation algorithm computes a value for the Pareto exponent of the size distribution ζ. The second step computes a discrete approximation of the distribution φ over a finite grid accounting for movements of particles in and out of the grid. In order to compute moments, I use the truncated distribution φ and apply the correction terms provided in

Gouin-Bonenfant and Toda 2018, which involve the Pareto exponent ζ.

A.2 Data

A.2.1 Variables construction

I now describe the methodology used to construct the main firm-level variables: value-added, employment, wage and capital stock. First, I use four variables which are constructed by summing line items from the corporate tax returns (T2 forms): `gross profits`, `wages and salaries`, `tangible capital assets`, and `intangible capital assets`. The last two variables (`tangible capital assets` and `intangible capital assets`) represent book values of assets net of accumulated depreciation. First, I construct a measure of gross value added at the firm which is consistent with the income approach in the *System of National Accounts*. Denote by $Y_{i,t}$ the output of firm i in year t by

$$Y_{i,t} \equiv \texttt{gross profits}_{i,t}$$

$$+\texttt{wages and salaries}_{i,t}.$$

This approach is consistent with the income approach to measuring GDP which, in the corporate sector, sums the income which accrues to firm owners (gross operating income) and the income which accrues to workers (worker compensation). Note that the following accounting identity

$$\texttt{gross profits} \equiv \texttt{revenue}_{i,t} - \texttt{cost of goods sold}_{i,t}$$

$$-\texttt{operating expense}_{i,t} - \texttt{wages and salaries}_{i,t}$$

implies that the value-added approach to computing value-added also holds at the firm-level:

$$Y_{i,t} \equiv \texttt{revenue}_{i,t} - \texttt{cost of goods sold}_{i,t}$$

$$-\texttt{operating expense}_{i,t}$$

The labor share at the firm LS (i.e. wages, salaries and benefits as a share of output) is computed as

$$LS_{i,t} \equiv \texttt{wages and salaries}/Y_{i,t}.$$

I winsorize the labor share at the 1% level in the upper tail. To be precise, I replace the values of $LS_{i,t}$ above the 99th percentile by the value of the 99th percentile (approximately 20).

Finally, I remove from the main sample firm-year observations that either have (1) negative value-added, and (3) missing values in employment, value-added, tangible capital assets, intangible capital assets or industry code. The resulting extract (from now on the *main sample*) covers roughly approximately 60% of employment and GDP amongst the industries covered and 40% of economy-wide employment and GDP over the 2000-2015 period.

Employment $N_{i,t}$ is obtained by averaging the number of employees declared on the monthly PD7 forms filed throughout the year. This procedure avoids attributing larger employment levels to firms with high turnover. If for instance, we had summed the number workers who worked at a firm throughout the year, we would be imputing a large workforce to firms with a high labor turnover (i.e. a high share of employees who only during part of the year). The average wage $w_{i,t}$ and labor productivity $LP_{i,t}$ are thus given by

$$w_{i,t} = \texttt{wages and salaries}_{i,t}/N_{i,t}, \quad LP_{i,t} = Y_{i,t}/N_{i,t}.$$

The value of the stocks of tangible and intangible capital assets are directly measured at book value net of accumulated depreciation on the firm's balance sheet. The indicator variable `foreign`

takes value one if the country of residence of the ultimate shareholder or group of shareholders is not Canada and takes the value zero otherwise. The indicator variable `subsidiary` takes the value one if and only if the firm is owned by another firm (i.e. the ownership threshold is 50%). Similarly, the indicator variable `parent` takes the value one if and only if the firm is the highest legal operating entity in an ownership structure. In order to be "the highest legal operating entity in an ownership structure", a firm must (1) have at least one subsidiary and (2) not be the subsidiary of another firm. Finally, the indicator variable `fdi` indicates the presence of inward or outward direct investment flows. The threshold is 10% meaning that the variable `fdi` takes value one if (1) the firm has at least 10% of the voting equity in a foreign firm or (2) the firm has 10% of its voting equity owned by a foreign firm.

Table A.1: Summary statistics (Main sample, 2000-2015)

Variable	Observations	Mean	Std. Dev.
revenue	3084182	1.01e+07	1.94e+08
cost of goods sold	3084182	5684233	1.38e+08
wages and salaries	3084182	1842980	3.16e+07
operating expenses	3084182	1562873	5.92e+07
intangible assets	3084182	600173.7	2.64e+07
tangible assets	3084182	2500054	9.15e+07
employment	3084182	38.63308	571.1065
wage	3084182	42845.83	59672.26
labor productivity	3084182	60014.69	215973.2
value-added	3084182	2809567	5.18e+07
R&D expenditures	3084182	53235.86	2778371
foreign	3084182	.0149553	.1213742
subsidiary	3084182	.0508939	.2197811
fdi	3084182	.0333246	.1794827

A.2.2 Sample Validation

I compute the aggregate labor share, I use data from the *National Accounts* (Statistics Canada table 380-0063) and apply the same methodology as Koh, Santaeulàlia-Llopis, and Zheng 2016, which assumes that components of income which are ambiguous (i.e. taxes net of subsidies and net mixed income) have a labor share equal to the aggregate labor share. This methodology implies the following formula:

$$LS_{\text{NAC}} \equiv \frac{\texttt{worker compensation}}{\texttt{worker compensation} + \texttt{gross operating surplus}}.$$

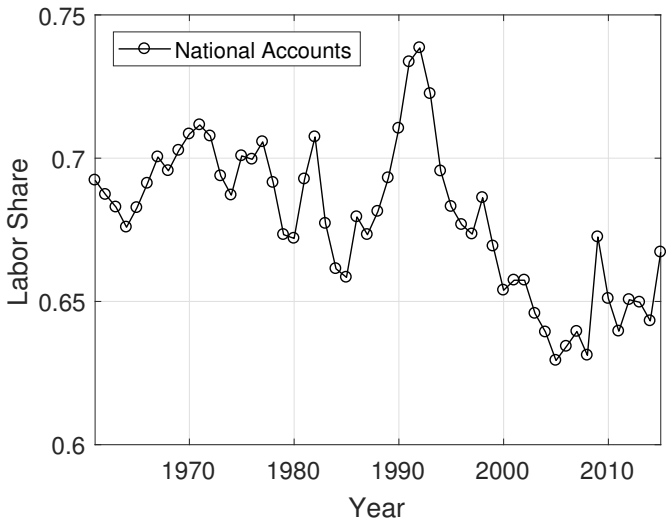

Figure A.1: Aggregate labor share (Canada, 1961-2015)

Figures A.2 and A.2 plot the labor share in the National Accounts and in the main sample. A few remarks are in order. First, the Canadian labor share in the National Accounts has sustained a large decline over the course of the 1990s and early-2000s but has then somewhat recovered over the course of the 2005-2015 period. Second, the aggregate labor share in the main sample has a similar level and trend as the one in the National account over the period where both datasets are available (2000-2015). Third, the labor share is a bit lower in the main main sample over the 2000-2004 period. Overall, the level and dynamics of the labor share is consistent with the

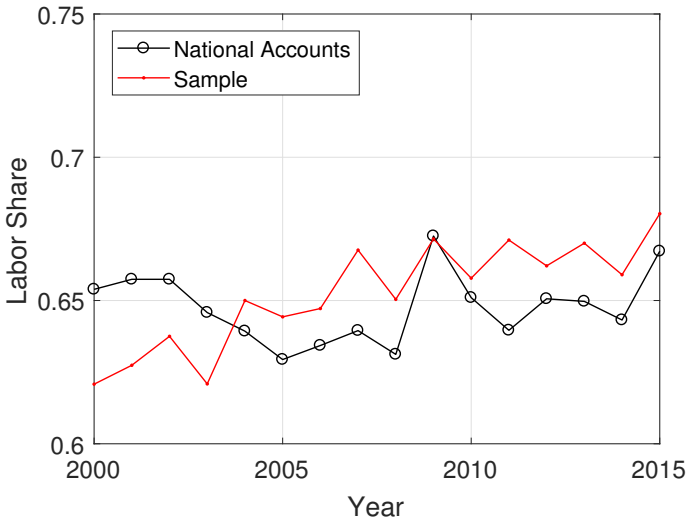

Figure A.2: Aggregate labor share (Canada, 2000-2015)

aggregate data.[3]

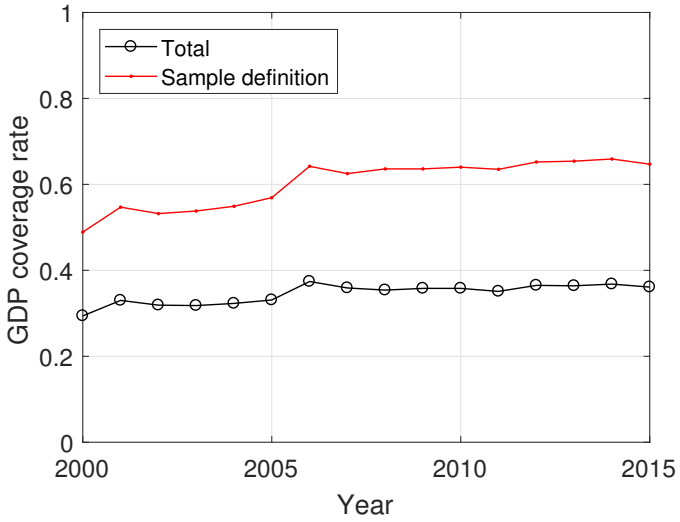

Figure A.3: Coverage rate of GDP

Figure A.3 plots the coverage rate of the main sample in terms of gross value added, benchmarked against aggregate data in the National Accounts. The "total" uses the aggregate data for all industries while the "sample definition" line corresponds to the aggregate data obtained

[3]Unfortunately, Table 383-0033 does not allow for the computation of the labor share at the industry level.

by excluding the some industries (Agriculture, Mining, Utilities, Education and Health - NAICS 11, 21, 22, 61, 62, 92), exactly as in the main sample. The "sample definition" is fairly high, at around 60%. As expected, the "total" coverage is lower, at around 40%.

