Accounting for Wealth Concentration

Simon Toussaint

ISBN: 978-94-91870-64-4
U.S.E. Dissertation Series
USE 079
Printed by: Ridderprint

This book was typeset by the author using LaTeX.

Accounting for Wealth Concentration

Vermogensconcentratie: Meten en Weten

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 14 februari 2025 des middags te 12.15 uur

door

Simon Jonathan Toussaint

geboren op 26 februari 1997
te Utrecht

Promotoren:	Prof. dr. C.N. Teulings
	Prof. dr. B.J.P. van Bavel
Beoordelingscommissie:	Prof. dr. M. Goos
	Prof. dr. B. Jacobs
	dr. C. Martínez-Toledano
	Prof. dr. D. Waldenström
	Prof. dr. J.L. van Zanden

Acknowledgements

And now, at last, it is done. This dissertation is not the end; for there are no endings, nor beginnings. But it is *an* ending. Since I first set foot in Adam Smith Hall, now over six years ago, I have learned a great many things, from a great many individuals. With deep gratitude, I would like to thank them; without their accompaniment, I could not have made this journey.

First, Bas and Coen. I remember Bas giving a guest lecture about his (then-new) book, *the Invisible Hand?*, in my introductory course in the Research Master. I was enthralled by his account of the long-run trajectories of wealth inequality and their societal consequences. Soon after, I emailed him to ask about doing my thesis with him. A bit bemused (this was two months into my first year, so my thesis wasn't due until a year and a half later), Bas accepted. Since Bas worked at a different department, I needed an additional supervisor from USE. I remain deeply grateful to Wolter Hassink for suggesting Coen. Since our first joint meeting, in January 2019, I have never looked back. As the thesis developed into a full PhD, I have continued to realize how lucky I was with them as supervisors. Both are great researchers, with a wealth of experience and insight. Their personalities complement each other very nicely, and their different backgrounds do so as well. A sincere thank you to both.

I owe additional thanks to Maarten Goos for his support for my PhD, as well as his kind advice over the years. I am lucky to have him on my committee. Further thanks to the other committee members (Bas Jacobs, Clara Martínez-Toledano, Daniel Waldenström, and Jan Luiten van Zanden); I have learned a lot from each of you over the years, and I am honored to be vetted by you.

A further big thanks to my co-authors Amaury, Michalis, and Tim. What started out as a "small project" in fall 2020 has turned into a not-so-small Chapter 1. I have learned a lot along the way and I am happy that I got to learn this with you. I also am happy to have made some inroads into the economic history community through this project. Besides the already-mentioned, I thank Bram van Besouw, Vincent Delabastita, Matthijs Korevaar, and Peter Koudijs for inspiring discussions.

One of the greatest joys of life as a researcher is traveling and meeting new and old friends. I learned an enormous amount from my stay at the LSE in 2023. A big thanks to Maarten de Ridder for hosting me and taking me on many coffee walks through the neighborhood. I further thank Per Krusell, Ben Moll, Ricardo Reis, Matthias Doepke, Wouter den Haan for inspiring conversations; and Tomer

Ifergane and the PhDs (in particular Stephan, Lukas, Marta, Marcia, Hugo, Tiago, Adi, Marcelo, and Adrien). Thanks also to Ryan and Tyler for making me feel at home in a big city.

I also met many inspiring people at various conferences and seminars. I am particularly thrilled to have gotten to know Daniel Albuquerque, Riccardo Cioffi, Thomas Douenne, Aart Gerritsen, Albert-Jan Hummel, Eva Janssens, and Wiemer Salverda; as well as the broader WID community, in particular Mona Baraké, Yonatan Berman, Matthias Haslberger, Viola Hilbert, Roberto Iacono, Theresa Neef, Louisa Roos, and David Sturrock. A special mention must go to my partner in crime, Wouter Leenders.

I chose to work on wealth inequality because I found it interesting, and I have had the fortune to find that policymakers agree. I thank the chair of the IBO Vermogensverdeling, Laura van Geest, for trusting me to add research to the report. I learned a lot from the process, and I wish to thank the secretariat (Jolanda Timmerman, Miriam Gielen, and Arieke Reiding), Rens Nissen, and the technical working group (Arjan Bruil, Elian Griffioen, Jeroen Nieuweboer, Mauro Mastrogiacomo) for their insights and support. Further thanks to the people at ESB and its podcast (Jasper Lukkezen, Elisa de Weerd, Yrla van de Ven, Sarah van Hugten), at CPB (Arjan Lejour, Céline van Essen, Egbert Jongen), and to all those who helped organise the inspirational Catshuis meeting with Minister Sigrid Kaag. I must especially thank my colleagues at CBS, in particular Mark, Arjan, Safan, and Jasper, for all their help and patience.

Doing a PhD is said to be a tough job, and it certainly wasn't always easy; but one of the greatest joys was to find friendship and community with my fellow PhDs. I have truly felt part of a group, and will deeply miss our social activities, job market peptalks, long walks, lunches, and all other small interactions that made these four years so precious. Without forgetting the others, I wish to particularly thank (in chronological order) Max, Emiel, Stefan, Ellen, Ruben, Victor, Anaya, Lara, Emilie, and Ima.

Throughout, music has been a great source of joy. I wish to thank the conductors and boards of various orchestras, and currently of my choir, for many inspiring rehearsals and concerts. Special thanks to Gonnie for all her singing and life advice. I am lucky to have met great friends along the way: Arjo, Dorien, Sophie, Hanna, Max, Magda, Daan, Marloes, Pauline, and all others from CM; Lauke; Thomas, Jorit, and Alexander; and especially my fellow academia-music combiners Jorgos and Janusz. I couldn't wish for better paranymphs.

My family has been an enduring source of joy and support. Thank you from the bottom of my heart to my dear parents, Chaja and Bert; to David, the best brother one could wish for; my wonderful family-in-law, Martin, Phia, Pepijn, and Fatou. Finally, I wish to thank the most important person in my life, who characteristically wants to remain anonymous. I couldn't have done this without you. Het werkstukje is eindelijk af.

Simon Toussaint
Utrecht, December 2024

Contents

List of Figures

List of Tables

Introduction

Wealth is the sum of assets minus the sum of liabilities, all components measured at market value. Wealth concentration is the process of wealth becoming more dominant, either among top wealth groups – so-called tail inequality – or in the economy at large. If the latter, we measure the growing importance of capital as a source of national income, compared to other production factors such as labor. This dissertation studies the empirics and interpretation of wealth concentration, both in the first and in the second sense.

As Hall (1967) argued half a century ago, wealth is a key variable for economic outcomes. Households choose between savings and consumption, and between renting an apartment and buying a house. Firms seek capital to finance their ideas. Governments finance their operations by issuing debt. Understanding the allocation of assets among households, firms and other agents is therefore key. Despite this importance, the subject of the wealth distribution was neglected for a long time in the literature, and its study was fragmented. On the empirical side, while wealth was of great interest to the political economists of the 19th century (e.g., Pareto 1896), research interest declined after World War II; some exceptions include Atkinson and Harrison (1978), Wilterdink (1984), and Wolff (1998).

On the theoretical side, research was fragmented between fields. Some theoretical models developed in the 20th century include random-growth models of wealth accumulation as in Champernowne (1953) and Wold and Whittle (1957), and models with heterogeneous savings rates as in Stiglitz (1969). Theoretical research started picking up in the late 1980s, when macroeconomic models with incomplete markets started to become computationally tractable to solve. The canonical models in this literature are Bewley (1987), Aiyagari (1994), Huggett (1993), and Krusell and Smith (1998).

Theoretical and empirical studies started to converge in the 2000s, when the study of wealth and its distribution received renewed attention. This attention followed the earlier surge in studies documenting rising wage and income inequality (Katz and Murphy 1992; Piketty 2003; Piketty and Saez 2003). The literature grew in particular following the influential publication of Piketty (2014). New empirical studies followed, showing rising wealth concentration in many cases (Kopczuk and Saez 2004; Saez and Zucman 2016; Alvaredo, Atkinson, and Morelli 2018). Simultaneously, various theoretical models of the wealth distribution emerged, seeking to explain the sizable concentration at the top (which proved difficult using the then-dominant Bewley-Aiyagari-Huggett framework). Examples include models of entrepreneurship (Quadrini 2000; Cagetti and De Nardi 2006), and models with stochastic capital returns (Benhabib, Bisin, and Zhu 2011), heterogeneous savings rates (Piketty and Zucman 2015), or stochastic discount factors (Toda 2019).

Despite this renewed attention, the study of wealth concentration is only just beginning. As I will discuss later in this Introduction, wealth is difficult to conceptualize and even more difficult to measure, and its distribution is often very poorly observed. This dissertation's empirical contributions confront the conceptualization of wealth, the measurement of private business wealth – a key component for understanding wealth concentration at the top – and the construction of reliable long-run data on wealth concentration. Theoretically, our understanding of the rapid fluctuations in top wealth concentration is incomplete. This dissertation also contributes to this front, by revisiting the canonical probability distribution thought to describe wealth, improving our understanding of cross-country differences in wealth concentration, and providing new evidence for a key mechanism underpinning the fast increase in wealth concentration, heterogeneous returns to wealth.

This dissertation consists of four chapters, which each both address empirical and theoretical open questions in the literature. Chapter 1 studies long-run trends in wealth concentration in the Netherlands, linking changes in the wealth-income ratio and top wealth shares to key moments of Dutch economic history. Chapters 2 and 3 revisit the key theoretical framework used to understand wealth inequality – the Pareto distribution – and develop models that are more consistent with the global distribution of billionaires. Chapter 4 takes on the measurement of private business wealth, the dominant component for the wealthiest households. Chapters 1 and 4 are more empirical in nature, while Chapters 2 and 3 are more theoretical. Nevertheless, theory and empirics go hand in hand in all four chapters, and the four chapters build on each other. All chapters share a preoccupation with robust measurement of (trends in) wealth concentration, and in trying to understand the empirics through the lens of economic theory. Only by proper accounting can wealth concentration be properly accounted for.

The rest of this Introduction serves to frame and introduce these contributions. I first consider the conceptualization of wealth, since this is not trivial, particularly in the Dutch context. Second, I discuss the empirics of wealth concentration and the contributions of this dissertation. Third, I summarize the key theoretical relations used in the literature, and how this dissertation challenges and refines them. I conclude with a chapter-by-chapter outlook and policy implications.

Conceptualization of Wealth

The definition of wealth used in the first sentence of this Introduction is standard and non-controversial (Zucman 2019). Yet, lurking beneath this simple definition are several non-trivial choices that confront the researcher. To clarify the controversies, it is helpful to express the definition in mathematical form:

$$W_{it} = \sum_{k \in \mathcal{K}} p^k_{it} A^k_{it} - \sum_{l \in \mathcal{L}} p^l_{it} D^l_{it}.$$

Here, W_{it} is wealth of agent i at time t, $\mathcal{K}$ is the set of assets, indexed by k, and $\mathcal{L}$ is the set of liabilities, indexed by l. Quantities of assets and liabilities are given by A^k_{it} and D^l_{it}, respectively, and the respective prices are p^k_{it} and p^l_{it}. Each aspect of this equation can and has been challenged. I list the most important ones. First, which i do we use? The natural choice for wealth accumulation is the household, since most consumption-savings decisions are made at that level. However, as with income, there are substantial differences in household composition over the wealth distribution. Moreover, household size has varied over time and varies between countries. This hinders empirical comparisons. The most common alternative is the adult individual. This is the unit chosen by the World Inequality Database (WID), for instance (Alvaredo et al. 2018). For single-person households, this choice is non-problematic; but for couples, things are more complicated. The WID currently assigns wealth held by a household in an equal split to each adult member, effectively assuming community of property. Clearly, this assumption can be strong and also varies significantly across countries and over time.

This dissertation mostly studies households. Chapters 1 and 4 exclusively focus on household measures of wealth concentration. Chapters 2 and 3, which use global billionaire rankings, nominally study individuals. However, for billionaires, the demarcation between individual and household wealth is highly arbitrary. This is particularly the case for family wealth, which is spread out over multiple heirs and generations. Distinguishing between individuals and households for billionaires is therefore fraught with difficulty, and this dissertation makes the pragmatic decision to take the data at face value.

A more fundamental controversy concerns the set of assets $\mathcal{K}$ (the choice of liabilities is less debated). What, exactly, is an asset? Here, there are two traditions in the literature studying wealth. One strand of literature, going back to Friedman (1957)'s permanent income hypothesis, views wealth as the net present value of lifetime consumption. Therefore, any economic object that will at some point generate consumption streams for the agent should be included in her asset set. This definition includes standard assets like stocks, deposits and housing, but also more nebulous concepts such as human capital (the net present value of future labor income). Indeed, government transfers, Social Security payouts, and other elements of the welfare state can all be capitalized and therefore should be included in wealth. Studies in this tradition include Feldstein (1976), Greenwald, Leombroni, Lustig, and Van Nieuwerburgh (2021), and Catherine, Miller, and Sarin (2023).

The alternative definition is employed by most statistical agencies and empirical researchers of wealth, and effectively sees assets as objects over which the agent has property rights. This implies that something is only an asset if an agent can decide what to do with it, and is residual claimant in the language of Grossman and Hart (1986). Standard assets like deposits and stocks fall under this definition. Human capital, however, is inalienable: It cannot be sold or otherwise transferred (Hart and Moore 1994; Andolfatto 2002). Likewise, claims to government payouts like Social Security cannot be traded, borrowed against, consumed early, or bequeathed by households. This suggests that human capital and Social Security should not be counted as assets for households under the property-rights view of wealth. Importantly, the tension is not an issue of liquidity. Houses are illiquid; selling them is subject to frictions. Yet, once owned, the owner can do with the house whatever she wishes. In contrast, human capital and Social Security are *legally* inalienable; they are impossible to market.

The tension between the two conceptualizations arises solely because markets are incomplete. In Walrasian settings, ownership is irrelevant (a consequence of the Coase theorem). Agents can buy perfectly replicating portfolios for any asset, thus mimicking ownership even if they do not have property rights. The definitions start to diverge if assets cannot be transferred across time or states of nature. Certainly, in incomplete-market settings, one *can* calculate net present values of future income streams. Yet if the agent does not have physical access to those capitalized income streams, there is a real question whether these values should count as assets. For some research settings, the answer to this question must be yes. When studying life-cycle problems, or questions of occupational choice, omitting future wage income or future Social Security streams would make no sense. On the other hand, when studying the response of households to adverse economic shocks, their use of wealth to influence political or social outcomes, or other questions affecting current-period outcomes, a property-rights view of wealth seems more warranted.

This dissertation does not resolve the tension, except to clarify its nature where most of the literature has not explicitly made these distinctions. Since all four chapters deal with empirics and applied theory, I am constrained by the restrictions the data impose. This means adopting a property-rights view of wealth, since neither human capital nor Social Security wealth is accurately measured in the data, let alone its distribution. A partial exception to this choice is Chapter 1. In studying long-run wealth concentration, I stick to the prevailing definitions of the System of National Accounts, as is also done by the WID. These definitions call for including capital-funded pensions in the set of assets. In the Dutch institutional setting, this definition is controversial, since unlike in other countries, the property rights over these pensions reside with the pension fund rather than the individual. Pensions cannot be accessed before retirement (unlike, e.g., the 401(k) in the United States), cannot be bequeathed or otherwise traded. Hence, property-rights views of wealth would not classify these pensions as household wealth, but permanent-income views naturally would. In Chapter 1, I take a middle road: I show aggregate household wealth both with and without pensions. There is no reliable long-run information on the distribution of pension claims, making it impossible to do the same for measures of wealth inequality. Hence, top wealth shares in Chapter 1 always exclude pension claims. Since the Dutch pension system dramatically grew in size since the 1990s, this omission does affect the trend of top wealth shares, in particular at the end of the period I study in Chapter 1.

A final difficulty concerns the price vector $\{p_{it}^k\}$. Market prices are observable for many assets, but certainly not for all. Houses are typically only recorded at cadastral value, for instance. More concerningly, private businesses have no observable market values, since they are not listed on the stock exchange. I call this concerning because this is the predominant wealth component of the wealthiest households, as my dissertation confirms. Without accurate prices for this component, there is a serious risk of mismeasurement in wealth concentration. This is one of the challenges this dissertation takes on, particularly in Chapter 4.

Empirics of Wealth Concentration

Having set out how I conceptualize wealth in this dissertation, let us move to what I mean by concentration. The macro view of wealth concentration is concerned with a rising dominance of capital in the aggregate economy. Since the seminal study by Piketty and Zucman (2014), the preferred measure to gauge this type of concentration is the *wealth-income ratio*: the ratio of aggregate (household or national) wealth W_t to national income Y_t. This measure is conceptually clear: it expresses the size of wealth in years of income, e.g., a wealth-income ratio of 6 indicates that aggregate wealth is worth six years of national income. A second advantage is that as a ratio of two nominal quantities, it is free of index-number problems, facilitating comparisons across time and space. A final advantage is that in closed economies, it coincides with the capital-output ratio familiar from neoclassical growth theory (Solow 1956), and hence also relates to the functional income distribution. Holding the aggregate return to wealth constant, a rising wealth-income ratio implies a rising capital income share (Barkai 2020).

In Chapter 1, I use this measure to study macro wealth concentration. I use household wealth rather than national wealth as my measure of aggregate wealth. This ensures that macro and micro perspectives on wealth concentration are more tightly linked. Naturally, the wealth-income ratio is not trivial to measure, since the default for measuring household wealth at the aggregate level – household balance sheets in the System of National Accounts – is available in Western economies only since the late 1990s. In Chapter 1, I employ three distinct methodologies to reconstruct aggregate household wealth in the Netherlands prior to 1995. The combination of different methods allows me to go back to 1854; moreover, it allows me to test various methods proposed in the literature. The wealth-income ratio in the Netherlands increased during the Industrial Revolution in the 1870s and 1880s, peaking around 900%, which is the largest wealth-income ratio recorded in the literature. Moreover, we are able to shed light on an understudied aspect of the dynamics of the wealth-income ratio prior to World War II, namely the role of colonial and foreign investments. I assemble the first series in the literature on colonial wealth. Using this series, I show that Dutch private investments in Indonesia and other foreign investments account for the high level of the Dutch wealth-income ratio relative to countries without colonial empires, and also explain the increase in the Dutch wealth-income ratio in the 1920s. After World War II, Dutch wealth-income ratios are more in line with international experience, declining precipitously until the late 1970s and increasing afterwards. Here, including pensions in the definition of wealth makes a difference, since I show that most of the increases in recent decades has been driven by pension wealth accumulation. Housing, which is the dominant asset in aggregate wealth accumulation in many other countries (e.g., Artola Blanco, Bauluz, and Martínez-Toledano 2020), plays only a secondary role. One reason for this is the role of policy: Homeownership grew quickly from the 1980s, in particular by very lenient conditions for mortgages, such as the interest-only mortgage – a mortgage construct where the principal does not have to be repaid before the end of the contract, but only the interest. Before 2013, mortgage interest was fully deductible from the income tax, incentivizing the adoption of interest-only mortgages and similar products which maximized the loan-to-value ratio, which was allowed to be up to 120% before 2018. This resulted in a large accumulation of mortgage debt, making the net contribution of housing secondary to private pensions.

For micro wealth concentration, there are three competing measures often used. The first is the Gini coefficient, which normalizes the surface under a society's Lorenz curve by the triangle under the line of equality (the Lorenz curve if all wealth were equally distributed). For non-negative variables, the Gini varies between 0 and 1. Wealth, however, can be negative; hence, the area under the Lorenz curve can theoretically be larger than the triangle, resulting in Ginis potentially larger than 1. A more serious concern is that the Gini puts large weight on the middle of the distribution, and is highly insensitive to changes in the tails. Thus, two societies can look completely different yet have identical Gini coefficients (Piketty 2014). For these reasons, the Gini will not be used in this dissertation. Instead, I employ the two most often used alternatives: top wealth shares and the Pareto coefficient. Top wealth shares are straightforward to understand: they measure the share of total wealth accruing to fractile p. I use the top 1% and top 0.1% shares of wealth as my preferred measures of wealth concentration. In Chapter 1, I show that these wealth shares followed the familiar U-shaped pattern also observed in the wealth-income ratio over the 20th century. Wealth concentration increased after the 2008 crisis, after house prices collapsed. This collapse decreased the wealth of the middle class, since housing is the predominant asset for the middle of the distribution. The wealthiest households, in contrast, hold most of their wealth in financial assets and private businesses, which did well relative to housing. The recent increase in housing prices has depressed top wealth shares again.

Chapter 1 establishes these trends using administrative data on household wealth. However, as noted before, several assets might be poorly measured even in the relatively high-quality Dutch data. Chapter 4 sets out to address the measurement of private business wealth, which is recorded at book value in official statistics. I use an econometric approach to estimate the market value of private business wealth. I build a model that structurally relates a firm's market value to its capital stock, akin to but more general than the role of Tobin's q in neoclassical investment models (e.g. Hayashi 1982). Using some initial estimates of private business wealth using capitalized firm profits, I regress these initial estimates on firms' capital stocks, using time-series restrictions as instruments to correct for endogeneity and measurement error. I use the fitted values of the resulting GMM estimation as my estimates of private business wealth. By linking firms to firm-owners in the data, I show that this procedure revises top wealth shares since 2008 substantially upward. The top 1% share increases by three to five percentage points on average, depending on the specification. The top 0.1% share increases by a similar magnitude, which means that in relative terms it increases much more. My results show that properly measuring wealth components can substantially alter our understanding of the empirics of wealth concentration.

The final measure of wealth concentration, the Pareto coefficient, directly relates to the theoretical contributions of this dissertation. These will be discussed next.

Theoretical Models of Wealth Concentration

Since Pareto (1896) observed that 80% of wealth accrued to the richest 20% of households, with 80% of the wealth of the richest 20% accruing to the richest 20% within the richest 20% and so on, the dominant model to understand wealth concentration has been the Pareto distribution. Above some threshold Ω, a Pareto-distributed variable $\underline{W}$ follows a power law, i.e., the probability of observing wealth in excess of some realization W is equal to the ratio of the realization to the lower bound, raised to some power $1/\alpha$:

$$\Pr\left[\underline{W} \geq W \mid W \geq \Omega\right] = (W/\Omega)^{-1/\alpha}, \quad W, \Omega > 0, \alpha \in (0,1).$$

The speed of the decay is governed by the Pareto coefficient α. The larger α, the slower the decay and hence the more tail inequality. When $\alpha = 1$, there are ten times as many households with 1 million as there are with 10 million, and ten times as many with 10 million as there are with 100 million, and so on. Pareto's 80-20 observation corresponds to an $\alpha = \ln 4/\ln 5 \approx 0.86$.

The Pareto distribution is simple to work with and appears to describe the top of the wealth distribution reasonably well. Therefore, most empirical and theoretical research takes the Pareto distribution as given and rarely questions whether it actually fits the data well. At the same time, Pareto imposes strong restrictions: Moments larger than $1/\alpha$ are undefined, and hazard rates of log wealth are constrained to be constant and equal to α. In Chapter 2, I investigate whether Pareto actually describes the data on top wealth. I develop a test to evaluate Paretianity, based on scaled ratios of log moments. These test statistics, $\mathcal{R}_k$, equal the kth log moment divided by k factorial times the first moment to the power k. For example, the statistic $\mathcal{R}_2$ equals the variance of log wealth (the second moment) divided by twice mean log wealth squared. These statistics should equal one under the null hypothesis that wealth is Pareto; this is our test. The statistics are easy to compute in the data for any moment k and have strong power in detecting deviations from Pareto. I derive their small-sample properties and show that even in very small samples, these statistics are informative. Using the *Forbes List of Billionaires* since 2001, I show that Pareto is rejected across years and regions.

Knowing that Pareto does not fit perfectly is one thing, but 'it takes a model to beat a model'. The second contribution of Chapter 2 is to show that a simple alternative distribution fits top wealth much better. This alternative is the (truncated) Weibull distribution. Relative to Pareto, Weibull adds an additional parameter, γ. This parameter governs the curvature of the hazard rate of log wealth. Whereas the hazard of log wealth is constant under Pareto – implying slowly decaying wealth

– log wealth decays exponentially fast under Weibull. The larger γ, the faster this decay. As $\gamma \to 0$, the model specializes back to Pareto.

Weibull is consistent with the values of the test statistics, which show persistent downward deviations from Pareto, with larger deviations for statistics based on higher-order moments. This is consistent with exponentially increasing hazard rates. Directly plotting the empirical hazard rates also confirms their exponential nature. A final piece of evidence comes from overidentifying restrictions. We derive the theoretical value of the test statistics $\mathcal{R}_k$ under Weibull, and show that they depend only on the *product* of α and γ, not on their separate values. Hence, we can calculate the theoretical predictions of $\mathcal{R}_k$ for several values of k given the empirical values of α and γ. If these theoretical values are close to the empirical test statistics, we have one moment fitting two equations, implying an overidentifying restriction.

Chapter 2 shows that these conclusions are not driven by measurement error, nor by choice of variable: the same patterns show up in the U.S. city size and firm size distribution. Hence, Weibull rather than Pareto seems a robust finding. This calls for theoretical models that can generate a Weibull distribution of wealth. The mechanics for generating Pareto distributions are well understood (Gabaix 2009; Benhabib and Bisin 2018): take some diffusion process of the form

$$\mathrm{d}W_t = \mu W_t \mathrm{d}t + \sigma W_t \mathrm{d}\mathcal{B}(t)$$

where $\mathcal{B}(t)$ is a standard Brownian motion, and μ and σ are parameters. To let this process converge to a stationary Pareto distribution, we need some 'friction' to prevent wealth from becoming unboundedly negative. Common choices in theoretical models include a borrowing limit in Bewley-Aiyagari-Huggett models, discount rate shocks, and birth and death processes in lifecycle models; see Gabaix (2009) and Moll, Rachel, and Restrepo (2022) for further examples.

Diffusion processes like the one above can be generalized to allow convergence to Weibull. In Chapter 2 and 3, I note an appealing alternative, based on network theory. Consider a random network with N nodes; between any two nodes there is an equal chance p of a link being generated. A Self-Avoiding Walk on this network is a path between nodes that does not visit a node twice. The length of SAWs converges to Gompertz, the log-transform of Weibull (Tishby, Biham, and Katzav 2016). Hence, this setup can serve as an explanation for Weibull.

Chapter 3 further builds on this idea. Here, I use this network structure to understand why there are persistent differences in billionaire numbers over time and across regions. The model views a billionaire as a firm-owner. The size of the firm is given by the length of a SAW on a random network. The total network size is given by a region's population size. Hence, the larger a region's population, the more connections are formed, and hence the more billionaires can emerge.

Moreover, since the distribution of path lengths also depends on the size of the network, a more populous region will have more ultra-rich billionaires relative to 'ordinary billionaires' than a small region will. This also facilitates the interpretation of the somewhat abstract concept of a SAW: We can think of the network as the space of ideas, and a firm as a succesful embodiment of adjacent ideas. Consider Apple. It grew by selling the Macintosh computer, then grew further by absorbing the idea of the iPhone. This idea was 'adjacent' to Apple's existing technology, making it possible to incorporate in the firm. An idea that is more distant from the firm's current location cannot be absorbed. The model implies that a larger population will generate more ideas and hence will have more scope for firms to absorb new ideas adjacent to their existing technology. This is fully consistent with endogenous and semi-endogenous models of economic growth (Romer 1990; Aghion and Howitt 1992; Jones 1995). An alternative interpretation views the network as an expression of customer size. Then, too, a larger market means firms grow larger, in the tradition of Melitz (2003).

The model is closed by studying entry into the network and hence the billionaire distribution. I view billionaires as a set of individuals who can capitalize their rental income into wealth by means of two factors: their region's nominal GDP per capita and a global wealth-income ratio. The first factor reflects the fact that a richer region will also have a more developed market and hence more scope for a firm-owner to grow rich. The second factor accounts for global investment conditions. These factors affect the value of two parameters governing entry: a conditional baseline hazard, and a conditional billionaire probability. The model delivers stark predictions. The elasticity of the hazard rate with respect to wealth – the parameter γ from Chapter 2 – should depend negatively on a region's population size, and on nothing else. The two parameters governing entry into the billionaire distribution should depend on a region's GDP per capita and a global wealth-income ratio with unit coefficients, and on nothing else.

All these predictions are tested and mostly hold. The first prediction easily holds: The log hazard rate elasticity negatively and exclusively depends on log population size. The other two parameters do not depend on log population, only on the macroeconomic factors. The coefficients are not quite equal to one, but close. Remarkably, the model does not need time fixed effects. This suggests that this simple macroeconomic framework is sufficient to explain all time-series variation in billionaire numbers. The model does require region fixed effects. This is unsurprising: Switzerland has more billionaires than it "ought to".

Most of the theoretical contributions are contained in Chapters 2 and 3, but Chapter 4 does contribute to a rapidly growing theoretical literature, on return heterogeneity. The diffusion processes with frictions discussed above do converge to right-skewed wealth distribution, but *too slow* to explain the rapid changes in top wealth shares seen in the data. In their seminal contribution, Gabaix, Lasry, Lions, and Moll (2016) formalize this claim and show how to modify these processes to allow for rapid changes in tail inequality. What is required to get fast dynamics is to allow returns to wealth to be positively correlated with the size of wealth, so that all else equal, the rich get richer. Gabaix, Lasry, Lions, and Moll (2016) note two types of modifications that achieve this. One, called type dependence, means that some types get higher returns than others, all else equal. This could be because these individuals are high-ability entrepreneurs, for instance (Jones and Kim 2018). Since their wealth grows faster, over time they will drift to the top of the distribution. The other way, scale dependence, introduces increasing returns to scale in wealth. This could be because of information frictions (Kacperczyk, Nosal, and Stevens 2019), scale economies in wealth management (Gerritsen, Jacobs, Spiritus, and Rusu 2024), or some other market incompleteness.

The theoretical literature has not converged to a preferred explanation yet, but empirical evidence for return heterogeneity is accumulating. Piketty (2014) noted that larger U.S. universities had higher returns on their endowments than smaller ones; Saez and Zucman (2016) found the same for foundations; Wolff (2017) noted that different returns between asset classes cause return heterogeneity over the distribution if households systematically differ in portfolio composition. More systematic evidence for comes from Norway (Fagereng, Guiso, Malacrino, and Pistaferri 2020) and Sweden (Bach, Calvet, and Sodini 2020), where the availability of administrative data on wealth and returns shows systematic heterogeneity across the wealth distribution, even within asset classes. While this heterogeneity is detected among all asset classes – even safe assets like deposits – the largest heterogeneity by far is within private business wealth. Given that this is the dominant asset component of the richest, heterogeneity in private business returns seems key to understand rapid changes in wealth concentration.

In Chapter 4 I revisit returns in this wealth component. I show that conclusions of return heterogeneity are highly sensitive to measurement error. This is because a return is the ratio of a cashflow to wealth. If wealth has measurement error, a regression of returns on wealth will be mechanically correlated and biased. The direction of the bias is ambiguous and depends on whether there is only measurement error in the denominator – wealth – or also in the numerator – cashflows. Therefore, it is crucial to accurately measure private business wealth to get the facts on return heterogeneity correct.

With the econometric procedure described above, I get error-free estimates of private business wealth. I use those to obtain updated measures of return heterogeneity, and contrast these measures with returns as estimated using unadjusted data. I find that there is sizable return heterogeneity, with a significant and steep gradient in wealth. This gradient is much steeper, moreover, than in the unadjusted data. This suggests that raw data, based on book values, significantly underestimate the extent of return heterogeneity in private business wealth. This Chapter is a first step towards a full empirical understanding of return heterogeneity, which will aid the theoretical literature seeking to understand the dynamics of inequality.

Dissertation Outline

In Chapter 1, joint work with Amaury de Vicq, Michail Moatsos, and Tim van der Valk, I analyze the evolution of aggregate household wealth, its composition, and top wealth shares since the mid-19th century for the Netherlands, a country which played a significant role in economic history. The main forces at play are the size and variation of colonial wealth up until WWII, and the introduction of a –particularly strong– pension system thereafter. We show that the wealth-income ratio followed the familiar U-shaped pattern over the 20th century. The Netherlands, however, had the largest wealth-income ratio on record, growing since the mid-1850s, driven by industrialization and booming private foreign investments, to a peak of 900% around 1880. In contrast to other countries, the wealth-income ratio remained high up until 1929. To understand these trends better, we construct the first series on colonial wealth and show that colonial and other foreign investment account for most of the gap with other countries in the pre-WWII period. The initial post-war decline in the ratio is driven by rapid income growth. The increase in the ratio since the 1970s has been mainly driven by the uniquely large capital-funded pension system. In contrast with other major countries, housing plays only a secondary role in net wealth accumulation due to significant mortgage debt. Methodologically, this chapter is the first to compare historical national accounts, estate multiplier, and wealth tax data approaches to construct aggregate wealth. I find that the estate multiplier is a good alternative to the historical national accounts benchmark, while the use of wealth tax data results in unrealistically low estimates.

Chapter 2, joint work with Coen Teulings, studies the shape of the global wealth distribution, using the *Forbes List of Billionaires*. I develop simple statistics based on ratios of log moments to test the default assumption of a Pareto distribution, which is strongly rejected. Hazard rates show that the log-transformed data instead follow a Gompertz distribution, which means that the data in levels follow a truncated-Weibull distribution. I further apply our model to the U.S. city size distribution and the U.S. firm size distribution. These distributions also show a rejection of Pareto in favor of (truncated-)Weibull. I discuss some theoretical and practical implications of my results.

Chapter 3, joint work with Coen Teulings, studies the fourfold increase in the ratio of billionaires to population since 2001, and its regional variation. I develop a model where wealth is proportional to the length of a Self-Avoiding Walk on a random network, which rationalizes the Gompertz distribution of log wealth in the data. The model predicts the elasticity of top inequality to depend solely on population size and the lower thresholds for wealth to depend solely and one-for-one on regional GDP per capita and a global wealth-income ratio. All predictions hold in the data. I find strong evidence that inequality, measured by the hazard rate of log wealth, is increasing in population size. Time fixed effects do not significantly improve the model fit, but regional effects do. Counterfactual exercises closely predict observed mean (log) wealth and billionaire numbers. The increases in billionaire numbers and mean (log) wealth are almost entirely driven by increases in GDP per capita. I interpret these results in the context of models where market size shapes firm (and hence wealth) concentration.

Chapter 4, which is single-authored, studies the value of non-listed businesses, which is a key input for many aggregate and distributional trends, yet is difficult to estimate. Measurement error in private business wealth estimates can severely bias regressions, such as for return heterogeneity. I propose an econometric adjustment, where I use capitalized profits as a measurement-error-contaminated measure of the market value. I use time-series restrictions on the measurement error as moment conditions in a GMM procedure, which delivers error-free fitted values for private business wealth. I apply my framework to Dutch administrative data linking firms and their owners. Across specifications, the adjusted top 1% and 0.1% wealth share increase by 3–5 percentage points on average. The adjusted top 1% shares peak at around 38%, and the top 0.1% shares peak at around 20%. Adjusted returns to firm wealth follow a steeper gradient along the wealth distribution than unadjusted returns.

Policy Implications

My research has several policy implications. My conceptual discussion of wealth and its measurement underscores how wealth is endogenous to the policy process: In the Dutch context, pensions are not part of households' property-rights wealth, but they are in many other countries. Policy can therefore affect not only the distribution of wealth, but even what counts as wealth. Chapter 1 demonstrates that policy has major impacts on household portfolio composition and wealth accumulation. There, I show how government policy has largely shaped the dynamics of wealth accumulation since World War II. Policy promoting homeownershi has induced households to allocate the majority of their wealth into this component. This poses serious concerns for the concentration of risk and the ability of households to absorb shocks.

A second conclusion emerges from Chapters 2 and 3. There, I show that wealth concentration at the extreme top is driven to a large extent by macroeconomic fundamentals. Population size increases tail inequality, as do increases in local and global market conditions. To a non-trivial extent, then, wealth concentration emerges endogenously in market economies (van Bavel 2016). At the same time, government policy can influence wealth concentration, as evidenced by the significant role for region fixed effects in Chapter 3. The level of wealth concentration is therefore to some extent a policy choice.

Chapter 4 has several policy implications. First, accurate measurement of private business wealth is key to understanding inequality. A more important conclusion concerns the findings on return heterogeneity. From a theoretical side, heterogeneous returns to wealth break the equivalence between capital income taxation and wealth taxation. As a result, it may be optimal to tax capital (income), overturning the classic Chamley-Judd result (Chamley 1986; Judd 1985). The literature has not yet settled on the optimal form of capital taxation: Guvenen et al. (2023) and Guvenen, Kambourov, Kuruscu, and Ocampo-Diaz (2024) argue for a tax on book-value wealth, while Gerritsen, Jacobs, Spiritus, and Rusu (2024) and Boar and Midrigan (2023) find that capital income taxes are optimal. Regardless of the form of taxation, the fact that returns truly are heterogeneous – and steeply so – has first-order implications for wealth taxation.

There are also practical reasons to be interested in return heterogeneity. The current Dutch tax system taxes financial assets with an effective wealth tax, by charging a progressive rate to a presumptive rate of return (Jacobs 2015). This system was deemed illegal by the Dutch Supreme Court in 2022, disabling a major part of Dutch government revenue. The current solution is to implement the taxation of realized capital returns. From a policy side, it is therefore imperative to accurately measure returns to wealth. My research using firm-level returns was possible due to the availability of granular data linking households to firms. Similar data do not exist for other assets in households' portfolios in the Netherlands, precluding the implementation of a capital income tax on actual returns. To get a complete picture of household-level returns – a key object for policymakers in years to come – it is therefore crucial to build a comprehensive household asset registry, like in Norway (Fagereng, Guiso, Malacrino, and Pistaferri 2020).

Chapter 1

Household Wealth and its Distribution in the Netherlands, 1854–2019

This chapter is based on Toussaint, de Vicq, Moatsos, and van der Valk (2022).