A.2.3 Detailed job flows

Table A.2: Detailed job flows.

Component	Data	Model
Job creation (continuers)	0.092	0.057
Job creation (entrants)	0.019	0.018
Job destruction (continuers)	0.082	0.039
Job destruction (exiters)	0.016	0.036

A.2.4 Averages by labor productivity deciles

Within each industry-year (where industries are defined as 2-digit NAICS sectors), I sort firms by labor productivity (i.e. value-added per worker). I then compute 9 thresholds and assign a "labor productivity decile" to each firm-year observation, so that each decile contains 10% of the firms. Within each decile, I compute the wage, labor productivity and capital per worker by computing an employment-weighted average of firm-level wages, labor productivity and capital per worker. Within each industry-year bin, I then normalize the average wage and labor productivity by their value in the first decile. Finally, I average over all industries within a year by applying employment weights and then compute a simple average of each variable over 2000-2015.

To compute the labor share and output share within each decile. I simply sum wages and value-added within each industry-year-decile bin. The labor share is constructed as the ratio of wages to value-added. The output share of a decile is compute as the share of a decile's

value-added within its industry-year bin. Finally, I average over all industries within a year by applying value-added weights and compute a simple average of each variable over 2000-2015.

A.2.5 Industry-level dataset

To construct the industry-level dataset, I use the main sample which covers most industries and most firms in Canada over the 2000-2015 and sort firms by labor productivity within industry-year bins and assign to each firm a productivity quintile. For example, the 20% most productive firms are in the 5^{th} quintile while the bottom 20% least productive firms are in the 1^{st} quintile. An industry is defined according to 3-digit NAICS definitions. I then compute measures of wage, labor productivity and output share within each decile exactly as in described in Appendix A.2.4. Within each industry-year bin, I also compute the labor share LS, the average (unweighted) firm labor share $\bar{LS}$ as well as the employment and output shares. I restrict the sample to industry-year observations which have at least 100 firms in every year and censor the observations of $\bar{LS}$. I am left with a balanced panel dataset covering 69 industries over the period 2000-2015

Appendix B

Appendix of Chapter 2

B.1 Asymptotic problem

In this appendix we describe how to derive the asymptotic homogeneous problem in an abstract dynamic programming setting. For the notation, we follow Ma and Stachurski (2018). Let

- X be a set called the *state space*;
- A be a set called the *action space*;
- $\Gamma : X \twoheadrightarrow A$ be a nonempty correspondence called the *feasible correspondence*;
- $g : X \times A \to X$ be a function called the *law of motion*;
- $\mathcal{V}$ be a subset of all functions from X to $\cup -\infty$ called the set of *candidate value functions*;
- $Q : X \times A \times \mathcal{V} \to R \cup -\infty$ be a map called the *state-action aggregator*.

Then we say that the value function $v \in \mathcal{V}$ satisfies the Bellman equation if

$$v(x) = \max_{a \in \Gamma(x)} Q(x, a, v(g(x, a))) \tag{B.1}$$

for all $x \in X$.

Definition 2 (Asymptotic homogeneity). We say that the dynamic programming problem is *asymptotically homogeneous* if has the following properties:

- $X = X_1 \times X_2$, where $_+ \subset X_1 \subset$;
- $\Gamma(x) = \Gamma_1(x_1,x_2) \times \Gamma_2(x_2)$, where $x = (x_1,x_2) \in X_1 \times X_2$ and $^d_+ \subset \Gamma_1(x_1,x_2) \subset^d$ for some d;
- $g(x,a) = g_1(x_1,x_2,a_1,a_2) \times g_2(x_2,a_2)$, where $x = (x_1,x_2) \in X_1 \times X_2$ and $(a_1,a_2) \in \Gamma_1(x_1,x_2) \times \Gamma_2(x_2)$;
- $\lim_{\lambda\to\infty} \frac{1}{\lambda}\Gamma_1(\lambda x_1,x_2) = \widetilde{\Gamma}_1(x_1,x_2)$ exists for $(x_1,x_2) \in X_1 \times X_2$;
- $\lim_{\lambda\to\infty} \frac{1}{\lambda} g_1(\lambda x_1,x_2,\lambda a_1,a_2) = \widetilde{g}_1(x_1,x_2,a_1,a_2)$ exists for $(x_1,x_2) \in X_1 \times X_2$ and $(a_1,a_2) \in \Gamma_1(x_1,x_2) \times \Gamma_2(x_2)$;
- $\lim_{\lambda\to\infty} \frac{1}{\lambda} Q(\lambda x_1,x_2,\lambda a_1,a_2,\lambda v) = \widetilde{Q}(x_1,x_2,a_1,a_2,v)$ exists.

Suppose that the dynamic programming problem is asymptotically homogeneous. Then

1. $\widetilde{\Gamma}_1$ is homogeneous of degree 1 in x_1: for any $\lambda > 0$ we have

$$\widetilde{\Gamma}_1(\lambda x_1,x_2) = \lambda\widetilde{\Gamma}_1(x_1,x_2).$$

2. $\widetilde{g}_1$ is homogeneous of degree 1 in (x_1,a_1): for any $\lambda > 0$ we have

$$\widetilde{g}_1(\lambda x_1,x_2,\lambda a_1,a_2) = \lambda\widetilde{g}_1(x_1,x_2,a_1,a_2).$$

3. $\widetilde{Q}$ is homogeneous of degree 1 in (x_1,a_1,v): for any $\lambda > 0$ we have

$$\widetilde{Q}(\lambda x_1,x_2,\lambda a_1,a_2,\lambda v) = \lambda\widetilde{Q}(x_1,x_2,a_1,a_2,v).$$

Proof. By the definition of $\widetilde{\Gamma}_1$, for any $\lambda > 0$ we have

$$\begin{aligned}\widetilde{\Gamma}_1(\lambda x_1, x_2) &= \lim_{\lambda' \to \infty} \frac{1}{\lambda'} \Gamma_1(\lambda' \lambda x_1, x_2) \\ &= \lambda \lim_{\lambda' \to \infty} \frac{1}{\lambda' \lambda} \Gamma_1(\lambda' \lambda x_1, x_2) = \lambda \widetilde{\Gamma}_1(x_1, x_2).\end{aligned}$$

The proofs of the other claims are similar. □

When the dynamic programming problem is asymptotically homogeneous, we define the asymptotic problem as follows.

Definition 3. Suppose that the dynamic programming problem is asymptotically homogeneous. Then the Bellman equation of the asymptotic problem corresponding to (B.1) is defined by

$$v(x_1, x_2) = \max_{(a_1, a_2) \in \widetilde{\Gamma}_1(x_1, x_2) \times \Gamma_2(x_2)} \widetilde{Q}(x_1, x_2, a_1, a_2, v(\widetilde{g}_1(x_1, x_2, a_1, a_2), g_2(x_2, a_2))). \tag{B.2}$$

The following lemma shows that we can reduce the dimension of the asymptotic problem by 1.

Suppose that the dynamic programming problem is asymptotically homogeneous. Consider the following "normalized" Bellman equation:

$$\widetilde{v}(x_2) = \max_{(a_1, a_2) \in \widetilde{\Gamma}_1(1, x_2) \times \Gamma_2(x_2)} \widetilde{Q}(1, x_2, a_1, a_2, \widetilde{g}_1(1, x_2, a_1, a_2) \widetilde{v}(g_2(x_2, a_2))). \tag{B.3}$$

If (B.3) has a solution $\widetilde{v}(x_2)$, then $v(x_1, x_2) = x_1 \widetilde{v}(x_2)$ is a solution to the asymptotic Bellman equation (B.2). Furthermore, letting $\widetilde{a} = (\widetilde{a}_1, \widetilde{a}_2)$ be the policy function of the normalized Bellman equation (B.3), the policy function $a = (a_1, a_2)$ of the asymptotic Bellman equation (B.2) is given by $a_1(x_1, x_2) = x_1 \widetilde{a}_1(x_2)$ and $a_2(x_1, x_2) = \widetilde{a}_2(x_2)$.

Proof. Immediate by multiplying both sides of (B.3) by $x_1 > 0$ and using the homogeneity of $\widetilde{\Gamma}_1, \widetilde{g}_1, \widetilde{Q}$ established in Lemma B.1. □

The following proposition shows that if a dynamic programming problem is asymptotically homogeneous, then the value function and policy functions are asymptotically linear.

Proposition 11. *Suppose that the dynamic programming problem is asymptotically homogeneous. Suppose that the Bellman equation* (B.1) *has a solution $v(x)$, and it can be computed by value function iteration starting from $v(x) \equiv 0$. Then under some regularity conditions, the value function and policy functions are asymptotically linear: we have*

$$\begin{aligned} v(x_1,x_2) &= x_1\widetilde{v}(x_2) + o(x_1), \\ a_1(x_1,x_2) &= x_1\widetilde{a}_1(x_2) + o(x_1), \\ a_2(x_1,x_2) &= \widetilde{a}_2(x_2) + o(x_1) \end{aligned}$$

as $x_1 \to \infty$, where $\widetilde{v}(x_2)$, $\widetilde{a}_1(x_2)$, and $\widetilde{a}_2(x_2)$ are defined as in the normalized Bellman equation (B.3).

Proof. Define the operator $T : \mathcal{V} \to \mathcal{V}$ by the right-hand side of (B.1). Let $v^{(0)} \equiv 0$ and $v^{(k)} = Tv^{(k-1)} = T^k 0$. Let us show by induction that

$$\lim_{\lambda\to\infty} \frac{1}{\lambda} v^{(k)}(\lambda x_1, x_2) = \widetilde{v}^{(k)}(x_1,x_2)$$

exists. If $k = 0$, the claim is trivial since $v^{(0)} \equiv 0$. Suppose the claim holds for some $k-1$. Then by Lemma B.1, we obtain

$$\begin{aligned} \frac{1}{\lambda} v^{(k)}(\lambda x_1, x_2) &= \frac{1}{\lambda}(Tv^{(k-1)})(\lambda x_1, x_2) \\ &= \max_{\substack{(a_1,a_2)\in \\ \frac{1}{\lambda}\Gamma_1(\lambda x_1,x_2)\times\Gamma_2(x_2)}} Q\left(\lambda x_1, x_2, \lambda a_1, a_2, v^{(k-1)}\left(\lambda \frac{1}{\lambda} g_1(\lambda x_1, x_2, \lambda a_1, a_2), g_2(x_2,a_2)\right)\right). \end{aligned}$$

Using the asymptotic homogeneity of Γ_1, g_1, Q established in Lemma B.1, the asymptotic homogeneity of $v^{(k-1)}$, and assuming that we can interchange the limit and maximization (, assuming enough conditions to apply the Maximum Theorem), it follows that $v^{(k)}$ is asymptotically homogeneous. Since by assumption $v^{(k)} \to v$ as $k \to \infty$ point-wise, assuming that the limit of

$k \to \infty$ and $\lambda \to \infty$ can be interchanged (which is the case if $v^{(k)}$ converges to v monotonically, which is often the case in particular applications), then v is asymptotically homogeneous in the sense that $\lim_{\lambda\to\infty} \frac{1}{\lambda} v(\lambda x_1, x_2)$ exists.

Now that asymptotic homogeneity of v is established, from (B.1) we obtain

$$v(\lambda x_1, x_2) = \max_{a \in \Gamma(\lambda x_1, x_2)} Q(\lambda x_1, x_2, a, v(g(\lambda x_1, x_2, a))).$$

Dividing both sides by $\lambda > 0$ and letting $\lambda \to \infty$, using the asymptotic homogeneity of Γ_1, g_1, Q, and v, we obtain the asymptotic Bellman equation (B.2). Thus if in particular (B.3) has a unique solution $\widetilde{v}(x_2)$, by Lemma B.1 it must be

$$\lim_{\lambda\to\infty} \frac{1}{\lambda} v(\lambda x_1, x_2) = x_1 \widetilde{v}(x_2).$$

Consequently, setting $x_1 = 1$ and $\lambda = x_1$, we obtain $v(x_1, x_2) = x_1 \widetilde{v}(x_2) + o(x_1)$. The proof for the policy functions is similar. □

B.2 Proofs

B.2.1 Proof of results in Section 2.3

First let us derive the asymptotic Euler equation (2.5). The Euler equation in the KS model is

$$c_t^{-\gamma} = \beta_s (1-p) R E[c_{t+1}^{-\gamma} | s].$$

Let $w_t = w$. Then $c_t = \bar{c}_s w$, and using the budget constraint, we obtain

$$w' = w_{t+1} = R(w_t - c_t) = R(1 - \bar{c}_s) w.$$

Noting that $c_{t+1} = \bar{c}_{s'} w'$ and combining the above equations, we obtain (2.5).

Next, let us show that the asymptotic Euler equation (2.5) has a (necessarily unique)

solution if and only if the spectral condition (2.6) holds. Setting $x_s = \bar{c}_s^{-\gamma}$, (2.5) can be rewritten as

$$x_s = \left(1 + \left((1-p)R^{1-\gamma}\beta_s \sum_{s'=1}^{S} p_{ss'}x_{s'}\right)^{1/\gamma}\right)^{\gamma}.$$

Setting $x = (x_1, \ldots, x_S)'$, we can express this equation as

$$x = \left(1 + (Kx)^{1/\gamma}\right)^{\gamma}, \tag{B.4}$$

where $K = (1-p)R^{1-\gamma}BP$ and $B = (\beta_1, \ldots, \beta_S)$. By the results in Borovička and Stachurski (2019), (B.4) has a positive solution if and only if $\rho(K) < 1$, in which case the solution is unique. Since $\rho(K) = (1-p)R^{1-\gamma}\rho(BP)$, a necessary and sufficient condition for the existence of a solution is the spectral condition (2.6).

Proof of Proposition 9. If $G_{t+1} \leq 1$ always, then by (2.8) we have

$$M_{ss'}(z) = E[G_{t+1}^z | s_t = s, s_{t+1} = s'] \leq 1$$

for all $z \geq 0$. Therefore $(1-p)\rho(P \odot M(z)) \leq (1-p)\rho(P) = 1-p < 1$, so (2.10) does not have a solution $z > 0$.

Suppose that $M(z)$ is finite for all $z > 0$ and P is irreducible. Define $A(z) = P \odot M(z)$. Define the $S \times S$ matrix $B(z)$ by $B_{ss}(z) = p_{ss}M_{ss}(z) > 0$ for the s satisfying the assumption $p_{ss}\Pr(G_{t+1} > 1 | s_t = s_{t+1} = s) > 0$, and 0 for all other entries. Then clearly $A(z) \geq B(z) \geq 0$ entry-wise, so

$$\infty > \rho(P \odot M(z)) = \rho(A(z)) \geq \rho(B(z)) = p_{ss}E[G_{t+1}^z | s_t = s_{t+1} = s] \to \infty$$

as $z \to \infty$. Since $(1-p)\rho(P \odot M(0)) = (1-p)\rho(P) = 1-p < 1$, by the intermediate value theorem there exists $z = \zeta > 0$ such that (2.10) holds. Uniqueness is proved in Beare and Toda (2017b). □

B.2.2 Proof of results in Section 2.4

We use the following notations. Let $\widetilde{\beta} = \beta(1-p)$ be the effective discount factor. For a vector $v = (v_1, \ldots, v_S)'$, let $v^{(\alpha)} = (v_1^\alpha, \ldots, v_S^\alpha)'$ be the vector of α-th powers and (v) the diagonal matrix whose s-th diagonal element is v_s.

The following proposition characterizes the solution to the capitalist's optimal consumption-savings problem.