1.1. Introduction

The Netherlands has historically been among the richest countries in the world, was once a colonial superpower, and played a major role in the development of modern capitalism ('t Hart, Jonker, and van Zanden 1997). The Netherlands, however, has been notably missing in studies which analyse the long-term evolution of household wealth (Piketty and Zucman 2014). This chapter amends this shortcoming by providing the first series of historical household balance sheets, consistent with existing balance sheets in the System of National Accounts for the Netherlands, from 1854 until 2019. We decompose these balance sheets into wealth components from 1880, tracking the relative importance of asset classes such as real estate, equity, bonds, liabilities and pension wealth. We track top wealth shares from 1894 onward, arriving at a comprehensive picture of the dynamics of household wealth over more than 125 years. Finally, we discuss various interpretations of these trends, and contrast them to the available international evidence.

Our main conclusions are the following: First, we find that the wealth-income ratio in the Netherlands followed the familiar U-shaped pattern observed in earlier studies, with a peak in the early 20th century, a subsequent decline until the 1970s, and an increase in recent decades. However, we find that the magnitude of the peak is substantially larger than those identified in other countries (Waldenström 2021). Specifically, we document that the wealth-income ratio peaked in excess of 900% at the end of the 19th century; moreover, the wealth-income ratio remained high during and after World War I, only declining after the Great Depression of 1929. This is in contrast to other countries in this period for which we have data, which either had consistently lower wealth-income ratios or featured strong declines during World War I (Waldenström 2021). After World War II, the wealth-income ratio declined precipitously until the 1970s, when it reached a trough of approximately 300%. Since the 1980s, the wealth-income ratio has increased again to reach 600% in 2019. Figure 1.1, panel (a), shows the overall evolution of the wealth-income ratio since 1854.

Second, our findings for top wealth shares largely echo those for total household wealth show a similar a clear U-shaped pattern. Figure 1.1, panel (b), depicts the top 1% wealth share from 1894 until 2019. We observe a peak in the top 1% share of household wealth peaked at close to 55% in the early 20th century, which was followed by a precipitous decline to as little as 10% in the 1970s and a subsequent increase to about 30% by the early 2010s.

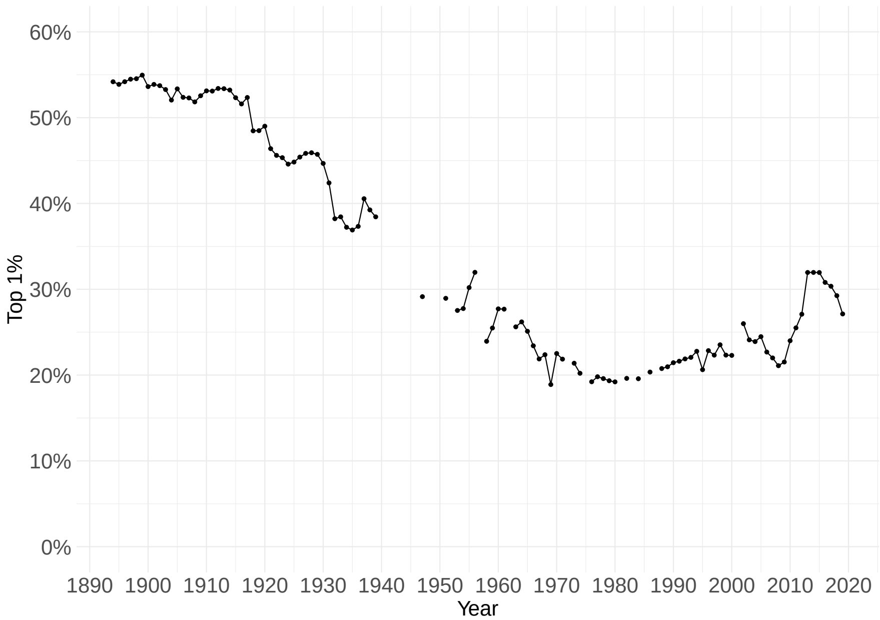
Third, our historical National Accounts also allow for the decomposition of aggregate wealth into wealth components. Grouping assets and liabilities in broad classes to ensure comparability over the full period, we find a stark decline in the importance of agricultural land prior to World War II, with financial assets dominating wealth composition. After World War II, financial assets declined in importance, while the Dutch pension system – which we term semiprivate wealth – rapidly became a dominant asset class for households, being worth 40% of the household portfolio in 2019. Moreover, housing also increased in importance from the 1980s onward. However, since the housing boom was financed by an enormous increase in mortgage debt, the net effect of housing accumulation on real wealth growth is modest. Instead, pension savings and capital gains are the most important drivers of wealth concentration since the 1980s.

Our contribution to the literature is threefold. First, we contribute to the growing and impactful literature taking stock of long run trends in top wealth shares and the long run evolution of household wealth. The seminal work in this series is Piketty and Zucman (2014), who estimate historical wealth-income ratios for eight large economies. We introduce data for the Netherlands which has so far been absent, and discuss our findings in relation to the early development of security markets and capitalism and its colonial past (van Zanden and van Riel 2000; 't Hart, Jonker, and van Zanden 1997). Historical estimates for Dutch household wealth do exist, including Boissevain (1891), and most notably Wilterdink (1984, 2015). Compared to these studies, we add a longer-run perspective, as well as an decomposition of aggregate wealth, we provide an international comparison, and we operationalize all three main methods for household balance sheet reconstruction (Figure 1.1). Compared to existing studies focusing on top wealth shares in the Netherlands, (notably Wilterdink 1984 for 1894–1974 and Salverda 2019 for 1993–2000 and 2006–2014), we provide a consistent treatment of household wealth as corresponding to the National Accounts, and employ a more flexible estimation method than strict Pareto interpolation, namely generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022).

Second, our findings challenge the claim by Waldenström (2021) that wealth-income ratios pre-World War II were smaller than previously thought. We are the first to construct a series on colonial wealth, which demonstrates that colonial and other foreign wealth in particular played an important role in explaining the observed divergence – in particular the peak of 900% in the wealth-income ratio in the 1880s, as well as the persistent large ratios post-World War I – between the Netherlands and most other countries. In addition, we argue that denominator effects also matter for cross-country differences in wealth-income ratios, since persistent differences in income levels and growth rates also affect its evolution. More generally, our results show the importance of careful accounting for foreign investment, colonial wealth, and other major trends in the late 19th century to explain global patterns in wealth inequality. Moreover, the sharp decline in the wealth-income

Figure 1.1: The Wealth-Income Ratio and the Top 1% Wealth Share, 1854–2019

(a) Wealth-Income Ratio



(b) Top 1% Share

Notes: Figure depicts the ratio of aggregate household wealth to net national income (1854–2019), and the top 1% wealth share (1894–2019).

ratio that happened in the 1950s – but not during the 1940s – is likely driven by the forced nationalization of Dutch enterprises by the newly independent Indonesian government, as well as the general upheaval associated with decolonization, and less likely to be related to the expansion of the welfare state.

Third, we compare historical national accounts, estate multiplier, and wealth tax data approaches to construct aggregate wealth. The first method, which we use as the benchmark, is the method used in most of the contemporary literature, including the seminal work by Piketty and Zucman (2014). The second method we employ is the estate multiplier approach (e.g. Kopczuk and Saez 2004) which estimates the aggregate stock of household wealth from tabulated inheritance tax records. This method was common for historical estimates of aggregate wealth and can be traced back to at least the late 19th century (e.g., Boissevain 1891). The third method builds upon available information from wealth tax data, and further assumes that wealth of those below the wealth tax threshold is reasonably approximated by a lognormal distribution. By comparing these two methods with our benchmark method, we improve our understanding concerning the accuracy of the levels and trends that these methods identify. In particular, we find that the estate multiplier method produces remarkably similar estimates to the benchmark historical national accounts pre-World War II; after the war, the two methods diverge. The lognormal extrapolation method underperforms relative to the two other methods throughout.

Chapter Outline: The rest of this chapter is structured as follows. Section 1.2 discusses existing work on the dynamics of wealth in other countries and in the Netherlands. In Section 1.3, we discuss our definitions of household wealth, analyzing the distinct role of pension wealth in greater detail. Moreover, we introduce the three methods we use to reconstruct aggregate household wealth, as well as the mapping from aggregate wealth to wealth shares. Section 1.4 presents an overview of the results, for aggregates, wealth shares, and wealth composition. In addition, in this section we provide a brief international comparison for all results. In Section 1.5, we discuss the dynamics of the wealth-income ratio before World War II in international perspective. Here, we focus on the role played by international and colonial investment, as well as numerator effects (i.e., differences in national income). Finally, in 1.6, we explore socio-economic developments in the Netherlands after World War II and show how the large accumulation of capital-funded pensions was the major determinant of the rise in wealth concentration since the 1980s. Section 1.7 concludes.

1.2. Literature Review

Studies on long-term wealth dynamics, including the share of top wealth, are relatively new. At least in part, this is because this type of research relies heavily on national stock accounts, instead of flow accounts; these are data which national statistical institutes only began to compile from the early 1990s onward.

The past few years have seen a notable resurgence of interest in the long-term evolution of size, composition and distributional patterns of private wealth mostly across the Western world, along with India and China. Seminal studies include Davies, Sandström, Shorrocks, and Wolff (2011) and the work by Piketty (2014) and Piketty and Zucman (2014). The latter work sparked a growing number of researchers to explore whether the U-shaped pattern observed by Piketty and Zucman (2014) also applied to countries with notable different institutional set-ups. Such studies include Waldenström (2017) for Sweden; Orthofer, Du Plessis, and Reid (2019) for South Africa; Piketty, Yang, and Zucman (2019) for China, Kumar (2024) for India; and most recently Artola Blanco, Bauluz, and Martínez-Toledano (2020) for Spain. In parallel, there is also a growing literature on top wealth shares. Key examples of this strand of research includes Saez and Zucman (2016) for the United States[1]; Garbinti, Goupille-Lebret, and Piketty (2020) for France; Alvaredo, Atkinson, and Morelli (2018) for the United Kingdom; Albers, Bartels, and Schularick (2022) for Germany; and Roine and Waldenström (2009) for Sweden.

For the Netherlands, studies on wealth dynamics are part of a once vibrant, but until relatively recently, almost entirely neglected scholarly tradition. The earliest studies on these matters can be traced back to at least the mid 19th century. Pareau (1864) for example was one of the first scholars which attempted to explore changing wealth patterns across several decades, comparing total private wealth in the 1830s and the 1860s to the respective size of the Dutch population. He argued that there has been a per capita decrease in national wealth throughout this period. Boissevain (1883, 1891, 1909, 1910) and Stuart (1888) were the first to provide somewhat reliable estimations of aggregate wealth for the 1880s and 1890s by relying on the estate multiplier method.

Following the events of the first World War, Bonger (1923) set out to measure total private wealth for the period between 1915 and 1920. His work criticizes the aforementioned method by Boissevain and Stuart, instead relying on an approach that is more akin to a reconstruction of the national accounts. His work was continued by Smeets (1932) and a few years later by van der Wijk (1939).

[1] However, see the discussion of their capitalization method and underlying assumptions (Kopczuk 2015; Smith, Zidar, and Zwick 2023; Saez and Zucman 2020).

In contrast to earlier decades, the period following the second World War until the 1980s was characterized by a notable absence of studies on long-term wealth dynamics. A major stimulus in increasing interest arose from the work by Wilterdink (1984) who documented and analysed long-term patterns of wealth dynamics based on wealth tax records from the 1890s until the 1970s, and provided a breakdown of top wealth shares. In his footsteps, the literature examining the Netherlands grew steadily, mostly focusing on periods that preceded Wilterdink's analysis. Verstegen (1996), for instance, looked at national wealth and income in the Netherlands between 1805 and 1910, whereas Bos (1990) sketched the capital holdings and status of the wealthiest members of society in the 19th century Netherlands. Recent contributions include Wilterdink (2015) who reflect on the decades after the 1980s, which were not covered yet by his previous work. The most recent noteworthy contribution to the study of the top wealth shares in the Netherlands was made by Salverda (2019) covering many of the years after 1993.

Van Bavel and Frankema (2017) argue that Wilterdink's estimates for the 1970s only constitute roughly 85 percent of Net National Income (NNI), which based on our findings is an underestimation by a factor of two and a half; this is a point also reiterated more recently by Coenen (2017). Building on all these existing studies for the Netherlands, we set out to take the next step forward and provide the first comprehensive data set on Dutch wealth since the 1850s, which is –in addition– implemented within a framework consistent with recent international studies.

1.3. Concepts, Sources, and Methods

1.3.1. Definitions

Our aim is to reconstruct household wealth, W_t, following the definition employed by the System of National Accounts, which is the total market value of assets minus liabilities. Assets include all financial and non-financial assets over which ownership rights can be enforced and which provide economic benefits to their owners. This definition includes most major wealth components, including housing, real estate, savings accounts, stocks and bonds, which can be accessed and capitalized by their households.

It is useful at this point to clearly spell out the concepts which we will estimate and pursue throughout the chapter. Denote national wealth by W_{nt}, government wealth by W_{gt}, corporate wealth by W_{ct}, household wealth by W_t, and foreign wealth by W_{ft} (where t denotes the calendar year). We have the standard accounting identity that national wealth equals the sum of the wealth of the four main economic sectors:

$$W_{nt} = W_t + W_{gt} + W_{ct} + W_{ft}. \tag{1.1}$$

Various other decompositions of national wealth exist; for instance, national wealth equals the sum of the capital stock and the net foreign asset position $W_{nt} = K_t + \text{NFA}_t$. Moreover, if the book value of equity equals the market value of equity (i.e., Tobin's q equals 1), corporate net worth is zero and national wealth equals private wealth plus government wealth. In general, corporate wealth need not be zero, as is discussed extensively in Piketty and Zucman (2014). Hence, we focus throughout on *household* wealth, with the understanding that clearly delineating between household and corporate wealth can be challenging in periods before official National Accounts balance sheets exist. In the Data Appendix, we provide an extensive discussion of the various steps we take to ensure consistency over the entire period.

Household wealth can be decomposed into financial assets A_t^f, non-financial assets A_t^{nf}, and liabilities D_t:

$$W_t = A_t^f + A_t^{nf} - D_t. \tag{1.2}$$

In this chapter, we will often focus on several interesting sub-components of these broad categories. We can decompose non-financial assets into housing – dwellings plus underlying land – as well as agricultural land and the remaining capital stock attributable to the household sector:

$$A_t^{nf} = A_t^h + A_t^l + A_t^k. \tag{1.3}$$

We can decompose financial assets into deposits, domestic securities (i.e., stocks and bonds), foreign securities – which includes claims to colonial investment before decolonization – and the value of life insurance and pension claims, which we term 'semiprivate wealth' following Wilterdink (1984):

$$A_t^f = A_t^d + A_t^{e,dom} + A_t^{e,for} + A_t^{sp}. \tag{1.4}$$

This final wealth component, which we refer to as semiprivate wealth following Wilterdink (1984), deserves some attention because its inclusion in the household wealth statistics is not trivial. The Dutch pension system consists of three 'pillars': (i.) universal retirement payouts, funded as a PAYGO scheme (*Algemene Ouderdomswet* or *AOW*); (ii.) employer-based pension funds, which every employee is required to contribute to; and (iii.) personal pension accounts. Component (i.) is not considered wealth. The main discussion revolves around the inclusion of pillars (ii.) and (iii.). Standard DINA guidelines prescribe that the capitalized values of these pension contributions should be included as wealth, reasoning from a life-cycle perspective. All existing Dutch data sources, on the other hand, have so far excluded pension wealth from wealth distribution statistics. The commonly given reason is that these pension assets are not freely disposable, are not bequeatheable and hence are more akin to claims on future income streams, like Social Security benefits (van Bavel and Frankema 2017). Other authors disagree with this assessment, pointing to the important substitution effects pension wealth has with regular savings (Caminada, Goudswaard, and Knoef 2014). The inclusion of pension wealth is not a trivial matter: Dutch pension funds are among the best funded internationally, with total capitalized contributions in excess of 200% of national income in recent years. Since pension contributions tend to be distributed more equally than other financial assets, including pension wealth also has profound effects on estimates of wealth inequality. The size of employer-mandated pension wealth relative to other assets makes the Netherlands a unique case in this regard.

In this chapter, we address the issue of pension wealth by defining it as semiprivate wealth and by presenting two series of aggregate wealth, one with and one without pension wealth; our top wealth shares series will be defined net of pension wealth. In our view, this gives the most transparent treatment of capitalized pension wealth, is consistent with international practice, and makes long-run series more meaningful, since the distribution of pension wealth has failed to be documented for the virtual entirety of our time frame.

1.3.2. Household Wealth Methods

We now turn to the discussion of the three distinct methods we employ to reconstruct aggregate wealth: (i) historical national accounts, (ii) the estate multiplier method, and (iii) lognormal extrapolation on wealth tax data. We describe each of these methods in turn, and briefly discuss the data sources we use for each method, and provide the details in the appendix. Table 1.1 provides a high-level overview of the different methods and data steps per time period.

1.3.2.1. Historical National Accounts

Our benchmark series reconstructs household balance sheets, with the aim of producing a series that is as consistent as possible with the current System of National Accounts (the 2008 version). For the post-1995 period, we can directly use the System of National Accounts' household balance sheets. Pre-1995, no official balance sheets exist; but an estimate compiled by the Netherlands Bureau of Economic Analysis (CPB) provides balance sheets since 1970, which we augment with wealth

Table 1.1: Overview of Methods and Data Sources

Purpose	Method	Years	Sources & Data steps
Aggregate Wealth	Historical National Accounts	1880–1938	Reconstruct manually
		1947–1969	Interpolate totals using savings flows
		1970–1994	Augment CPB estimates
		1995–2019	System of National Accounts
	Estate Multiplier	1854–1981	Death duties + multiplier estimates
	Lognormal extrapolation	1894–1993	Extrapolate from wealth tax returns
Wealth Shares	Generalized Pareto Interpolation	1894–1993	Wealth tax returns
		1993–2019	Tabulated full distribution

components not covered in these estimates. Pre-1970, we manually reconstruct household balance sheets. From 1880 through 1938 we use a variety of sources to build full balance sheets, whilst for 1947 to 1969 we have to interpolate stocks of wealth based on observed saving flows and estimated capital gains. Hence, we break this series into four methodologically distinct chunks: (i) 1880–1938; (ii) 1947–1969; (iii) 1970–1994; and (iv) 1995–2019.[2]

For 1880–1938, we use a variety of sources to manually reconstruct household balance sheets. To estimate the value of non-financial assets in the Netherlands for this period we relied first and foremost on property tax assessments. For housing –the principal component of non-financial assets– total housing area was published by the Central Bureau of Statistics (CBS 1880-1938). Moreover, we have an official Statistics Netherlands estimate of total housing for the year 1913. Hence, we obtain an estimate of the market value of housing per square meter for 1913. We index this price ratio for other years using the housing price index by Korevaar, Francke, and Eichholtz (2021), which is the most representative historical index for the Netherlands. In each year, the total value of housing then becomes the housing area multiplied by the indexed average housing price.[3] For the other main component of non-financial assets – agricultural land – we employ a similar procedure. Here, we

[2] In Appendix 1.A, we provide more detailed discussions for each wealth component separately.
[3] We confirm in Appendix 1.A that this method matches official balance sheets in 1938 remarkably well.

take the total volume of agricultural land from van der Bie (2001) based on Knibbe (1993), and then multiply this volume by the agricultural land price estimates by Luijt and Voskuilen (2009). The final major component of non-financial assets – the fixed capital stock – is based on detailed estimates by Smits, Horlings, and van Zanden (2000) and Groote, Albers, and De Jong (1996), adjusted by us to capture the part of the capital stock attributable to the household sector.

Moving to financial assets, the principal sources used to estimate the value of deposits and currency are balance sheet information made available by the Dutch central bank De Nederlandsche Bank (2000) as well as various publications by Statistics Netherlands van der Bie (2001). The total value of privately-held Dutch government bonds is directly observed in the Annual Statistics. For the value of domestic stocks, we start with the total Dutch stock market capitalization, available in van der Bie (2001). We then subtract the value held by foreigners and add the value of Dutch investments abroad using a capitalization of the net primary income from abroad, observed in Smits, Horlings, and van Zanden (2000) and den Bakker (2019). We capitalize these dividends using a "world average" dividend yield of the five financial markets most relevant for Dutch investors: Belgium, France, Germany, the UK, and the US. Data for all dividend yields are from Jordà et al. (2019). Finally, funded occupational pension entitlements and private insurance savings are based on the technical reserves as recorded by the Dutch Central Bank. This gives us estimated household balance sheets from 1880 until 1937. For the year 1938, we have an official household balance sheet from Statistics Netherlands, which we use to calibrate the estimates for the years prior to 1938.

After 1938, the earliest reliable year we can recover is 1947, when Statistics Netherlands also published a balance sheet for the household sector in its National Accounts. From this year on through 1969, we do not have the sources to reconstruct stocks of wealth, which only begin again in 1970, with the exception of some intervening years. However, since we do observe savings flows of wealth for the entire period, as well as an estimate of total private wealth in 1947 and in 1970, we can interpolate stocks of wealth for the years in between. Specifically, we can think of the accumulation equation of wealth W going from period t to period $t+1$ as being determined by both savings s and capital gains q:

$$W_{t+1} = (1 + q_{t+1})(1 + s_t) W_t. \tag{1.5}$$

We observe the initial wealth stock, W_{1947}, annual saving flows s_t, and a final wealth stock W_{1970}. Hence, we can residually estimate an average capital gains rate

$q_t = q$ using a recursive estimation process.[4] Once we have total stocks of wealth, we can estimate wealth composition for these years using available individual series on stocks of housing wealth, agricultural land, fixed capital stock, financial assets, and pension wealth to arrive at plausible estimates of the wealth components.[5]

From 1970, we have an estimate of household balance sheets by the Netherlands Bureau of Economic Analysis (CPB). In Appendix 1.A.5, we compare all wealth components in these balance sheets with external sources, finding that they are remarkably similar for the most part. We augment the balance sheet with life insurance wealth, agricultural land, and the fixed capital stock.

Finally, from 1995 until 2019, we simply use the most recent version of the System of National Accounts (the 2015 revision), maintained by Statistics Netherlands. To preserve consistency throughout the time series, we report end-of-year values for all years.

1.3.2.2. Estate Multiplier Method

The second method we employ to estimate aggregate wealth is the estate multiplier method, which builds on succession tax data. The succession tax was introduced in the Netherlands in 1818, and until 1877 only indirect heirs with an estate value higher than 300 guilders were taxed. From 1879, the tax widened in scope to include both direct and indirect heirs.[6] The administrative process creating this source has been described in de Vicq and Peeters (2020). In short, death duties list all assets – with no exemptions– a person possessed at the moment of death. The listing was made by representatives of the deceased (heirs, legatees, custodians, etc.) valued at market prices and verified by local officials via income and wealth taxation documents, and the cadastre.

Dutch statistical agencies published inheritance tax returns records in an aggregate form since the mid-1850s, reporting wealth information that refers to the decedent population using market prices (Gelderblom, Jonker, Peeters, and de Vicq 2023). The exact level of detail in the wealth reports depends on the period (see section 1.B.1 in the Appendix). In the literature, the usual caveat regarding this kind of data is that the threshold for including an estate in the inheritance tax returns records is relatively high. However, this does not hold in the case of the Netherlands

---

[4] I.e., we fix a value $q_t = \overline{q}$ for all years, calculate final wealth as $W_{1970} = \prod_{t=1947}^{1970}(1+\overline{q})(1+s_t)W_t$, and compare this estimate with actual wealth in 1970, stopping once the difference is negligible. See Piketty and Zucman (2014, Online Appendix K) for more details.

[5] The year 1969 is missing from the National Accounts, and is hence linearly interpolated from the values for 1968 and 1970.

[6] The exact thresholds and various changes introduced with later legislation are summarized in Table 1.5 in the appendix.

since the threshold approximates the annual earnings of an unskilled worker, and it can be argued that it is covering a broader range of the full wealth distribution, instead of a high upper tail (Gelderblom, Jonker, Peeters, and de Vicq 2023). Table 1.7 in the appendix demonstrates this using data from 1880–1980, showing that the coverage is around 10-20% in the earlier period, and 30-40% in the more recent period. This can be contrasted with the low 1–3% coverage found in the US (e.g., Kopczuk and Saez 2004), although it is somewhat lower than the coverage of in modern day data for Italy that comes around 61% (Acciari, Alvaredo, and Morelli 2024).

The idea behind the estate multiplier method is to use the available wealth totals from the inheritance tax (i.e. the sample), to estimate the total wealth in the entire population. This method works by multiplying the total wealth captured by the inheritance tax with the ratio of the deceased individuals covered by this tax over the surviving population in that year. Given that the sample in the annual inheritance tax returns, and the entire population, have different age-wise mortality rates (since age and size of the estate are positively associated and size of the estate is not orthogonal to age), a correction needs to be applied in the form of adjusted mortality rates. Typically, age-wise mortality rates for the entire population are available from the national statistical agencies. However, we need additional data to estimate the mortality rates for the individuals present in the inheritance tax records. One solution is to turn to companies that keep records of life insurance holders, and their respective mortality rates (e.g. see Lampman 1962). Other researchers have used social class mortality multipliers based on occupational classes (Alvaredo, Atkinson, and Morelli (2018) and Atkinson and Harrison (1978)), or use more sophisticated wealth mortality gradient using the relationship between mortality and housing wealth distribution (Acciari, Alvaredo, and Morelli 2024). In our approach we use mortality rates that are estimated based on the detailed individual level data from the Tafel V-bis annual ledger for 1921, which has been made available by Gelderblom, Jonker, Peeters, and de Vicq (2023). In Appendix 1.B.1 we explain the details of Tafel V-bis and how we apply this entire procedure.

1.3.2.3. Wealth Tax Method

The wealth tax was instituted in 1893. The first collections of statistics appeared for the fiscal year 1894, and the data series continues until 1993. Initially, the wealth tax only applied to fortunes with a value of 13,000 guilders and above.[7] Households above these thresholds typically comprised about 5-7% of the population. In 2001,

[7] This threshold was later set to 15,000 in 1915, then dropped to 10,000 in 1947, to subsequently be increased to 50,000 in 1957, 100,000 in 1970, and 200,000 guilders in 1983.

the historical wealth tax was replaced by the current tax system, which taxes – at least ostensibly – capital income rather than net wealth. The historical wealth tax did not apply to households, but to natural persons, making direct comparisons with later tax statistics slightly difficult. Married couples were treated as a single natural person for tax purposes. The tax basis covered financial assets (including both listed and nonlisted stocks and other forms of pass-through business wealth), deposits and cash, real estate, items of transport such as horse-carts and cars, claims to life insurance, and from 1918 onward also jewelry and precious metals. It did not cover pension claims, artworks, or consumer durables. The assessment was in principle on market value basis, although this was not always easy to establish for all wealth components (Wilterdink 1984).

We have tabulated wealth tax data from 1894–1993, with a few interruptions, particularly around the second World War. In the Dutch literature on national wealth the key reference is Wilterdink (1984), for which Potharst (2022) has developed (and published separately much later) a method for estimating μ and σ of lognormal distributions from censored tabulated data on wealth. The basic assumption they use is that the sum of wealth below the tax threshold, is equal to that of a lognormal approximation of the wealth distribution, again below the tax threshold, with $W_t \sim \log \mathcal{N}(\mu_t, \sigma_t^2)$ and μ and σ estimated from the available wealth tax data. By adding this estimated component to the wealth captured by the wealth tax data we arrive to an estimation of total wealth. The issue then becomes how to estimate μ_t and σ_t^2 using the available information on the top brackets from the wealth tax data. In the method developed by Potharst, each data point takes the form of a triple (k, n_k, μ_k), where k is the lower threshold of the bracket (e.g., 1,000 guilders), n_k is the number of individuals in that bracket, and μ_k is the bracket's average wealth. We obtain our estimate for μ and σ^2 by minimizing the distance between a lognormal distribution and the parameters of the low threshold of each bracket and the frequencies of observations per bracket (see Appendix 1.C).

1.3.3. Wealth Shares

Once we have estimates for aggregate household wealth, we have a denominator for top wealth shares. For the numerator – i.e., total wealth per percentile – we rely on wealth tax statistics for 1894–1993, and for the wealth distribution statistics by Statistics Netherlands from 1993 onward. The wealth tax statistics were published in tabulated form in *Jaarcijfers voor Nederland*; we can therefore employ the semiparametric interpolation method developed by Blanchet, Fournier, and Piketty (2022) to estimate top wealth shares. This method, known as generalized Pareto interpolation, takes as inputs for each bracket k a bracket lower threshold q_k, its corresponding percentile p_k and the bracket average μ_k. It then interpolates the entire distribution based on the given inputs and the (known) mean population wealth $\overline{\mu}$ (previously estimated from the Historical National Accounts). Since we do not possess information about the distribution below the lowest threshold of the wealth tax – usually the bottom 95% or so – we hesitate to report the interpolation results for anything but the upper 5% of the distribution.[8] A caveat here is that wealth as captured by wealth tax does not include some important components such as pension wealth and physical capital stock. To avoid wealth shares being mechanically biased downwards by this mismatch between wealth tax data and historical national accounts, we scale up the total wealth in each bracket by the average ratio between the historical national accounts method and the lognormal extrapolation. The lognormal extrapolation, which is based on the wealth tax statistics, may then be viewed as an estimate of aggregate taxable wealth; by scaling up wealth in each bracket, we implicitly assume that the wealth that is not captured in the wealth tax statistics is spread relatively uniformly across the distribution. Although this is not an entirely appealing assumption, it is the best possible solution to the fact that distributional statistics in historical data are scarcely available.

Post-1993, Statistics Netherlands starts reporting tabulated data for the full wealth distribution. These data are taken from a variety of fiscal sources, including the wealth tax, wealth surveys, and in recent years also mandatory disclosures of deposits by banks and other financial institutions. In this chapter, we only report the upper percentiles of these Distributional Accounts, to preserve consistency with the historical series.

It is well-known that the post-1993 data, while significantly better than the wealth tax records that preceded it, significantly increase in quality from 2011 onward, when Statistics Netherlands started drawing on the universe of tax filers and received automatic disclosure of various wealth components. In this chapter, we simply take wealth components as reported by Statistics Netherlands at face value; but doubtlessly, future research can and should critically adjust distributional statistics to produce fully time-consistent Distributional Financial Accounts.

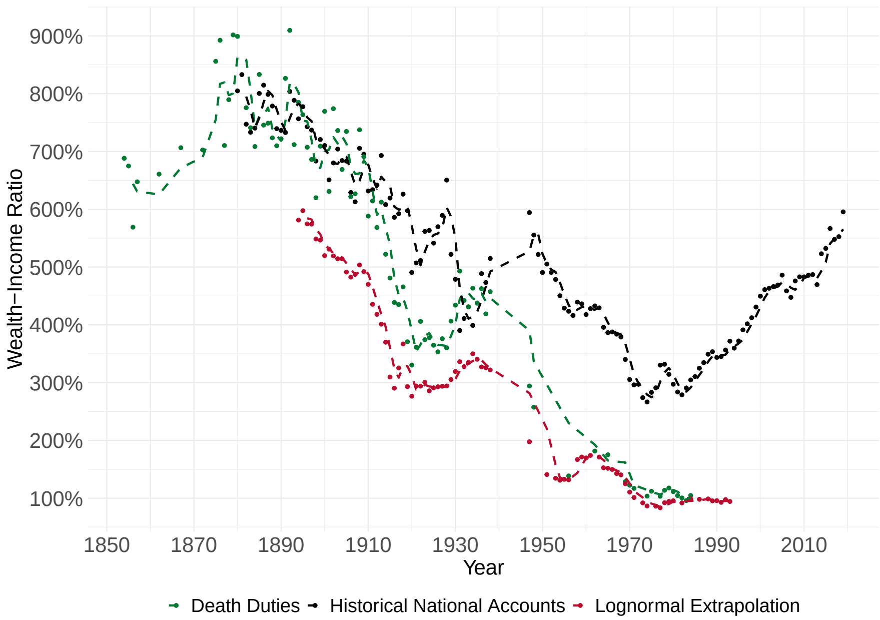
1.4. Results

1.4.1. Aggregate Wealth: Comparing the Three Methods

Figure 1.2 presents the ratio of household wealth to national income for the period 1854–2019.

[8]This concern is noted by Blanchet, Fournier, and Piketty (2022) as well, who caution that the interpolation works best if at least some lower quantiles are also covered.

Figure 1.2: Wealth-Income Ratios per Method



Notes: Figure shows the ratio of household wealth to net national income, using our three main methods: (i) Historical National Accounts (the benchmark), (ii) Estate Multiplier methods using death duties, and (iii) Lognormal extrapolation from wealth tax data.

From the 1880s until the 1920s, the historical national account and estate multiplier methods are remarkably similar, and diverge afterwards. The wealth tax method results in lower wealth-income ratios in the same period, as noted before by van Bavel and Frankema (2017). After the 1950s the estate multiplier and the wealth tax methods give very similar estimates which are well below the historical national account method. This is likely due to the increasing importance of pension claims, which are covered in the historical national accounts but not in the other two methods.[9] Another reason often identified is the increased unreliability of both wealth tax records and estate tax records (Wilterdink 1984; van Bavel and Frankema 2017), which might bias results based on those sources downwards.

[9]In Figure 1.19, we add semiprivate wealth to the baseline series of Lognormal Extrpolation and Estate Tax Multiplier methods, and show that this does reduce the gap with the historical national accounts quite substantially from the 1980s onward.

The first lesson we can draw from this comparison is that while the three methods differ in their estimated levels, they do identify similar trends. This similarity is further strengthened in the post 1950s period once pensions are added to the estate multiplier and the wealth tax estimates (see Figure 1.19).

A second conclusion is that the estate multiplier method is highly similar to the historical national accounts prior to the 1920s, which makes it a useful substitute for countries or years where historical national accounts sources may be unavailable. After the 1920s, the historical national accounts are much larger than the estate tax multiplier. As we will discuss in more detail in Section 1.5.1, this was a period when capital gains on colonial wealth holdings skyrocketed; hence, it might be possible that death duties did not fully reflect these asset revaluation effects.

One interesting trend in the data is the increase in the wealth-income ratio between 1930 and 1950 across all three series. This trend is most prominent in the historical national accounts, and is unlikely to be driven by measurement error, since we have an official balance sheet both for 1938 and 1947, which indeed show that household wealth almost doubled in the intervening decade. The increase in the wealth-income ratio prior to 1938 is denominator effects since national income declined strongly in the aftermath of the Great Depression and World War II, whereas private wealth declined relatively less. Of course, the more-than-doubling of national wealth between 1938 (from 28 billion guilders) and 1947 (more than 70 billion) is in nominal terms. Even in real terms, however, wealth remained more or less constant or even grew slightly. According to these official estimations the destruction caused by the war did not reduce the value of private wealth as much as it did elsewhere (Piketty and Zucman 2014).

After World War II, aggregate wealth grew less rapidly than national income, resulting in the familiar decline in the wealth-income ratio observed elsewhere (Piketty and Zucman 2014. Similar to other countries, the trough occurs somewhere in the 1970s, at a depth of almost 300% of national income. Afterwards, private wealth grew in importance again relative to national income, speeding up in the 1990s and reaching a post World War II period maximum of 600% in 2019. In the next section, we decompose household wealth to analyze the underlying trends.

1.4.2. Wealth Composition

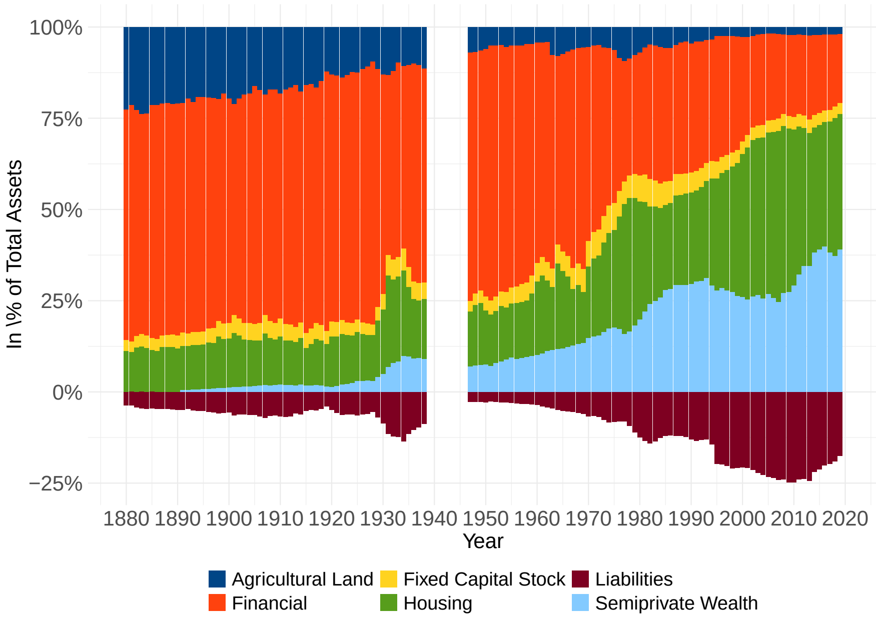
Figure 1.3 illustrates the decomposition of household wealth from 1880 until today. We focus on the following broad categories: Financial assets, agricultural land, housing, other capital stock, liabilities, and semiprivate wealth. In panel (a), all asset classes are expressed as a percentage of the value of total assets, and in panel (b), we express the same series in percentage of the net national income, to give an idea of the magnitudes of each wealth component.