Proposition 12. *Let $z = (z_1, \ldots, z_S)'$ be the vector of gross excess returns. A solution to the optimal consumption-savings problem exists if and only if*

$$\widetilde{\beta} R_f^{1-\gamma} \rho((z^{(1-\gamma)})P) < 1. \tag{B.5}$$

Under this condition, the value function and optimal consumption rule are

$$V_s(w) = b_s \frac{w^{1-\gamma}}{1-\gamma}, \tag{B.6a}$$

$$c_s(w) = b_s^{-1/\gamma} w, \tag{B.6b}$$

where $b = (b_1, \ldots, b_S)' \gg 0$ is the smallest solution to the system of nonlinear equations

$$b_s = \left(1 + (\widetilde{\beta}(z_s R_f)^{1-\gamma} E[b_{s'}|s])^{1/\gamma}\right)^\gamma, \quad s = 1, \ldots, S. \tag{B.7}$$

Proof. Immediate from Toda (2019, Proposition 1). □

Let us simplify the equilibrium condition (2.34) by exploiting the linearity in Proposition 12. Using the budget constraint (2.33) and the optimal consumption rule (B.6b), the individual wealth dynamics is

$$w' = z_s R_f (1 - b_s^{-1/\gamma}) w =: G_s w.$$

Letting W_s be the aggregate wealth held by agents in state s, by accounting we obtain

$$W_{s'} = p\pi_{s'} w_0 + (1-p)\sum_{s=1}^{S} p_{ss'} G_s W_s.$$

Letting $\pi = (\pi_1, \ldots, \pi_S)'$, $G = (G_1, \ldots, G_S)'$, and $W = (W_1, \ldots, W_S)'$, in matrix form this becomes

$$W = pw_0\pi + (1-p)P'(G)W \iff W = pw_0(I-(1-p)P'G)^{-1}\pi.$$

Let $m_s = b_s^{-1/\gamma} \in (0,1)$ be the marginal propensity to consume out of wealth in state s and $m = (m_1, \ldots, m_S)'$. Then the vector of saving rates is given by $1-m$, where $1 = (1, \ldots, 1)'$ is the vector of ones. Using this, the aggregate capital supply is given by

$$K = (1-m)'W = pw_0(1-m)'(I-(1-p)P'G)^{-1}\pi, \tag{B.8}$$

assuming $(1-p)\rho(P'G) < 1$. (If this inequality is violated, we just set $K = \infty$.) On the other hand, by (2.32) the aggregate capital demand is

$$K = \left(\frac{R_f - 1 + \delta}{A\alpha}\right)^{\frac{1}{\alpha-1}}. \tag{B.9}$$

Equating (B.8) and (B.9), the market clearing condition (2.34) becomes

$$0 = f(R_f) := pw_0(1-m)'(I-(1-p)P'G)^{-1}\pi - \left(\frac{R_f - 1 + \delta}{A\alpha}\right)^{\frac{1}{\alpha-1}}. \tag{B.10}$$

The following theorem shows that a stationary equilibrium exists and that the stationary wealth distribution has a Pareto upper tail.

Theorem 1. A stationary equilibrium exists if and only if there exists $R > 1-\delta$ such that

$$\widetilde{\beta} R^{1-\gamma} \rho((z^{(1-\gamma)})P) < 1 \tag{B.11}$$

and $f(R) < 0$, where f is given by (B.10). The equilibrium is unique if $\gamma < 1$. If in addition $p_{ss} > 0$ for all s and $G_s > 1$ for some s, then the stationary wealth distribution has a Pareto upper tail with exponent $\zeta > 1$ that satisfies

$$(1-p)\rho(PG^{(\zeta)}) = 1. \tag{B.12}$$

Proof. The existence of equilibrium follows from a continuity argument similar to Toda (2019, Theorem 3), which we only sketch for space considerations. The condition (B.11) ensures that (B.5) holds for $R_f > R$ sufficiently close to R. Then we can show that the individual optimization problem has a solution and the aggregate wealth is finite for some range $R_f \in [R, \bar{R})$, and that the aggregate wealth (as well as supply of capital) diverges to ∞ as $R_f \uparrow \bar{R}$. Since $f(R) < 0$ by assumption and $f(\bar{R}) = \infty$, by the intermediate value theorem there exists $R_f \in (R, \bar{R})$ that satisfies the market clearing condition (B.10). Uniqueness of equilibrium when $\gamma < 1$ follows by the exact same argument as in Toda (2019, Theorem 3).

The Pareto tail result follows from the general theorem in Beare and Toda (2017b) and a similar argument to Toda (2019, Theorem 4). □

Numerically solving for the equilibrium is straightforward. Given the guess of the interest rate $R_f > R$, solve for the fixed point $b = (b_s)$ using (B.7), and solve (B.10) to obtain the equilibrium risk-free rate.

B.2.3 Proof of results in Section 2.5

Proof of Proposition 10. The maximization problem (2.47) is equivalent to

$$\max_{0 \le \theta \le \bar{\theta}_s} \frac{1}{1-\gamma} E[(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta)^{1-\gamma} | s].$$

Let $f(\theta)$ be the objective function of this problem. Since by assumption $z_S + \varepsilon_1 < 1$ and $z_1 < \cdots < z_S$, we have $z_s + \varepsilon_1 < 1$ for all s. Therefore $\bar{\theta}_s = \frac{\widetilde{R}_f}{\widetilde{R}_f - \widetilde{R}_{s1}} > 0$ and

$$f'(\theta) = E[(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta)^{-\gamma}(\widetilde{R}_{sj} - \widetilde{R}_f) | s] \to -\infty$$

as $\theta \uparrow \bar{\theta}_s$. Furthermore,

$$f''(\theta) = -[(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta)^{-\gamma-1}(\widetilde{R}_{sj} - \widetilde{R}_f)^2|s] < 0$$

for $\theta \in (0, \bar{\theta}_s)$, so f is strictly concave. Therefore there exists a unique θ_s^* that maximizes (2.47), and hence ρ_s is well-defined.

Assume $\varepsilon \neq 1$. By the discussion in the text, the Bellman equation of the asymptotic problem is

$$v_s(w) = \max_{\substack{0\leq\theta\leq\bar{\theta}_s \\ 0\leq c\leq w}} \left((1-\beta)c^{1-1/\varepsilon} + \beta E[(v_{s'}(R(\theta)(w-c)))^{1-\gamma}|s]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}},$$

where the upper bounds on c, θ ensure that $w' \geq 0$ and

$$R(\theta) = (1-\tau_w)(\widetilde{R}_f(1-\theta) + \widetilde{R}_{sj}\theta)$$

is the gross portfolio return. By homogeneity, the value function must be of the form $v_s(w) = b_s w$. Substituting into the Bellman equation, we obtain

$$b_s w = \max_{\substack{0\leq\theta\leq\bar{\theta}_s \\ 0\leq c\leq w}} \left((1-\beta)c^{1-1/\varepsilon} + \beta(w-c)^{1-1/\varepsilon} E[(b_{s'}R(\theta))^{1-\gamma}|s]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}.$$

Noting that $R(\theta)$ does not depend on s' and j is independent of s', using the definition of ρ_s in (2.47), we can rewrite this as

$$b_s w = \max_{0\leq c\leq w} \left((1-\beta)c^{1-1/\varepsilon} + \beta\rho_s^{1-1/\varepsilon}(w-c)^{1-1/\varepsilon} E[b_{s'}^{1-\gamma}|s]^{\frac{1-1/\varepsilon}{1-\gamma}} \right)^{\frac{1}{1-1/\varepsilon}}.$$

For notational simplicity let $\kappa_s = \rho_s E[b_{s'}^{1-\gamma}|s]^{\frac{1}{1-\gamma}}$. Then the above problem becomes equivalent to

$$\max_c \frac{1}{1-1/\varepsilon} \left((1-\beta)c^{1-1/\varepsilon} + \beta\kappa_s^{1-1/\varepsilon}(w-c)^{1-1/\varepsilon} \right).$$

Clearly this is a strictly concave function in c. Taking the first-order condition and solving for c, we obtain

$$c = \frac{(1-\beta)^{\varepsilon}}{(1-\beta)^{\varepsilon} + \beta^{\varepsilon}\kappa_s^{\varepsilon-1}} w.$$

Substituting into the Bellman equation, after some algebra we obtain

$$b_s = \left((1-\beta)^{\varepsilon} + \beta^{\varepsilon}\rho_s^{\varepsilon-1} E[b_{s'}^{1-\gamma}|s]^{\frac{\varepsilon-1}{1-\gamma}}\right)^{\frac{1}{\varepsilon-1}},$$

which is (2.49). The optimal consumption rule then simplifies to $c = (1-\beta)^{\varepsilon} b_s^{1-\varepsilon}$ and we obtain the optimal investment rule using $\theta = \frac{I}{w-c}$.

To complete the proof it remains to show that the system of nonlinear equations (2.49) has a solution. For this purpose let us write $\sigma = \frac{1-\gamma}{\varepsilon-1}$ and $x_s = b_s^{1-\gamma}$. Then we can rewrite (2.49) as

$$x_s = \left((1-\beta)^{\varepsilon} + (\beta^{\varepsilon\sigma}\rho_s^{1-\gamma} e[x_{s'}|s])^{1/\sigma}\right)^{\sigma},$$

which is equivalent to

$$x = ((1-\beta)^{\varepsilon} + (Kx)^{1/\sigma})^{\sigma}$$

for $x = (x_1, \ldots, x_S)'$ and $K = \beta^{\varepsilon\sigma}\left(\rho_1^{1-\gamma}, \ldots, \rho_S^{1-\gamma}\right) P$. Since this equation is essentially identical to Equation (24) in Borovička and Stachurski (2019), by their Theorem 3.1, a necessary and sufficient condition for the existence of a unique fixed point is $\rho(K)^{1/\sigma} < 1$, which is equivalent to (2.48).