The period between 1880 and 1938 is first and foremost characterised by a sharp decline in the relative value of agricultural land compared to net national income, declining from almost 200% in 1880 to 60% in 1938. This coincides with trends in several other countries (Piketty and Zucman 2014), and is driven to a large extent by falling agricultural prices as a result of global competition and technological improvement (Knibbe 2014). A second major finding is the strong importance of financial assets net of semiprivate wealth; this category dominates all others until the late 1970s. As we will explore in Section 1.5.1, this is mostly driven by securities, both domestic and foreign. The Netherlands developed a stock market in the 17th century, and this early financialization persisted over time. Furthermore, by the start of our series in 1880, the Industrial Revolution had taken off in the Netherlands, which led to a boom in industrial corporations seeking equity (van Zanden and van Riel 2000). Finally, and very significantly, foreign securities played an outsized role in the Dutch economy from the 1870s onward, with investments in American steel and railroad companies as well as Austrian and Russian bonds making up a major part of Dutch investors' portfolios. Notably, the Dutch colony of Indonesia also opened up for private investment in the late 1870s, and would become a major source of household wealth by the early 1900s.

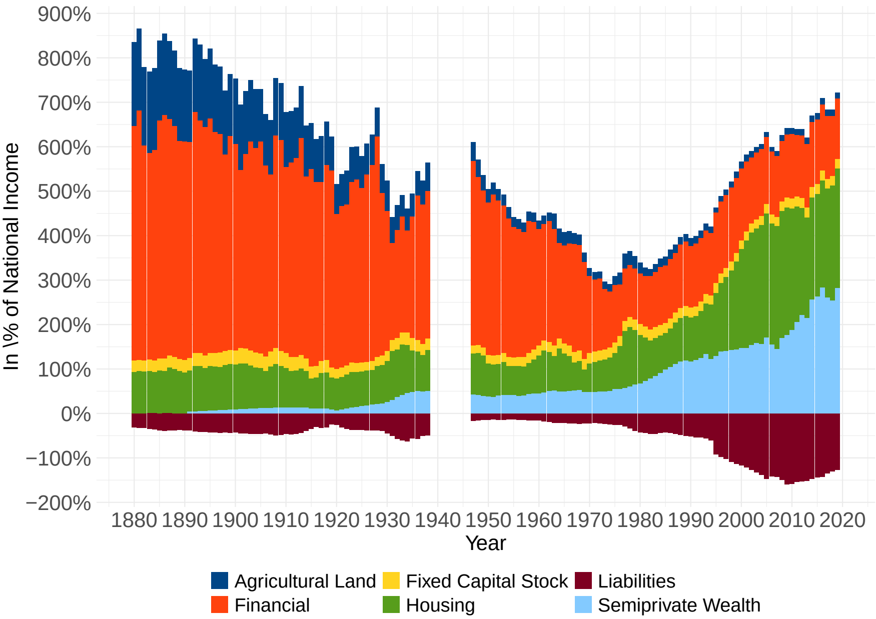
The other wealth components we identify play smaller roles prior to World War II. Housing is relatively constant at around 100% of national income, or 10-15% of household assets. Liabilities build up quite significantly before World War II. This is mostly due to an increase in mortgage debt (de Vries 1976).

After World War II, we notice several similarities and other differences. Semiprivate wealth remains relatively constant, both as a share of wealth and as a share of national income. The same holds true for all non-financial assets – housing, land and the capital stock. In contrast, the major reason for the decline in the wealth-income ratio in this period is due to a relative decrease in the value of financial assets. The Bretton Woods system featured capital controls and other measures which limited capital mobility and the ability of securities to be traded freely. The Dutch stock market grows rather slowly from 1947 until 1980, only picking up afterward. Furthermore, foreign investment was much more limited in this period than before World War II. This is especially true for investments in Indonesia following Indonesian independence. We observe a sharp drop in the relative value of securities in the late 1950s, which corresponds precisely with the peak of nationalization of Dutch firms remaining in Indonesia at the time: a process also referred to as

Figure 1.3: Wealth Composition, 1880–2019



(a) as % of Total Assets



(b) as % of National Income

Notes: Figure shows the composition of household wealth, with each asset expressed as a proportion of total assets (panel (a), and as a proportion of national income (panel (b)). All wealth composition data are from the Historical National Accounts method.

Indonesianization.[10]

One of the most striking trends in the post-1970s period is the dynamic in the relative value of housing. Starting from a precipitous decline in relative value in the early 1980s, the value of housing then continued to grow almost uninterrupted until the mid-2010s. This increase in the value of private real estate was mirrored by a sharp increase in home mortgages; which, as a percentage of net national income, nearly quadrupled throughout this period, reflecting the growing financialization of the Dutch economy. A particular institution which contributed to the buildup in mortgage debt was the interest-only mortgage, which was enormously popular until the Great Financial Crisis of 2008 (Bernstein and Koudijs 2024). To a lesser extent this is also reflected in the stock market. While the holdings of both domestic and foreign securities decreased throughout the subsequent oil crises in the 1970s, it gradually increased in the following years. This growth came to a standstill due to the Dot-com bubble in the early-2000s and more recently the Global Financial Crisis of 2007-2008. The relative value of other asset categories seem notably stagnant throughout this period, which is similar to most of the 1880–1938 period.

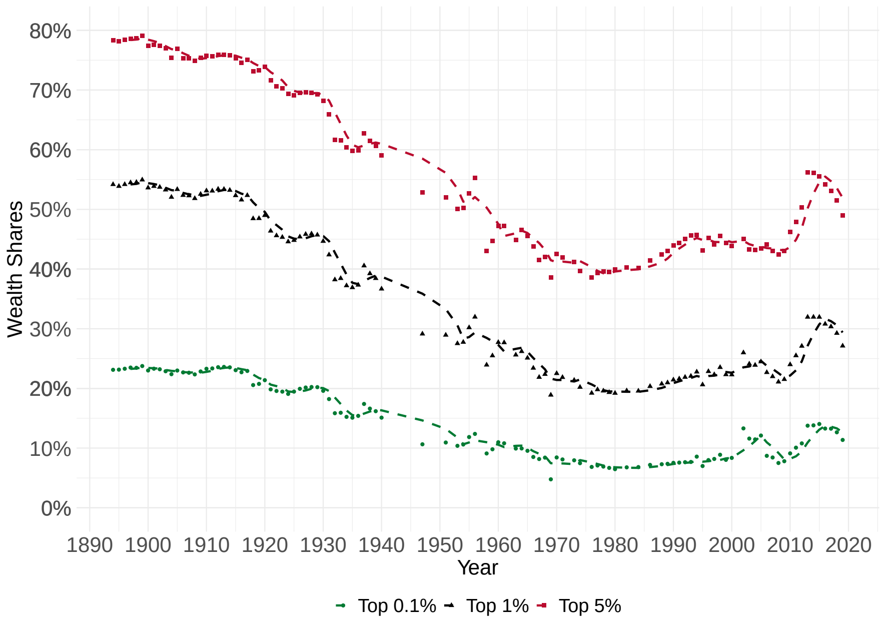
Notably for the Netherlands, it is important to point out the rapid growth of semiprivate wealth after World War II. Pension wealth in particular expanded exponentially from the 1980s. Pension funds increasingly matured with a growing number of households that paid pension contributions for the full duration of their career. Falling discount rates play their role in more recent years by inflating the net present value of these household claims on pension funds. Whereas it was once all but a minor component of private wealth, it now slightly exceeds housing as the most important asset of households.

1.4.3. Wealth Shares

We now turn to our results for top wealth shares. Figure 1.4 shows the evolution of the top 5%, 1%, and 0.1% share from 1894–2019. As noted before, the data for this series are drawn from the wealth tax records and the results are estimated using generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022). The individual data points are connected by a three-year moving average trend line. Since the distribution of pension wealth is historically unavailable, this wealth component is excluded in these figures.

[10] 'Indonesianization' (indonesianisasi) is defined as the process set in motion by the Indonesian government to realise an 'economic decolonization'. It consisted of various measures aiming at the transfer of property and economic functions held by foreigners or foreign businesses (primarily Dutch), or residents viewed as foreigners (primarily of Indonesians of Chinese descent), to what the government considered to be indigenous Indonesians (Thee 2012).

Figure 1.4: Evolution of Top Wealth Shares, 1894–2019



Notes: Wealth shares calculated using generalized Pareto interpolation and the wealth tax records. Three-year moving average shown in dashed lines.

We observe a familiar U-shaped pattern in top wealth shares over the 20th century. The top 1% share peaked at close to 60% around 1900, before declining to as little as 20% by the end of the 1970s. From the late 1980s, the top shares start increasing again, peaking at a 1% share of around 35% in 2015. Several caveats about this trend should be noted. First, a lack of data availability of the full distribution before 1993 precludes robust conclusions about the evolving role of lower percentiles, which makes it difficult to verify the trends observed in these top wealth shares. Second, it is generally accepted that the wealth tax records become increasingly uninformative from the 1970s onward (Wilterdink 1984; van Bavel and Frankema 2017). This is due to several factors, including increased thresholds (up to 200,000 guilders by the late 1970s), increased exemptions of business capital from the 1980s onward, as well as – according to Wilterdink (2015) – a weakening of tax morale, and an observable increase in the reliance on offshore tax havens such as Curaçao to avoid taxation (van Beurden and Jonker 2021). These factors are difficult to quantify precisely; however, in the raw output we observe a marked jump in top wealth shares around 1993 – the first year for which Statistics Netherlands is able to recover the full distribution with some confidence – particularly for the 5% share.[11] The unavailability of reliable sources before 1993 makes it difficult to determine which part of the decline in wealth shares prior to 1993 is due to worsening data quality and which part is a true decline in wealth concentration. In this chapter, we opt for a pragmatic approach to address this jump. We assume that wealth tax data is more or less reliable until 1964, when a major tax reform occurred (Wilterdink 1984, 2015). We assume the same for the data from 1993 onward. In the intervening years, we gradually adjust the raw output to make the series match the improved data in 1993. See Appendix 1.C.2 for further details.

Data availability vastly improves after 1993, allowing us to reconstruct the full distribution of wealth with far higher confidence than before. Nevertheless, despite this slight break in the series, the overall trajectory pre- and post-1993 is clear, with a clear upward trend that continues to this day, interrupted by a contraction in top wealth shares after the recessions in 2001 and 2008. After the peak in wealth concentration in 2015, the top 1% share fell somewhat until 2019. As we analyze in Section 6, this is likely explained by the strong increase in housing prices. As housing is the middle class's most important asset, this increases their wealth relative to the wealthiest 1% whose wealth is mainly in closely held business equity and financial assets. This mechanical contraction in top wealth shares resembles the "race between the housing market and the stock market" analyzed in the United States by Kuhn, Schularick, and Steins (2020).

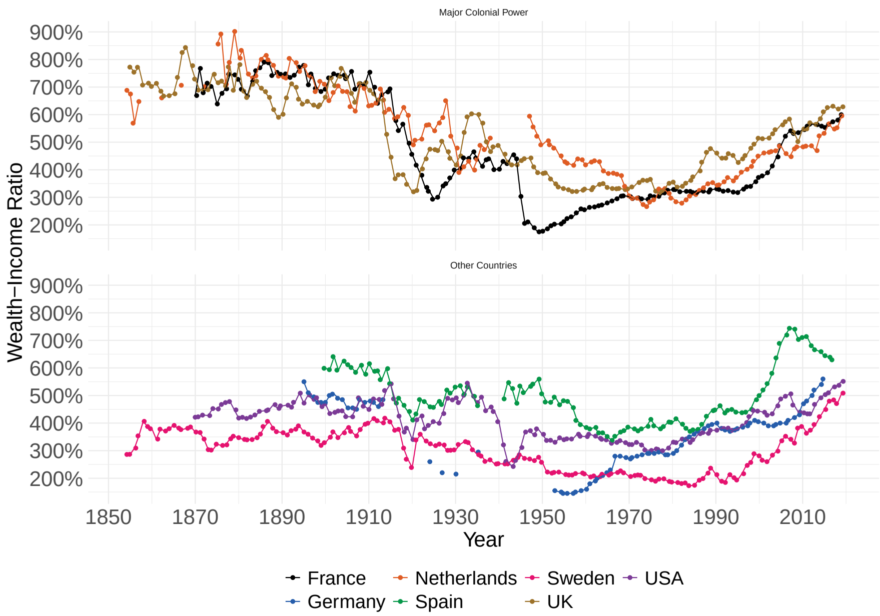
1.4.4. International Comparison

Figure 1.5 looks at private wealth-income ratios for several western countries for which such long-run evidence exists: the United Kingdom, France, Germany, Sweden, Spain, and the United States. We use the series for Spain by Artola Blanco, Bauluz, and Martínez-Toledano (2020), as well as the estimates by Albers, Bartels, and Schularick (2022) for Germany and the series by Waldenström (2017) for Sweden; all other estimates are from Piketty and Zucman (2014). The Spanish, German, and Swedish series are all quite substantially lower before World War I than the series reported by Piketty and Zucman (2014). This fact has led Waldenström (2021) to state that wealth-income ratios were much lower pre-World War I than previously thought.[12]

[11] See Figure 1.18 in the Appendix.

[12] Waldenström (2021) also uses the UK series by Madsen (2019), which arrives at substantially different estimates than Piketty and Zucman (2014) based on cumulated investments starting in the 1200s, as well as upward revisions of national income. We use the Piketty-Zucman series in Figure 1.5 for methodological comparability with the other series.

Figure 1.5: Wealth-Income Ratio, International Comparison



Notes: Figure shows the evolution of the ratio of private wealth to net national income for the Netherlands, Germany (from Albers, Bartels, and Schularick (2022)), Spain (Artola Blanco, Bauluz, and Martínez-Toledano 2020), Sweden (Waldenström 2017); as well as the series for the UK, France, and the US from Piketty and Zucman (2014). The top panel shows France, the UK, and the Netherlands, which we have termed "major colonial powers"; all the other countries are shown in the bottom panel.

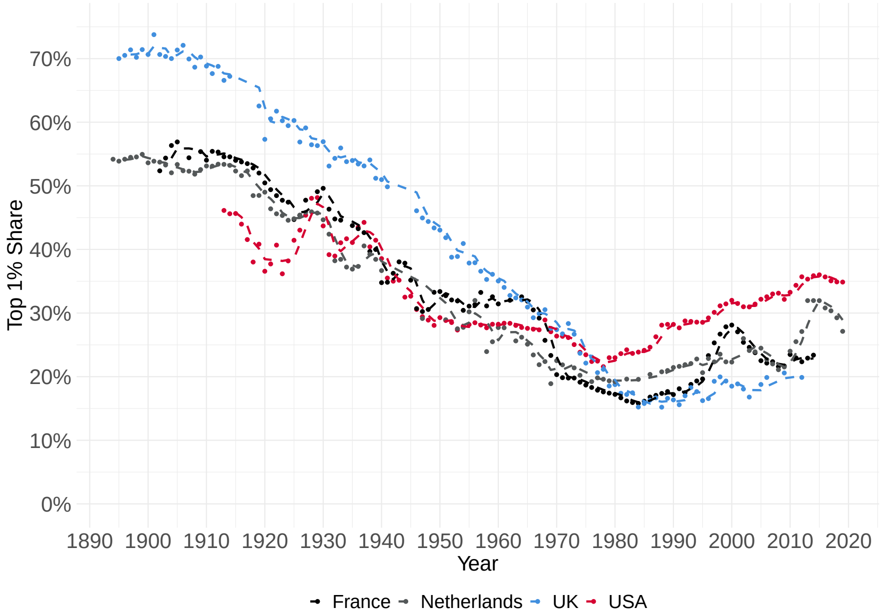
Our evidence, however, provides important nuance to Waldenström's claim. We notice that the Dutch series is among the highest observed, peaking in excess of 900% in the early 1880s, and generally being remarkably close to the French and British series until the late 1910s. By the 1920s, the Netherlands has by some distance the highest wealth-income ratio, only to converge to the other series after the 1929 crash. Hence, on the basis of these series, we can group pre-World War I countries into two groups: a high wealth-income ratio group, which consists of France, the Netherlands, and the United Kingdom; and the other countries. Not coincidentally, the first group consists of major colonial powers, whereas the other countries had no or relatively negligible colonies during this period. Understanding the differences between these groups, and between the Netherlands and these other colonial powers in particular, will be at the center of our analysis in Section 1.5.

For all countries, ratios continued to decrease after the World War II until the 1980s, when all countries experienced a sharp increase from about 300 percent in 1970 to about 400–600 percent today. This general trend obscures country-specific variation. In Europe, the trajectories for France and the United Kingdom are comparable: the private wealth–income ratio rose from about 300 percent in 1970 to about 550 percent in recent times. In Germany and Sweden, the increase was more notable, but the wealth-income ratio remained lower overall, never exceeding 400 percent up until the early 2000s. In the last decade, Sweden however experienced a remarkable increase; with the wealth–income ratio now exceeding 500 percent (Piketty and Zucman 2014: 1277-1279; Waldenström 2017: 291-293). Focusing on how the Netherlands fared compared with the other countries, two findings stand out. First, the Netherlands' wealth-income ratios are low compared to all other countries from roughly the mid 1960s until the mid 1980s, always hovering around 300 percent. A second finding is that the Netherlands appears to follow a similar trend of a sharply increasing wealth-income ratio in recent decades, most similar to the patterns observed for Germany and Sweden.

In sum, comparing the main trends for the Netherlands with other countries for which long-term series on private wealth-income ratios are available shows first and foremost that the patterns are relatively similar. Nevertheless, somewhat provocatively one could label the Netherlands as the land of the extremes. It is characterized by both the highest ratio from the mid-19th century until the early 20th century, as well as one of the lowest ratios from the mid-1960s until the mid-1980s.

Next, we turn to the international comparison of top wealth shares. Figure 1.6 shows the long-run evolution of the top 1% share in France, the Netherlands, the United Kingdom, and the United States. We observe that the trends for the top 1% share are much more similar across all countries. The United States, which had lower levels of inequality than the other countries pre-World War II, converges in the 1930s. The United Kingdom has markedly larger wealth shares than the other countries until the 1950s; it is an interesting open question why this should be so. From the 1980s, the United States start to notably diverge, with higher top wealth shares than the other countries. Note, however, that the Netherlands approaches the U.S. in recent years, both hovering between 30 and 40%, although the trends in recent years are diverging with the Netherlands demonstrating a decrease. This observed volatility is to a large extent explained by the major role of housing for aggregate wealth, and most importantly the middle class. As can be observed from Figure 1.3 the trajectory of total housing value coincides with the movement of Top 1% wealth shares in the post 2010 period.

Figure 1.6: Top 1% Wealth Share, International Comparison



Notes: Figure shows the wealth share of the top 1% for France, the Netherlands, the UK, and the US. Data are from the World Inequality Database.

Since the underlying data for top wealth shares are more fragile than those for aggregate wealth, we view that conclusions from these long-run trends should be drawn sparingly.[13] However, it is clear that top wealth shares have increased since the 1980s, reaching levels not previously seen since the 1950s.

1.5. The Rise and Decline of Wealth pre-World War II

In this Section, we dive deeper into the dynamics of the Dutch wealth-income ratio before World War II. Specifically, we are interested in two questions: Why was the wealth-income ratio so high in the Netherlands? And why was the wealth-income

[13] Apart from the concerns about data quality, which we have discussed extensively, there are different estimation methodologies associated with different *kinds* of data. It is quite plausible that a data series estimated using estate multiplier techniques, such as the British data (Alvaredo, Atkinson, and Morelli 2018), would differ structurally from series estimated with income capitalization methods, such as the U.S. series (Saez and Zucman 2016). See also the discussions on this issue by Kopczuk (2015), among others.

ratio considerably lower in non-colonial powers? We will argue that the historical context provides compelling answers to both questions. In our analysis, we focus on two aspects. First, we note that the extent of Dutch foreign and colonial investment was unprecedented, particularly from the 1910s onward, and that this can account for a large part of the divergence with most other countries. Second, we also note that *denominator* effects matter: cross-country differences in the wealth-income ratio reflect not only differences in wealth levels but also in national income levels. We will argue that this explains, for instance, why the British wealth-income ratio was lower than the Netherlands prior to the 1930s, since its national income was significantly higher.

1.5.1. Explaining the Trends: Foreign and Colonial Asset Holdings

The upheavals of the late eighteenth century and the occupation of the French had inaugurated a period of economic stagnation in the Netherlands lasting some fifty years. Recovery after the restoration of independence in 1813 proved relatively slow, until the mid-1820s. The lack of an economic impetus meant that most companies did not need to raise substantial amounts of capital (Jonker 1996). At this time, colonial investment by private individuals was also limited. Indonesia – the largest and most important Dutch colony by far[14] – had been administered by the Dutch East India Company until its dissolution around 1800, and had been contested by France and Britain in the Napoleonic era. By 1830, the Dutch had re-established some control over the archipelago, and developed a new colonial policy to finance the Dutch mounting sovereign debt: the *Cultuurstelsel* (Cultivation System). This was a system of forced labor, mainly on the island of Java, where farmers were forced to grow cash crops like coffee and sugar.[15] These cash crops were sold far below market prices to the *Nederlandsche Handel-Maatschappij*, a trade corporation which had a government-granted monopoly. As a result of the monopoly of the Handel-Maatschappij, private investment in Indonesia was absent in this period. Only when the Cultivation System was phased out in the 1870s,

[14] During the period we study in our chapter, the Netherlands also held other colonies, notably Suriname and the Dutch Caribbean, which became independent (with some exceptions) in the 1970s. We do not attempt to estimate colonial possessions in these "West Indies" due to data limitations. After the abolition of slavery in the 1860s, these colonies are unlikely to have contributed significantly to household wealth, making this exclusion on pragmatic grounds defensible. We leave the important task of quantifying the value of the Western colonies pre- and post-abolition to future research (see Koudijs, de Jong, and Kooijmans (2022) for research in this direction.)

[15] See van Zanden and Marks (2013) for a general economic history of Indonesia, and Dell and Olken (2020) for a recent economic analysis of the Cultivation System.

did private colonial investments really commence. The period from 1880 until World War II is generally referred to as the "liberal period" in the colonial history of Indonesia, and is the period we study most intensively. In 1945, Indonesia declared independence, which was recognized by the Netherlands in 1949, following a brutal colonial war. In the late 1950s, Indonesian President Sukarno forcibly nationalized remaining Dutch enterprises, without compensating the shareholders (Thee 2012). This marks the end of colonial asset holdings by Dutch private households.

Our main period of study in this section, the liberal period of 1880–1942, coincides with burgeoning industrialization in the Netherlands. It is also at the same time that investment in other foreign countries accelerates (van Zanden and van Riel 2000). Chief among these other foreign investments were industrial corporations in the United States, particularly steel and railroad companies. Bosch (1948) estimates the total value of Dutch investments in the United States in 1908 at 1.5 billion guilders, or close to 100% of national income.

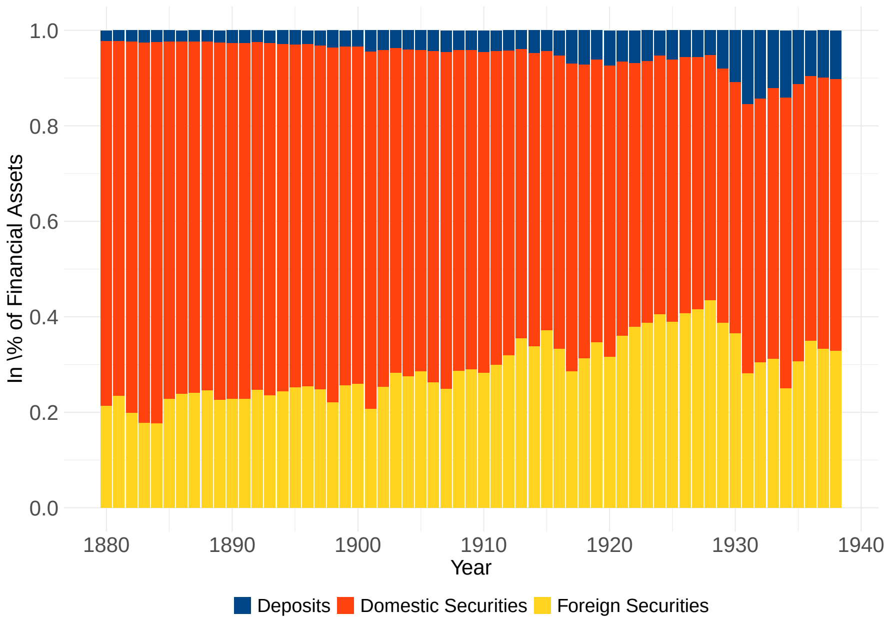
The aforementioned trends are made visible by Figure 1.7 and Figure 1.8, which show the composition of total financial assets (minus semiprivate wealth) from 1880 until 1938. Both series are based on observed dividends, which were carefully recorded on an annual basis by the Statistics Netherlands throughout various historical publications (Smits, Horlings, and van Zanden 2000; den Bakker 2019). We capitalize these dividends using information on dividend yields from Jordà et al. (2019), to arrive at total wealth invested abroad.[16]

We observe that foreign investment was substantial from the beginning of our series in 1880. Prior to the 1900s, most of foreign wealth was non-colonial, likely invested in American corporations (Bosch 1948), as well as in Austrian and Russian bonds (de Vries 1976). Colonial dividends were small in this period, but increased dramatically during the long era of economic expansion which started in 1895 and ended in 1914 (van Zanden and van Riel 2000). Although the Cultivation System had been abolished, cash crops still amounted for the majority of wealth generated in Indonesia. From the 1890s, oil reserves were also found on the islands of Sumatra and Borneo, fueling the rise of the corporation now known as Shell. The expansion of colonial investment is clearly visible in Figure 1.8.

Non-colonial foreign investment collapsed during the 1910s. The impacts of World War I, as well as the accompanying hyperinflation, likely played an important role here. In contrast, is that colonial investment only continued to expand during this period. After the end of World War I, non-colonial investment crashed, reaching a nadir of around 10% in 1920; whereas colonial investment peaked to reach a zenith

[16]Note that, to the best of our knowledge, ours is the first series on private colonial wealth. In Appendix 1.A.3.4, we discuss this methodology and show how it is robust to other specifications. Moreover, we show how our figures closely match various estimates in the historical literature.

Figure 1.7: Composition of Financial Assets, 1880–1938

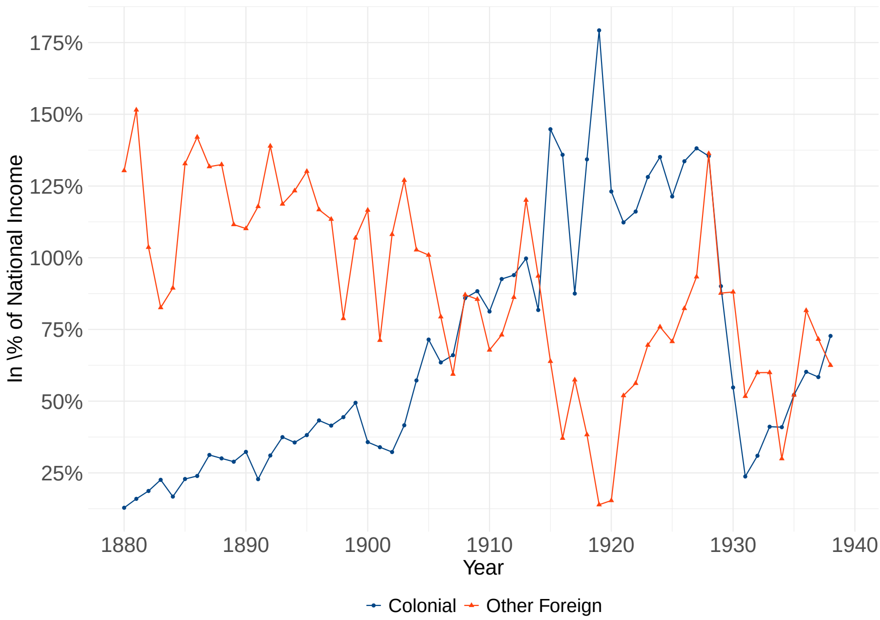


Notes: Figure shows the composition of financial directly held assets, from 1880 until 1938. All data are based on the Historical National Accounts method.

of approximately 175%, although with substantial volatility. Disregarding this temporary peak, both colonial and non-colonial investment trended upward again after the early-1920s. This is mainly driven by a boom in colonial commodity prices. The British had tried to introduce import restrictions on many cash crops to boost the post-war economy, the so-called Stevenson Plan. The Netherlands ignored this plan and continued to export large amounts of cash crops and petroleum during the 1920s. The reduced supply by other producers resulted in high demand for Indonesian commodities, boosting profits and prices to unprecedented levels (Buelens and Frankema 2016). At its late-1920s peak, colonial investment was worth approximately 140% of national income. A recovery in investment in the United States and other countries had also led to enormous growth in non-colonial foreign investment, resulting in a total foreign wealth peak of almost 300% of national income in 1928.

The Great Depression hit foreign investment hard; Indonesian corporations recorded losses of around 20% of their total value (van der Eng 1998), and investment in the United States also collapsed (Bosch 1948). By the late 1930s, some of these losses had been undone, with the final value of Dutch total private foreign

Figure 1.8: Foreign and Colonial Investment, in % of National Income



Notes: Figure shows the evolution of (net) foreign and colonial investment from 1880 until 1938. Both series are expressed as percent of net national income.

investment in 1938 being close to 140% of national income.

Note that all these figures, which are based on colonial dividends and other returns to capital investment, are an estimate of the *direct* impact of Indonesia and foreign investment on household wealth. In several ways, this represents a lower bound on the overall importance, given that the value of many *domestic* firms also depended indirectly on the colonies. The textile industry in the rural region of Twente, for instance, would derive much of its value from the availability of a large sales market in Indonesia. We do not attempt to quantify such general-equilibrium effects of colonial investment here, but note that precisely such a calculation has been made back in 1945 by Derksen and Tinbergen (1945). Their estimate puts the total contribution of Indonesia to Dutch national income in 1938 at almost 14% (or about 765 million guilders at the time). In the same article, they estimate the total value of Dutch private investment in Indonesia at almost 4 billion guilders, which equals our estimate.

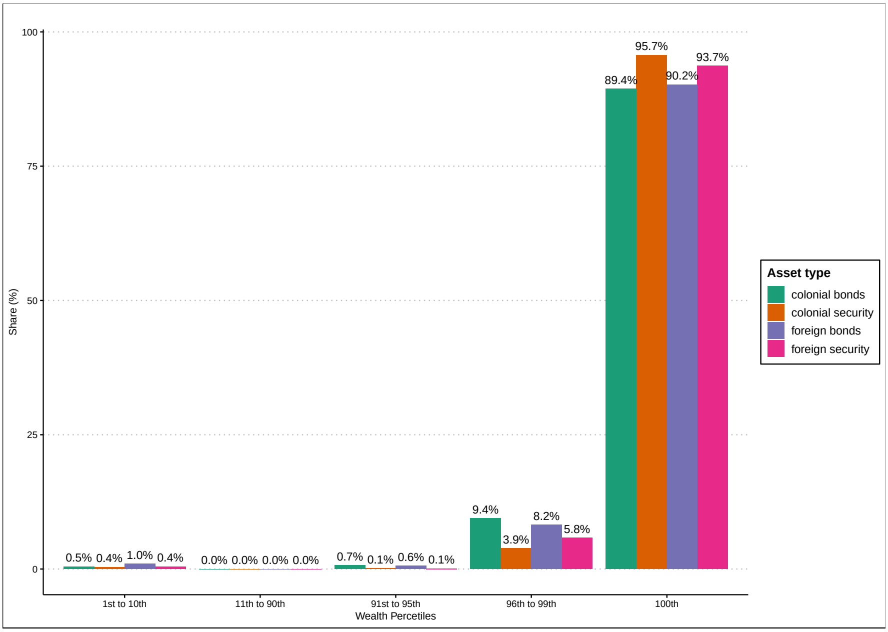
In an attempt to more properly assess the long-term impact of colonial investments on the wealth distribution, and to estimate the share of households invested in foreign and colonial securities, we rely on a dataset recently made available by Gelderblom, Jonker, Peeters, and de Vicq (2023). This dataset is based on a stratified sample of 2,325 death duties for 1921. It was drawn from an annual ledger (the so-called Tafel V-bis) containing the personal information and net worth of approximately 24,000 individuals considered for assessment in the inheritance tax, out of a total of 77,000 individuals who died in 1921.

This year was chosen because the inheritance tax records completed after 1927 are not yet publicly available due to privacy restrictions and because bank assets to GDP peaked at about 70 per cent in 1920-1921 period just before the sharp drop following the financial crisis which hit the sector in 1922 Gelderblom, Jonker, Peeters, and de Vicq (2023). Put simply, 1921 may be considered as the zenith of pre-World War II financialization in the Netherlands. More importantly for our purposes, it also nears the peak in the share of investments in colonial assets as a percentage of the net national income. Given that these tax records are entirely hand written and therefore enormously time consuming to process, we are unable to replicate their efforts for additional years.

In total, their publicly available database contains over 75,000 unique entries of assets and liabilities on the individual level. By relying on the full description, transcribed from the original death duty, we further distinguished colonial from foreign and domestic securities, allowing for a full construction of the investment portfolio on the individual level. Given that the aforementioned ledger contains information on individuals with as little as 20 guilders, despite the official threshold being 1,000 guilders, it is fair to assume that the remaining 53,000 individuals who died in 1921, but were not listed on this ledger had close to zero wealth. Based on this assumption we then proceeded to calculate the distribution of colonial and financial securities across the wealth distribution. Our results are presented in Figure 1.9. We find that the distribution of both foreign and colonial securities is highly skewed, with the top 1% holding up to 95.7% of the total value of colonial assets. The possession of bonds, both foreign and colonial, were slightly less concentrated, but only marginally so.

How do these statistics compare to other countries? To the best of our knowledge, no direct estimates of colonial investment in total wealth are available for other major colonial powers. Nor do we have any information on how colonial (and foreign) securities were distributed amongst households. Instead, we will compare net private foreign asset positions of three major colonial powers for which these are available: the Netherlands, France, and the UK. The result is given in Figure 1.10. In panel (a), we compare the net private foreign asset positions of these three countries. Note that for France, only isolated benchmark years exist, whereas for the UK and the Netherlands, a continuous series is available. The Netherlands and

Figure 1.9: Colonial Asset Distribution, 1921

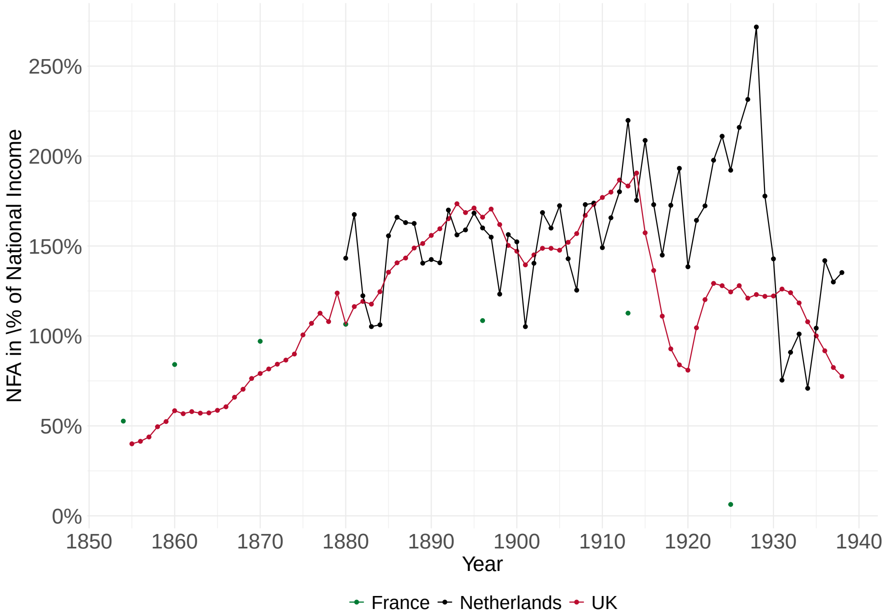


Notes: Figure shows the allocation of colonial and foreign stocks and bonds over the wealth distribution in the 1921 Death Duties.

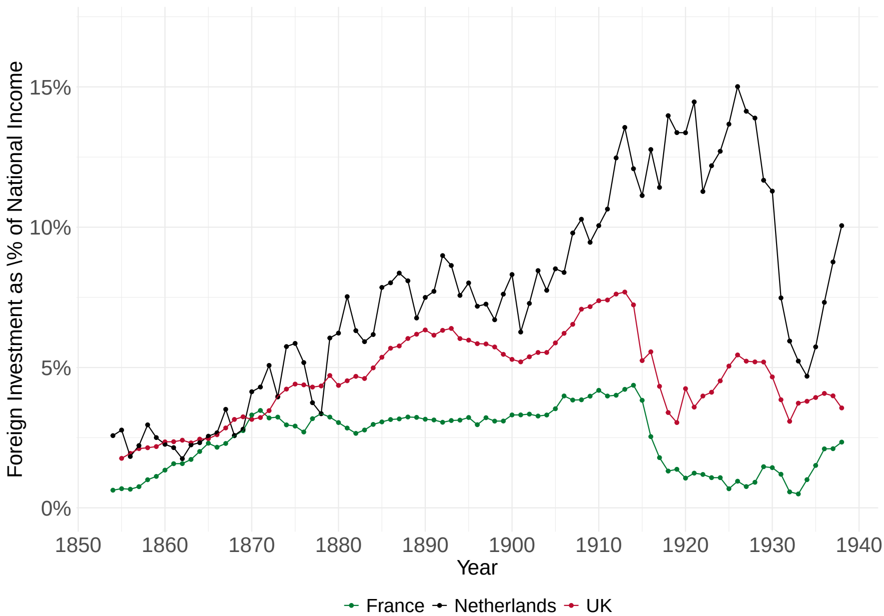
the United Kingdom had comparable levels of foreign wealth until 1910, but started to diverge dramatically afterward, with the Netherlands increasing its foreign investment substantially, whereas the United Kingdom's foreign investment plummeted. France's foreign investment was much lower throughout this period.

The Dutch series seems much more volatile than the other two series, which are estimated using different methods (see Piketty and Zucman (2014) for details). One might wonder how robust this feature is. Panel (b) shows that this is actually not driven by the estimation method, since the flows of foreign income are also much more volatile for the Netherlands than for the United Kingdom or France. It is an interesting open question why this should be the case; it is possible that the nature of colonial relations in the British and French empires resulted in more price stability, or that the Dutch investments were more tied to the world's financial markets and hence more volatile. Also note that for foreign income, we can go back further than our starting point of 1880 for wealth. As a result, we see clear evidence that foreign investment in the Netherlands increased substantially in the second half of the 19th century, with foreign income growing from 2% of national income to almost 15% in the 1920s.

Figure 1.10: Foreign Wealth and Income as % of National Income, International Comparison



(a) Foreign Wealth



(b) Foreign Income

Notes: Figure shows the evolution of foreign wealth and foreign income for France, the Netherlands, and the United Kingdom. Data for France and the United Kingdom are from Piketty and Zucman (2014).

Hence, we can conclude that foreign investment, and especially colonial investment, was to a large extent responsible for the different trajectory of the Netherlands compared to other countries. Given that the British and French empires were vastly larger in size, it is remarkable that the value of investment relative to national income of the Netherlands is comparable to those two empires and even exceeds their value from the 1910s onward. Although the extent of colonial exploitation in Indonesia is contested (Buelens and Frankema 2016; van der Eng 1998; Gordon 2010), it seems safe to say that it was exceedingly important for Dutch private wealth accumulation, particularly from 1900 onward, as is also confirmed by Figure 1.9.

1.5.2. Explaining the High Level: Income and Capital Income

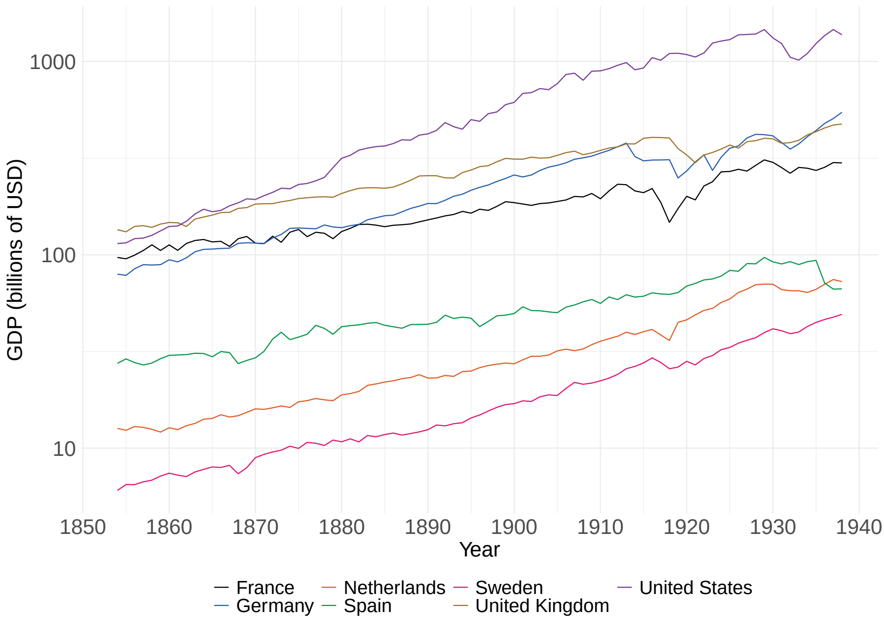
While foreign investment may account for the trends pre-World War II, it does not entirely explain the differences in *levels* between countries. In Figure 1.5, we could see that by 1880, the Dutch wealth-income ratio already was at a high level, even though foreign investment would continue to grow in magnitude afterwards. How can we explain the differences in levels between the Netherlands and other countries?

In this section, we advance two arguments. First, we note that denominator effects matter greatly when studying ratios such as the wealth-income ratio. The lag between the income levels of, say, Britain and France and the Netherlands already indicates the relative low size of the denominator for the Netherlands. Second, we argue that the long history of financial development in the Netherlands resulted in a uniquely financialized country, where therefore the value of financial investment was more significant than elsewhere.

We start with cross-country income differences. This aspect of studying wealth-income ratios has so far been mostly neglected in the literature, but it actually matters significantly for interpreting observed trends like those in Figure 1.5. One reason for using ratios as object of study is that they are independent of price levels, facilitating cross-country comparisons. When we start directly comparing income, we do have to take price differences into account. The most detailed and careful project doing so is the Maddison Project (Bolt and van Zanden 2020), which collects detailed historical GDP estimates for many countries and presents them in a single database, historically using the 1990 International Conference on Prices (ICP) weights to account for price level differences. We use their real GDP figures, which provide a very good proxy for NNI, for all countries covered in Figure 1.5 from 1850 until World War II; the result is presented in Figure 1.11.

Differences in real GDP were substantial across countries for the entire period; in fact, the divergence increased significantly until World War I. The Netherlands

Figure 1.11: Real GDP, 1854–1938



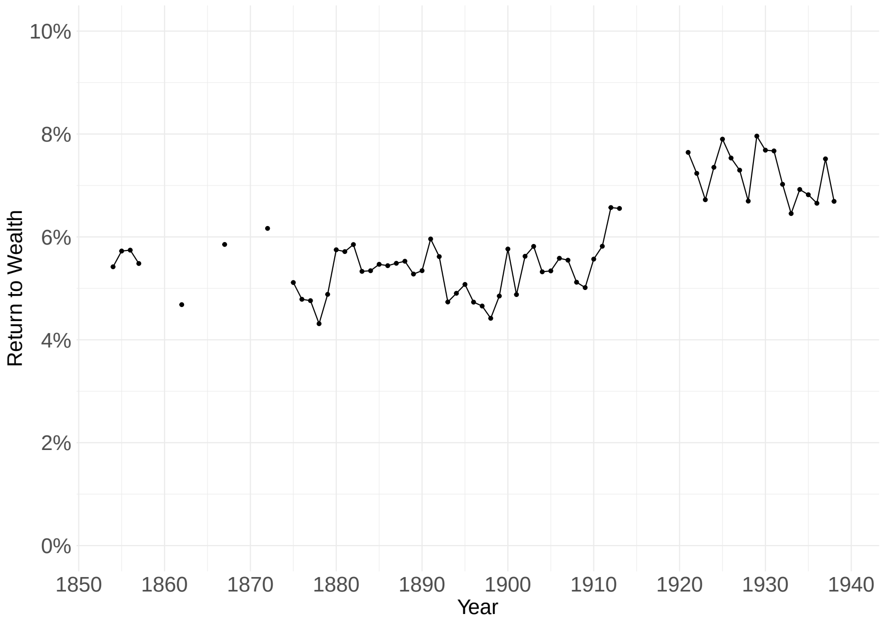
Notes: Figure shows in log scale GDP for the Netherlands, France, Germany, Sweden, Spain, the UK, and the US. Data are from the Maddison Project (Bolt and van Zanden 2020), and all values are in billions of international 1990 USD.

lags behind all countries except Sweden. This observation has several implications for our interpretation of wealth-income ratio differences. First, denominator effects can explain a large part of the difference between the wealth-income ratio *levels*. The United Kingdom's GDP in 1880 was an order of magnitude larger than the Dutch level, hence the conclusion becomes immediate that income differences play a major role. This is also the core of Madsen (2019)'s revisions of the original Piketty and Zucman (2014) series for the United Kingdom: most of his revisions deal with the fact that national income was underestimated in the Piketty-Zucman estimates; clearly, denominator effects matter.

Conversely, the fact that France and the Netherlands had comparable wealth-income ratios for most of the late 19th century, despite the fact that France's GDP was significantly larger than the Netherlands', suggests that wealth was much more important for the Netherlands relative to France. The same holds for the other countries in the series, with the exception of Sweden which has a lower real GDP.

The reader might wonder at this stage how it could be possible that the Netherlands simultaneously had comparatively high wealth and relatively low income; wouldn't that imply an unrealistically low return to wealth? Using our reconstruction of historical national accounts, Figure 1.12 shows that this is not the case, by plotting the return to wealth from 1854 until 1938. The return to wealth is calculated directly as $r_t := \alpha_t Y_t / W_t$, where α_t is the capital share of national income, which we have calculated ourselves for this period (see Appendix 1.D for details). Figure 1.12 shows that the return to wealth was rather stable and sizable, hovering around 6% pre-World War I. After World War I, the return increases to around 7% in the 1920s.

Figure 1.12: Return to Wealth, 1854–1938



Notes: Figure plots the return to household wealth, using the Historical National Accounts. See Appendix 1.D for details.

This indicates that capital played an outsized role in the Dutch economy. This striking pattern can best be explained by path-dependency. Starting in the 15th century, the Netherlands — especially its Western regions – were already highly urbanized, with an economy focusing primarily on trade. Unlike most of its neighbouring countries, the most influential class at the time consisted of merchants, the so-called "Regents", and not landowners. Trade allowed the merchant class to accumulate massive amounts of wealth, which they used to consolidate their political power (Wilterdink 2015; Jonker 1996).

Indeed, much of their wealth was allocated to foreign investment opportunities, facilitated due to Amsterdam's distinctive position as an international trade center, but also in public bonds. Investing in government bonds linked Regent's private benefits to their public pursuits in a most direct way (Wilterdink 2015; 't Hart, Jonker, and van Zanden 1997). This political-economic setup allowed for fiscal and financial structures, which further entrenched the beneficiary position of merchants: (i) the customs rates were low; furthermore, (ii) taxes were mostly levied on consumption and property rather than trade and direct investments ('t Hart, Jonker, and van Zanden 1997).

Following the end of The Dutch Golden Age in the 18th century, the Netherlands increasingly started to play second fiddle compared to France and especially Great Britain, which had become the prime mercantile nation, supported by an ever-expanding colonial empire. By then, the initial advantages of the Dutch Republic, most notably its urbanization and its well developed financial and commercial systems, no longer outweighed its inherent disadvantages such as its limited area, population, and relatively fragmented political system. Throughout the 18th century, national income tended to stagnate while national private wealth expanded steadily. Despite of this, the city of Amsterdam remained one of the leading financial markets of the time. Dutch capital penetrated far into European countries and their colonies ('t Hart, Jonker, and van Zanden 1997).

Due to the large reserve of accumulated funds, the Dutch elite managed to safeguard —or even strengthen – their privileges throughout the 19th century. Relying on their strong capital basis, many of them successfully responded to the decline of the Netherlands as a leading mercantile power by shifting their business from merchants to bankers, commissioners in bills of exchange, insurers, and stockbrokers. Per these developments, a further shift was noted from mercantile finance to investments in foreign equity and bonds (Jonker 1996). The climate for a thriving capital market remained highly profitable, and private wealth continued to grow faster than national income (Wilterdink 2015).

By the 1860s, the Netherlands started to industrialize, several decades after its southern neighbour Belgium and most other surrounding countries. The reasons for this comparatively laggard development are still widely debated. Even on the exact timing of the Dutch industrialization, there appears to be no real consensus (van Zanden and van Riel 2000; Philips 2020).

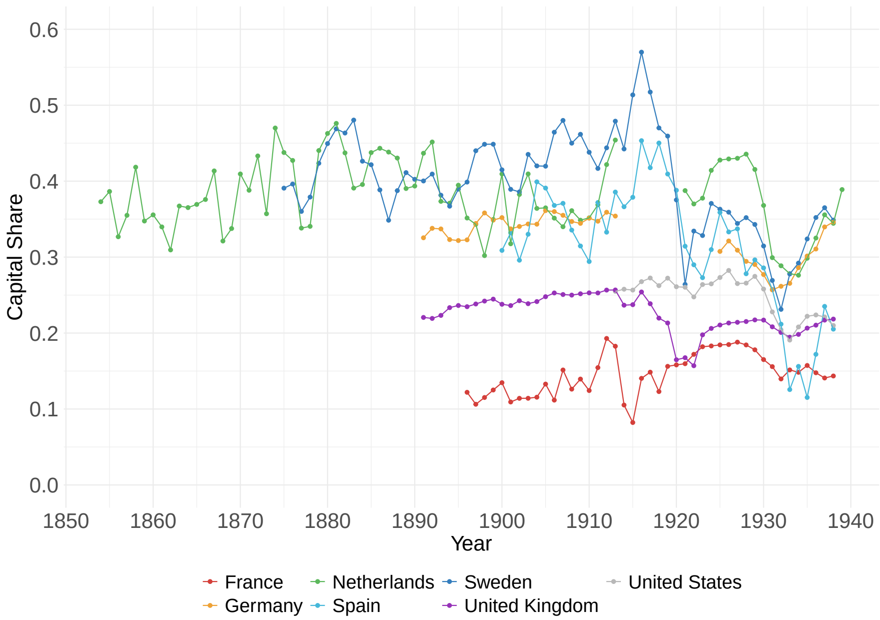
Sketching the outlines of this debate takes us far beyond the purpose of this article. However, what is of importance here is that the take-off of Dutch industrialization was fully underway by the late-19th and early-20th century. Simply looking at the nearly exponential rate in which the number of joint-stock companies (i.e., in 1850 there were only 137 joint-stock companies, by 1894 this had risen to over 1,700) rose throughout this period is exemplary for this trend (de Vicq 2021).

The industrial expansion of the Dutch economy also allowed for new manners in which to accumulate riches. Consequently, a new class of industrialists joined the ranks of the Dutch elite, while the existing elite managed to preserve their wealth by expanding their mostly foreign investment portfolio with new domestic investment opportunities (Wilterdink 2015). The growing number of wealthy industrialists had a clear impact on the wealth-income ratio, which continued to grow in the decades following the 1860s; reaching an apex of more than 900 percent of national income by the dawn of the 20th century.

This importance of capital to the national economy is confirmed in Figure 1.13, which plots the capital share of national income for all seven countries in consideration. We have estimated the capital share for the Netherlands ourselves; data for all other series are from Bengtsson and Waldenström (2018).

We observe that the Dutch capital share was among the highest on the record from the 1850s until the 1930s, only being outpaced by Sweden in some years. The British and French capital shares were significantly lower. This corroborates our focus on the denominator as a major source of cross-country differences: These large countries both had a larger weight of labor in the economy than the Netherlands, whose labor share would remain low until the post-World War II era. The Dutch economy was dominated by capital to a far greater extent than most other countries, with the possible exception of Sweden. Sweden, however, had much smaller wealth-income ratios than any other country on record, suggesting that it was in effect a developing country for most of this period (Waldenström 2017).

Figure 1.13: Capital Share of National Income, 1854–1938



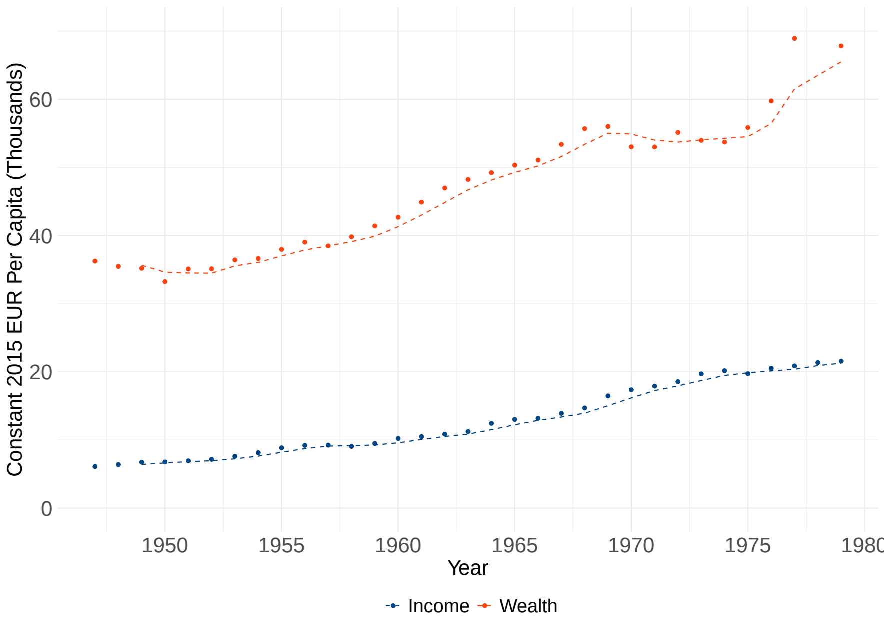
Notes: Figure shows the capital share of national income, net of depreciation. Data for countries other than the Netherlands are from Bengtsson and Waldenström (2018). Details on the Dutch capital share can be found in Appendix 1.D.

1.6. Post-World War II: Pension Savings, Corporate Savings, Housing Capital Gains

In this Section, we explore the decline and subsequent increase in the wealth-income ratio after 1947. Specifically, we will focus on the role played by savings and capital gains in the observed real growth rates, and we will argue that the dominance of capital gains in the Netherlands reflects the institutional design of its pension system and its housing market.

After World War II, a large expansion of domestic investments contributed to the continuous growth of the Dutch economy, characterized by an annual increase of nearly 3 to 4 percent in national income in the 1950s and 1960s. This is seen in Figure 1.14, which shows real wealth and income growth from 1947 until 1980. Comparing the increase of national income across this period to the nominal growth in wealth of the upper classes, it becomes apparent that the growth of wealth has lagged considerably behind. Further, contributing to this trend was the noteworthy shift from private wealth to so-called semiprivate wealth detailed earlier (Wilterdink 2015). Because of all these reinforcing factors, the Netherlands experienced its equivalent of The Great Compression (Goldin and Margo 1992). Between the 1950s and the 1970s, the wealth-income ratio dropped to unprecedented levels, with an all-time low of approximately 300 percent in the early 1970s.

Figure 1.14: Real Wealth and Income Growth, 1947–1980



Notes: Figure shows the evolution of household wealth and national income per capita, from 1947–1980. Both wealth and income are expressed in constant 2015 EUR.

Things started to change by the 1980s. This is clearly seen in Figure 1.15, where we observe that real wealth skyrockets, whereas real income grows at much slower rates and even seems to stagnate since the beginning of the 21st century. Section 1.4.2 already alluded to the important role played by pension assets in the rise of household wealth from the 1980s. Unlike other nations, housing wealth plays a more minor role in these developments as much of the increase in housing assets is matched by a similar increase in mortgages. The latter was supported by a liberal system of mortgage finance starting from the 1960s and the limited incentives for households to build up wealth given their already substantial pension claims (van der Valk 2019).

Figure 1.15: Real Wealth and Income Growth, 1980–2019

Notes: Figure shows the evolution of household wealth and national income per capita, from 1980–2019. Both wealth and income are expressed in constant 2015 EUR.

1.6.1. Savings vs. Capital Gains

How can we make sense of these large upswings and downturns in the wealth-income ratio? A useful accounting decomposition is to split real wealth growth into a savings component and a capital gains component, using Equation 1.5. We observe real wealth growth $g_t := \Delta W_t / W_t - 1$, as well as savings rates out of wealth s_t; hence, we can residually calculate real capital gains or losses as $q_t = (1 + g_t)/(1 + s_t) - 1$. Starting with Piketty and Zucman (2014), the subsequent literature has used this identity to identify an important role for capital gains in driving the recent boom in wealth-income ratios (e.g., Artola Blanco, Bauluz, and Martínez-Toledano 2020; Baselgia and Martínez 2023). When we try to do the same for the Netherlands, however, we first run into a conceptual difficulty: Which savings rate is the correct one to use?

Starting from national accounting identities, *national* saving is the sum of government and private saving; private saving, moreover, can be decomposed into savings by households, financial institutions, and non-financial corporations. Most studies which have used accounting equation 1.5 so far have focused on *national* wealth. Then, the choice is clear: use the national savings rate. However, when we measure household wealth, it is not always clear which savings rate to use. Financial institution savings are mostly savings by pension funds; these savings conceptually are to the benefit of households. Likewise, corporate savings – retained earnings – can also be accrued to households to the extent that they own shares in the corporate sector. Obviously, the more expansive the savings concept, the lower the residually estimated capital gains will be, with the potential to radically alter conclusions about wealth accumulation. This point is acknowledged and discussed thoughtfully by Piketty and Zucman (2014) and Baselgia and Martínez (2023), but is especially important for the Dutch context, given the massive role that pension savings and retained earnings play. This point is illustrated in Table 1.2, which shows the decomposition of real wealth growth into its savings and capital gains components, using increasingly expansive definitions of savings for the period 1947–2019.

Table 1.2: Savings and Capital Gains by Savings Concept, 1947–2019

		H		**H+F**		**H+F+C**	
Period	g	s	q	s	q	s	q
1947–2019	**3.4%**	**1.6%**	**1.8%**	**2.3%**	**1.1%**	**3.8%**	**-0.4%**
1947–1959	2.7%	0.6%	2.1%	1.4%	1.3%	2.5%	0.2%
1960–1969	4.9%	1.7%	3.2%	2.7%	2.1%	4.4%	0.4%
1970–1979	3.1%	3.2%	0.0%	3.5%	-0.4%	5.6%	-2.4%
1980–1989	2.7%	3.1%	-0.4%	3.8%	-1.1%	5.3%	-2.5%
1990–1999	6.4%	2.3%	4.0%	3.2%	3.0%	4.9%	1.4%
2000–2009	2.8%	0.6%	2.2%	1.1%	1.7%	3.3%	-0.4%
2010–2019	3.8%	1.2%	2.6%	1.9%	1.9%	3.5%	0.3%

Notes: Table shows geometric average real wealth growth rates (g) for the respective periods in each row. This average growth rate is then decomposed into a savings (s) and capital gains component (q) using Equation 1.5. The savings concept used in each column is indicated at the top of the table, respectively: Households, Households + Financial Institutions, Households + Financial Institutions + Corporations.

Looking at the first row of Table 1.2, we observe that household wealth growth was sizable and positive, averaging 3.4% per year from 1947 until 2019. It is also clear that a sizable portion of this wealth growth is attributable to savings. This is true even when we use the narrowest savings concept, household savings; even with that conceptualization, savings accounted for $1.6/3.4 \approx 50\%$ of all household wealth growth. If we move to more expansive savings concepts, the role of capital gains mechanically declines even further. If we add savings by pension funds and other financial institutions, capital gains only contribute 1.1 percentage point on average to real wealth growth. If we add retained earnings, capital gains turn *negative*, indicating that total private savings account for more than 100% of real wealth growth in the Netherlands. Again, we do not take a stance on which savings concept to use. For one thing, if we want to include corporate and financial saving, we presumably also need to estimate the balance sheets of the corporate and financial sector. However, we stress that these issues are not clear-cut and have first-order impacts on the interpretations of wealth accumulation.

An even clearer picture emerges once we move to the rows which show decadal averages. Here, it is clear that household savings really took off from the 1970s, with capital gains being zero or negative even in the narrowest definition from 1970 until 1990. The 1990s witnessed a large boom in household wealth, which is mainly driven by capital gains. Again, using more expansive savings concepts decreases the role of capital gains even in these years. We can conclude that retained earnings and financial institution savings matter significantly for the development of private wealth.

From 1995, we can get a grasp on the sources underlying these savings and capital gains components. From that year on, Statistics Netherlands publishes modern household balance sheets that include all wealth components, as well as volume and price mutations happening during each year to these components. Hence, we can decompose the household wealth growth rate into savings and capital gains components per major wealth component. We do so in Table 1.3, where we focus on four major categories: financial wealth (net of financial debt), semiprivate wealth (pensions and life insurance), housing wealth (net of mortgage debt), and non-financial wealth. This exercise bears some similarity to the work of Bauluz, Novokmet, and Schularick (2022) and Baselgia and Martínez (2023), and others. Important to note here is that we do not estimate synthetic savings and capital gains over the wealth distribution, as is done by Saez and Zucman (2016) and the subsequent literature. In contrast, we are only concerned with decomposing aggregate household wealth growth into the relative contributions of each component.[17]

[17] While the distributional decomposition is interesting, it is not feasible with the data at our disposal, since the distribution of pension claims is not known and would require stringent assump-

Table 1.3: Decomposition of Household Wealth Growth by Wealth Component, 1995–2019

	1995–2019	1995–1999	2000–2004	2005–2009	2010–2014	2015–2019
Real Wealth Growth	**3.8%**	**8.4%**	**4.0%**	**2.7%**	**1.2%**	**7.1%**
- Financial	0.5%	2.5%	-0.3%	0.2%	0.0%	0.6%
- Semiprivate	2.0%	2.3%	1.3%	1.6%	3.9%	3.0%
- Housing	1.2%	3.0%	3.0%	0.5%	-2.6%	3.3%
- Non-Financial	0.2%	0.5%	0.1%	0.4%	0.0%	0.2%
Due to Savings	**1.3%**	**2.3%**	**1.4%**	**1.2%**	**1.4%**	**1.6%**
- Financial	0.6%	1.7%	1.0%	0.6%	0.1%	0.4%
- Semiprivate	1.1%	2.0%	1.5%	1.1%	1.0%	0.8%
- Housing	-0.5%	-1.5%	-1.2%	-0.7%	0.2%	0.3%
- Non-Financial	0.1%	0.1%	0.2%	0.2%	0.1%	0.1%
Due to Capital Gains	**2.5%**	**5.9%**	**2.6%**	**1.5%**	**-0.2%**	**5.4%**
- Financial	-0.1%	0.8%	-1.3%	-0.4%	-0.1%	0.2%
- Semiprivate	0.9%	0.3%	-0.1%	0.5%	2.8%	2.2%
- Housing	1.7%	4.6%	4.3%	1.1%	-2.7%	3.0%
- Non-Financial	0.1%	0.4%	-0.1%	0.2%	-0.1%	0.1%

Notes: Table shows the decomposition of average real wealth growth per five-year period from 1995 until 2019, split into savings and capital gains components per major wealth component. All wealth components are measured net of debt.

By focusing on five-year intervals, we can shed further light on dynamics happening in the last 25 years. We notice that the positive real wealth growth since 1995 has been very heterogeneous over time, with the years following the Great Financial Crisis seeing little more than a percentage point of growth per year. In contrast, the late 1990s were characterized by growth rates of over 8% annually. By decomposing these growth trends across savings and capital gains and across wealth components, we observe various patterns. First, the major contributor to wealth growth since 1995 has been semiprivate wealth – pensions and life insurance. This component accounted for slightly more than half of all real wealth growth, with housing playing a secondary role. Other financial and non-financial assets matter far less for wealth growth in general, with the exception of the late 1990s, when financial assets were the most important contributor to wealth growth leading up to the Dot-Com bubble.

Semiprivate wealth added to wealth growth both via savings and via capital gains, with those channels being roughly equal in magnitude across the whole

tions to impute. Hence, we focus on the aggregate decomposition and leave distributional decompositions for future work.

period but showing significant fluctuations in between. By contrast, housing actually dominates semiprivate wealth in most five-year intervals; however, housing experienced a stark slowdown in growth from 2005 onward, turning negative in the 2010-2014 interval. This period, associated with the collapse of housing prices following the Great Recession, clearly repressed real wealth growth of households. Even in the other periods, housing's contribution to wealth growth almost exclusively originates from capital gains, with housing savings being negative or very weakly positive throughout. This is due to the enormous accumulation of mortgage debt by households; the value of mortgages, in excess of 100% of national income in 2019, was stimulated by various government policies from the 1980s onward. One of those policies was the internationally unique institution of the interest-only mortgage, which did not require any amortization, and was only regulated following the Great Financial Crisis (Bernstein and Koudijs 2024). Another contributing feature of the Dutch institutional design was that households could until recently borrow in excess of the value of their home, up to 130% in the 1990s (van der Valk 2019). We can conclude from this exercise that pension wealth is the dominant factor explaining real wealth growth of Dutch households. Note that if we were to use more expansive savings concepts, this conclusion would be strengthened further.

1.7. Conclusion

Following the seminal work by Piketty and Zucman (2014), this chapter analysed the historical development of aggregate wealth-income ratios for the Netherlands from 1854 until 2019; a country that was notably missing in their analysis. In addition, we decompose total private wealth into various components, tracking the relative value of financial and non-financial asset categories for 140 years (from 1880 until 2019). Furthermore, we track top wealth shares from 1894 onward. Finally, we discuss various interpretations of these trends, and contrast them to the available international evidence.

We find that while the private wealth-income ratio in the Netherlands followed the familiar U-shaped pattern observed in earlier studies, the highs and lows were more outspoken compared to most other countries for which such long-term evidence is available. In comparison with other industrialised countries, the Netherlands, experienced periods with some of the highest as well as one of the lowest private wealth-income ratio: from a ratio in excess of 900% at the turn of the 20th century, to a ratio as low as 300% in the 1970s. Likewise, the top 1% share of household wealth peaked at about 55% in the early 20th century, which was followed by a precipitous decline to about 20% in the 1970s and a subsequent increase to about 30% by the early 2010s.

The main empirical contribution of this chapter is to expand the existing evidence on long-term wealth dynamics for large, at times closed, economies with evidence on a small and very open economy throughout. A novelty of this chapter is that it provides empirical evidence that the very high wealth-income ratio was at least in large part due to a significant proportion of private wealth invested in colonial and non-colonial foreign securities, which was predominantly held at the top of the wealth distribution. This finding further highlights the significance of colonial empires in explaining (global) wealth dynamics and thus makes an important contribution to this ongoing debate (Chancel and Piketty 2021).

Methodologically, we exploit the rich availability of data sources for the Netherlands, and simultaneously use (i) historical national accounts; (ii) estate multiplier methods; (iii) lognormal extrapolation from wealth tax data. We find that historical balance sheets and the estate multiplier produce remarkably similar results for the 19th century and early 20th century. Hence, the estate multiplier, if one is not too far from the benchmark year, is likely to be a reasonable method to employ in cases where the necessary data to reconstruct historical balance sheets are unavailable. The wealth tax extrapolation method performs less satisfactorily, and is more sensitive to the quality of the underlying wealth tax data.

Our findings can inform policymakers about the level of wealth concentration, by placing recent figures in a historical perspective. Moreover, our decomposition of aggregate wealth highlights the important institutional determinants of household portfolio choice. Policy choices since the 1980s aimed at stimulating homeownership, like the mortgage interest deduction, are likely to have contributed to the rise in mortgage debt and housing prices. Likewise, the tax-exempt treatment of pension wealth will have contributed to its increase in relative importance in household portfolios. These results illustrate the impact of careful policy design to stimulate household private savings while also not encouraging overreliance on debt. Policy choices on the composition and distribution of household wealth also matter for macroeconomic stability, the level of the interest rate, and other key macroeconomic variables, and therefore the findings presented here, are relevant regardless of preferences for wealth redistribution.

1.A. Historical National Accounts

1.A.1. Concepts and Preliminaries

This section details the construction of our benchmark series on aggregate household wealth, using the historical national accounts method. As detailed in the main text, we work with standard accounting definitions of household wealth, W_t. In the following subsections, we will focus on the following decompositions of household wealth. Household wealth is the sum of financial and non-financial assets minus liabilities:

$$W_t := A_t^f + A_t^{nf} - D_t.$$

Financial assets can be divided into deposits, securities, and semiprivate wealth (pension + life insurance):

$$A_t^f := A_t^d + A_t^s + A_t^{sp}.$$

Non-financial assets can be divided into land, housing, and the fixed capital stock (including small remainder items):

$$A_t^{nf} := A_t^l + A_t^h + A_t^k.$$

Liabilities, finally, can be defined as the sum of housing debt (mortgages) and financial debt:

$$D_t := D_t^h + D_t^f.$$

We start with non-financial assets, before moving to financial assets, and we conclude with liabilities. At the end of this section, we also discuss the comparisons of the CPB balance sheets since 1970 to alternative sources. Table 1.4 provides a high-level summary of all steps per component per period.

1.A.2. Non-Financial Assets

1.A.2.1. Housing

The value of housing is the sum of the value of dwellings and the land underlying dwellings. For the construction value of dwellings, we have perpetual inventory method (PIM) estimates from Smits, Horlings, and van Zanden (2000) for 1807–1913 and from Groote, Albers, and De Jong (1996) for 1900–1994. Although the assumptions differ a bit between these two sources, they are broadly comparable and yield almost identical estimates.

Table 1.4: Historical National Accounts, Sources and Methods per Component

Category	Sub-component	Years	Sources & Data Steps
Non-Financial Assets	Housing	1880 – 1947	Total volume of housing + price index, benchmarked to 1913 total value
		1947 – 1969	Housing stock (% owner-occupied) + price index, benchmarked to 2011 total value
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
	Agricultural Land	1880 – 1994	Agricultural area + price index for farmland
		1995 – 2019	National Accounts
	Livestock	1880 – 1938	Annual statistics (+ interpolations) for cattle + assumed value for horses from wealth tax
		1947 – 1994	Total number of livestock + agricultural land price index, benchmarked to 1958 total value
		1995 – 2019	National Accounts
	Other Fixed Capital Stock	1880 – 1968	Existing total capital stock estimats, adj. to capture part attributable to households
		1969 – 2019	National Accounts
	Inventories	1880 – 1968	Fixed percentage of total capital stock
		1969 – 2019	National Accounts
Financial Assets	Deposits	1880 – 1938	Total deposits at banks, benchmarked by 1938 balance sheet for split household/corporate
		1947 – 1969	Statistics Netherlands series, benchmarked by 1938 balance sheet
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
	Currency	1880 – 2019	Same sources and procedure per subperiod as for deposits
	Pension and life insurance	1880 – 1969	Central bank statistics
		1970 – 1994	CPB estimates (pension); Central bank estimates, adjusted to match trends
		1995 – 2019	National Accounts
	Bonds	1880 – 1938	Value of privately held treasury bonds (benchmarked to 1938 balance sheet)
		1947 – 1969	Total value of treasury bonds
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
	Listed Stocks	1880 – 1938	Total value of stock market index, + capit. value of net foreign income
		1947 – 1969	As above
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
	Nonlisted stocks & Other	1880 – 1938	Residual estimate, benchmarked to 1938 balance sheet
		1947 – 1969	As above
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
Liabilities	Mortgages	1880 – 1969	Annual statistics, adj. for part of mortgages attributable to households
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts
	Other Liabilities	1880 – 1969	Annual statistics
		1970 – 1994	CPB estimates
		1995 – 2019	National Accounts

Notes: See Appendix 1.A for details; the order of the subsections follows the order in the table.

Unfortunately, no estimates exist of the value of land underlying dwellings. Hence, we do not use the PIM estimates until the official National Accounts start in 1995. Before 1995, we construct a series on the market value of housing. Before 1947, we use a series on the total volume of housing available, which was published in the Annual Statistics for the Netherlands (Jaarcijfers voor Nederland) in various editions. Moreover, in Centraal Bureau voor de Statistiek (1947) there is an estimate of the total value of housing wealth in 1913 of 2.2 billion guilders; combining these sources gives us an estimate of housing wealth per square meter. We index this estimate to the housing price index of Korevaar, Francke, and Eichholtz (2021).

This series produces estimates of housing which correspond exactly to the value of housing recorded on the National Accounts balance sheet in 1938, namely 5 billion guilders. The value of housing for 1947, which also recorded in the same balance sheet, seems implausibly high, by contrast, at 16 billion guilders, whereas our estimates give it approximately half that value. Although World War 2 plausibly caused many changes in the structure of the economy, it seems implausible that the value of the housing stock would triple in nominal terms, especially when taking into account wartime destructions. Hence, while we think that the 1938 figure is plausible and a nice match with our series, we do not use the 1947 figure.

From 1947 until 1969, we construct housing wealth in the following way. We define housing assets in year t, HW_t, as the product of the average housing price P_t, the housing stock HS_t and the share of owner-occupied housing DS_t. The average housing price is equal to

$$P_t = \mathrm{PI}_t \cdot \left(\frac{\mathrm{HW}_{2011}}{\mathrm{HS}_{2011}} \cdot \frac{1}{\mathrm{DS}_{2011}} \right), \tag{1.6}$$

where PI_t is a price index equalling 1 in 2011 and the expression in brackets is the average house price in 2011. Data on the housing stock comes from Statistics Netherlands, the share of owner occupied housing from Haffner, Hoekstra, Oxley, and van der Heijden (2009), and the housing price index from Korevaar, Francke, and Eichholtz (2021)[18]. The market value of housing assets in 2011 is taken from the wealth distribution statistics of Statistics Netherlands.

From 1970 until 1994, we use the value of housing recorded in the CPB balance sheets. Our own estimates prior to 1970 match very closely with the CPB balance sheets.

[18]This housing price index is based on the entire Amsterdam housing market; hence, while not perfectly representative of the aggregate housing index, it is more representative than the narrow *Herengracht-index* that forms the basis for the housing index in Jordà et al. (2019).

From 1995, we use the value of dwellings and land underlying dwellings recorded in the National Accounts. The difference between the market-value series before 1995 and the PIM series after 1995 results in a slight jump of around 20% of national income.

1.A.2.2. Agricultural Land

For non-residential land (which is predominantly farmland), we obtain a total value by multiplying estimated total area with estimated average prices. We use volume data on the area of agricultural land from van der Bie (2001), which is based on the work of Knibbe (1993). For land prices, we rely on work by Luijt and Voskuilen (2009). Their data series gives estimations of the price of farms and farmland from 1952. For the period before 1952, it only provides estimations of the total value of farms. We calculate the ratio between farms and farmland throughout the 1950s and take the average for this period. This ratio (of 1.4) is then applied to estimate the value of farmland for the period 1880 and 1938. We interpolate some of the missing years. The resulting series for the value of agricultural land is very comparable to that of Knibbe (2014), who basically uses the same data and methods as we do, but does not adjust the pre-1952 series. As a result, his series show higher values of land pre-1952, but these values also capture implicitly the value of the farm buildings, capital stock and other attributes that would be reflected in the farm price. Since these aspects are better attributed to other wealth components, we feel our adjustment is closer to the likely value of land.

After 1995, we use the value recorded in the Non-Financial Accounts by Statistics Netherlands.

1.A.2.3. Livestock

The number and total value of cattle is readily available in the Annual Statistics for the Netherlands. We interpolated some of the missing years between 1880–1897; between 1908–1913; between 1922–1925; and between 1925–1930. Having the total value of cattle and the number of cattle at our disposal for most year, we were able to calculate the value of a single cattle. We took the average of this individual price (172 guilders) and assumed a horse would be approximately 5 times more expensive (862 guilders). Since we were able to retrieve the number of horses held by individuals based on their tax record, we were thus able to estimate the total value of all horses.