Finally we briefly comment on the case $\varepsilon = 1$. Although this case requires a separate treatment, it turns out that the equations are valid by taking the limit $\varepsilon \to 1$. To show (2.49), define $g(\varepsilon) = \log((1-\beta)^{\varepsilon} + \beta^{\varepsilon}\kappa^{\varepsilon-1})$ for $\kappa > 0$. Then as $\varepsilon \to 1$ we obtain

$$\log((1-\beta)^{\varepsilon} + \beta^{\varepsilon}\kappa^{\varepsilon-1})^{\frac{1}{\varepsilon-1}} = \frac{g(\varepsilon)}{\varepsilon-1} = \frac{g(\varepsilon)-g(1)}{\varepsilon-1} \to g'(1).$$

But since

$$g'(\varepsilon) = \frac{(1-\beta)^{\varepsilon}\log(1-\beta) + \beta^{\varepsilon}\kappa^{\varepsilon-1}\log(\beta\kappa)}{(1-\beta)^{\varepsilon} + \beta^{\varepsilon}\kappa^{\varepsilon-1}} \to (1-\beta)\log(1-\beta) + \beta\log(\beta\kappa)$$

as $\varepsilon \to 1$, it follows that

$$((1-\beta)^{\varepsilon} + \beta^{\varepsilon}\kappa^{\varepsilon-1})^{\frac{1}{\varepsilon-1}} \to (1-\beta)^{1-\beta}(\beta\kappa)^{\beta},$$

which explains (2.49) for $\varepsilon = 1$. The existence and uniqueness of a positive solution can be proved by taking the logarithm of (2.49) and applying a contraction mapping argument to $x = \log b$. □

Proof of Lemma 3. By Proposition 10 and the budget constraint of the asymptotic problem, we obtain the the law of motion

$$w' = (1 - (1-\beta)^{\varepsilon} b_s^{1-\varepsilon})(\widetilde{R}_f(1-\theta_s^*) + \widetilde{R}_{sj}\theta_s^*)w.$$

Taking the expectation conditional on s and using the definition of G_s in (2.51), we obtain $E[w'|s] = G_s w$. By the same derivation as (B.8), a necessary condition for aggregate wealth to be finite is $\rho(P'G) < 1$. Since G is diagonal, we obtain

$$\rho(PG) = \rho((G)'P') = \rho((G)P') = \rho(P'G) < 1. \quad \square$$

Proof of Lemma 2.5.2. Since by assumption $G_{sJ} > 1$ for some s, we have $M_s(z) \to \infty$ as $z \to \infty$ for this s. Since by assumption $p_{ss} > 0$ for all s, it follows that $\rho(PD(z)) \to \infty$ as $z \to \infty$. Since $D(1) = G$, by (2.52) we obtain $\rho(PD(1)) = \rho(PG) < 1$. By the intermediate value theorem, there exists $\zeta > 1$ such that $\rho(PD(\zeta)) = 1$. Uniqueness follows from the convexity of $\rho(PD(z))$ established in Beare and Toda (2017b). The Pareto tail result follows from (2.10) with $p = 0$. □

B.3 Solution accuracy of Aiyagari model

In this appendix we evaluate the solution accuracy of the Aiyagari model in Section 2.4.1. We consider the evenly- and exponentially-spaced grids as well as simulation.

B.3.1 Evenly-spaced grid

We first consider an N-point evenly-spaced grid on $(0, \bar{w}]$, where we set the truncation point to $\bar{w} = 10, 20, 40$ and the number of points to $N = 100, 200, 400$. Therefore the wealth grid is $nd_{n=1}^{N}$, where $d = \bar{w}/N$ is the distance between grid points.[1] Table B.1 shows the relative error $\widehat{K}/K - 1$ in the aggregate capital.

Table B.1: Relative error (%) in aggregate capital for the truncation and Pareto extrapolation methods with an evenly-spaced grid.

Method:	Truncation			Pareto extrapolation		
$\bar{w}$	$N = 100$	200	400	100	200	400
10	-40.00	-39.69	-39.62	0.214	0.105	0.052
20	-33.58	-32.88	-32.63	0.430	0.097	0.043
40	-26.99	-27.56	-27.03	3.588	0.331	0.046

Note: N: number of grid points; $\bar{w}$: wealth truncation point.

We can make a few observations from Table B.1. First, the conventional truncation method is extremely poor at calculating the aggregate capital: the relative error is about 27–40% depending on the specification. On the other hand, the Pareto extrapolation method is astonishingly more accurate. Second, for the truncation method, choosing a larger truncation point $\bar{w}$ somewhat improves the accuracy, probably because it misses less of the upper tail. On the other hand, the accuracy in the Pareto extrapolation method is not necessarily monotonic in $\bar{w}$. There seems to be a trade-off between less truncation (larger $\bar{w}$) and higher density of grid points (smaller $d = \bar{w}/N$).

Figure B.1 shows the stationary wealth distribution using $\bar{w} = 10$ and $N = 100$. The two methods are indistinguishable except at the upper tail. To study the tail behavior, Figure B.2 plots the tail probability in a log-log scale. As we can see, the graphs show a straight line pattern, which is consistent with the theoretical Pareto distribution. However, the graph for the truncation method becomes concave towards the upper tail, which implies that it underestimates the tail

[1] Note that we exclude 0 from the wealth grid because in equilibrium agents never hit 0 due to the Inada condition.

probability. On the other hand, the Pareto extrapolation method shows a straight line pattern including the very top tail.

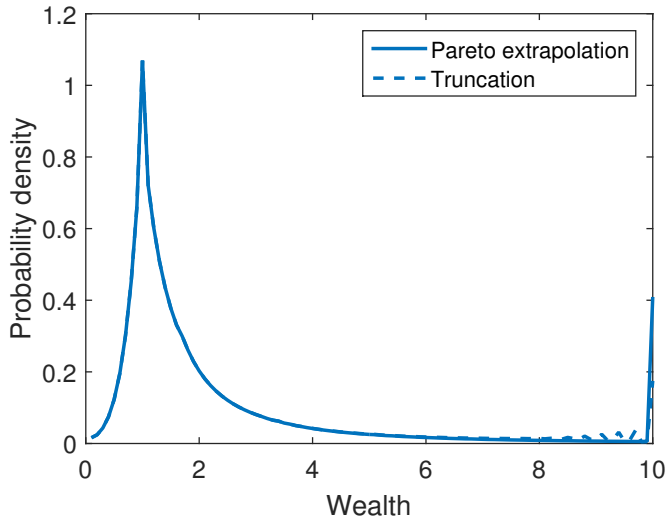

Figure B.1: Stationary wealth distribution.

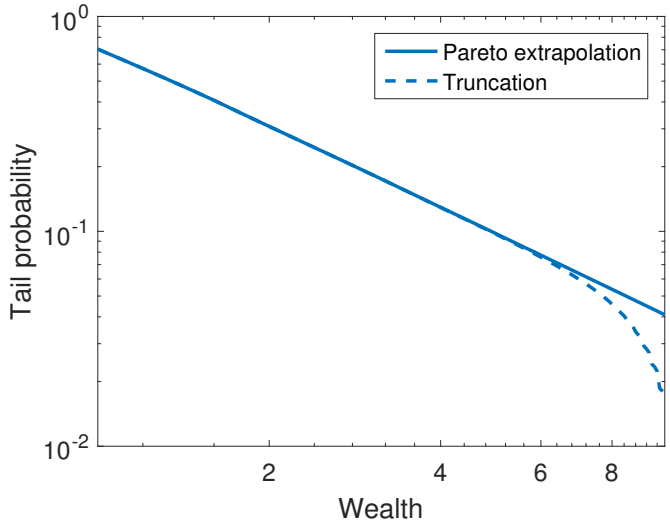

Figure B.2: Log-log plot of wealth distribution.

These seemingly small differences have an enormous impact on aggregate quantities, as we have seen in Table B.1. To assess the robustness, Figure B.3 shows the aggregate capital for the semi-analytical solution as well as the Pareto extrapolation and truncation solutions when we

change the initial wealth in the range $w_0 \in [0.2, 5]$. For all cases we set $\bar{w} = 10$ and $N = 100$. The horizontal axis shows the corresponding equilibrium Pareto exponent. The graph for the Pareto extrapolation method is indistinguishable from the semi-analytical solution except when the Pareto exponent is very close to 1 (Zipf's law), in which case the truncation method is especially poor.

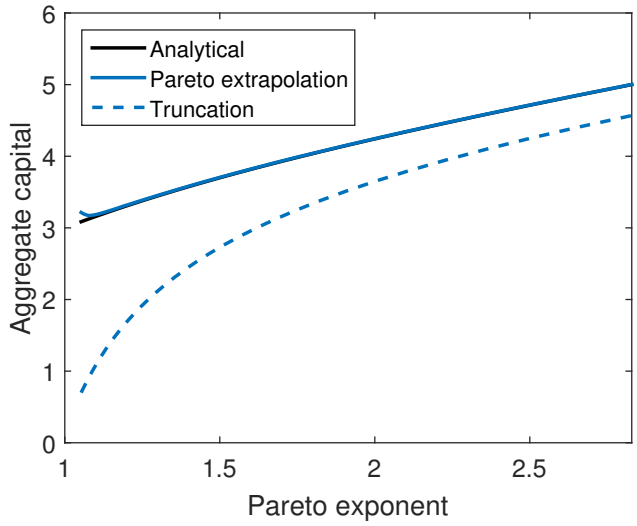

Figure B.3: Aggregate capital.