After World War 2, we rely on the total number of cattle presented in van der Bie (2001). The total value of livestock is put at 3 billion guilders in 1958; hence, we obtain an estimate of the average value of cattle for that year. For the remaining years, we assume that this value follows the development of agricultural land prices, so we index the average value in 1958 to our agricultural land price series. After 1995, we use the National Accounts, which do not explicitly include a post for livestock; hence, it appears in our series in the residual capital stock (i.e., the part of the non-financial accounts that isn't one of the main items mentioned in the rest of this section).

1.A.2.4. Other Fixed Capital Stock

This remainder item mainly includes nonresidential dwellings, roads and public buildings, machinery and equipment, as well as smaller components. We start by estimating the total capital stock. From 1854 until 1900, we use the capital stock estimates from Smits, Horlings, and van Zanden (2000). From 1900 until 1969, we use the estimates by Groote, Albers, and De Jong (1996). Inspection of the two series in the overlapping years (1900–1913) reveals that they track each other very closely, despite differing in the underlying assumptions made on the lifecycle of the components.

From 1969, we can use official National Accounts balance sheets. It is also in this year that we therefore can allocate the total capital stock into the different sectors. This exercise reveals that households consistently owned about 40% of the private capital stock, with the remaining 60% being allocated to the corporate sector. Moreover, the series from Groote, Albers, and De Jong (1996) – which runs until 1994 – is almost identical to the National Accounts series on the private capital stock. Hence, we assume that in all years prior to 1969, the historical estimates also correspond to the private capital stock, and we allocate 40% of these estimates to the household sector.

For the years 1938, 1947, and 1948, we also have National Accounts estimates of the total capital stock. However, these are not easy to interpret. For instance, the total capital stock (including housing and land) is put at 20 billion guilders in 1938, but the total 'freely disposable' capital stock attributable to households is put at 11.5 billion, with 2.9 billion belonging to the government and the remainder being 'covered' by mortgages, shares and other claims. Since we would nowadays place mortgages on the liabilities side – which isn't done in the balance sheet – we increase the private capital stock that is 'disposable' for households by the amount of mortgages. The adjusted National Accounts totals are almost identical to the series by Groote, Albers, and De Jong (1996) (after inclusion of land and livestock). The difference between our capital stock and the adjusted National Accounts total is assumed to be the part of the capital stock that belongs on the corporate balance sheet.

1.A.2.5. Inventories

We have an estimate in the 1938 National Accounts of 2.6 billion guilders for total inventories. This number seems implausibly high, implying that almost 10% of private wealth was in inventories; it is more likely that this figure also includes part of the capital stock. No direct sources exist before this year; only indirect estimates of changes in inventories by Smits, Horlings, and van Zanden (2000) and den Bakker (2019), who both estimate this change as the residual between national savings and gross fixed capital formation. This imprecision leads to implausibly high values of inventories if taken at face value (values well over 200% of national income before 1900), meaning these numbers cannot be used.

Instead, we opt for an indirect approach. The first balance sheet of non-financial assets, in 1969, puts the value of inventories at around 8% of the capital stock. We use this percentage and apply it to our estimates of the capital stock before 1969. This adjustment works fairly well, with a smooth series of inventories from 1880 until 1994. After 1995, we use the National Accounts. In several respects, this is a conservative adjustment, since the share of the capital stock attributable to inventories monotonically declines after 1969, from 8% to 2.2% in 2019. Had we extrapolated this trend backwards, we would have arrived at implausibly high values for inventories in the past. We prefer to make more conservative extrapolations, and stick with 8% before 1969.

1.A.3. Financial Assets

Following the System of National Accounts, financial assets include deposits and currency, shares and mutual funds, bonds, individual pension, and insurance savings.

1.A.3.1. Deposits

We begin by reconstructing the total amount of deposits from 1880 until 1970. The principal sources material used to estimate the value of these asset classes are (i.) the Statistical Publication by the Dutch Central Bank, which reported on the balance sheet information of commercial banking institutions as well as saving banks and cooperatives banks from 1900 onwards (DNB 1987, 2000); and (ii.) the previously mentioned Annual Statistics for the Netherlands. For Saving Banks, the data on the total value of deposits held by these institutions was readily available in the Annual Statistics for the Netherland from 1885 onward. Thus, leaving a gap in the period between 1880–1885. For these years we however knew the amount of saving banks there were active; so we looked at the average deposits held by saving banks in

1880 and 1885 and interpolated this data based on the number of banks between 1881 and 1884. For Farmers' Cooperatives, we relied on Westrate (1948: 374-376). This memorial book, published to celebrate the 50-years jubilee of Cooperative Banks reported the value of deposits held by this type of banks from 1899 onward. For Postal Savings Banks, we relied on the Annual Statistics for the Netherlands. This data was readily available from 1885 onward. The data for Commercial Banks, was retrieved from the Statistical Publication by the Dutch Central Bank. This data was however only estimated for the entire commercial banking sector for the years 1903, 1908, 1913, 1918, 1923, 1928, 1933, and 1938. We therefore collected the deposits from the 3 largest banks from 1880 to 1900 and interpolated this data to calculate the deposits held by all commercial banks. We did the same to fill in the gaps between 1900 and 1908, but in this case, we relied on the data for the 5 largest banks as published in the Statistical Publication by the Dutch Central Bank. We then cross-referenced this estimation of all deposits held by commercial banks by comparing it to a newly collected dataset of approximately 140 individual commercial banks (De Vicq and Peeters 2022). This results in a series for aggregate deposits from 1880 until 1938. In 1938, we cross-check the amount in deposits with the official National Accounts balance sheets. The numbers align reasonably well; our stock of deposits is 3.3 billion guilders, whereas the official balance sheet gives a sum total of 4.4 billion. However, only 1.9 billion of these deposits should be ascribed to the household sector; the rest shows up on the balance sheet of corporations, the government, the insurance sector, and the foreign sector. Hence, for 1938, we take the official number for households as given, and for all years prior to 1938, we divide our series by the ratio of the series in 1938 to the official number ($3.3/1.9 \approx 1.7$).

On the website of Statistics Netherlands[19], we also find a series for total deposits, starting in 1900 and with continuous values from 1935. Inspection of this series yields that it is a bit higher than the official National Accounts total, 2.9 billion instead of 1.9. The same holds for the values of the balance sheets for 1947 and 1948. Hence, we downweight this series by the ratio of the series in 1938 ($2.9/1.9 \approx 1.5$).

After 1970, we use the deposits total noted in the CPB balance sheet. This amount is quite a bit higher than the adjusted deposits series, with a jump of about 17 billion guilders. However, if we adjust the previous series using the CPB balance sheet (or take it at face value), we lose consistency with the balance sheets in 1938, 1947 and 1948. Hence, we decided to preserve consistency with the earlier official balance sheets, and accept the (small) trend break that occurs in 1970. The scale of the discontinuity is relatively minor, around 20% of national income.

After 1995, we use the National Accounts.

[19]Link: https://opendata.cbs.nl/#/CBS/nl/dataset/37758/table?dl=6E2C5.

1.A.3.2. Currency

The value of coins and printed money was consistently published in the Annual Statistics for the Netherlands. We only had to interpolate some missing data points in the years between 1881 and 1884. We cross-referenced our figures with similar estimations made by Kymmel (1992).

As with deposits, we use the value recorded in the 1938 balance sheet to adjust this series, to obtain the amount of currency held by households. We also use the same sources and procedures for currencies after 1947 as with deposits, outlined above. However, the trend break is much less severe in 1970, only 1.7 billion guilders. After 1970, we use the CPB balance sheet, and after 1995 the National Accounts.

1.A.3.3. Pension and insurance funds

The value of funded occupational pension entitlements and private insurance savings are based on the technical reserves as recorded in the Statistical publication by the Dutch Central Bank. We simply transcribed the results from this publication, as this data was readily available.

1.A.3.4. Securities

Domestic Government Bonds The value held by privately owned treasury bonds was listed in the Annual Statistics for the Netherlands from 1891 onward; similar series appear before 1891 as well. There seems to be no trend-break in 1891, hence we use the full series from 1880 until 1938. We confirm that the value of privately held bonds corresponds almost identically to the value listed in the 1938 official balance sheets: 3.2 billion guilders.

After World War 2, we no longer possess direct information on the value of privately-held bonds, hence we take the entire value of Dutch bonds until 1970. From 1970, we use the bond holdings listed in the CPB balance sheets. Remarkably, despite using the full value of bonds from 1947 onward, we find no trend break when we switch to the CPB balance sheets in 1970, with only a slight jump from 29 to 38 billion guilders. Of course, the bondholdings recorded in the CPB balance sheets would plausibly also include corporate bonds and bonds from other countries; yet for our purposes this does not matter, since we are only interested in reconstructing total securities holdings by households, which thus seem to be captured quite accurately, especially when combined with the other components, which are detailed in the next sections. From 1995, we use the official National Accounts.

Listed Domestic and Foreign Stocks We take as benchmark the nominal value of the Amsterdam Stock Exchange, for which an index can be found in van der Bie (2001). This index gives us the value of all Dutch firms listed on the exchange, taking into account dividend issuance and other factors which mechanically depress the share price. The index is available from 1890 until the present; since the index appears to be remarkably stable (around an index value of 21, where the 1983 stock market capitalization is set to 100) for the 1890s and 1900s, we assume it is also stable in the 1880s and extrapolate the 1890s average backwards to get a stock market index from the 1880s until the present. The total stock market value of Dutch firms as a percentage of national income aligns reasonably with the estimates presented by Kuvshinov and Zimmermann (2022), who use a different source for their Dutch pre-World War 2 series, namely Moore (2012) for the period 1900–1913; they do use the official index by Statistics Netherlands for the post-war period and seem unaware of the pre-war index estimates presented in van der Bie (2001). Comparing Moore's data with the official index reveals that Moore's estimates align reasonably well with the index until about 1910, after which Moore's series vastly outpaces the index. Retaining the index preserves both consistency throughout the period, and is a more conservative estimate of household equity holdings.

To arrive at the total value of listed stocks held by Dutch households, we need to (a) add the value of foreign stocks held by Dutch households, and (b) subtract the value of domestic stocks owned by foreigners. While direct information on asset holdings is not available, we can arrive at a reasonable figure in the following way. We know the net primary income received, which is calculated by Smits, Horlings, and van Zanden (2000) until 1913 and by Statistics Netherlands for most years afterwards[20]. Then, we capitalize these dividend streams $y_t^{f,div}$ using an average dividend yield $r_t^{f,div}$, using

$$y_t^{f,div} := r_t^{f,div} \cdot W_t^f .$$

Note that for our purposes, the dividend yield is the correct variable by which to capitalize the dividend streams and not the total return, i.e., dividends plus capital gains. This is because we are interested in the nominal value of equity at time t, and not in its real value, corrected for price revaluations.

For each year, we take the average of the dividend yields of Berlin, Brussels, London, New York, and Paris; the stock exchanges which were by far the most important for Dutch non-colonial foreign investment in terms of cross-listed equities

[20] The years 1918-1920 are missing and are linearly interpolated from the values of 1917 and 1921, which reasonably captures the decline in foreign capital income due to the upheavals of the war, the Russian Revolution, and other large international shocks.

and bonds (Moore 2012). Data on dividend yields for those exchanges are taken from Jordà et al. (2019). The five series are generally quite close to each other, with an average standard deviation over the 1880–1938 period of slightly more than 1%. During crisis years, such as 1917, the standard deviation increases; nevertheless, we feel that the average dividend yield gives a reasonable capitalization factor even in volatile years, particularly when considering that the resulting series is quite smooth; sticking with only one series would likely result in much more artificial volatility in the equity series.

When capitalizing net dividends from abroad, we implicitly capitalize Dutch dividends paid to foreigners with the same yield as foreign dividends paid to Dutch households. Unfortunately, no systematic information exists on Dutch dividend yields for the entire pre-war period; the only available series covers 1900–1917, and is included in Jordà et al. (2019). Inspection of dividend yields in this series reveals that Dutch dividends did obtain higher yields than the average yield we have chosen for this period. This outperformance is on average 1.6 percentage points. However, the volatility is large, with a standard deviation of 1 percentage point. In short, we are not certain that Dutch dividends systematically performed better than foreign dividends even in this limited sample. This conclusion also holds for each of the five different series individually: none of them seem systematically over- or underperformed by the Dutch series, and the Dutch series does not track any of them particularly closely. Hence, we stick with the averaged series throughout, noting that this likely represents an *underestimation* of Dutch equity holdings, since we capitalize dividends paid to foreigners by a too large factor.

Our series on colonial dividends, covered in Smits, Horlings, and van Zanden (2000) and den Bakker (2019), is mainly based on the work of Korthals Altes (1986), who carefully reconstructs the Indonesian balance of payments from 1822 until 1939.[21] Unfortunately, neither Korthals Altes nor anyone else has provided consistent estimates of the dividend *yield* of Indonesian stocks. Bosch (1948) presents estimates based on a sample of firms on the Batavia stock exchange, but his series are problematic since he does not weight stocks by market capitalization, nor does he compute geometric averages of monthly yields, but only simple averages. As a result, his series of returns are much too high. van der Eng (1998) presents several estimates, but notes that these are mostly unweighted as well, and agrees with our assessment

[21] As noted in Smits, Horlings, and van Zanden (2000), no sources exist on equity holdings in the other Dutch colonies, which were the Dutch Antilles and Surinam. Particularly the inclusion of Surinam would be interesting, since the abolition of slavery in 1863 and the subsequent compensation of Surinamese slaveholders would be an important aspect of the Dutch private wealth distribution to cover; we hope that future research uncovers methods to estimate these colonial holdings.

that this likely results in upward-biased dividend yields; his critical conclusion is that dividend yields were probably only 2 percentage points higher than bond yields on average, which would put Indonesian dividend yields closely in line with our calculated world average. Buelens and Frankema (2016) present average rates of return and dividend yields for 1919-1958, for a sample of 17 firms which were listed on the Belgian stock exchange. They find geometric average dividend yields of 2.7% for the 1919–1928 period, and 1.3% for the 1929–1938 period. These averages are *lower* than our estimates of the world dividend yield by several percentage points. Since it is unclear how representative their sample is of the whole, we stick with capitalizing Indonesian dividend yields with the world dividend yield; this results in a more conservative series of colonial wealth, if anything.

The resulting series of colonial and other foreign asset holdings align very well with existing estimates. The colonial holdings in 1938 are estimated at approximately 4 billion guilders (70% of national income), which is exactly Tinbergen's estimate and very close to a number of other estimates covered in Bosch (1948). Moreover, we estimate the total amount of listed equity in 1914 at 6.5 billion guilders, which is very close to the 6 billion given by de Vries (1976); moreover, his estimates of domestic shares in that year, at 1.7 billion, are almost identical to ours; his figures for foreign equity – which do not include colonies – is given at around 3 billion, which is close to our 3.5 billion estimate.

Bosch (1948), meanwhile, critically analyzes several estimates of Dutch investments in the United States, one of the major destinations of foreign investment. He suggests a total investment in the U.S. of 1.5 billion guilders in 1908, 1.5-1.7 billion in 1914, 1 billion in 1919, 600 million in 1924, 1.1 billion in 1929, 1.2 billion in 1935, and 1.5-1.75 billion in 1939. All these numbers are well within the range of possibility in our series, consistently suggesting an American share in total non-colonial investment of around 50%, which is very plausible. The only year where this doesn't align well is 1908, where Bosch's estimate is almost our entire estimate for foreign wealth. Since it would be implausible in our view for U.S. investment to remain stagnant for the entire 1908-1914 period, this suggests to us that his 1908 figure might be an overestimate.

Ideally, we would also need to subtract ownership by financial institutions such as banks and insurance corporations. Unfortunately, information on asset holdings of these institutions is very limited in nature, but appears not to be quantitatively significant; a 1942 census of the 42 major mortgage banks, for example, only found equity holdings equalling 14 million guilders in 1938, which is a tiny fraction of the total size of equity (Amsterdamsch Effectenblad 1942). This view is confirmed by the 1938 balance sheet, which shows combined equity holdings of the financial sector of around 300 million guilders, which is less than 1% of the total estimated value of the corporate sector. A final confirmation for a slightly earlier year comes from de Vries (1976), who mentions that equity holdings by institutions accounted for only 0.3% of the total value in 1935. All these facts combined give us confidence that our total equity series is quite robust.

Comparing this series with the 1938 National Accounts is not trivial, since the National Accounts list several items on the household balance sheet which added together resemble our measure of total equity. If we add up all these items – which include shares, bonds, and mortgages of corporations held by households as assets, shares and bonds of banks and other financial corporations, and investments in the rest of the world – we arrive at a figure of 14.2 billion guilders, which corresponds very closely to our aggregate figure for all equity excluding government bonds of 14.5 billion guilders. As mentioned in the previous section, government bonds also align very closely with the National Accounts figure. Our estimate of 14.5 billion includes the residually estimated nonlisted stocks, to which we turn next.

After 1947 until 1970, we continue with our basic approach as outlined above; however, since we already know total wealth for all these years (whether directly our by our interpolation), we are less concerned with precise estimations of each wealth component, but rather to get the relative magnitudes right. Nevertheless, our series aligns remarkably closely to the CPB balance sheet, which we start using in 1970. The balance sheet lists a total value of listed equity of 29 billion guilders in 1970, which is a bit lower than the 1969 combined value of domestic and foreign equity of 43 billion; however, when combined with the slightly higher estimate of household-owned bonds, the total value of equity – excluding non-listed stocks – remains remarkably stable, going from 70% of national income in 1969 to 60% in 1970.

After 1995, we use the official National Accounts.

Nonlisted Stocks and other Financial Wealth Nonlisted stocks are the most difficult item to estimate, since there are no official sources for them until 1970. Dutch corporate law did not make a distinction between corporate forms until the early 1970s, when a 'closed' corporation with limited liability was established, the *besloten vennootschap*. Until the 1970s, the main corporate form was the "nameless" corporation, *naamloze vennootschap*, which could be either listed or non-listed. Remarkably, there exist no official figures on the total number of corporations until 1930, shortly after the first official law on this corporate form was established. Hence, for the vast majority of our period, we do not possess any additional information about even the number of corporations, let alone their balance sheets. Hence, we resort to estimating this wealth component residually until 1970. All other wealth components are accounted for, as described in this appendix. Hence, any remaining difference between an official balance sheet and our series must be attributable to non-listed equity. We use the 1938 balance sheet as our benchmark for total wealth, where we add to the household balance sheet our series of semiprivate wealth, which were noted in the 1938 balance sheet as "potential private wealth" on the balance sheet of the insurance sector; we also subtract the residual part of non-financial assets that we attribute to the corporate sector, as explained in section 1.A.2.4. The residual between our series in 1938 and the official balance sheet is approximately 6.3 billion guilders, or 120% of national income. This strikes us as a reasonable figure, especially when considering that this figure will also include smaller financial items which we also do not estimate separately, such as corporate bonds and shares in non-listed financial institutions. Remarkably, our figure for nonlisted stocks aligns very well with the figures in Bosch Kemper (1950), who mentions a total paid-up capital by all corporations – listed and unlisted – which is very close to our series.

Since there is no additional information before 1938 on non-listed firms that we might use, we simply peg the value of non-listed firms to the total value of equity, and extrapolate this ratio backwards until 1880. This is the least intrusive assumption we could think of to measure non-listed equity before 1938.

After 1947 until 1970, we continue with this residual approach, until 1970, when we use the information in the CPB balance sheet for households, which notes 'aanmerkelijk belang' (significant ownership), i.e., whether a household owns more than 5% of shares in a firm. Almost always, significant ownership pertains to non-listed firms; nowadays, the majority of significant ownerships are in closed limited liability corporations (*besloten vennootschappen*). Our residually estimated series aligns quite well with the 1970 balance sheet, being a bit higher (88 billion guilders in 1969 versus 65 billion in 1970). This difference – amounting to 20% of national income – is unlikely to significantly affect our results.

After 1995, we use the total value of equity holdings, which include non-listed corporations.

1.A.4. Liabilities

The total value of private liabilities was based on the total value of private mortgages (loop der hypotheken/ openstaande inschrijvingen), which are made available in the Annual Statistics for the Netherlands. Not all mortgages can be ascribed to the household sector; a large fraction is attributable to the corporate sector instead. We have a first breakdown of mortgage debt in 1970, when we both have the historical sources described above and the balance sheet compiled by the CPB. We take the average ratio of household mortgages to total mortgages and apply this pre-1970.

The result is a consistent series of mortgage debt; while we might miss some fluctuations by taking this ratio, we have no indication that there were dramatic trends in mortgage debt prior to the 1970s that would qualitatively challenge our results.

We added the number of loans issued by cooperative banks, help banks and credit unions. Unlike commercial banks, these banks were know to issue private, consumer loans. We retrieved this data from Westrate (1948), de Vicq and van Bochove (2024, 2023) and de Vicq (2022) respectively.

After 1970, we use the CPB balance sheets for both mortgage debt and other liabilities, until 1995, when we switch to the National Accounts. There is a small trend break in 1995, as the Financial Accounts record more liabilities for the household sector than the CPB estimates, on the order of 30% of national income.

1.A.5. Total Household Wealth, 1947–1969

We have balance sheets for the years 1946–1952, available in the National Accounts of 1954, which we show in Figure . These balance sheets also show the single estimate for 1938, which we use to calibrate our manual reconstruction of the national accounts from 1880–1938, detailed in the previous section. These balance sheets, although they do not decompose total wealth into components, do include estimates for total household wealth.

After 1952, the National Accounts no longer feature balance sheets regularly. Two exceptions exist: For 1958, we have a breakdown of national wealth, from which we can subtract the value of government assets to arrive at private wealth; and 1960, where the total size of national wealth is mentioned. We assume that government wealth is the same proportion of national wealth in 1960 as in 1958, and subtract this estimated government wealth to arrive at household wealth for 1960. In sum, we have estimates from National Accounts for household wealth for 1946–1952, 1958, and 1960. As discussed in the main text and in the next section, we also have estimates of balance sheets from 1970 onward. We then interpolate all missing years using the multiplicative decomposition (1.5), where we residually estimate an average capital gains rate q such that the known endpoints (1958, 1960, and 1970, respectively) are reached. We define private savings as the sum of household, corporation, and financial institution saving. After having estimated the endpoints of each year using this method, we average W_{t-1} and W_t to reach middle-of-year estimates, as is consistent with DINA practice. This means that we have to disregard the data point for 1946, as its value is subsumed in the averaged value for 1947.

The main source for our balance sheets are the balance sheets constructed by the CPB Netherlands Bureau of Economics Analysis, which they published as an appendix to their 2013 *Macro Economische Verkenningen* (Macroeconomic

Figure 1.16: Household Wealth Estimates in National Accounts, 1938 and 1946–1952

96

P. NATIONALE REKENINGEN

159. Nationale Balans, 31 December 1)

	1938	1946	1947	1948	1949	1950	1951	1952
	× mld gld.							
Activa								
Bedrijven en banken:								
Kapitaalgoederen	20,5	44,9	48,3	52,3	57,6	62,8	77,9	78,6
Buitenland:								
Buitenlands saldo	8,5	8,1	5,9	5,5	4,2	4,8	6,6	8,9
Nationaal vermogen	29,0	53,0	54,2	57,8	61,8	67,6	84,5	87,5
	× mld gld.							
Passiva								
Verzekeringsfondsen:								
Potentiëel privaatvermogen .	3,5	6,1	6,5	7,0	7,6	8,3	9,1	10,0
Verbruikers:								
Particulier vermogen	29,9	63,7	65,2	66,9	71,6	74,6	88,3	88,6
Overheid:								
Overheidsvermogen 2)	– 4,4	– 16,8	– 17,5	– 16,1	– 17,4	– 15,3	– 12,9	– 11,1
Nationaal vermogen	29,0	53,0	54,2	57,8	61,8	67,6	84,5	87,5

1) Zie voor toelichting „Statistische en econometrische onderzoekingen" jrg. 9, no. 1 en de daarin genoemde publicaties. 2) Deze negatieve bedragen representeren het verschil tussen activa en passiva van de Overheid.

160. Nationaal vermogen per 31 December 1)

	1938	1946	1947	1948	1949	1950	1951	1952
	× mld gld.							
Grond	4,6	11,0	11,5	12,0	12,7	13,5	16,0	16,1
Voorraden	2,6	0,7	1,4	2,1	2,8	4,5	6,8	5,5
Overige kapitaalgoederen	13,3	33,2	35,4	38,2	42,1	44,8	55,1	57,0
Buitenlands saldo	8,5	8,1	5,9	5,5	4,2	4,8	6,6	8,9
Nationaal vermogen	29,0	53,0	54,2	57,8	61,8	67,6	84,5	87,5

1) Zie noot 1 bij tabel 159.

Explorations, MEV). These balance sheets include financial assets, deposits, housing, other real estate, business wealth, and pension claims. We verify that all estimates of these wealth components correspond closely with estimates from other sources, such as various series by Statistics Netherlands and De Nederlandsche Bank. All series mentioned so far match very closely with National Accounts totals. Stocks and bonds are a bit noisier, but the results are still very comparable.

The largest difficulty with the 1970–1994 balance sheets lies in life insurance, which is not included in the MEV balance sheets. There are three sources: Long-run data from DNB on life insurers' technical reserves, data from Statistics Netherlands on life insurers' technical reserves, and the data from the National Accounts. The first source is the only one available for the entire period, but is also one that diverges widely from the other two. Hence, we opt for the following approach: For 1970–1974, we adjust the DNB series such that it merges perfectly with the Statistics Netherlands series in 1975, which we use until 1994. This adjustment ensures that the life insurance series do not counterfactually exceed the values reported by Statistics Netherlands or the National Accounts, which ensures consistency.

1.B. Estate Multiplier Methods

1.B.1. Death Duties Data

Table 1.5: Exemptions for the Succession Tax in the Netherlands since 1818

Period	Direct accession line	Indirect accession line
<1878	Exempt	Exempt when below 300 guilders
1878–1896	Exempt when below 1,000 guilders	Exempt when below 300 guilders
1897–1910	Exempt when below 1,500 guilders	Exempt when below 500 guilders
>1911	Exempt when below 1,000 guilders	Exempt when below 300 guilders

Note: Table shows the exemptions that applied to the inheritance tax in various years. 'Direct accession line' refers to direct family of the decedent; 'indirect accession line' refers to other family.

The procedure that the authorities went through to identify the amount of tax to be paid has as follows: If after a formal application by the successors of an estate, it was found that the estate's net worth was likely higher than the appropriate threshold, then a detailed evaluation called "Memorie van Successie" was drawn up. The net worth of those estates – along with the all other deceased individuals with net worth lower than the threshold – were listed alphabetically in Tafel V-bis, which functioned as an annual ledger for the more detailed "Memorie van Successie". To ensure a high tax morale the authorities maintained "[p]enalties for fraud and evasion were about twice the due tax plus any costs" (Gelderblom, Jonker, Peeters, and de Vicq (2023)). Several crosschecks where in the disposal of the tax inspectors. The actual value of the estate's land, deposits, and other investments where relatively easy to be verified. In contrast, various types of debt where significantly more difficult to validate (Gelderblom, Jonker, Peeters, and de Vicq (2023)).

Table 1.6: Succession Tax Data Thresholds

Period	# Brackets	Thresholds
1882–1910	20	300; 500; 1,000; 1,500; 2,000; 3,000; 5,000; 7,500; 10,000; 15,000; 20,000; 25,000; 30,000; 40,000; 50,000; 75,000; 100,000; 150,000; 200,000; 300,000; 500,000; >500,000
1900–1955	8	100; 1,000; 2,000; 5,000; 10,000; 25,000; 50,000; 200,000; >200,000
1956–1984	10	100; 1,000; 2,000; 5,000; 10,000; 25,000; 50,000; 100,000; 200,000; 500,000; >500,000

In terms of population coverage, during the early period of 1854–1878 it is only the indirect heirs that were subject to the estate tax. However, we do not have data on how many indirect heirs (which are the actual filers) are included in the reported aggregate wealth totals. The number of filers is important in making the conversion from the amount of wealth identified by the death duties tax to the aggregate wealth in the economy. But for the years 1878–1910 we do have the data split between direct and indirect heirs, so we can extrapolate backward and approximate the number of (indirect heir) filers for the 1854–1878 period, as shown in table 1.7. In this we are assuming that the ratio of indirect heir filers is equal to that from 1880 (and therefore the coverage remains fixed at 5.27%). The data from 1878 onward are available in tabulated form, with more granular thresholds, as shown in table 1.6 above.

The specific Tafel V-bis that was processed and made available by Gelderblom, Jonker, Peeters, and de Vicq (2023) contains all individuals that died in 1921, that had a wealth above the tax threshold, and also includes their demographic profile (age and gender), and their total wealth valuation. A limitation of using this source to estimate an estate multiplier is that it is we need to use the same multiplier for all the years. To address this limitation we devise a method based on the ratio of

Table 1.7: Coverage of the Death Duties Tax in the Netherlands 1850–1980

Year	Total Deceased	Filers	Coverage
1850	69,377	*3,656*	*5.27%*
1860	84,382	*4,447*	*5.27%*
1870	95,289	*5,022*	*5.27%*
1880	95,282	9,508	9.98%
1890	93,246	10,090	10.82%
1900	92,043	11,101	12.06%
1910	79,984	10,712	13.39%
1920	81,525	13,623	16.71%
1930	71,682	14,382	20.06%
1940	87,722	18,251	20.81%
1948	72,459	19,602	27.05%
1956	85,000	37,119	43.67%
1963	96,000	35,874	37.37%
1970	110,000	38,167	34.70%
1975	114,000	48,398	42.45%
1980	114,000	43,410	38.08%

Notes: Selected Years. 1850–1870 are estimated, see the text for details.

the estimated estate multiplier for 1921 and the naive estate multiplier (which is the one obtaining by naively assuming that there is no mortality rate differential between the rich and the general population). We estimate this naive multiplier as the ratio of total population size over the total number of deaths. We index the series by dividing with the naive multiplier for 1921. Multiplying the 1921 multiplier that we estimate based on the Tafel V-bis data with the indexed series of these naive multipliers we produce a dynamic series of alternative dynamic estate multipliers that consider the changes in the population dynamics. The last step in this procedure is to take the average of the fixed and the alternative series to obtain our final estate multiplier series. This step is based on the observation that the bias from the fixed multiplier and the bias introduced from the dynamic alternative series move in opposite directions. Both methods and the rationale behind these biases are discussed next.

Given the availability of detailed wealth and age data on the individual level for 1921 from Gelderblom, Jonker, Peeters, and de Vicq 2023, we will use them to estimate the aggregate estate multiplier for 1921. Our problem, however, is that for 1921, although we have the population that died at various wealth and age groupings, we do not know what is the corresponding size of each of those groupings in the general population. For example, we do know that in 1921, there are say 3275 50-60 year old with wealth more than 15K who died, but we do not know in 1920 how many were the living 50-60 years old with wealth more than 15K. Therefore, we do not have a proper denominator to estimate the mortality rates for each wealth group. In the solution described in the next paragraph we are able to estimate an average mortality rate for the rich as a whole, and through that arrive at an aggregate estate multiplier.

To overcome the lack of proper denominator problem, we work as follows: from a different source (see next section) we have the wealth tax data from 1920 (which is the reference year for the death rates of 1921). We have these wealth tax data for the population as a whole distributed in various wealth buckets. But, we have no information with respect to their age distribution per bucket. To address this we combine the wealth tax data buckets with the 1921 data in the following way: for each of the wealth buckets we get the age distribution from the 1921 Tafel-V bis data, by splitting our complete 1921 inheritance tax data into the same buckets that the wealth tax data are provided with. Doing so for all buckets in 1921 we obtain an estimate for the age-wealth distribution. We then re-combine the 1921 data using the wealth buckets used in 1920 in order to get an estimate of the 1920 age structure of the wealth tax data. In this we assume that the individual sample in the death duties for 1921 is not substantially different compared to 1920. Unfortunately thought we do not have enough data to populate all the age-wealth groups required from the wealth tax data buckets. We therefore gather all the rich in one group (>15,000 guilders),[22] and we thus obtain the mortality rate of the rich in general, as we are not able to distinguish the age wise distribution of the death rate of the rich.

Our end goal here is to estimate an aggregate estate multiplier that can be applied upon the total death duties wealth, since we do not have the inheritance data spit across age groups. To estimate this aggregate estate multiplier for 1921 we will exploit the mortality rate for the rich that is estimated based on the procedure in the previous paragraph. A byproduct of the procedure is the age distribution of the rich (we already have the total number of rich we have for 1920 based on the wealth tax). We then multiply each age group in that distribution with the mortality rate of the general population, to get the number of the rich that would exist should the rich and the general population have the same mortality rates (R_t). The ratio of the actual number of rich over R_t is the ratio of the average mortality rate differential between the rich and the general population (M_d). We now divide the mortality rates from 1921 with this ratio M_d to create the adjusted mortality rate table. We

[22] The aggregate from the official data is 451,912,000 but in the Tafel V-bis data for 1921 we have 412,440,216. We therefore multiply all inheritance entries with this ratio.

populate the adjusted mortality rates table with the expected wealth for each age group based on the data from Tafel V-bis, which gives us the average wealth per age group. Dividing each average wealth per age group cell with the adjusted mortality rate for each age group, gives us the total wealth in the population for each age group. Summing all these total wealth estimates produces our estimate for total wealth in the population for 1921. Finally, taking the ratio of the estimated total wealth in the population over the total wealth captured by the estate tax gives us the estimate for the aggregate estate multiplier for 1921.

For the period 1854–1878 period, where only indirect heirs are taxed and registered, we work in two steps: first, we use one multiplier from the years 1878–1910 for which we have the data split between direct and indirect heirs, to bring the indirect–heir–only data to a comparable level with the post-1878 period; second, we multiply by the same mortality rate multiplier as we do for all other years.

Applying the estate multiplier from 1921 on all years assumes that the age differential between the rich and the entire population is constant across all years. To incorporate a dynamic element in our estate multiplier estimates we introduce the concept of the naive estate multiplier. The naive estate multiplier for year t is the ratio of all the living in year $t - 1$ over all that died in year t. The naive estate multiplier assumes that there is no differential in terms of mortality rates between the rich and the general population. Taking the ratio of the naive estate multiplier over the estimated multiplier for 1921, we isolate the differential between the two in one year of reference. We then index the entire series of naive estate multipliers. This way we have another series of estate multipliers that can be used under the assumption that the ratio between the *true* multiplier (meaning the multiplier that we would obtain from ideal data) and the naive multiplier is constant across the years. This implies that the mortality rate differential between the rich and the rest of the population is constant across the years. But we know from other sources, that there are evidence that do not support this claim, and indicate that this differential is probably diminishing in time (Kopczuk and Saez (2004)). Therefore the adjustment required to move from the available total mortality rates to the (estimated) mortality rates of the rich becomes smaller. This means that the results of our fixed multiplier series and the naive multiplier series will move relatively as a product of time, and this is captured well in figure 1.17). The two methods provide reasonably close estimates for the post 1911 period, but diverge seriously in the earlier years. One reason of this divergence may be the substantial increase of the crude death rate in the earlier period. On average, during the 1910–1979 period the crude death rate is around 9.5% while in the 1850–1909 period it is 21.7% Petersen 1960. Most of the divergence takes place during the 1890–1910 period, which also corresponds to the period where the bulk of the divergence builds-up as shown in the figure above. Given this large deviation in the earlier period between the two series, we take the average of the two series as our final estate multiplier series, so that we reasonably lower the probability that our results are driven from the surge in the crude death rates prior to our benchmark year.

Figure 1.17: Estate Multiplier Estimates per Method

Notes: Figure shows total wealth estimates based on the estate multiplier method, using three alternatives: (a) the fixed aggregate estate multiplier for 1921, (b) the dynamic aggregate estate multiplier, and (c) the average of (a) and (b).

1.C. Wealth Tax and Wealth Distribution

In this appendix, we describe the wealth tax data and associated methods. First, we describe the lognormal extrapolation method, as developed by Wilterdink (1984) and Potharst (2022). Then, we discuss our methods and sources for the wealth distribution.

1.C.1. Lognormal Extrapolation Method

We use the tabulated figures produced in *Jaarcijfers voor Nederland*, which became its English equivalent *Statistical Yearbook* in the 1970s and which provide values from 1894–1993. The following Table 1.8 reports the structure of brackets over the years.

Table 1.8: Wealth Tax Data Thresholds for Different Periods.

Period	# Brackets	Thresholds (in 1,000 NLG)
1894–1914	18	13; 15; 20; 30; 40; 50; 75; 100; 150; 200; 300; 500; 750; 1,000; 1,500; 2,000; 5,000; 10,000; >10,000
1915–1924	17	15; 20; 30; 40; 50; 75; 100; 150; 200; 300; 500; 750; 1,000; 1,500; 2,000; 5,000; 10,000; >10,000
1925–1941	9	16; 30; 50; 100; 200; 300; 500; 1,000; >1,000
1942–1956	11	<10; 10; 15; 20; 30; 50; 100; 200; 300; 500; 1,000; >1,000
1957–1969	7	<50; 100; 200; 300; 500; 1,000; >1,000
1970–1973	16	100; 150; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 1,500; 2,000; 3,000; 5,000; 10,000; >10,000
1974–1975	10	100; 150; 200; 300; 500; 1,000; 1,500; 2,000; 5,000; 10,000; >10,000
1976–1982	6	100; 150; 200; 300; 500; 1,000; >1,000
1983–1993	6	200; 300; 400; 500; 750; 1,000; >1,000

The method applied by Wilterdink (1984) and developed by Potharst (2022) uses information on the thresholds of each wealth bracket to estimate a lognormal distribution. Essentially, the method estimates the overall mean μ and variance σ^2 by minimizing the squared distance between the observed percentile-bracket average pairs of each bracket, and the theoretical lognormal distribution. Once we have an estimated mean and variance, we can integrate over the density to arrive at an estimate of total wealth. Then the estimated total wealth above the lowest wealth threshold that it is captured by the wealth tax data is substituted by the actual total wealth contained in the wealth tax tabulations (although the difference between the estimated and the data is relatively small with the theoretical being on average 0.5% –and a standard deviation of 4%– lower than the data across the entire period). We refer the reader to Potharst (2022) for further details.

1.C.2. Wealth Distribution

Our series on top wealth shares can be split in two parts: the pre-1993 and post-1993 period. Pre-1993, we use the wealth tax tabulations extensively discussed in the previous section. Using total wealth derived in the lognormal distribution method as denominator, we then calculate top wealth shares using generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022); we use the `gpinter` interface on `www.wid.world` for our interpolations. The gpinter algorithm takes as inputs for each bracket k a bracket lower threshold q_k, its corresponding percentile p_k and the bracket average μ_k. It then interpolates the entire distribution based on the given inputs and the (known) mean population wealth $\overline{\mu}$, using the fact that at each bracket threshold we can exactly calculate a local inverted Pareto coefficient $b(p) = \mathrm{E}\left[W \mid W > q_k\right]$. The points of the distribution between the known thresholds are then interpolated using quintic splines.

Blanchet, Fournier, and Piketty (2022) caution that the algorithm works best when one or more of the thresholds is placed at the bottom of the distribution. For the pre-1993 data, this is not possible, since the wealth tax returns only cover the upper tail. Hence, our results might be sensitive to the exact functional form of the wealth distribution, which is not known.

After 1993, Statistics Netherlands publishes full wealth distributions of taxable wealth (i.e., excluding semiprivate wealth). These tabulations cover the entire distribution for every year from 1993 until 2019, with the exception of 2001; we interpolate the 2001 values linearly using the values in 2000 and 2002. We should note that the distributional statistics generally improve in quality over time; in particular, data quality increases drastically in 2006, when Statistics Netherlands starts to cover the universe of households instead of a representative sample; and in 2011, when many difficult-to-measure wealth components like unsecured credit and small deposits become integrally measured. Hence, data from 2006 (and especially 2011) are considered to be close to error-free, subject to the usual caveats about tax avoidance and evasion, One particular concern about the Dutch wealth distribution is the importance of private business equity for the top of the distribution (Toussaint, van Bavel, Salverda, and Teulings 2020). This wealth component is taxed preferentially, and is difficult to assess at market-value due to the closely-held nature of the businesses. Since only realized capital gains are taxed, fluctuations in flows are also difficult to measure. Although Statistics Netherlands has recently improved its measurement of private business equity, concerns about the potential underestimation of this wealth component remain. In this chapter, we simply take the distributional statistics published by Statistics Netherlands as given, and do not adjust them to account for increasing quality over time or potential underreporting of some components. Future work will doubtlessly improve upon these decisions.

The wealth tax data get gradually worse over the years, with a noted drop in quality coming after a major tax reform in 1964, which increased the threshold to 100,000 guilders and increased exemptions for various wealth components (Wilterdink 1984). Hence, the quality of the estimates based on the wealth tax steadily decreases, in particular after 1963. This is evident when looking at Figure 1.18. We see a major jump happening in 1993 – the first year of the new Statistics Netherlands estimates. This jump is particularly pronounced for the 5% share. Inspection of the `gpinter` output reveals that the algorithm puts increased weight on the bottom 50% share, and a steadily declining weight on the top-10% share.[23] There is no easy way to fix this issue, since we do not know exactly how much of the decline after 1963 is due to a genuine decrease in inequality, and how much due to declining data quality.

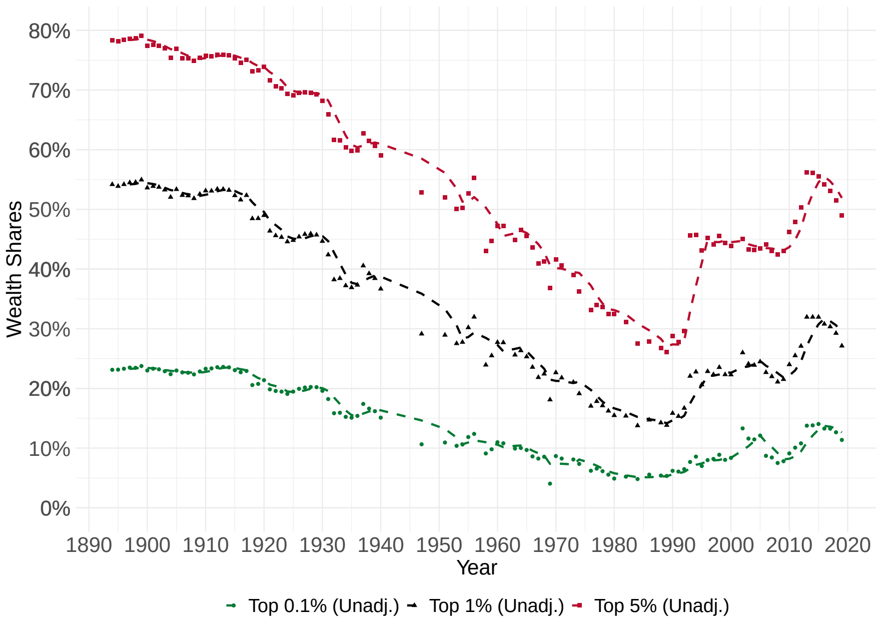
There is one factor in our favor, which is that for the year 1993, we have both a wealth tax estimate and a full tabulated estimate from Statistics Netherlands. We use this year to benchmark the divergence between the two series as follows. We first run `gpinter` on both series separately, giving a full interpolated wealth distribution, with wealth shares for every percentile from the bottom to the top. Then, we adjust all percentiles from the wealth tax series output between 1963 and 1993, using the following weighting scheme:

$$
\begin{aligned}
S_t^*(p) &= \omega_t S_{1993}^{\mathrm{CBS}}(p) + (1-\omega_t) S_t^{\mathrm{WT}}(p), \qquad (1.7)\\
\omega_t &:= \frac{t-1963}{1993-1963}, \quad t \in \{1963, \ldots, 1993\}.
\end{aligned}
$$

Here, $S(p)$ is the wealth share of percentile (or fractile) p, superscript CBS refers to the Statistics Netherlands series, and superscript WT refers to the wealth tax series. This linearly interpolated adjustment puts increasing weight on the 1993 data point as t approaches 1993; this allows us to retain any genuine trend in wealth shares between 1963 and 1993 whilst removing the jump in 1993 by construction. Note that since we adjust all percentiles, the wealth shares output of `gpinter` is still consistent in that all wealth shares sum to one.

[23] Note that we do not report either of these statistics, since in our view the wealth tax data are generally only reliable for the top 5% (since the range covered by the tax is at most 7% or so in a given year). Nevertheless, `gpinter` outputs the entire wealth distribution, from the bottom to the top.

Figure 1.18: Unadjusted Wealth Shares



1.D. Population, Income, Prices, and Savings

1.D.1. Population

We use population figures from van der Bie (2001). Note that for the wealth tax extrapolations and wealth share estimates, we have to consider *tax units* as a reference, i.e., all adults above 20, where married couples are treated as a single unit (Wilterdink 1984). We follow Wilterdink (1984) and simply subtract the total number of married women from the population total above 20. To be clear, these are estimates for the total number of tax units, and are not used for any per capita estimates of wealth or income throughout the chapter.

1.D.2. Prices

Throughout, we use the Statistics Netherlands Consumer Price Index, as it appears in van der Bie (2001), which is based on the work of Smits, Horlings, and van Zanden (2000) before 1913 and Statistics Netherlands and den Bakker (2019) thereafter. We normalize the index such that 2015 equals 100.

1.D.3. Income and Factor Shares

We measure net national income at market prices. Data are from van der Bie (2001), which are based on Smits, Horlings, and van Zanden (2000) for the period up to 1913, and Statistics Netherlands and den Bakker (2019) thereafter.

For the labor and capital shares of income pre-World War 2, we also rely on Smits, Horlings, and van Zanden (2000) and den Bakker (2019). Smits, Horlings, and van Zanden (2000) carefully present separate series for wages, capital income, mixed income, and indirect taxes, which sum up to national income; but they do not calculate the share of mixed income (called 'profits' in their series) that is attributable to labor. den Bakker (2019) does do this, by presenting both wage shares and labor income shares for the period 1921–1939; however, his denominator is GDP, not national income. Hence, we opt for the following approach. Before 1913, we use the data by Smits, Horlings, and van Zanden (2000), allocating 30% of mixed income to labor. This is the average of the allocation in the den Bakker (2019) series, and seems reasonable, given that we know that labor only started to obtain importance at the turn of the 20th century. For 1921–1939, we use the series by den Bakker (2019), but add net primary income from abroad, also from the same source, to add up to national income. All primary income from abroad is allocated to capital, as Den Bakker mentions that labor remittances were negligible in this period and hence also considers all primary income to be capital income. This results in a consistent labor and capital share out of national income for the entire 1854–1939 period, with the exception of the years 1914-1920. The obtained labor and capital share are broadly comparable to those estimated in Bengtsson and Waldenström (2018), who do not use the novel estimates by den Bakker (2019) and do not attempt to estimate factor shares in the pre-1913 data.

Once we have factor shares, we can also estimate the return to wealth r_t, using

$$r_t := \frac{\alpha Y_t}{W_t}, \tag{1.8}$$

where α is the capital share of national income.

1.D.4. Savings

For 1854–1913, we use Smits, Horlings, and van Zanden (2000). National saving is defined as gross fixed capital accumulation minus depreciation plus exports minus imports. Helpfully, Smits et al. also construct estimates of government investment, which we take as a measure of government saving. Hence, private saving equals national saving minus government saving. No further decomposition of the private saving rate is possible for this period.

For 1921–1939, we have data on savings from den Bakker (2019), who constructs estimates of national saving, government saving, and corporate retained earnings, and obtains household saving as a residual.

From 1947 onward, we use savings rates reported by Statistics Netherlands, in various editions of *Jaarcijfers voor Nederland*. We define private savings as the sum of household, corporate and financial institution savings, but we also report these estimates separately. From 1969, we use the official National Accounts.

For table 1.3, we use the volume and price changes recorded in the Financial and Non-Financial Accounts of the household sector (including nonprofits) from 1995.

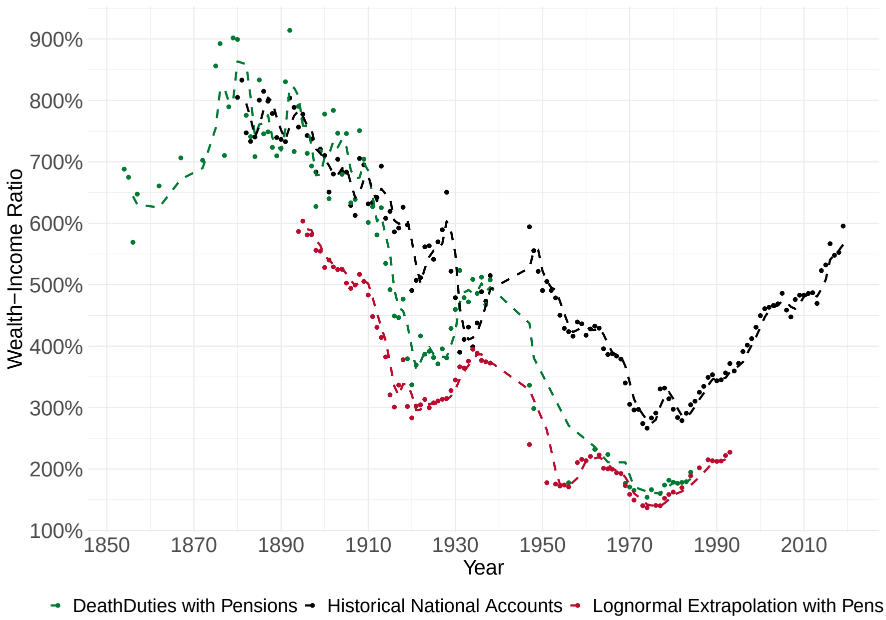
1.E. Robustness Checks

Table 1.9: Comparisons with Historical Estimates of Household Wealth

Year	HNA	Estate	Wealth Tax	Historical	Source
1854-1857		4,894		5,723	Stuart, Ons Maatschappelijk Vermogen, 1888
1855		4,894		3,800	Verstegen, National Wealth and Income
1865		6,138		5,150	Verstegen, National Wealth and Income
1872		7,877		5,260	*Ons Nationaal Vermogen, 1875*
1875 - 1877		10,069		3,021	Gleichman, Memorie van Toelichting, 1879
1879		9,460		11,166	Vissering, Memorie van Toelichting, 1879
1880	9,113	10,181		10,280	Boissevain, De Omvang, 1883
1879-1883	9,674			9,822	Boissevain, De Omvang, 1884
1879-1882	9,113	10,181		8,397	Fabricant & Maarschalk, International capital, 1952
1879-1882	9,113	10,181		8,262	Van Zanden, Income and wealth inequality, 1995
1879-1882	9,113	10,181		8,702	Verstegen, National Wealth and Income, 1996
1879-1882	9,113	10,181		11,318	Stuart, Ons Maatschappelijk Vermogen, 1888
1883-1886	8,633	8,261		10,992	Stuart, Ons Maatschappelijk Vermogen, 1888
1888-1890	9,227	8,855		11,902	Boissevain, De Omvang, 1891
1908-1912	12,711	11,837	9,460	12,120	Derksen, Berekening van het nationaal vermogen, 1934
1908-1912	12,711	11,837	9,460	11,558	Fabricant & Maarschalk, International capital, 1952
1908-1912	12,711	11,837	9,460	11,240	Van Zanden, Income and wealth inequality, 1995
1908-1912	12,711	11,837	9,460	12,860	Verstegen, National Wealth and Income, 1996
1915	19,987	15,520	9,995	7,453	Bonger, Vermogen en Inkomen, 1923
1916	21,635	16,197	10,721	7,995	Bonger, Vermogen en Inkomen, 1923
1917	21,713	15,952	11,927	9,058	Bonger, Vermogen en Inkomen, 1923
1918	25,670	19,091	15,043	11,602	Bonger, Vermogen en Inkomen, 1923
1919	33,036	20,499	16,205	12,711	Bonger, Vermogen en Inkomen, 1923
1920	30,831	20,763	17,379	13,855	Bonger, Vermogen en Inkomen, 1923
1927	36,825	23,490	18,357	14,073	Subcommisie, 1927

Note: All values in millions of nominal NLG. Columns 2–4 give our estimates, respectively using Historical National Accounts, the Estate Tax multiplier, and the Wealth Tax extrapolation. Column 5 notes historical estimates, and column 6 gives their source. For historical estimates that cover a range, we use our midpoint estimate for that year (e.g., the year 1881 for 1879–1883).

Figure 1.19: Wealth-Income Ratios per Method, with Semiprivate Wealth



Notes: Figure shows all three main methods to reconstruct aggregate household wealth, but with semiprivate wealth added to the estimates from the death duties and the lognormal extrapolation.

Chapter 2

Top Wealth Is Distributed Weibull, Not Pareto

This chapter is based on Teulings and Toussaint (2023).

2.1. Introduction

Right-skewed distributions pervade many aspects of economic life (Gabaix 2009, 2016).[1] The standard instrument for analyzing these phenomena is the Pareto distribution, of which the characteristics are well known for a long time (Pareto 1896; van der Wijk 1939). This distribution has one parameter governing the fatness of its right tail, the *tail index* α^{-1}. The higher α, the more mass is concentrated at the extreme end of the distribution. However, there is increasing theoretical and empirical evidence that Pareto provides a poor fit to the data in many applications. Blanchet, Fournier, and Piketty (2022) show that the predictions of Pareto for mean wealth and income among the upper tail are heavily at odds with reality. In the city size literature, Eeckhout (2004, 2009) and Rossi-Hansberg and Wright (2007) have challenged the Pareto assumption, arguing for respectively lognormal and a non-specified log-concave distribution function. For firm size, Jones (2023) argues that even thin-tailed firm productivity distributions can give rise to exponential growth, given the combinatorial nature of endogenous growth. So far, however, Pareto remains the default assumption in theoretical and empirical analysis.

Most empirical research on the right tail uses the well known tool of the log rank regression: the relation between the log rank in a sample of the rich/cities/firms and log wealth (for the rich) or log size (for cities and firms) should be linear for a Pareto distribution, see Rosen and Resnick (1980) for an early application. The use of this tool, however, is unfortunate. The log rank is an order statistic. Its construction requires the researcher to order the data by the magnitude of the variable of interest. This introduces correlation between observations, which invalidates many standard techniques and causes the OLS coefficients to be biased. Gabaix and Ibragimov (2011) set out to correct the log rank regression for this bias. This complication, however, can easily be avoided by applying maximum likelihood estimation. If the null of a Pareto distribution is correct, the maximum likelihood estimator of the tail index is most efficient. Moreover, it is extremely simple: the mean of log wealth, see Hill (1975). However, since the predictions of the Pareto model for mean wealth are very bad, see our discussion of Blanchet, Fournier, and Piketty (2022) above, the maximum likelihood estimator has been discredited. Hence, Blanchet, Fournier, and Piketty (2022) give up the idea of characterizing the actual distribution by some simple functional form. They argue that top income and wealth shares in the

[1] For example, heavy-tailed distributions feature prominently in the distributions of city-size (e.g. Gabaix 1999, Eeckhout 2004,Rossi-Hansberg and Wright 2007), firm-size (Luttmer 2011; Autor et al. 2020; Jones 2023), CEO salaries (Gabaix and Landier 2008), income and wealth (e.g. Atkinson, Piketty, and Saez 2011; Vermeulen 2018), and on financial markets (e.g. Huisman, Koedijk, Kool, and Palm 2001).

United States and France exhibit a tail index *curve*, with one parameter for every bracket in the distribution rather than a single tail index parameter.

This chapter provides empirical support for a more positive view. We have two main contributions. First, we develop a simple set of tests for the null hypothesis of Paretianity. If a variable is Pareto, its log is exponential. Since all moments of the exponential distribution are well-defined – unlike for Pareto, where only moments up to α^{-1} are defined – working with logs has distinct advantages. Our test statistics work as follows. For exponential distributions, the mean is a maximum likelihood estimator of α. Furthermore, the kth higher moment equals k factorial times α to the kth power. Hence, we can scale higher moments by appropriate powers of the sample mean to obtain a ratio statistic $\mathcal{R}_k$ that is independent from α and is asymptotically equal to unity.

We apply two versions of this test (based on the second and third moment respectively) to the data from the *Forbes List of Billionaires* from 2001 to 2021. Our data unequivocally reject the Pareto null for all years and for all 18 sub-regions of the world that we distinguish in our data. We show that this rejection cannot be driven by measurement error, since measurement error scales up the density proportionately. In addition, we derive the small-sample properties of $\mathcal{R}_k$ and show that small-sample bias cannot account for these results. Remarkably, both test statistics have very similar values for all years and all sub-regions, being consistently less than unity, the more so for the test statistic based on the third rather than the second moment. This suggests some systematic pattern in the data not covered by the Pareto prior. In particular, the data seem to be less skewed to the right. These findings foster the hope that there might be a simple alternative for Pareto that describes the data well.

Our second contribution is to develop a suitable alternative. We hypothesize that the actual distribution is not Pareto but *truncated-Weibull*. While the Pareto distribution is invariant to the choice of the lower bound, the Weibull distribution is not. The research therefore has to estimate the truncation point of the part of the distribution that is fitted to the data. Where the log of a Pareto variable is distributed exponentially, the log of a Weibull variable follows a *Gompertz* or counter-Gumbel distribution.[2] Like for Weibull, we use the truncated rather than the full Gompertz distribution; for the sake of brevity, we drop the word "truncated" from now on. The Gompertz distribution is not often used in economics, but it is a standard tool for modelling life expectancy in demography. The practical difference between Pareto and Weibull can be understood most easily by analysing the distributions

[2] When $\underline{W}$ is Weibull, $\ln \underline{W}$ is distributed Gompertz and $-\ln \underline{W}$ is distributed Gumbel. There is some confusion in the literature about the definition of the Gompertz distribution, where Gompertz is sometimes defined as Gumbel. See our discussion in footnote 9.

of their logs, that is: exponential and Gompertz respectively. The exponential distribution is characterized by a constant hazard rate, equal to α^{-1}, implying that for every unit increase in a particular lower bound of log wealth, the number of people that is even richer than that lower bound decreases by a constant percentage. The Gompertz distribution, instead, is characterized by not a constant but an increasing hazard above some lower bound, equal to $\alpha^{-1}e^{\gamma\omega}$, where γ is a positive parameter and ω is this lower bound: for every .01 increase in the lower bound, the hazard rises by γ%. As $\gamma \to 0$, we recover the exponential distribution with a constant hazard rate.

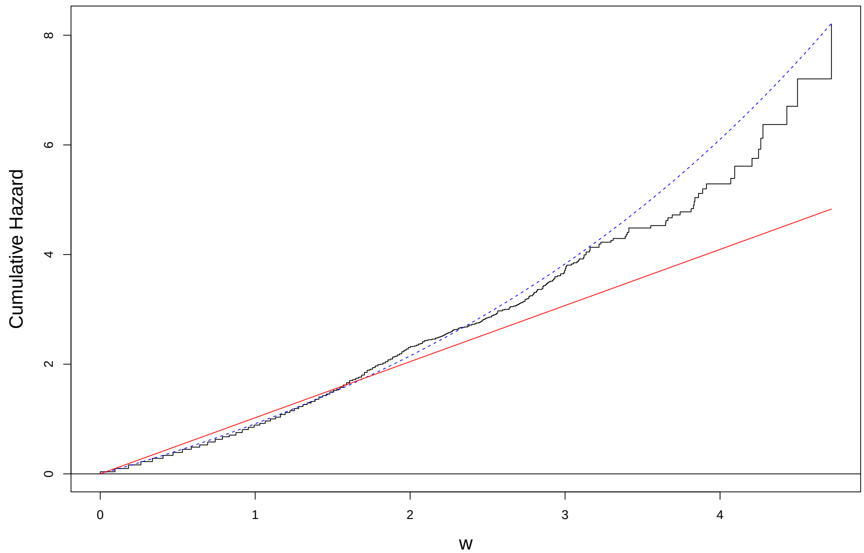
The hazard rate of the log-transform can be directly computed from the data and compared to the parametric hazards of some candidate distributions. Figure 2.1 demonstrates this for the global distribution of billionaires in 2018. The black line is the empirical hazard, estimated with a simple Kaplan-Meier estimator. The colored lines represent parametric cumulative hazards for the log-transform for the exponential (red) and Gompertz (blue) distribution. The exponential distribution has a constant hazard α^{-1} and hence a linearly increasing cumulative hazard $\alpha^{-1}w$. The prediction of a constantly increasing cumulative hazard of the exponential distribution is strongly at odds with the data. Instead, the exponentially increasing (cumulative) hazard of Gompertz fits the data well.

Figure 2.1 also reveals why Weibull has gone hitherto unnoticed: the divergence between exponential and Gompertz only occurs far in the upper tail. Visual inference based on log-log plots is therefore prone to miss these departures from Paretianity; in contrast, our test statistics $\mathcal{R}_k$ and the hazard-based plots provide sharp evidence.

Whenever Pareto has previously been challenged, the response has been that one has to go further into the right to arrive at the exponential tail (the log of Pareto), where the hazard rate would converge to a constant. Figure 2.1 shows that this response is mistaken. The hazard rate never converges but continues to diverge to infinity for higher levels of log wealth. In fact, the hazard rate of the Gompertz distribution (the log of Weibull) diverges even faster than that of the Normal distribution (the log of lognormal): the Gompertz hazard diverges exponentially, while the hazard rate of the Normal distribution diverges only linearly asymptotically.

We estimate the parameters of the Weibull distribution by maximum likelihood and find a rather stable value of $\gamma \approx 0.25$. Weibull has another advantage over Pareto: all its moments exist for all values of the parameters. We use our model to predict mean wealth for all sub-regions, using a common value of γ for all subregions. Where Pareto fails miserably, the Weibull distribution provides an almost perfect fit.

Figure 2.1: Empirical and Parametric Hazard Rates, Billionaire Log Wealth Distribution



Notes: Figure plots the Kaplan-Meier estimator of the empirical cumulative hazard of the billionaire log wealth distribution, using the 2018 *Forbes List of Billionaires* and pooling all observations. The blue dotted line is a fitted Gompertz hazard, and solid red is an exponential hazard.

We provide further evidence based on overidentifying restrictions. Our test statistics should equal 1 if wealth were Pareto. In contrast, if wealth is Weibull, $\mathcal{R}_k$ is a declining function of the product $\alpha\gamma$. Since the statistics only depend on the product of the parameters and not their separate values, we have an overidentification test: Compute the theoretical value of $\mathcal{R}_2$ and $\mathcal{R}_3$ for each observation, based on their estimated α and γ, and compare these with their actual test statistics. The theoretical values match the empirical test statistics extremely closely, which is strong evidence in favor of Weibull.

Our results raise the question whether they are typical for the Forbes data on billionaires' wealth or for the wealth distribution, or that they apply more generally to phenomena that have been presumed to be Paretian. We show evidence for two other variables: city and firm size. The shape of the U.S. city size distribution is subject of a long-standing debate. The default of Pareto (specifically Zipf) is assumed by, among others, Krugman (1996) and Gabaix (1999), who use the 135 largest Metropolitan Statistical Areas. Eeckhout (2004, 2009) challenges this result. We find that even for the upper tail, Pareto is strongly rejected by our test-statistics and that there is solid support for Weibull. The same applies to the U.S. firm size distribution. Again, this distribution is typically assumed to be Pareto (e.g., Axtell 2001; Luttmer 2011). We use data gathered by Kwon, Ma, and Zimmermann (2024). They use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022) to go from tabulations to a full distribution. This tool can also be used to create synthetic samples of the top 1% of firms. Again, we find overwhelming evidence against Pareto and in favor of Weibull.

Weibull can easily replace Pareto in many economic models. We discuss some avenues for further research. For example, the Gompertz distribution figures in stochastic networks. The length of a *self-avoiding walk* along a graph (that is, once a node is visited it cannot be revisited), is distributed Gompertz (Tishby, Biham, and Katzav 2016). This model structure has interesting parallels with the variables we study; for instance, city size is determined by area already used by existing cities.

Related Literature: This chapter is related to three strands of literature. First, our chapter adds to the literature on tail index estimation. This literature, which arose as a consequence of the development of Extreme Value Theory (e.g., Gumbel 1958; Balkema and de Haan 1974; Pickands 1975), is highly multidisciplinary and extensive in scope, and we will not be able to do it justice here; see Bingham, Goldie, and Teugels (1989) and Beirlant, Goegebeur, Segers, and Teugels (2006). Within the tail index estimation literature – reviewed in Fedotenkov (2020) – our work relates most closely to estimation on the basis of moments of the log-density function (Hill 1975; Dekkers, Einmahl, and de Haan 1989).

Second, we contribute to the study of heavy-tailed distributions within the field of economics. Examples of heavy-tailed distributions abound (Gabaix 2009, 2016), ranging from the distributions of income and wealth (Atkinson, Piketty, and Saez 2011; Vermeulen 2018), city size (Gabaix 1999), firm size (Luttmer 2011; Autor et al. 2020), and the cross-section of stock returns (Huisman, Koedijk, Kool, and Palm 2001). A lot of theoretical literature, however, simply assumes the Pareto distribution without properly testing this assumption, or looks at a log rank – log size plot to conclude that a particular distribution looks Pareto-ish. Within this literature, our work relates to studies which question the Pareto assumption. Examples include Eeckhout (2004) and Rossi-Hansberg and Wright (2007) for city size, Jones (2023) and Kondo, Lewis, and Stella (2023) for firm size, and Blanchet, Fournier, and Piketty (2022) for income and wealth. Our work also relates to papers which use the heavy-tailedness of a distribution as a basis to construct causal estimators of macroeconomic aggregates, such as Gabaix and Koijen (2024).

Finally, this chapter relates to economic studies on the dynamics of size distributions. A theoretical basis for studying the dynamics of top income and wealth is given by Gabaix, Lasry, Lions, and Moll (2016), who show that the Bewley-Aiyagari-Huggett random-growth models of wealth accumulation (Benhabib and Bisin 2018) fail to deliver Pareto tails with the same speed as observed in the data. Their framework has spurred much theoretical and empirical work. Similarly, Luttmer (2011) shows that only deviations from Gibrat's Law (i.e., firm growth is not independent of firm size) can generate a Pareto-shaped distribution of firms within a reasonable timeframe. Perhaps closest to our focus on billionaires is Gomez (2023), who uses tools from stochastic calculus to decompose the growth of the American Forbes 400 wealth share into growth by incumbents, growth by new entrants, and entry/exit effects; and Blanchet (2022), who also uses empirical data on income and wealth to study stochastic properties of economic models. In Chapter 3, we use our Weibull framework to study the dynamics of billionaire numbers since 2000, finding that a simple model can account for most of the time-series and cross-sectional variation (Teulings and Toussaint 2024). Since the number of billionaires is more sensitive to variation in the lower bound under Weibull, an increase in wealth in a given region (and hence a decline in the effective lower bound) will lead to a larger increase in billionaire numbers than predicted by Pareto.

Chapter Outline: The rest of this chapter is structured as follows. Section 2.2 presents our Pareto framework. In Section 2.3, we discuss our main data, the *Forbes List of Billionaires*. Section 2.4 present the empirical results for the test of the Pareto assumption. Section 2.5 develops the Weibull distribution. In Section 2.6, we test this distribution empirically. In Section 2.7, we further apply our framework to cities and firms. In Section 2.8, we discuss some theoretical implications of our results. Section 2.9 concludes.

2.2. Pareto: Theory and Estimation

2.2.1. Theory

We consider a random variable $\underline{X} \geq \Omega$, where $\Omega > 0$ is a parameter (stochastic variables will be underlined). We assume Ω to be known. For example, when using data from the *Forbes List of Billionaires*, Ω is a billion USD.[3] The complementary

[3] If the lower bound is unknown, one can use the lowest observation on $\underline{X}$ in the data as an estimate for Ω and drop this observation afterwards.

distribution function of the Pareto distribution and its moments are given by:

$$\Pr\left[\underline{X} \geq X \mid X \geq \Omega\right] = (X/\Omega)^{-1/\alpha}, \tag{2.1}$$

$$\mathrm{E}\left[\underline{X}^k \mid \underline{X} \geq \Omega\right] = \frac{1}{1-\alpha k}\Omega^r, \quad 0 < k < \alpha^{-1}, \tag{2.2}$$

where $\alpha > 0$ is a parameter; α^{-1} is commonly referred to as the tail index or Pareto coefficient (Jones 2015).[4] If $\alpha = 1$, the Pareto distribution specializes to the Zipf distribution. Equation (2.1) is easy to interpret, since it predicts the probability of a "large" observation to scale with size like a power law. This simple rule, together with its apparent fit of the distribution of the right tail of many empirical phenomena, has contributed to the popularity of Pareto.

Equation (2.2) reveals a problem that comes with Pareto: its moments $\mathrm{E}\left[\underline{X}^k\right]$ for $k \geq \alpha^{-1}$ do not exist. Since α^{-1} is estimated to be between 1 and 2 in many economic applications[5], this implies that the variance does not exist, let alone higher moments. For many phenomena, even the expectation does not exist. This makes the interpretation of α problematic.

It is therefore more convenient to work with the log-transform of $\underline{X}$, as we shall do from now on. Define $\underline{W} := \underline{X}/\Omega$ as the variable of interest divided by its lower bound, or using lower cases for logs, $\underline{w} := \ln \underline{W} = \underline{x} - \omega$ where $\underline{x} := \ln \underline{X}$ and $\omega := \ln \Omega$. Since $\underline{X} \geq \Omega$, $\underline{W} \geq 1$ and hence $\underline{w} \geq 0$; the realisations w can be directly calculated from the data. The log transform of a Pareto distributed random variable is the exponentially distributed:

$$\Pr\left[\underline{w} \geq w\right] = \mathrm{e}^{-w/\alpha}, \tag{2.3}$$

$$\mathrm{E}\left[\underline{w}^k\right] = \alpha^k \Gamma(k+1) = \alpha^k k!, \quad r > 0, \tag{2.4}$$

where $\Gamma(\cdot)$ is the Gamma function, and where we imply that k has an integer value whenever we write $k!$. Where only a limited set of moments for $k < \alpha^{-1}$ exists for the Pareto distribution, all moments exist for the distribution of its log transform, making it particularly suitable for testing.

2.2.2. Estimation and Testing

As a diagnostic to test whether a distribution looks sufficiently Pareto-ish, the log-log (or Zipf) plot is often used. Take the log of equation (2.1), and equate

[4] It is also common to work with $\zeta = \alpha^{-1}$ (e.g., Gabaix 2009, 2016). Both notations have advantages and disadvantages. Our choice to work with α rather than its inverse is motivated by the fact that the maximum likelihood estimator of α is unbiased and hence that of α^{-1} is not.

[5] $\alpha^{-1} \approx 1.5$ for most wealth distributions (Vermeulen 2018), and $\alpha^{-1} = 1.05$ for the U.S. firm size distribution (Luttmer 2011), and is close to 1 for city sizes (Eeckhout 2004).

the complementary distribution function with the empirical distribution's order statistic ranking. Since

$$\text{plim } [\ln \text{Rank}[w]] = \ln \Pr\left[\underline{w} \geq w\right] = -\alpha^{-1} w,$$

we obtain the regression equation

$$\ln \text{Rank}[w] = \alpha^{-1} - \alpha^{-1} w + \varepsilon, \tag{2.5}$$

where ε is an error term. When plotted on a double-logarithmic scale, equation (2.5) predicts the data to form a straight line, with a slope given by $-\alpha^{-1}$. A plot reasonably close to a straight line is typically interpreted as evidence in favour of the Pareto distribution and the regression coefficient is interpreted as an estimate of α^{-1}. Clearly, this does not constitute a formal test of Paretianity; Eeckhout (2009) argues that visual inference based on the log-log plot is dangerous since the double-logarithmic scale distorts observations at the tails, causing many distributions to look almost linear. More importantly, the calculation of the left hand variable $\ln \text{Rank}[w]$ requires ordering the data. Hence, the error terms ε are not i.i.d., invalidating standard diagnostic tests for OLS. Futhermore, this procedure is known to introduce small-sample bias. Hence, the common solution, recommended by Gabaix and Ibragimov (2011), is to use $\ln \text{Rank}[w - 1/2]$ as the dependent variable.

The common alternative to the regression approach is to use the method of moments. Equation (2.4) for $k = 1$ implies that the mean $\overline{w}$ for the full sample of billionaires can be used as a method-of-moments estimator of α (a bar above a random variable denotes its sample mean).[6] In fact, $\overline{w}$ is the maximum likelihood estimator of α (Hill 1975).[7] Since $\widehat{\alpha}^{\text{Hill}}$ is also the maximum likelihood estimator, it is efficient. Since it is linear in $\overline{w}$, it is unbiased. These features make the moment approach preferable to the regression approach *if* the model is correctly specified. If the data-generating process is not Pareto, in contrast, maximum likelihood gives very bad results, in particular when used to predict expected wealth $\text{E}\left[\underline{W}\right] = (1 - \alpha)^{-1}$ for the case where α is close to unity, or even below unity is biased. However, a rank order regression is likely to yield the same problem.

Test Statistics for Pareto: Equation (2.4) can be used for the construction of

[6]The sample mean of a stochastic variable is a stochast itself, so we should write $\underline{\overline{w}}$. We save on notation by omitting the underbar.

[7]This is a consequence of the well-known fact that for distributions in the exponential family, moment estimators are equivalent to maximum-likelihood estimators (van der Vaart 2000).

simple statistics to test the null hypothesis that the distribution of $\underline{w}$ is indeed exponential (and by implication: the distribution of $\underline{W}$ is Pareto). Our proposed test statistics read:

$$\mathcal{R}_k := \frac{\mathrm{E}\left[\underline{w}^k\right]}{k!\mathrm{E}\left[\underline{w}\right]^k}, \quad \widehat{\mathcal{R}_k} := \frac{\overline{w^k}}{k!\overline{w}^k}. \tag{2.6}$$

Our statistics $\widehat{\mathcal{R}_k}$ can be interpreted as normalizations of higher non-central sample moments of $\underline{w}$ by corresponding powers of the sample mean. We refer to $\widehat{\mathcal{R}_k}$ for $k = 2$ and 3 to as the normalized variance and skewness, respectively. Appendix 2.A derives the following general formula for the expectation and the variance of $\widehat{\mathcal{R}_k}$:

Proposition 2.1. *Assume $\underline{w}$ is exponentially distributed. Then, for any integer $k \geq 1$, $\mathcal{R}_k$ defined in equation (2.6) satisfies:*

$$\mathcal{R}_k = 1,$$

while $\mathrm{E}\left[\widehat{\mathcal{R}_k}\right]$ *has expectation and variance*

$$\begin{aligned}
\mathrm{E}\left[\widehat{\mathcal{R}_k}\right] &= 1 + N^{-1}\left(\frac{(2k)!}{k!^2} - \frac{3k^2 - k + 2}{2}\right) + O\left(N^{-2}\right) \\
\mathrm{Var}\left[\widehat{\mathcal{R}_k}\right] &= N^{-1}\left(\frac{(2k)!}{k!^2} - k^2 - 1\right) + O\left(N^{-2}\right),
\end{aligned}$$

where N is the sample size.