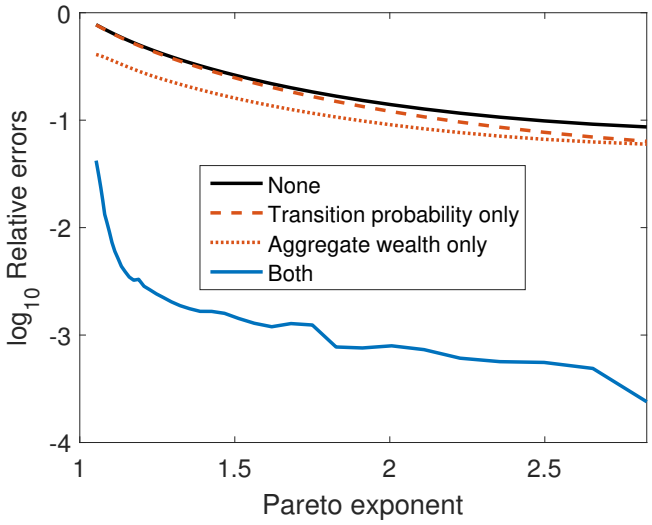

Figure B.4: Relative errors.

Figure B.4 shows the relative error $\left|\widehat{K}/K-1\right|$ in a semi log scale. For this exercise, we consider four solution methods that correspond to using/not using Pareto extrapolation when constructing the transition probability matrix and/or calculating aggregate capital. For example, "None" and "Both" in Figure B.4 correspond to the truncation and Pareto extrapolation solutions, respectively. According to the figure, using Pareto extrapolation for only one step (either constructing the transition probability matrix or calculating aggregate capital) improves the accuracy only slightly, and correcting the aggregate capital matters more. However, combining both increases the solution accuracy dramatically.

The intuition for this (surprising) result is as follows. According to (2.22), the two sources of errors introduced by the truncation method (incorrect transition probability matrix *and* incorrect aggregate wealth held by agents at the top grid point) interact with each other. With the truncation method, the last term $\pi_{sN}\frac{\zeta}{\zeta-1}w_N$ in (2.22) becomes $\widetilde{\pi}_{sN}w_N$, with typically $\widetilde{\pi}_{sN}<\pi_{sN}$. Therefore, errors in π_{sN} are inflated by a factor $\frac{\zeta}{\zeta-1}$, which is large if $\zeta>1$ is small.

B.3.2 Exponentially-spaced grid

One may argue that the poor performance of the truncation method in Table B.1 is due to the fact that the truncation point $\bar{w}=10,20,40$ is relatively small compared to the aggregate capital $K=3.4231$. What if we take $\bar{w}$ much larger, say a million? Then we can no longer use evenly-spaced grids because there will be too few points to cover the bottom of the wealth distribution. Therefore we need to consider an exponentially-spaced grid.

In more general models, the state variable may become negative (, asset holdings), which causes a problem for constructing an exponentially-spaced grid because we cannot take the logarithm of a negative number. Suppose we would like to construct an N-point exponential grid on a given interval (a,b). A natural idea to deal with such a case is as follows. [Constructing the exponential grid]

1. Choose a shift parameter $s>-a$.

2. Construct an N-point evenly-spaced grid on $(\log(a+s), \log(b+s))$.

3. Take the exponential.

4. Subtract s.

The remaining question is how to choose the shift parameter s. Suppose we would like to specify the median grid point as $c \in (a,b)$. Since the median of the evenly-spaced grid on $(\log(a+s), \log(b+s))$ is $\frac{1}{2}(\log(a+s)+\log(b+s))$, we need to take $s > -a$ such that

$$\begin{aligned}
& c = \mathbb{E}\left(\frac{1}{2}(\log(a+s)+\log(b+s))\right) - s \\
\iff & c+s = \sqrt{(a+s)(b+s)} \\
\iff & (c+s)^2 = (a+s)(b+s) \\
\iff & c^2 + 2cs + s^2 = ab + (a+b)s + s^2 \\
\iff & s = \frac{c^2 - ab}{a+b-2c}.
\end{aligned}$$

Note that in this case

$$s + a = \frac{c^2 - ab}{a+b-2c} + a = \frac{(c-a)^2}{a+b-2c},$$

so $s+a$ is positive if and only if $c < \frac{a+b}{2}$. Therefore, for any $c \in \left(a, \frac{a+b}{2}\right)$, it is possible to construct such a grid. The remaining question is how to choose c, but we can use information from the problem we want to solve. Note that by construction, half the grid points will lie on the interval (a,c). Therefore we should choose the number c such that c is a "typical" value for the state variable (, initial capital, aggregate capital in a representative-agent model, etc.).

Returning to our particular Aiyagari model, we choose the shift parameter s such that the median grid point corresponds to the capital in a representative-agent model, which in our case is $K_{\text{RA}} = ((1/(\beta(1-p)) - 1 + \delta)/(A\alpha))^{\frac{1}{\alpha-1}} = 4.5577$. Since an exponential grid allows us to use far fewer points than an evenly-spaced grid, we consider $N = 25, 50, 100$. (We have checked

that increasing N further also increases the accuracy.) For the truncation point, we consider $\bar{w} = 10^1, 10^2, 10^3, 10^4, 10^5, 10^6$. Table B.2 shows the results.

Table B.2: Relative error (%) in aggregate capital for the truncation and Pareto extrapolation methods with an exponentially-spaced grid.

Method:	Truncation			Pareto extrapolation		
$\bar{w}$	$N = 25$	50	100	25	50	100
10^1	-41.90	-40.77	-39.92	2.643	0.528	0.276
10^2	-27.54	-23.36	-21.08	-1.697	-0.186	0.588
10^3	-20.85	-15.27	-12.32	-2.423	-0.785	-0.002
10^4	-17.12	-10.78	-7.62	-2.553	-0.905	-0.184
10^5	-14.93	-8.06	-4.93	-2.546	-0.879	-0.222
10^6	-13.19	-6.27	-3.32	-2.445	-0.807	-0.210

Note: N: number of grid points; $\bar{w}$: wealth truncation point.

The accuracy of the truncation method somewhat improves by using the exponential grid with a large truncation point $\bar{w}$ and many points. However, even with a large truncation like a million and $N = 100$ grid points, the error still exceeds 3%. (Increasing N further decreases the error, but only slightly for the truncation method.) Again, the Pareto extrapolation method is overwhelmingly more accurate.

B.3.3 Simulation

We conduct simulations by recursively computing the wealth of I agents over T periods using the semi-analytical solution for the consumption policies and the risk-free rate. We initialize the wealth distribution at $t = 0$ by setting $w_0 = 1$ for all agents. For every simulation, we compute the cross-sectional mean of the wealth distribution $\widehat{K}$ in the terminal period T, which corresponds to the aggregate capital K in the model.[2] Given that $\widehat{K}$ is a random variable, we compute the relative error for $B = 10{,}000$ independent simulations and take the average to obtain the "mean relative error" defined by $\frac{1}{B}\sum_{b=1}^{B}\left|\widehat{K}_b/K - 1\right|$, where K is the true aggregate capital and $\widehat{K}_b$ is the

[2]Since there is a unit continuum of agents in the model, the average wealth is equal to the aggregate wealth.

numerical aggregate capital from simulation b. Finally, we compute a measure of the dispersion of results across simulations defined by $\left|\widehat{K}_{p95}/\widehat{K}_{p5}-1\right|$, where $\widehat{K}_{p5},\widehat{K}_{p95}$ denote the 5th and 95th percentiles of $\widehat{K}$ across simulations. Table B.3 shows the results for different combinations of sample size I and simulation length T.

Table B.3: Solution accuracy of the simulation method in the Aiyagari model.

T	I	Relative error (%)	Dispersion (%)	Time (sec)
1,000	10^3	21.93	92.97	0.07
	10^4	10.97	47.99	0.38
	10^5	6.70	27.53	3.13
	10^6	6.69	16.33	41.23
10,000	10^3	23.76	96.68	0.39
	10^4	20.42	50.07	3.76
	10^5	9.69	27.47	31.43
	10^6	6.64	16.07	414.30

Note: T: simulation length in model-years; I: number of simulated agents; Relative error: $\frac{1}{B}\sum_{b=1}^{B}\left|\widehat{K}_b/K-1\right|$, where $\widehat{K}_b$ is the aggregate capital in simulation b and K is the true value from the analytical solution; Dispersion: $\left|\widehat{K}_{p95}/\widehat{K}_{p5}-1\right|$, where $\widehat{K}_{p5},\widehat{K}_{p95}$ are the 5th and 95th percentiles of aggregate capital across $B=10{,}000$ simulations; Time: computing time of one simulation (one b) in seconds.