Proof. See Appendix 2.A. □

$\widehat{\mathcal{R}_k}$ converges asymptotically to unity for all k. In this chapter, we specifically work with $\mathcal{R}_2$ and $\mathcal{R}_3$. Their expectation and variance read:

$$\begin{aligned}
\mathrm{E}\left[\widehat{\mathcal{R}_2}\right] &= 1 + O\left(N^{-2}\right) & \mathrm{Var}\left[\widehat{\mathcal{R}_2}\right] &= N^{-1} + O\left(N^{-2}\right) \\
\mathrm{E}\left[\widehat{\mathcal{R}_3}\right] &= 1 + 7N^{-1} + O\left(N^{-2}\right) & \mathrm{Var}\left[\widehat{\mathcal{R}_3}\right] &= 10N^{-1} + O\left(N^{-2}\right)
\end{aligned} \tag{2.7}$$

Apart from being unbiased up to terms of order N^{-2}, the variance of $\widehat{\mathcal{R}_2}$ is much smaller than that of $\widehat{\mathcal{R}_3}$. Hence, $\widehat{\mathcal{R}_2}$ is a more powerful test statistic than $\widehat{\mathcal{R}_3}$. Nevertheless, both $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ will prove to be useful for purpose of this chapter. Our test of Pareto boils down to a two-sided test for the value of $\widehat{\mathcal{R}_k}$ for $k \geq 2$, where we reject Pareto if $\widehat{\mathcal{R}_k}$ significantly differs from unity.

What about measurement error? The data on top wealth are likely to be subject to substantial measurement error. We now show that classical measurement error affects neither the tail index α^{-1} of the distribution of top wealth, nor our test statistics. To show this, let $\underline{v}$ be a normally distributed measurement error with zero mean and variance of σ^2 and Cov $[\underline{w}, \underline{v}] = 0$, so that observed log wealth $\widetilde{w}$ satisfies $\widetilde{w} := w + v$. The density function $f(\widetilde{w})$ of observed log wealth satisfies for large w (the upper tail)

$$f(\widetilde{w}) = \int_{-\infty}^{\infty} \sigma^{-1} e^{-(\widetilde{w}-v)/\alpha} \phi\left(\frac{v}{\sigma}\right) dv = e^{-\widetilde{w}/\alpha} \int_{-\infty}^{\infty} \sigma^{-1} e^{v/\alpha} \phi\left(\frac{v}{\sigma}\right) dv$$
$$= e^{(\sigma/\alpha)^2/2} e^{-\widetilde{w}/\alpha}, \quad (2.8)$$

where $\phi(\cdot)$ denotes the standard-normal density function. The tail index of actual and observed log wealth is therefore the same, though the density function of the latter is scaled up by a proportional constant $e^{(\sigma/\alpha)^2/2} > 1$.

Intuitively, measurement error causes some individuals' log wealth to be overreported and other individuals' wealth to be underreported. Focus on one particular level of observed log wealth $\widetilde{w}$ and on one particular level of the absolute value of the measurement error $|v|$. Since the density function of w is declining, there are more people with actual log wealth $w = \widetilde{w} - |v|$ whose wealth gets overreported to be $\widetilde{w}$, than there are people with log wealth $w = \widetilde{w} + |v|$ whose wealth gets underreported. Hence, the density of the right tail of the distribution of observed log wealth is increased by some proportional factor compared to the distribution of actual log wealth. However, since the ratio of underreporters to overreporters is the same for all $\widetilde{w}$ due to the exponential distribution of actual log wealth w, this constant of proportionality is the same for all $\widetilde{w}$. Hence, the moment ratios $\widehat{\mathcal{R}_k}$ remain unaffected. Measurement error does therefore not affect our test procedure.[8]

However, the common procedure of using rank-size regressions as in equation (2.5) *will* be biased under measurement error. First, mismeasured w causes the OLS estimate for α^{-1} to be attenuated toward zero. A subtler problem is that this method depends on order statistics, and in effect assumes these are known perfectly. In empirical applications, however, measurement error introduces uncertainty about precise rankings, making order statistics an object to be estimated rather than something known *a priori* (Mogstad, Romano, Shaikh, and Wilhelm 2024). This introduces additional uncertainty into the estimator.

[8]Note that this argument applies only to the extreme right tail of the distribution of wealth. As soon as $\widetilde{w} - |v|$ falls below the lower support of $\underline{w}$, the argument no longer applies. However, the literature typically assumes that the Pareto tail would start at a threshold much below a billion USD, perhaps at several million USD (e.g., Albers, Bartels, and Schularick (2022) assume it starts at the 99th percentile), so measurement error will not affect the distribution of billionaires' wealth.

2.3. Data

For our main application, we use the *Forbes List of Billionaires* for the years 2001–2021. The dataset provides the names of billionaires, net worth, country of origin, age and citizenship. We classify billionaires according to their citizenship.

Forbes calculates net worth at the individual level, but aggregates family wealth, unless each family member has USD 1 billion or more after the split. On the one hand, as observed by Piketty (2014), this is likely to create an upward bias on individual fortunes around the threshold. On the other hand, given the difficulty for *Forbes* to estimate wealth components that are not publicly observed, some fortunes may well be biased downward. Moreover, they use available documentation and sometimes data provided by billionaires themselves to estimate their net worth. The number of billionaires and their wealth is likely to be underestimated in less developed countries or for wealth derived from nefarious activities. *Ex ante*, it is unclear which direction the measurement error goes; "rounding up" to 1 billion may put too much weight on the bottom of the list, whereas the difficulty of measuring liabilities and other poorly observable wealth components may overstate wealth for fortunes at the top. We follow the existing literature which uses rich lists like Forbes, as other sources are likely to underestimate the number of billionaires (Vermeulen 2016; Novokmet, Piketty, and Zucman 2018; Piketty, Yang, and Zucman 2019; Gomez 2023).

We cluster countries first in regions and then separate out sub-regions from some of these regions. The guiding principle for this clustering is to merge countries that are geographically connected and close in terms of GDP per capita, and have sufficient billionaire numbers to make the estimation of our statistics precise. With this in mind we define nine regions, which together cover all countries. The sub-region classification subdivides regions in coherent geographical units consisting of one or a small number of countries. As a rough threshold for the minimum number of billionaires for a country or a group of countries to be considered as a sub-region we use 40 billionaires in 2019. Countries that cannot easily be included in a sub-region are excluded from the sub-region classification. Hence, where region-classification encompasses the whole world, the sub-region classification excludes some countries. Therefore, a region can have more billionaires than the sum of its constituent sub-regions. Equation (2.6) implies that, under the assumption of top wealth being Pareto, a minimum of 40 billionaires in 2019 yields a maximal variance for our estimate of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ of 0.025 and 0.25 respectively; likewise, the upward bias of $\widehat{\mathcal{R}_3}$ is at most 0.175. Since the total number of billionaires has gone up over time, the number of billlionaires and hence the precision of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ are lower for the early years in our data. The regions China and India consist of one country and are therefore also classified as sub-region. The region Rest of the World has fewer than 40 billionaires in 2019. Moreover, this region is rather heterogeneous, with some really poor countries in Sub-Saharan Africa, as well as Afghanistan and Bangladesh, but also some middle income countries like South Africa. We therefore exclude it from our analysis from now on. We end up with 18 sub-regions. We use this sub-region classification for our regressions. Table 2.1 gives an overview of these regions and sub-regions.

Table 2.1 provides summary statistics for all regions and sub-regions. We report average numbers for $\widehat{\mathcal{R}_2}$, $\widehat{\mathcal{R}_3}$, and $\overline{w}$, and the average number of billionaires N (both raw and normalized by total population in millions). The summary statistics immediately reveal some striking facts. First, mean log wealth in Europe exceeds unity, implying that $\mathrm{E}\left[\underline{W}\right]$ does not exist for the Pareto distribution. Second, both $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ are consistently smaller than one in all regions and sub-regions, contradicting the prediction of a Pareto distribution. These estimates might be affected by a small sample bias, but equation (2.7) shows the bias for $\widehat{\mathcal{R}_2}$ to be of order $O\left(N^{-2}\right)$ only and for $\widehat{\mathcal{R}_3}$ to be upward rather than downward, making the low reported values for both statistics even more striking. Given the small number of billionaires in many of these sub-region/year observations (in particular for early years), it might be hard to reject the null $\widehat{\mathcal{R}_k} = 1$ for an individual observation. However, since we have multiple observations, we can ask: what is the likelihood that we observe this *distribution* of statistics? This we do in the next subsection.

Table 2.1: Region Classification & Descriptive Statistics

Region Classification		Statistics (Average 2001–2021)				
(Sub–)Region	Area	$\widehat{\mathcal{R}_2}$	$\widehat{\mathcal{R}_3}$	$\overline{w}$	N/L	N
North America	US and Canada	**0.866**	**0.688**	**0.92**	**1.35**	**472**
– U.S.		0.87	0.697	0.923	1.41	443
– Canada		0.797	0.564	0.894	0.832	29.2
Europe	excl. former USSR but incl. Baltics	**0.781**	**0.516**	**1.02**	**0.595**	**264**
– Germany		0.722	0.43	1.12	0.901	74.1
– British Islands	U.K. + Ireland	0.793	0.525	0.828	0.589	40.6
– Scandinavia	Sweden + Denmark + Norway + Finland	0.759	0.461	1.17	1.17	30.6
– France	incl. Monaco	0.779	0.502	1.27	0.402	26.5
– Alps	Switzerland + Austria + Liechtenstein	0.658	0.342	1.05	1.49	25
– Italy		0.792	0.535	1.01	0.405	24.1
China	excl. Taiwan, incl. Hong Kong	**0.929**	**0.771**	**0.794**	**0.135**	**188**
East Asia	Asia East of India and South-East of China; incl. Australia	**0.799**	**0.533**	**0.798**	**0.175**	**135**
– Southeast Asia	Thailand + Malaysia + Singapore	0.741	0.447	0.923	0.3	31.1
– Asian Islands	Taiwan + Philippines + Indonesia	0.766	0.49	0.727	0.101	38.5
– South Korea		0.843	0.63	0.673	0.368	18.7
– Japan		0.819	0.562	0.875	0.209	26.6
– Australia		0.81	0.537	0.736	0.802	19
India		**0.824**	**0.582**	**0.964**	**0.043**	**56.1**
Central Eurasia	former USSR except Baltics	**0.851**	**0.61**	**0.919**	**0.371**	**78**
– Russia		0.843	0.597	0.956	0.478	68.8
South America	incl. middle America and Mexico	**0.821**	**0.597**	**0.964**	**0.0978**	**59.8**
– Brazil		0.809	0.56	0.861	0.148	30
Middle East	Middle East incl. Turkey and Egypt excl. Iran	**0.858**	**0.635**	**0.765**	**0.568**	**60.4**
– Israel + Turkey		0.891	0.652	0.585	0.43	36.2
Rest of World	mainly Africa excl. Egypt, incl. Iran, Afghanistan, Pakistan, Bangladesh	**0.743**	**0.436**	**0.961**	**0.00556**	**11.8**
World		**0.856**	**0.647**	**0.896**	**0.183**	**1325**

Notes: Regions may include billionaires from countries not part of their constituent sub-regions. China and India count both as regions and sub-regions. N = total number of billionaires; L is total population in millions; $\overline{w}$ = mean log wealth; $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ are the normalized variance and skewness of mean log wealth from equation (2.6).

2.4. Testing the Model

This section formally tests the Pareto assumption discussed in Section 2.2. The results are reported in Table 2.2. We have 18 sub-regions × 21 years = 378 observations available for estimation. From equation (2.7), the variance of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ is predicted to be proportional to N^{-1}, the inverse number of billionaires in a sub-region. Hence, the model exhibits heteroskedasticity and OLS is inefficient. We correct for this by using weighted least squares (WLS) with $\sqrt{N}$ as weights.

Table 2.2: WLS Regressions, Pareto Test

Dependent Variables:	$\widehat{\mathcal{R}_2}$			$\widehat{\mathcal{R}_3}$		
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.82***	0.86***	0.86***	0.59***	0.66***	0.65***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics						
Observations	378	75	67	378	75	67
RMSE	0.274	0.309	0.270	0.451	0.540	0.467
Theoretical RMSE	1	1		$\sqrt{10}$	$\sqrt{10}$	

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

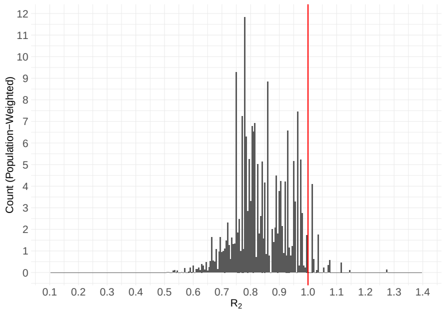
Notes: Table reports regressions using the square root number of billionaires in a sub-region–year observation as weights. Parentheses underneath point estimates denote standard errors clustered by sub-region and year. Specifications (2) and (5) drop sub-region–year observations with fewer than 64 billionaires; specifications (3) and (6) further drop observations in the top and bottom 5% of the dependent variable. The theoretical root mean squared error is derived from Equation (2.7).

We clearly reject the null that $\widehat{\mathcal{R}_2} = \widehat{\mathcal{R}_3} = 1$. The intercepts are highly significantly different from unity. This remains the case if we do the following robustness checks: drop observations with fewer than 64 billionaires to reduce the small-sample bias in $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ (column (2) and (5)) and further drop the bottom and top 5% of observations for $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ as a robust regression (column (3) and (6)). We also report the theoretical root mean squared error (RMSE) derived from equation (2.7), except for column (3) and (6) where the selective dropping of observations invalidates the prediction for the RMSE. If all variation in $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ were sampling variation, the observed RMSE should be equal to the theoretical

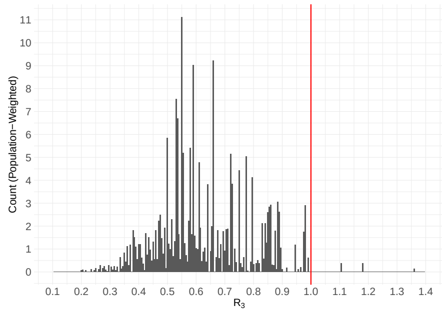
RMSE. This is clearly not the case. Using the sample with observations with $N \geq 64$, the observed RMSE is 31% of the theoretical RMSE for $\widehat{\mathcal{R}_2}$ and $0.54/\sqrt{10} \approx 17\%$ for $\widehat{\mathcal{R}_3}$. This is a further indication that the Pareto distribution does not fit the data.

Figure 2.2 shows $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ for all observations: the great majority is smaller than unity. This holds true when we restrict attention to observations with more than 64 billionaires and (obviously) when we further drop the top and bottom 5% of observations. The main takeaway from Table 2.2 and Figure 2.2 is that Pareto distribution is strongly rejected across all sub-regions and years.

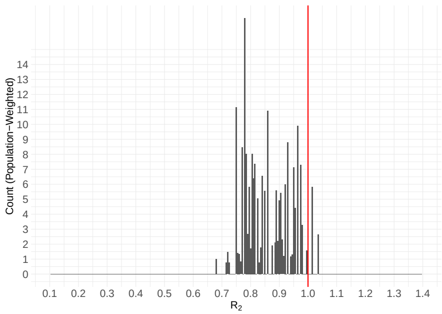
Figure 2.2: Distribution of Test Statistics $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$



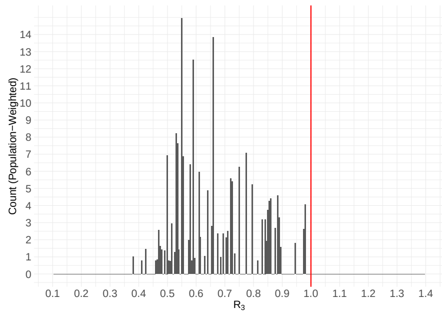
(a) $\widehat{\mathcal{R}_2}$, Full Sample



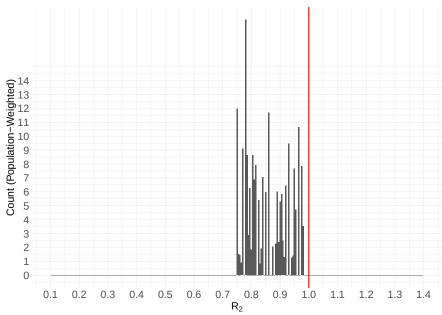
(b) $\widehat{\mathcal{R}_3}$, Full Sample



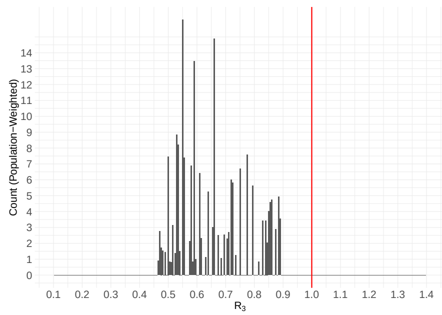
(c) $\widehat{\mathcal{R}_2}$, $N \geq 64$



(d) $\widehat{\mathcal{R}_3}$, $N \geq 64$



(e) $\widehat{\mathcal{R}_2}$, $N \geq 64$, trimmed



(f) $\widehat{\mathcal{R}_3}$, $N \geq 64$, trimmed

Notes: Figures show the distribution of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$, with all observation-years pooled. The red line indicates the predicted value under the null that wealth is Pareto-distributed, i.e., $\widehat{\mathcal{R}_2} = \widehat{\mathcal{R}_3} = 1$. All counts are weighted by the number of billionaires. Panels (a) and (b) use the full sample, panels (c) and (d) drop observations with fewer than 64 billionaires, and (e) and (f) further drop the top and bottom 5% of observations from (c) and (d).

2.5. The Alternative: Weibull

2.5.1. Theory

We have documented that our test statistics $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ strongly reject the null of a Pareto distribution. Moreover, the value of these statistics are very similar across observations, $\widehat{\mathcal{R}_2}$ being strongly clustered around 0.85 and and $\widehat{\mathcal{R}_3}$ around 0.65, see Table 2.1 and Figure 2.2. This suggests that there is a common pattern in the deviation from Paretianety and fosters the hope that our results not merely reject the hypothesis of a Pareto distribution but that there is an alternative distribution that describes the data more accurately.

We claim that the Weibull distribution gives a better description of top wealth than Pareto. Where the log of a Pareto distributed variable follows an exponential distribution, the log of a Weibull distributed variable follows a Gompertz distribution.[9] Since our interest regards the distribution of top-wealth, we focus on the right tail of both distributions, the part above a lower bound $\underline{W} \geq 1$ in levels or $\underline{w} \geq 0$ in logs. Hence, it would be more precise to refer to the lower truncated Weibull and Gompertz distributions. In the interest of parsimony, we omit this further qualification in the subsequent discussion. Like it was more convenient to work with the exponential distribution for log wealth $\underline{w} \geq 0$ rather than the Pareto distribution for wealth $\underline{W} \geq 1$, it is more convenient to work with the Gompertz distribution for $\underline{w}$ rather than the Weibull distribution for $\underline{W}$.

The Gompertz distribution for $\underline{w} \geq 0$ reads:

$$\Pr\left[\underline{w} \geq w\right] = \exp\left(-\frac{e^{\gamma w} - 1}{\alpha\gamma}\right), \tag{2.9}$$

where $\gamma > 0$ is a parameter. Since $\lim_{\gamma\to 0} \frac{e^{\gamma w}-1}{\alpha\gamma} = w/\alpha$, we recover the exponential distribution for $\gamma \to 0$, compare equation (2.3).

The difference between the exponential and the Gompertz distribution is understood most easily by comparing their hazard rates

$$H(w) := \Pr\left[\underline{w} = w\right] / \Pr\left[\underline{w} \geq w\right].$$

[9] The Gompertz distribution is identical to the counter-Gumbel distribution: if $\underline{w}$ is Gompertz, $-\underline{w}$ is Gumbel. There is some confusion in the literature about the definition of the Gompertz distribution, where Gompertz is sometimes defined as Gumbel; see Kleiber and Kotz (2003), following Ahuja and Nash (1967). If $\underline{w}$ is Gumbel, $\underline{W}$ is inverse-Weibull or Frèchet. The confusion is permeated on Wikipedia, which gives our definition for Gompertz but then claims that the exponent of Gompertz is inverse-Weibull, citing Kleiber and Kotz (2003). We follow the definition used in demography, which implies $\underline{W}$ to be (truncated-)Weibull. We thank Christian Kleiber for helpful conversations on this point.

Table 2.3 lists the hazard rates of four distributions: next to the exponential and Gompertz distribution, we also list the Weibull and Normal distribution. All four distributions are parameterized such that the hazard rate for $w = 0$ is equal to α^{-1}. While the hazard is constant for the exponential distribution, it is increasing for Gompertz. The latter hazard diverges to infinity for $w \to \infty$. For every unit increase ($\mathrm{d}w$) in wealth, the relative increase of the hazard is $\gamma \mathrm{d}w$. Conditional on being part of the group billionaires with a log wealth larger than or equal to w, the probability of leaving the group of billionaires with an even higher wealth than w is increasing in w, reducing the probability of finding a billionaire with even higher wealth. The Gompertz distribution therefore has a thinner tail than the exponential distribution.

Table 2.3: Hazard Rates for Four Distributions

Distribution	Complement Distribution $\Pr\left[\underline{w} \geq w\right]$	Density $\Pr\left[\underline{w} = w\right]$	Hazard Rate $H(w)$
Exponential	$\mathrm{e}^{-w/\alpha}$	$\alpha^{-1}\mathrm{e}^{-w/\alpha}$	α^{-1}
Gompertz	$\exp\left(-\frac{\mathrm{e}^{\gamma w}-1}{\alpha\gamma}\right)$	$\alpha^{-1}\exp\left(\gamma w - \frac{\mathrm{e}^{\gamma w}-1}{\alpha\gamma}\right)$	$\alpha^{-1}\mathrm{e}^{\gamma w}$
Weibull	$\exp\left(-\frac{(1+w)^{\gamma+1}-1}{\alpha(\gamma+1)}\right)$	$\alpha^{-1}(1+w)^{\gamma}\Pr\left[\underline{w} \geq w\right]$	$\alpha^{-1}(1+w)^{\gamma}$
Normal	$\Phi\left(-\frac{1+w}{\alpha\psi_\alpha}\right)$	$\left(\alpha\psi_\alpha\right)^{-1}\phi\left(\frac{1+w}{\alpha\psi_\alpha}\right)$	$\left(\alpha\psi_\alpha\right)^{-1}\frac{\phi\left(\frac{1+w}{\alpha\psi_\alpha}\right)}{\Phi\left(-\frac{1+w}{\alpha\psi_\alpha}\right)}$

Notes: $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution and density functions respectively of the standard normal distribution; ψ_α solves $\psi_\alpha\Phi\left(-\left(\alpha\psi_\alpha\right)^{-1}\right) = \phi\left(\left(\alpha\psi_\alpha\right)^{-1}\right)$; $\psi_1 = 1.330$.

As discussed, the distributions are parameterized such that the hazard rate is equal to α^{-1} for the lower bound $\underline{w} = 0$, or equivalently, $\underline{x} = \omega$. Since the hazard is constant for the exponential distribution, the choice of the value of ω does not matter for the value of α. For Gompertz, as for the other two distributions, the hazard rate varies with log wealth relative to its lower bound $\underline{w} := \underline{x} - \omega$. Hence, the choice of the lower bound ω matters for the value of α. To see this, presume the distribution of top log wealth $\underline{x}$ to be Gompertz. Let α_0 be the value of α when we set $\omega = 0$, so $\underline{w} = \underline{x}$. Now consider what happens when we decide to study the distribution of the wealth of bi-billionaires instead of billionaires, so $\omega = \ln 2$ instead of $\omega = \ln 1 = 0$. Then, the hazard rate at the lower bound would equal $\alpha_0^{-1}\mathrm{e}^{\gamma \ln 2}$ rather than α_0^{-1}. Hence, a general expression for the relation between α

and ω reads

$$\ln \alpha_\omega = \ln \alpha_0 - \gamma\omega. \tag{2.10}$$

While γ is a structural parameter, the value of the parameter α depends on the lower bound of our data.

Generalizing this argument, one might hypothesize that the sub-region/time observations of the distribution of billionaires follow a Gompertz distribution with a common γ parameter. But then α must differ between these observations, since one has to dive much deeper into the right tail to encounter a billionaire, in 2001 rather than 2021 (just due to growth and inflation), or for that matter, in India rather than the U.S. (with 0.043 and 1.41 billionaires per million respectively, see Table 2.1). The parameter α can therefore be expected to be lower in 2001 rather than 2021 and lower in India rather than the United States.

For the sake of comparison, we also show the hazard rates of the Weibull and Normal distribution in Table 2.3.[10] Similar to Gompertz, the hazards of both distributions are increasing and diverge to infinity for $w \to \infty$, however, at a slower rate. For the Weibull distribution, this can be seen directly from the hazard rate, since while $e^{\gamma w}$ and $(1+w)^\gamma$ are equal for $w = 0$, $e^{\gamma w}$ exceeds $(1+w)^\gamma$ for any $w > 0$ and $e^{\gamma w}/(1+w)^\gamma$ diverges to infinity for $w \to 0$ for any $\gamma > 0$.

For the Normal distribution, this is more difficult to see since there is no closed form solution for the hazard rate. Figure 2.3 plots the hazard rates for $\alpha = 1$ and for $\gamma = 1/2, 1/3$, and $1/4$ (the latter for Gompertz only).[11] For $\gamma > 1/2$, the hazard increases more quickly for the Gompertz rather than the Normal distribution. For $\gamma < 1/3$, the hazard increase more rapidly for Normal initially, but eventually the hazard for Gompertz will dominate for any $\gamma > 0$.[12] The Gompertz distribution therefore has an even thinner extreme right tail than the Normal distribution. The fact that this feature shows up only in the extreme right tail for $\gamma < 1/2$ might explain why our empirical claim that top log wealth is Gompertz rather than exponential has gone unnoticed in the empirical literature so far.

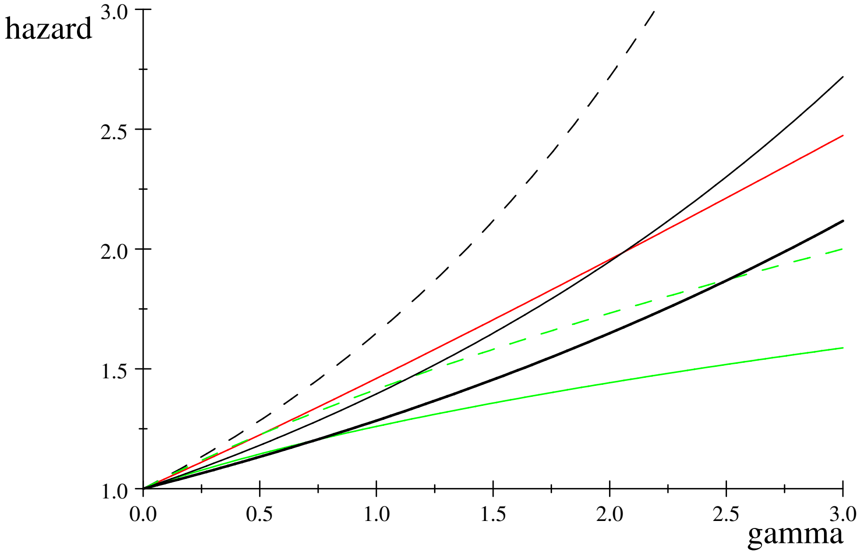
Moments of Weibull and Gompertz: We gather our results on the moments of Weibull and Gompertz in the following proposition, compare Wingo (1989).

[10] When wealth follows a Weibull distribution, log wealth follows a Gompertz distribution. In Table 2.3, we analyze the case where *log wealth* rather than *wealth* follows a Weibull distribution. Similarly, when log wealth follows a Normal distribution, the distribution of wealth is log Normal .

[11] For the Normal distribution, we must introduce an additional parameter ψ_α that is a function of α that achieves that $H(0) = \alpha^{-1}$, see the footnote to Table 2.3 for details.

[12] This follows from the fact that the hazard of the Normal distribution converges to an increasing linear function since $\lim_{x\to\infty} \frac{\phi(x)}{x\Phi(-x)} = 1$. An increasing exponential function will eventually always dominate an increasing linear function.

Figure 2.3: Variation of Hazard Rates with γ



Notes: Black lines = Gompertz, green = Weibull, red = Normal; dashed lines: $\gamma = 1/2$, continuous lines: $\gamma = 1/3$, fat line: $\gamma = 1/4$ (Gompertz only).

Proposition 2.2. *Assume $\underline{W}$ is Weibull distributed and hence $\underline{w}$ is Gompertz with the complement distribution function given in equation (2.9). Then, the moments for $k > 0$ of $\underline{W}$ and $\underline{w}$ read*

$$\mathrm{E}\left[\underline{W}^k\right] = (\alpha\gamma)^{k/\gamma} \mathrm{e}^{(\alpha\gamma)^{-1}} \Gamma\left(1 + k/\gamma, (\alpha\gamma)^{-1}\right), \tag{2.11}$$

$$\mathrm{E}\left[\underline{w}^k\right] = \gamma^{-k} h(\alpha\gamma, k), \tag{2.12}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function and where

$$h(\alpha\gamma, k) := k\mathrm{e}^{(\alpha\gamma)^{-1}} \int_0^\infty q^{k-1} \exp\left(-(\alpha\gamma)^{-1} \mathrm{e}^q\right) \mathrm{d}q,$$
$$h(\alpha\gamma, 1) = \mathrm{e}^{(\alpha\gamma)^{-1}} \mathrm{Ei}\left((\alpha\gamma)^{-1}\right),$$

where $q := \gamma w$ and Ei $(\cdot)$ *is the exponential integral.*

Proof. See Appendix 2.B. □

While the moments of Gompertz are not available in closed form, the mean can be easily computed with routines available in all common statistical software packages and higher moments can be calculated by simple numerical integration. Recall that moments of $\underline{W}$ higher than α^{-1} do not exist when $\underline{w}$ is exponential and hence $\underline{W}$ is Pareto. In contrast, when $\underline{w}$ is Gompertz and hence $\underline{W}$ is Weibull, all moments of $\underline{W}$ exist.

We can use equations (2.11) and (2.12) for the calculation of the asymptotic expection of the $\widehat{\mathcal{R}_k}$-statistics as defined in equation (2.3), which are a function $\mathcal{R}_k\left(\alpha\gamma\right)$ of the parameter product $\alpha\gamma$ for the Gompertz distribution:

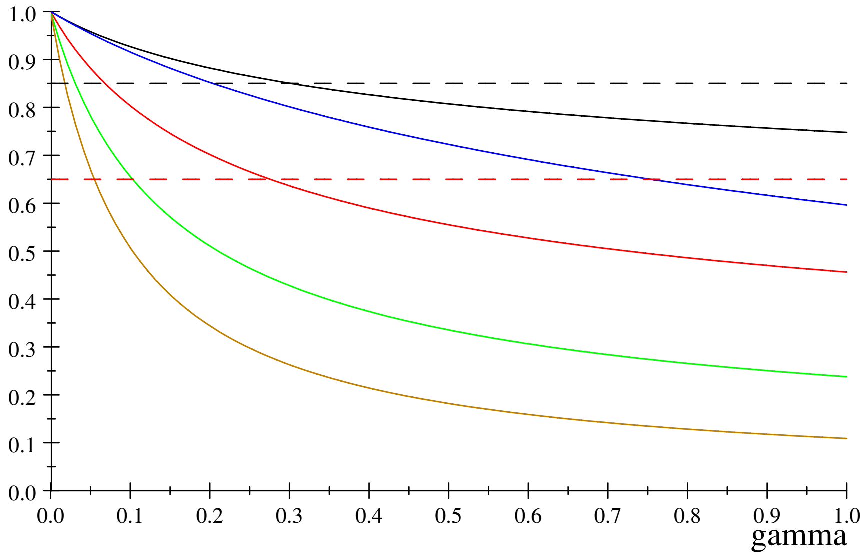
$$\mathcal{R}_k\left(\alpha\gamma\right) = \frac{h\left(\alpha\gamma, k\right)}{k!h\left(\alpha\gamma, 1\right)^k} = \frac{\int_0^\infty q^{k-1}\exp\left(-\left(\alpha\gamma\right)^{-1}\mathrm{e}^q\right)\mathrm{d}q}{(k-1)!\mathrm{e}^{(\alpha\gamma)^{k-1}}\mathrm{Ei}\left(\left(\alpha\gamma\right)^{-1}\right)^k}. \tag{2.13}$$

Figure 2.4 plots the values of E $\left[\underline{w}\right]$ (blue) and $\mathcal{R}_k(\alpha\gamma)$ for k from 2 until 5 (black, red, green, and beige respectively). For $\gamma \rightarrow 0$ and hence $\alpha\gamma \rightarrow 0$, the Gompertz distribution converges to the exponential distribution. Indeed, E $\left[\underline{w}\right]$ and $\mathcal{R}_k(\gamma)$ are equal to unity in this case, the same as for the exponential distribution. For $\alpha\gamma > 0$, $\mathcal{R}_k(\alpha\gamma) < 1$ and $\mathcal{R}_k(\alpha\gamma) > \mathcal{R}_{k+1}(\alpha\gamma)$. We also plot the mean value of $\widehat{\mathcal{R}_2} \approx 0.85$ and $\widehat{\mathcal{R}_3} \approx 0.65$ from Table 2.1. Figure 2.4 documents that these values are both consistent with a value of $\alpha\gamma$ between 0.25 and 0.30. Note that these values depend on the *product* of α and γ. Hence, the availability of data on both $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ provides over-identifying restrictions that we can use to test the null hypothesis of top log wealth beign Gompertz. The consistency of both estimates for $\alpha\gamma$ suggest that the null hypothesis fits the data well. We return to this issue in Section 2.6.

While the hazard rate will continue to diverge to infinity irrespective how far one dives into the right tail of the distribution (that is: for an arbitrarily high lower bound ω) and therefore never converges to a constant, as predicted by the Pareto distribution, $\mathcal{R}_k\left(\alpha\gamma\right)$ does converge to its unit value obtained under Pareto. An ever higher lower bound ω implies an ever lower value of α, see equation (2.10), and $\lim_{\alpha\rightarrow 0}\mathcal{R}_k\left(\alpha\gamma\right) = 1$, the value of $\mathcal{R}_k$ under Pareto. The reason for this apparent contradiction is that the increasing values of the hazard rate become increasingly irrelevant as the initial value of the hazard rate is already high, so that by time the hazard rate has gone up noticeably, most of the remaining rich have already left the pool of even richer people. While there is no convergence in the extreme right tail of Weibull to Pareto for hazard rates, there is therefore convergence in the truncated moments.

Figure 2.5 plots the expected level of wealth E $\left[\underline{W}\right]$ as a function of γ, for $\alpha = 0.80$ (red) and $\alpha = 1$ (black). Since log wealth is Gompertz, the level of wealth

Figure 2.4: Variation of Mean Log Wealth and $\mathcal{R}_k(\gamma)$ with γ



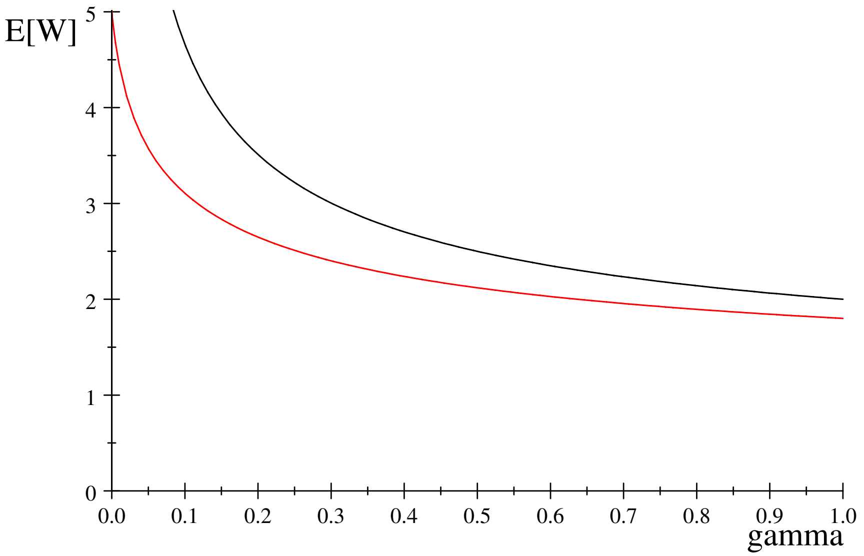
Notes: Values for $\alpha = 1$. Blue = mean log wealth, black = $\mathcal{R}_2$, red = $\mathcal{R}_3$, green = $\mathcal{R}_4$, and beige = $\mathcal{R}_5$. Horizontal dashed lines represent the empirical mean values of $\mathcal{R}_2$ (black, 0.85) and $\mathcal{R}_3$ (red, 0.65).

is Weibull. Since Gompertz converges to exponential for $\gamma \longrightarrow 0$, Weibull converges to Pareto. For Pareto, $\mathrm{E}\left[\underline{W}\right] = (1 - \alpha)^{-1}$, which is 5 for $\alpha = 0.80$ and infinite for $\alpha = 1$. Figure 2.5 confirms that $\mathrm{E}\left[\underline{W}\right]$ converges to these values when γ converges to zero. However, Figure 2.5 also shows that $\mathrm{E}\left[\underline{W}\right]$ is highly sensitive to even slight variations in γ for any $\gamma < 0.2$, in particular for values of α close to or above unity.