A few remarks are in order. First, the simulation method performs poorly on average and the results are very dispersed across simulations. Second, increasing the number of agents I helps reduce both the mean relative error and the dispersion. The gains in accuracy associated with increasing I by an order of magnitude tend to be small, consistent with the fact that the sample mean of a fat-tailed distribution converges very slowly to the population mean. In fact, in our model the Pareto exponent is $\zeta=1.28$, and the relative errors for sample sizes $I=10^4,10^6$ in Table B.3, which are about 11% and 7% respectively, are about the same order of magnitude as the error order 11.9% and 4.1% in the column for $\zeta=1.3$ in Table 2.1. Third, increasing the simulation length beyond $T=1{,}000$ does not seem to improve accuracy, suggesting that the simulated wealth distribution has already converged after 1,000 periods.

The last column of Table B.3 reports the computing time (without parallelization) associated with producing a *single* simulation using a machine equipped with an `Intel Xeon E3-1245 3.5GHz` processor and 16GB of memory. While increasing the sample size I by a factor of ten is associated with small accuracy gains, it implies a tenfold increase in the computing time. Given the inaccuracy and the large dispersion of results across simulations, we conclude that the simulation method is not a viable option for solving models with fat-tailed wealth distributions.

B.4 Dynamic programming in the MBA model

B.4.1 Euler and asset pricing equations

First, we derive the Euler and asset pricing equations. Noting that $\widetilde{R}_{sj}$ is strictly increasing in j, the borrowing constraint (2.40) can bind only in state $j = 1$. Therefore the Bellman equation (2.41) is equivalent to

$$\frac{1}{1-1/\varepsilon} v_s(w)^{1-1/\varepsilon} = \max_{c,I \geq 0} \frac{1}{1-1/\varepsilon} \left((1-\beta) c^{1-1/\varepsilon} + [v_{s'}(w')^{1-\gamma} | s]^{\frac{1-1/\varepsilon}{1-\gamma}} \right), \tag{B.13}$$

where

$$w' = \widetilde{R}_f (w + (1-\tau_h)\omega h_s - I - c) + \widetilde{R}_{s1} I \geq w.$$

Let $\mathcal{L}_s(c, I, w)$ be the Lagrangian and $\lambda_s(w), \mu_s(w)$ be the corresponding Lagrange multipliers for the borrowing constraint and the nonnegativity constraint on investment:

$$\begin{aligned} \mathcal{L}_s(c, I, w) = \frac{1}{1-1/\varepsilon} \Bigg(& (1-\beta) c^{1-1/\varepsilon} + [v_{s'}(w')^{1-\gamma} | s]^{\frac{1-1/\varepsilon}{1-\gamma}} \Bigg) \\ & + \lambda_s(w) \left(\widetilde{R}_f (w + (1-\tau_h)\omega h_s - I - c) + \widetilde{R}_{s1} I - w \right) + \mu_s(w) I. \end{aligned}$$

The first-order condition for consumption is given by

$$(1-\beta)c_s(w)^{-1/\varepsilon} = \beta\widetilde{R}_f E[v_{s'}(w')^{1-\gamma}|s]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} E[v_{s'}(w')^{-\gamma}v'_{s'}(w')|s] + \lambda_s(w)\widetilde{R}_f. \tag{B.14}$$

Differentiating both sides of the Bellman equation (B.13) by w, it follows from the Envelope Theorem that

$$v_s(w)^{-1/\varepsilon}v'_s(w) = \beta\widetilde{R}_f E[v_{s'}(w')^{1-\gamma}|s]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} E[v_{s'}(w')^{-\gamma}v'_{s'}(w')|s] + \lambda_s(w)\widetilde{R}_f. \tag{B.15}$$

By (B.14) and (B.15), we obtain the following expression for the derivative of the value function

$$v'_s(w) = (1-\beta)\left(\frac{c_s(w)}{v_s(w)}\right)^{-1/\varepsilon}. \tag{B.16}$$

Substituting (B.16) into (B.14), we obtain the consumption Euler equation

$$c_s(w)^{-1/\varepsilon} = \beta\widetilde{R}_f E[\left(\frac{v_{s'}(w')}{v_{s'}(w')^{1-\gamma}|s^{\frac{1}{1-\gamma}}}\right)^{1/\varepsilon-\gamma} c_{s'}(w')^{-1/\varepsilon}|s] + \frac{\lambda_s(w)\widetilde{R}_f}{1-\beta}. \tag{B.17}$$

The first-order condition for investment is given by

$$\beta E[v_{s'}(w')^{1-\gamma}|s]^{\frac{\gamma-1/\varepsilon}{1-\gamma}} E[v_{s'}(w')^{-\gamma}v'_{s'}(w')(\widetilde{R}_{sj}-\widetilde{R}_f)|s] + \lambda_s(w)(\widetilde{R}_{s1}-\widetilde{R}_f) + \mu_s(w) = 0. \tag{B.18}$$

Using the expression for $v'_s(w)$ from (B.16) and rearranging, we obtain the following asset pricing equation

$$E[\left(\frac{v_{s'}(w')}{v_{s'}(w')^{1-\gamma}|s^{\frac{1}{1-\gamma}}}\right)^{1/\varepsilon-\gamma} c_{s'}(w')^{-1/\varepsilon}(\widetilde{R}_{sj}-\widetilde{R}_f)|s] = -\frac{\lambda_s(w)(\widetilde{R}_{s1}-\widetilde{R}_f)+\mu_s(w)}{\beta(1-\beta)}. \tag{B.19}$$

B.4.2 Policy function and value function iteration

The strategy is to start with guesses for the value function $v_s^{\text{old}}(w)$ and policy functions $c_s^{\text{old}}(w)$, $I_s^{\text{old}}(w)$ and update by doing the following steps for each individual state (s,w). Equipped with the asymptotic solution to the individual problem $\bar{c}_s, \bar{I}_s, \bar{v}_s{}_{s=1}^{S}$, we choose the following initial guesses

$$c_s(w) = \bar{c}_s(w-w), \quad I_s(w) = \bar{I}_s(w-w), \quad v_s(w) = \bar{I}_s(w-w).$$

Then, we iterate over steps (1), (2), and (3) until convergence.

1. **Consumption decision**: First, use the old decision rules $c_s^{\text{old}}(w)$, $I_s^{\text{old}}(w)$ to construct next period's wealth $w'_{sj}(w)$ and the upper bound on consumption $c_s^{\text{ub}}(w)$:

$$\begin{aligned} w'_{sj}(w) &= \widetilde{R}_f \left(w + (1-\tau_h)\omega h_s - I_s^{\text{old}}(w) - c_s^{\text{old}}(w) \right) + \widetilde{R}_{sj} I_s^{\text{old}}(w), \\ c_s^{\text{ub}}(w) &= w + (1-\tau_h)\omega h_s - w/\widetilde{R}_f. \end{aligned}$$

   Then, update the consumption using the Euler equation (B.17) as if the borrowing constraint was slack.

$$c_s^*(w) = \left(\beta \widetilde{R}_f E \left(\frac{v_{s'}^{\text{old}}(w'_{sj}(w))}{E[v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma}|s]^{\frac{1}{1-\gamma}}} \right)^{1/\varepsilon-\gamma} c_{s'}^{\text{old}}(w'_{sj}(w))^{-1/\varepsilon} | s \right)^{-\varepsilon}.$$

   Finally, impose the upper bound on consumption so that $c_s^{\text{new}}(w) = \min c_s^*(w), c_s^{\text{ub}}(w)$ and compute the Lagrange multiplier

$$\lambda_s^{\text{new}}(w) = \frac{1-\beta}{\widetilde{R}_f} \left(c_s^{\text{new}}(w)^{-1/\varepsilon} - c_s^*(w)^{-1/\varepsilon} \right).$$

2. **Investment decision**: This step applies only to types s such that $E[\widetilde{R}_{sj}|s] > \widetilde{R}_f$, for other

types set $I_s(w) = 0$. Using $c_s^{\text{new}}(w)$, construct the following function

$$\Psi_s(w,I) = E[\left(\frac{v_{s'}^{\text{old}}(w'_{sj}(w,I))}{v_{s'}^{\text{old}}(w'_{sj}(w,I))^{1-\gamma}|s^{\frac{1}{1-\gamma}}}\right)^{1/\varepsilon-\gamma} c_{s'}^{\text{new}}(w'_{sj}(w,I))^{-1/\varepsilon}(\widetilde{R}_{sj}-\widetilde{R}_f)|s] + \frac{\lambda_s^{\text{new}}(w)(\widetilde{R}_{s1}-\widetilde{R}_f)}{\beta(1-\beta)},$$

where

$$w'_{sj}(w,I) = \widetilde{R}_f\,(w+(1-\tau_h)\omega h_s - I - c_s^{\text{new}}(w)) + \widetilde{R}_{sj}I.$$

To solve for the optimal investment, compute the root I^* that solves $\Psi_s(w,I^*) = 0$ and set $I_s(w) = I^*$. If a root does not exist and $\Psi_s(w,0) < 0$, set $I_s(w) = 0$.

3. **Value function**: Construct next period's wealth using the new policy functions

$$w'_{sj}(w) = \widetilde{R}_f\,(w+(1-\tau_h)\omega h_s - I_s^{\text{new}}(w) - c_s^{\text{new}}(w)) + \widetilde{R}_{sj}I_s^{\text{new}}(w)$$

and update the value function until convergence using the following formula

$$v_s^{\text{new}}(w) = \begin{cases} \left((1-\beta)c_s^{\text{new}}(w)^{1-1/\varepsilon} + \beta E[v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma}|s]^{\frac{1-1/\varepsilon}{1-\gamma}}\right)^{\frac{1}{1-1/\varepsilon}}, & (\varepsilon \neq 1) \\ c_s^{\text{new}}(w)^{1-\beta}\left(v_{s'}^{\text{old}}(w'_{sj}(w))^{1-\gamma}|s^{\frac{1}{1-\gamma}}\right)^{\beta}. & (\varepsilon = 1) \end{cases}$$

Appendix C

Appendix of Chapter 3

C.1 Definition of hierarchy levels

Table C.1: Organizational hierarchy and job titles (management).

Code	Job title	Scope of contribution
110	President	Is the head of the organization and is the highest employer representative.
111	Vice-President	Collaborates with the executive team to determine the strategic direction of the organization and of her sector. Is primarily responsible for her area of activity and makes recommendations concerning the organizations direction. Ensures that recommendations are implemented organization-wide or in her sector. Approves requests for human, material and financial resources in accordance with her sectors needs. Acts as a representative of the employer on matters related to her area of activity.
112	Director	Assumes a decision-making role so as to achieve the objectives for her area of activity established by the executive team. Develops and issues recommendations and contributes to her sectors short-term strategic development. Ensures annual planning and the management of the human, financial and material resources needed for her sectors operations. Determines and prioritizes the use of resources. Is accountable for establishing and adhering to the budget and for expected results.
113	Manager	Supervises the members of her team. Participates in establishing budgets, action plans, objectives and strategies for her sector. Proposes and ensures the implementation of guidelines issued by more senior levels. Ensures that targets are met within established timelines and using allocated resources. Intervenes and recommends appropriate solutions when new and complex problems arise. In the absence of an expert, could act as the reference person for her field of expertise.
114	Supervisor	Coordinates and assigns tasks to the members of his or her team. Ensures compliance with guidelines issued by senior levels. Ensures that assigned objectives are achieved within the specified time frame and with the allocated resources. Intervenes and recommends appropriate solutions when new and complex issues arise.

C.2 Data

Table C.2: Organizational hierarchy and job titles (professionals).

Code	Job title	Scope of contribution
211	Professional Expert	Acts as the reference for the organization within her area of activity and is responsible for innovative direction. Makes recommendations on the development and completion of operationally complex and conceptually important projects that require a detailed and comprehensive analysis and understanding of the area of activity as well as the organization. Generally provides functional leadership within her team.
212	Professional III	As a senior level professional, proposes solutions for improving or optimizing processes, programs and procedures. Helps resolve complex problems requiring a comprehensive and detailed analysis and understanding of all variables. Collaborates with various internal and external stakeholders. Contributes to knowledge transfer by providing functional guidance and coaching to the teams less experienced employees.
213	Professional II	Assumes a first line advisory role with internal and external stakeholders on projects related to her field of expertise. Suggests ideas and uses her analytical skills and in-depth knowledge of her field of expertise to help solve problems. Develops tools, policies and practices linked to her area of activity.
214	Professional I	Provides support to more experienced professionals from her area of activity. Conducts research and preliminary analyses and develops early drafts of tools, policies and practices. Actively participates in implementing tools linked to her field of expertise.

C.3 Proofs

C.3.1 Lemma 5

First, I obtain the expression for N_k (3.30) by substituting $q_k = q$ in (3.16). To obtain the expression for the total payroll $\sum_{k\geq 0} w_k N_k$, I first substitute the expression for N_k (Equation 3.30) and $w_k = w_0 e^{\pi k}$

$$\sum_{k\geq 0} w_k N_k = \sum_{k\geq 0} w_k \frac{\psi}{\psi+q}\left(\frac{q}{q+\psi}\right)^k e^{\pi k} = w_0 \frac{\psi}{\psi+q}\sum_{k\geq 0}\left(\frac{q}{q+\psi}e^{\pi}\right)^k.$$

Table C.3: Organizational hierarchy and job titles (technical and support).

Code	Job title	Scope of contribution
311	Technical III	Assumes a technical expert role and helps solve technically complex problems. Influences practices, processes and procedures from a technical perspective and helps develop them to optimize their quality and efficiency. Collaborates with various internal and external stakeholders. Contributes to knowledge transfer by providing functional guidance and coaching to less experienced employees from her team.
312	Technical II	Assumes a collaborator role in completing technical activities and helps solve technical problems. Helps develop and improve work processes and procedures. May contribute to knowledge transfer by providing functional guidance and coaching to less experienced employees from her team.
313	Technical I	Provides general technical support, carrying out activities in accordance with established policies and procedures. Assists more experienced technicians from her area of activity.
411	Support II	Performs administrative or operational activities linked to the organizations strategic activities in a high performance context in terms of quality. Collaborates with internal and external stakeholders. Could act as a resource person for her team to help solve simple and common problems or distribute tasks to other support team members.
412	Support I	Performs administrative or operational activities in accordance with specific directives and procedures. Resolves simple and common problems and completes assigned tasks.

Table C.4: Variance decomposition of log annual earnings.

Fixed effects	K	R^2	Properties of residuals	
			Variance	Pareto exp.
None	-	-	0.21	3.7
Age and education	19	0.16	0.18	3.8
Occupation	47	0.31	0.15	4.1

Notes: Sample size is $N = 1,790,823$; "K" is the number of fixed effects; "Pareto exponent" is the maximum-likelihood estimator of the Pareto exponent using the largest 10% observations.

Then, under that assumption that $\frac{q}{q+\psi}e^{\pi} < 1$, we have that

$$\sum_{k\geq 0}\left(\frac{q}{q+\psi}e^{\pi}\right)^{k} = \frac{1}{1-\frac{q}{q+\psi}e^{\pi}} = \frac{q+\psi}{q+\psi-qe^{\pi}}.$$

Putting the pieces together, I obtain (3.31)

$$\sum_{k\geq 0} w_k N_k = w_0 \frac{\psi}{\psi+q}\frac{q+\psi}{q+\psi-qe^{\pi}} = w_0 \frac{\psi}{\psi-q(e^{\pi}-1)}$$

I now show that the optimal level of effort a_k is given by (3.32) for all $k \geq 0$. The first step is to derive an expression for v_k. I start by substituting $q_k = q$ and $w_k = w_0 e^{\pi k}$ in (3.26).

$$\left(r+\psi+\frac{\sigma-1}{\sigma}q\right)v_k = w_0 + \pi k + \frac{\sigma-1}{\sigma}qv_{k+1}$$

Difference the equation (forward difference), I obtain

$$\left(r+\psi+\frac{\sigma-1}{\sigma}q\right)v_k = \pi + \frac{\sigma-1}{\sigma}qv_{k+1}.$$

Notice that this is a deterministic difference equation with constant increments. By setting $v_{k+1} = v_k = v$, I obtain

$$v = \frac{\pi}{r+\psi}.$$

I then obtain the desired result by substituting $v_k = v$ and $q_k = q$ in (3.25)

$$a_k = \frac{q\pi}{(r+\psi)\sigma\beta}$$

I now derive the expression for total output Y. I start by substituting the solution for a in the definition of output and re-arranging.

$$Y = \sum_{k\geq 1}(a_k N_k)^{1-\theta}(a_{k-1}N_{k-1})^{\theta} = a\sum_{k\geq 1} N_k^{1-\theta}N_{k-1}^{\theta} = a\sum_{k\geq 1} N_k\left(\frac{N_{k-1}}{N_k}\right)^{\theta}$$

Using (3.30), I obtain and using the fact that $\sum_{k\geq 0} N_k = 1$, I obtain the desired result.

$$Y = a\left(\frac{\psi+q}{\psi}\right)^{\theta}\sum_{k\geq 1} N_k = a\left(\frac{\psi+q}{q}\right)^{\theta}(1-N_0) = a\left(\frac{\psi+q}{q}\right)^{\theta}\frac{q}{q+\psi} = a\left(\frac{q}{\psi+q}\right)^{1-\theta}$$

Finally, to obtain the expression for the constraint in terms of w_0 (3.34), I first substitute the expressions for v and a in the Bellman equation

$$(r+\psi)v_0 = \log w_0 + \frac{\sigma-1}{\sigma}\frac{q\pi}{r+\psi}$$

The constrain that $v_0 \geq \mathcal{V}$ is thus equivalent to

$$\frac{\log w_0 + \frac{\sigma-1}{\sigma}\frac{q\pi}{r+\psi}}{r+\psi} \geq \mathcal{V} \iff w_0 \geq e^{(r+\psi)\mathcal{V} - \frac{\sigma-1}{\sigma}\frac{q\pi}{r+\psi}}.$$