2.5.2. Estimation

The parameters α and γ of the Gompertz distribution can be estimated with the method of moments, like in Wingo (1989). We opt to estimate these parameters by maximum likelihood using the density function of w. The derivation of the likelihood function is in Appendix 2.C. The procedure yields an explicit expression for the estimator $\widehat{\alpha}$ as a function of the parameter γ and the data on w and it yields

Figure 2.5: Variation of Mean Wealth with γ



Notes: Values for $\alpha = 0.8$ (red) and $\alpha = 1$ (black).

an implicit expression for $\widehat{\gamma}$:

$$\widehat{\alpha} = \gamma^{-1}\left(\overline{e^{\gamma w}} - 1\right), \tag{2.14}$$

$$\text{SE}(\widehat{\alpha}) = \widehat{\alpha}\sqrt{N}^{-1},$$

$$N^{-1}\log \mathcal{L}(\gamma) = \ln\gamma - \ln\left(\overline{e^{\gamma w}} - 1\right) + \gamma\overline{w}. \tag{2.15}$$

$$\widehat{\gamma}^{-1} = \frac{\overline{we^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} - \overline{w},$$

$$\text{SE}(\widehat{\gamma}) = \sqrt{N\left(\frac{\overline{(w^2+1)\,e^{\widehat{\gamma}w}} - 2\overline{w}\,\overline{we^{\widehat{\gamma}w}}}{\overline{e^{\widehat{\gamma}w}} - 1} + \overline{w}^2\right)}^{-1}.$$

The first two lines are the expression for $\widehat{\alpha}$ and its standard error, the third line is the concentrated log likelihood, while the fourth and fifth line give the implicit condition for $\widehat{\gamma}$ and its standard error. The latter is an explicit function of $\widehat{\gamma}$ and $\overline{w}$. For $\widehat{\gamma} = 0.25$ and $\overline{w} = 1$, $\text{SE}(\widehat{\gamma}) \approx 0.166\sqrt{N}^{-1}$, so that $\widehat{\gamma}$ can be estimated reliably for $N \geq 64$. The fact that we have a simple expression for $\widehat{\alpha}$ as a function of γ and that we can therefore eliminate α from the likelihood is convenient. As discussed before, the value of α depends on the choice of the lower bound ω. One might presume that the data for various years or subregions are characterized by a common γ but by different values of α, depending how far one has to go into the right tail of in the distribution to arrive at a billionaire. One can first estimate a common $\widehat{\gamma}$ for all sub-region/year observations with $N \geq 64$ and then use the first line to calculate a separate $\widehat{\alpha}$ for each observation setting γ to this common $\widehat{\gamma}$. The restriction of a common γ can be tested using the concentrated likelihood for a likelihood ratio test.

An alternative, discussed in Appendix 2.D, is to use maximum likelihood estimation based on the hazard rates $H(w) := \Pr\left[\underline{w} = w\right] / \Pr\left[\underline{w} \geq w\right]$, where $\Pr\left[\underline{w} \geq w\right]$ is calculated from the data. Using the likelihood of hazard rates rather than w itself comes with a disadvantage: the calculation of an empirical estimate for $\Pr\left[\underline{w} \geq w\right]$ requires ordering the data, similar to the log rank regression. In fact, $\Pr\left[\underline{w} \geq w\right]$ is the inverse of an order statistic. The observations on the hazard are therefore not mutually independent. Maximum likelihood estimation based on w rather than $H(w)$ is therefore more efficient.

When we use a log rank regression instead we obtain:

$$\begin{aligned}\ln \text{Rank}\,[w] = \ln \Pr\left[\underline{w} \geq w\right] &= \left(\alpha\gamma\right)^{-1}\left(1 - e^{\gamma w}\right) \\ &= -w/\alpha - \frac{1}{2}\frac{\gamma}{\alpha} w^2 + O\left(w^3\right). \qquad (2.16)\end{aligned}$$

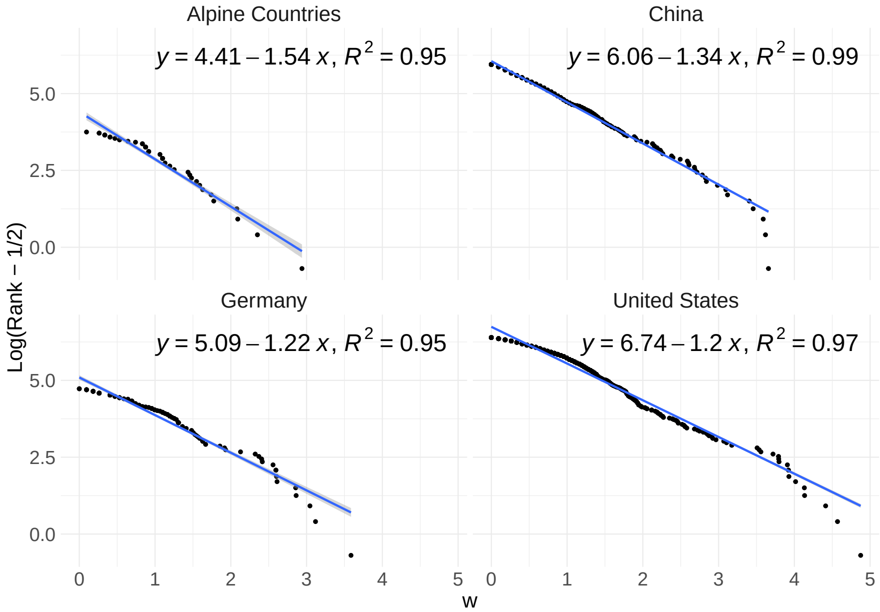
The quadratic term is close to zero for small γ and given the limited range of w in most empirical settings (w rarely exceeds 4 in the data on top wealth) the second order term is likely to be insignificant. This explains the ubiquitous linearities found in log-log plots in the literature.

We can directly show this spurious linearity in our setting. Figure 2.6 shows the results of four rank-size regressions, for the year 2019. The four regions studied are the Alpine Countries (Austria, Switzerland and Liechtenstein), China, Germany, and the United States. We chose these regions to illustrate that the illusory nature of rank-size regressions is an issue across economic and institutional environments and independent of population size.

Figure 2.6 reveals that all four cases have rank-size plots which are almost perfectly linear. This is despite the fact that all cases also have estimated values for γ which are significantly different from 0. The more populous regions of the United States and China have lower values of γ[13] and are therefore closer to the

[13] This is a systematic finding: more populous regions have higher tail inequality. In Chapter 3, we explore this issue further.

Figure 2.6: Spurious Log-Rank – Log-Size Regressions



Notes: Gabaix and Ibragimov (2011) regressions for the year 2019. $\widehat{\gamma} = 0.885$, (Alpine Countries), 0.129 (China), 0.503 (Germany), and 0.304 (United States), all statistically significant.

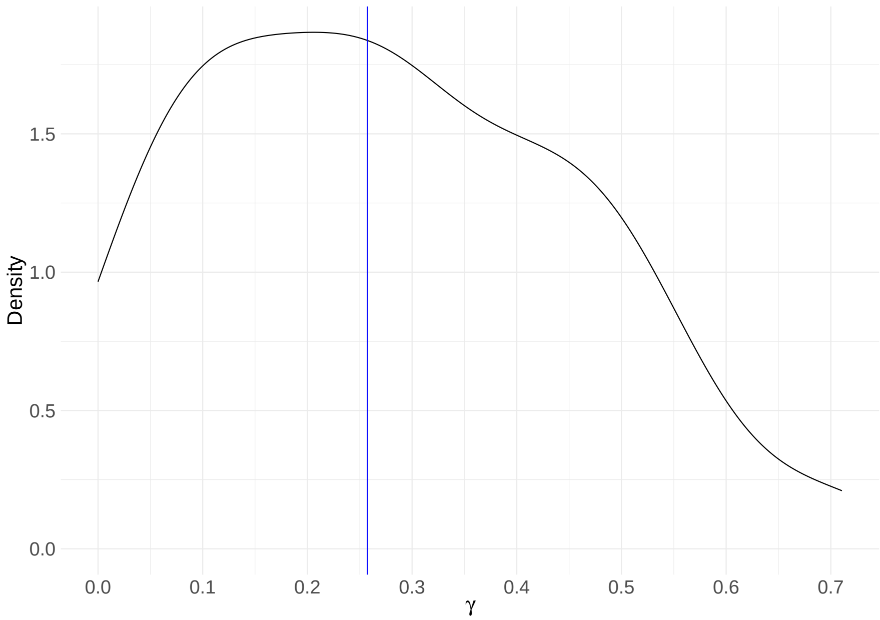
Pareto benchmark. Correspondingly, their rank-size plots are closer to the perfect straight line than those of the Alpine Countries and Germany. Nevertheless, we significantly reject Pareto and accept Weibull in all cases. This cautions against the widespread use of rank-size regressions as "evidence" for Pareto.

2.6. Testing Weibull

This section implements the estimators derived in the previous section. We use equation (2.14) to estimate $\widehat{\gamma}$ and $\widehat{\alpha}$ for the 75 observations with at least 64 billionaires. Figure 2.7 shows the distribution of $\widehat{\gamma}$. Recall that $\gamma = 0$ recovers the Pareto distribution. $\widehat{\gamma}$ is positive[14] for all but three observations. The vertical blue line shows the median value $\widehat{\gamma}^{\text{Med}} \approx 0.257$.

[14]Numerically, all estimates are positive, but we treat any observation for which $\widehat{\gamma} < 10^{-6}$ as $\widehat{\gamma} = 0$.

Figure 2.7: Distribution of γ Across Sub-Region–Years



Notes: Figure plots the density of all values of γ estimated on the sub-region–year observations of the *Forbes List of Billionaires* with at least 64 billionaires. The vertical blue line indicates the median value $\gamma^{\text{Med}} \approx 0.257$.

Equation (2.13) shows that $\mathcal{R}_k(\alpha\gamma)$ is a declining function of $\alpha\gamma$, or in logs of $\ln \alpha + \ln \gamma$. As a rough test of this prediction, we regress $\widehat{\mathcal{R}_k}$ for $k = 2, 3$ on $\ln \widehat{\alpha}$ and $\ln \widehat{\gamma}$. The test is only a rough test since the estimates on $\ln \widehat{\alpha}$ and $\ln \widehat{\gamma}$ are contaminated by estimation error. Hence, their coefficients are attenuated toward zero. Equation (2.13) predicts that the coefficients on both variables should be negative and equal for $\widehat{\mathcal{R}_2}$ and for $\widehat{\mathcal{R}_3}$. Figure 2.4 shows that $\mathcal{R}_3(\alpha\gamma)$ declines more strongly with $\alpha\gamma$ than $\mathcal{R}_2(\alpha\gamma)$, so that the coefficients are predicted to be larger in absolute value for $\widehat{\mathcal{R}_3}$ rather than $\widehat{\mathcal{R}_2}$. The results from this test are reported in Table 2.4 for observations with at least 64 billionaires. We drop the top and bottom 5% of observations for robustness.[15]

All predictions are confirmed by the data. The coefficients on $\ln \widehat{\alpha}$ and $\ln \widehat{\gamma}$ are all negative and significant and the coefficients are indeed more negative for $\widehat{\mathcal{R}_3}$.

[15] These observations are ($\widehat{\gamma}$ in parentheses): The U.S. in 2003 ($< 10^{-6}$) and 2005 ($< 10^{-6}$), China in 2011 ($< 10^{-6}$) and 2012 (0.009), Germany in 2014 (0.709) and 2018 (0.582), the British Islands in 2021 (0.711) and Brazil in 2021 (0.611).

Table 2.4: Weibull Predictions

Dependent Variables:	$\widehat{\mathcal{R}_2}$	$\widehat{\mathcal{R}_3}$	
Model:	(1)	(2)	(3)
Constant	0.781***	0.524***	-0.788***
	(0.023)	(0.026)	(0.061)
$\ln\widehat{\alpha}$	-0.115**	-0.207***	
	(0.037)	(0.031)	
$\ln\widehat{\gamma}$	-0.058***	-0.101***	
	(0.015)	(0.013)	
$\widehat{\mathcal{R}_2}$			1.68***
			(0.081)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics			
Observations	67	67	67
R^2	0.941	0.986	0.950
Adjusted R^2	0.939	0.986	0.949
RMSE	0.066	0.054	0.105
F-Stat	1.212	5.780	

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Clustered (Sub-Region & Year) standard-errors in parentheses. The regressions are estimated on observations with at least 64 billionaires, with the top and bottom 5% dropped. The bottom line reports F-tests on the restriction that $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$ have equal coefficients.

The restriction that the coefficient on $\ln \widehat{\alpha}$ and $\ln \widehat{\gamma}$ are equal is accepted for both $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$, as evidenced by the insignificant F-statistics. Moreover, the fact that both the coefficient and standard error for $\ln \alpha$ exceeds those for $\ln \gamma$ is predicted by our model.[16]

Similarly, since both $\mathcal{R}_2(\alpha\gamma)$ and $\mathcal{R}_3(\alpha\gamma)$ are increasing functions of the same parameter $\alpha\gamma$, $\mathcal{R}_3$ can be expressed as a function of $\mathcal{R}_2$. Ignoring measurement error and sampling variation, $\widehat{\mathcal{R}_2}$ should therefore be a perfect predictor of $\widehat{\mathcal{R}_3}$. The correlation coefficient between both statistics is indeed close to one (0.97 to be precise). Moreover, we run the regression $\widehat{\mathcal{R}_3} = \beta_0 + \beta_1 \widehat{\mathcal{R}_2}$, reported in column (3). Figure 2.4 suggests that β_1 should be about 1.5 and Table 2.1 suggests that $\overline{\widehat{\mathcal{R}}_{,3}} = \beta_0 + \beta_1 \overline{\widehat{\mathcal{R}}_{,2}} = \beta_0 + 0.85\beta_1 = 0.65$. The estimation results closely fit this observation. Conditional on the empirical value $\widehat{\mathcal{R}_2}$, the prediction for $\widehat{\mathcal{R}_3}$ matches its empirical value almost perfectly.

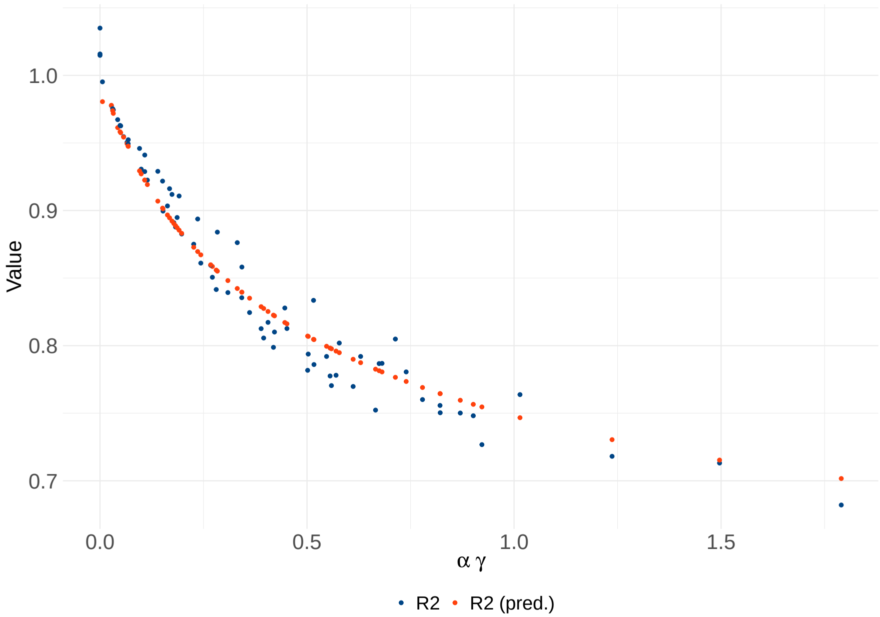
We now directly test the over-identifying restrictions implied by equation (2.13). For each observation with at least 64 billionaires, we calculate the theoretical values of $\mathcal{R}_2$ and $\mathcal{R}_3$, based on the estimated values of $\widehat{\alpha}$ and $\widehat{\gamma}$. We then contrast these theoretical values with the values in the data.

Figure 2.8 shows the results. Since both panels depend on the same value $\alpha\gamma$, a close fit in both panels provides an over-identifying test. The Figure shows that this test is easily accepted. The data for both $\mathcal{R}_2$ and $\mathcal{R}_3$ are extremely closely in line with their theoretical values. This provides direct evidence in favor of Weibull as a well-fitting distribution.

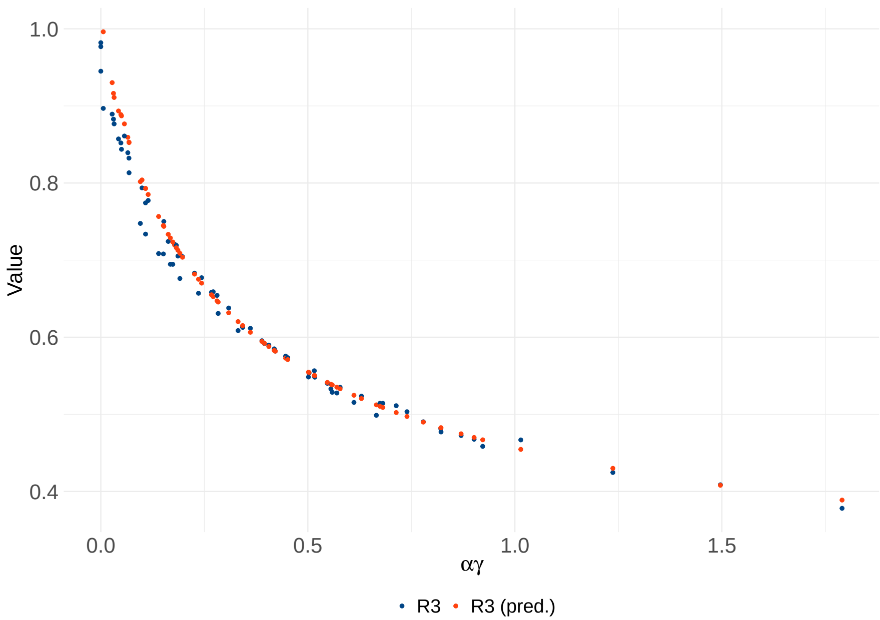
Our final test runs a horse race between Weibull and Pareto for the prediction of mean wealth in 2018. For the Weibull distribution, we assume that all observations have a common value for $\widehat{\gamma}$, so that both Weibull and Pareto have a single parameter $\widehat{\alpha}$ to fit the data for each observation. We use the median value $\widehat{\gamma}^{\text{Med}} = 0.257$. This horse race provides a fair comparison between both distributions, since we have a single parameter for each observation for both distributions. We test the restriction of the common γ equal to $\widehat{\gamma}^{\text{Med}}$ by means of a likelihood ratio test. We sum the concentrated likelihood (2.15) across all 75 observations, and compare this to the sum of restricted likelihoods where we impose $\gamma = \widehat{\gamma}^{\text{Med}}$ on all observations. The likelihood ratio statistic (75 degrees of freedom) is 202, rejecting the restriction of a common γ. Using a common γ therefore works against the Weibull distribution, as we impose a restriction which does not fit the data perfectly. As we shall see, this is of minor concern for the quality of predictions of the Weibull model.

[16] For small $\widehat{\gamma}$, equation (2.14) implies $\text{SE}(\widehat{\gamma}) \approx \widehat{\gamma}\sqrt{2N}^{-1}$ and hence $\text{SE}(\ln \widehat{\alpha}) = \sqrt{N}^{-1} > \sqrt{2N}^{-1} \approx \text{SE}(\ln \widehat{\gamma})$. In our regression sample, $\text{Var}(\widehat{\alpha}) \approx 0.098$ and $\text{Var}(\widehat{\gamma}) \approx 0.023$, conforming to our predictions.

Figure 2.8: Predicted versus Realized Values of $\mathcal{R}_2$ and $\mathcal{R}_3$



(a) $\mathcal{R}_2$



(b) $\mathcal{R}_3$

Notes: Red dots are the theoretical values for an observation based on equation (2.13), given its estimated $\widehat{\alpha}$ and $\widehat{\gamma}$. Blue dots show the empirical values.

We then use equation (2.14) to estimate a value of α specific for each observation. For the Pareto estimate of α, we use $\widehat{\alpha}^{\text{Hill}} = \overline{w}$, see Equation (2.4). Table 2.5 reports actual and estimated mean wealth among billionaires for Weibull and Pareto for all sub-regions, together with the estimated values of α for both distributions.

Table 2.5: Predicted vs. Realised Values of Mean Billionaire Wealth, 2018

	Mean Wealth			$\widehat{\alpha}$	
Sub-Region	Data	Weibull	Pareto	Weibull	Pareto
United States	5.29	5.65	∞	1.48	1.16
Canada	3.23	3.25	5.96	1.02	0.832
Germany	4.7	5.76	∞	1.5	1.18
British Isles	3.83	4.35	∞	1.26	1.01
Scandinavia	3.51	4.15	227	1.22	0.996
France	7.44	8.46	∞	1.83	1.37
Alpine Countries	3.8	4.8	∞	1.34	1.1
Italy	3.96	4.39	∞	1.26	1.02
China	3.3	3.12	4.97	0.983	0.799
South-East Asia	3.52	3.85	16.7	1.16	0.94
Asia Islands	2.85	2.86	4.17	0.912	0.76
South Korea	2.88	2.8	3.8	0.894	0.737
Japan	3.95	3.88	12.4	1.16	0.919
Australia	2.74	2.77	3.86	0.886	0.741
India	3.7	3.97	24.1	1.18	0.959
Russia	4.05	4.13	21.3	1.21	0.953
Brazil	4.2	4.51	∞	1.29	1.03
Israel+Turkey	2.17	2.25	2.64	0.716	0.622

Notes: Sub-regions are defined in Table 2.1. The Weibull prediction is made using Equation (2.14), where we take the median maximum likelihood estimate of $\widehat{\gamma}$ ($\widehat{\gamma}^{\text{Med}} \approx 0.257$), and where we calculate α using equation (2.14). The Pareto prediction is made using Equation (2.4), with $\overline{w}$ as the maximum likelihood estimator of α.

Actual mean wealth varies widely across sub-regions. In the sub-region Israel + Turkey, billionaires have a mean wealth of 2.1 billion, whereas in the U.S., this is almost 5.3 billion. One super-rich billionaire can have an enormous influence on these statistics; for instance, France's mean wealth of 7.4 is heavily affected by the wealth of LVMH owner Bernard Arnault (estimated to be about 72 billion in 2018).

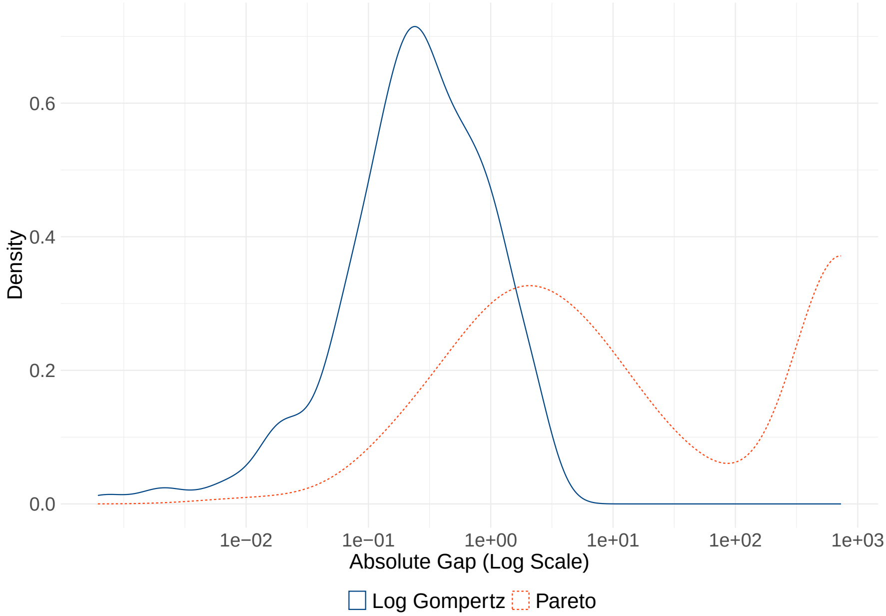
The mean wealth predicted by the Weibull model traces actual mean wealth remarkably well. Four predictions are almost identical to the real value (Canada, Asia Islands, Australia, Israel+Turkey), and a further nine estimates are within half a point of the real value. The remaining five estimates, moreover, are also reasonably close, with the largest gap being for Germany (1.06 points).

Now compare the prediction of the Weibull and the Pareto model. Table 2.5 shows $\widehat{\alpha}$ to be above 1 for seven sub-regions, including the United States. This immediately results in infinite values for these seven regions. We note a further five cases where $\widehat{\alpha}$ is close to 1, such as Scandinavia. Here, we obtain finite values, but these are obviously so large as to be meaningless. This leaves us with six reasonable estimates. None of these, however, are closer to the real value than truncated-Weibull. We conclude that Pareto is not a useful model to predict mean wealth, whereas Weibull performs well.

This conclusion is not driven by our choice of year. We calculate the absolute value of the gap between real and predicted values for all observation-years, for both Weibull and Pareto. Figure 2.9 plots the density of these gaps on a log scale. For visual purposes, infinite values for $E[W]$ have been set to the maximal finite value of 733.

Figure 2.9 shows that truncated-Weibull performs very well, with the vast majority of gaps being less than 1 in absolute value and a significant number even less than 0.1. This stands in stark contrast to the Pareto predictions. Although Pareto can result in reasonable estimates, the heavy right tail shows a nontrivial fraction of observations where the predictions gaps are between 10 and 100. The jump in the right tail further illustrates the fragility of Pareto; 129 out of 378 observations are predicted to have infinite mean wealth.

Figure 2.9: Absolute Prediction Gaps for Mean Wealth, Truncated-Weibull vs. Pareto



Notes: Figure shows the absolute gap between real mean wealth and the predictions made by truncated-Weibull and Pareto. Infinite values for predicted mean wealth have been set to the maximal predicted finite value, 733.

2.7. Additional Applications

2.7.1. Cities

Might our conclusion that top wealth is distributed Weibull rather than Pareto apply to other phenomena? We check this by applying our testing procedure to two other distributions, U.S. city and U.S. firm size. There has been considerable debate about whether city size is distributed Pareto or something thinner-tailed. Proponents for Pareto, such as Krugman (1996) and Gabaix (1999), typically find evidence for this relationship by observing a linear slope in a log-log plot of the largest 135 or so metropolitan statistical areas (MSAs). Eeckhout (2004, 2009) criticizes this approach on both statistical and substantial grounds. His statistical criticism, much like ours in Section 2.2, consists of pointing out that a log-log plot distorts observations at the tails of the distribution, and that formal statistical tests such as a Kolmogorov-Smirnov test cannot distinguish between Pareto and lognormal at this range. Substantively, he argues that the MSAs – which must have at least 50,000 inhabitants and typically do not consist of integrated economic units – are an imperfect measure of urban density, and argues for lognormality on the basis of the entire distribution of places, including small towns.

Here we take a different approach. We return to the distribution of MSAs, but show that even for this subset of the data, the Pareto hypothesis fails. For the sake of comparison with Eeckhout (2004), we use the 2000 U.S. Census in the 2001 vintage. Our results are reported in Table 2.6.

Table 2.6: Statistics for the U.S. City Size Distribution

Sample	n	ω	$\overline{w}$	$\overline{w^2}$	$\overline{w^3}$	$\mathcal{R}_2$	$\mathcal{R}_3$
Full	280	10.8	1.93	4.97	15.8	0.668	0.368
Top 135	135	12.6	1.1	2.07	5.06	0.857	0.636
Top 100	100	12.9	1.06	1.89	4.35	0.848	0.615
Top 50	50	13.8	0.851	1.26	2.42	0.871	0.654

Notes: Data are for the 2000 U.S. Census, 2001 vintage. n is sample size, ω is the log lower bound, $w := x - \omega$ is log city size minus log lower bound, overlines refer to sample averages, and the statistics $\mathcal{R}_2$ and $\mathcal{R}_3$ are as defined in the main text.

The $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ statistics are nowhere close to equalling 1, regardless of our sample selection choice. It is often argued that the Pareto hypothesis applies only to the extreme upper tail (e.g., Gabaix (2009) and Jones and Kim (2018)). However, we see that even at narrower slices of the data Pareto is rejected. In particular, for the top 135 MSAs – the traditional cutoff in city-size analyses (Krugman 1996; Eeckhout 2004) – the $\widehat{\mathcal{R}_k}$ statistics are again far from unity. Remarkably, $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ hover around the same values found for billionaire wealth, namely $\widehat{\mathcal{R}_2} \approx 0.85$ and $\widehat{\mathcal{R}_3} \approx 0.65$.

We apply our maximum likelihood procedures to estimate the parameters of Weibull distribution at various lower bounds. The results are reported in Table 2.7. All estimates for $\widehat{\alpha}$ and $\widehat{\gamma}$ are positive and significant. The likelihood clearly indicates that the model fits better to the top 135 MSAs than to the full distribution, which is underscored by the stability of $\widehat{\gamma}$ across the bottom three rows. The estimates for $\widehat{\alpha}$ decrease with the threshold. We test our prediction that this decline is a function of the lower bound and γ, using equation (2.10). The predictions work well, in particular when applied to the top 135 MSAs; the predictions made for the values of α for the top 100 and top 50 conform almost identically to their estimated values. In sum, γ remains constant while α varies; this further underscores our discussion in Section 2.5.

Table 2.7: Gompertz Estimation, City Size

					Predicted α	
Sample	γ	S.E.	α	S.E.	$\alpha_0 = 3.94$	$\alpha_0 = 1.82$
Full	0.452***	0.016	3.94***	0.236	–	–
Top 135	0.219***	0.013	1.82***	0.118	2.656	–
Top 100	0.252***	0.017	1.69***	0.135	2.321	1.687
Top 50	0.252***	0.024	1.24***	0.147	1.845	1.345

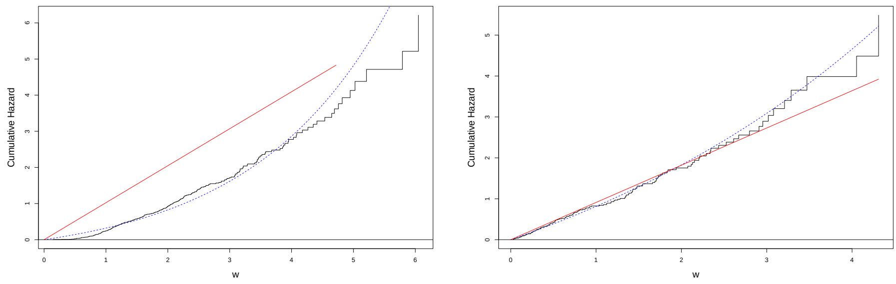
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*
Notes: Parameters γ and α estimated with maximum likelihood; see Appendix 2.C for details. The predictions for α are made with equation (2.10), where the respective lower bounds ω are taken from Table 2.6. The first prediction assumes the prediction to apply to the full sample, the second prediction takes the 136th MSA as lower bound.

We also report the visual fit of the Gompertz model, using hazard rates. This is done in Figure 2.10, which plots the empirical cumulative hazard (estimated with a nonparametric Kaplan-Meier regression) against the predicted fit from a Gompertz distribution and of an exponential distribution.

The first thing to note is that the Gompertz model approximates the data remarkably well. This is especially true for the top 135 MSAs (panel (a)), but the fit for the full sample is also very close until the final observations (panel (b)). Our results in Table 2.7 and Figure 2.10 clearly indicate that Gompertz fits the hazard of the log city size distribution very well, with only mild parameter instability when moving to the extreme upper tail.

In contrast, Figure 2.10 also shows the hazard rate for an exponential distribution, in red. It is clear that the exponential hazard fails to match the empirical hazards. The Weibull hazard fits the data well, in particular for the top 135 MSAs.

Figure 2.10: Cumulative Hazard Rates for the U.S. City Size Distribution



(a) Full Sample (b) Top 135 MSAs

Notes: Figure plots the empirical cumulative hazard rate (black) against the fitted Gompertz hazard (blue, dotted) and an exponential hazard with $\alpha = 1$ (red). Panel (a) fits the full sample of Metropolitan Statistical Areas, panel (b) only the 135 largest MSAs. Note that w is defined relative to a different lower bound in both panels.

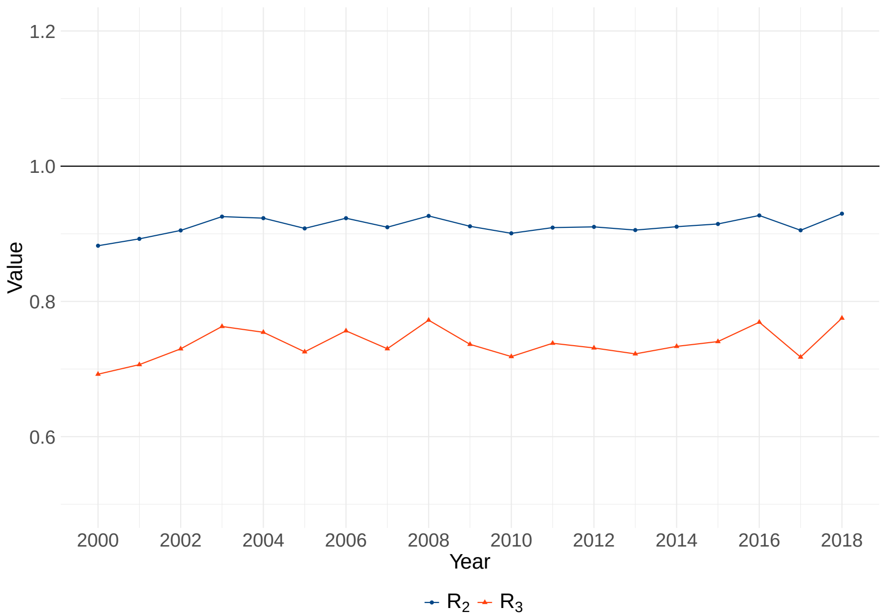
2.7.2. Firm Size

Contrary to our previous two cases, which suffer from relatively small sample sizes, we draw inference from a large (synthetic) sample of U.S. firms. We use data on the tabulated firm size distribution in the U.S. since 1930, compiled by Kwon, Ma, and Zimmermann (2024). These tabulations report the number of firms and average firm size in each bracket. The authors use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022, `gpinter`) to study the increase in top firm size shares. This interpolation method takes as inputs the bracket lower bound, the percentile corresponding to that lower bound, and the bracket average, and interpolates the entire distribution based on a semiparametric approximation to the Lorenz curve.

We use Kwon et al.'s tabulated datasets, and use the `gpinter` interface on `www.wid.world` to generate the interpolations. We use asset value as our measure of firm size. Once we have these interpolations, we can use the interface to sample from this interpolation. Effectively, this procedure assumes the interpolation to be the correct data-generating process, and proceeds to draw synthetic samples of a given sample size which are representative of the interpolated data. As a result, assuming that the interpolated data are a good approximation to the true distribution, we can effectively study the upper tail of the firm size distribution with arbitrary precision. Hence, we do not report standard errors.

We draw a million observations from the full distribution for all years between 2000–2018.[17] We keep the top 1% of the distribution, resulting in 10,000 observations per year. The results are plotted in Figure 2.11.

Figure 2.11: Tests for Pareto, U.S. Firm Size Distribution



Notes: $\underline{\mathcal{R}}_2$ and $\underline{\mathcal{R}}_3$ are defined in equation (2.6). Each point represents a sample of 10,000 drawn from the top 1% of the U.S. firm size distribution using generalized Pareto interpolation; see the main text for details.

We observe that the value of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ are constant over time. The values of $\widehat{\mathcal{R}_2}$, while not far from 1, are significantly different and hover around 0.9. The values for $\widehat{\mathcal{R}_3}$, moreover, are clearly lower than for $\widehat{\mathcal{R}_2}$ and are about 0.75. These values are consistent with the plots of $\mathcal{R}_2$ and $\mathcal{R}_3$ in Figure 2.4 for $\alpha\gamma \approx 0.1$. We conclude that firm size too does not follow Pareto and fits Weibull.

The Gompertz coefficients are reported in Table 2.8, where we again use maximum likelihood. We observe stable coefficients for both parameters. The values for γ are much lower than in our other two applications. Firm size is therefore

[17] Drawing this many observations for all years since 1930 is computationally burdensome; we have drawn samples of 10,000 from all years, with very similar results which are available on request.

closer to the Pareto benchmark than wealth or city size. $\alpha\gamma \approx 0.1$ in most years, conforming to the predictions made based on Figure 2.4.

Table 2.8: Gompertz Coefficients, U.S. Firm Size

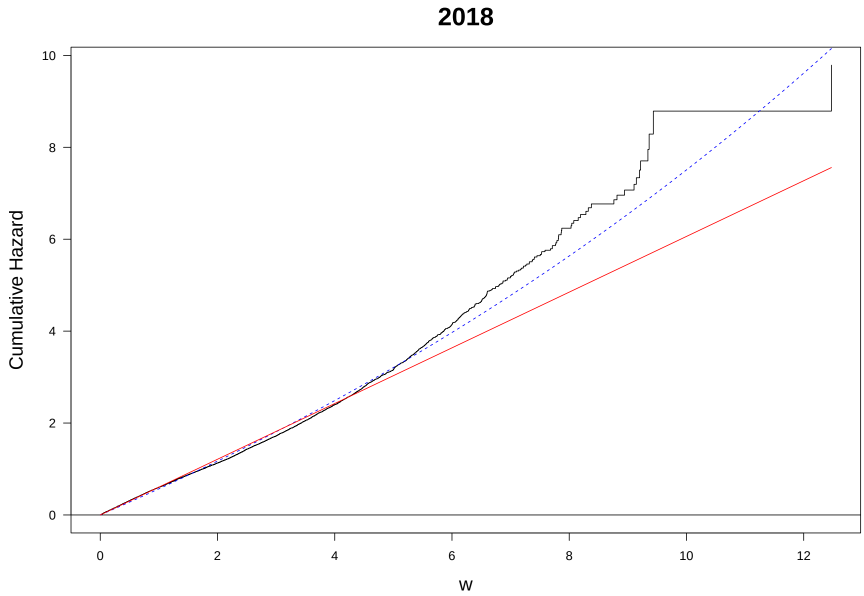
Year	γ	α
2000	0.11	1.8
2001	0.099	1.9
2002	0.084	1.9
2003	0.064	1.8
2004	0.067	1.8
2005	0.08	1.9
2006	0.066	1.8
2007	0.079	1.9
2008	0.062	1.8
2009	0.077	1.9
2010	0.086	2
2011	0.077	1.9
2012	0.079	1.9
2013	0.083	1.9
2014	0.077	1.9
2015	0.075	1.9
2016	0.062	1.8
2017	0.085	1.9
2018	0.059	1.8

Notes: Parameters γ and α estimated with maximum likelihood; see Appendix 2.C for details. We use generalized Pareto interpolation (Blanchet, Fournier, and Piketty 2022) to draw a sample of size 10,000 from the top 1% of the firm size distribution for each year, with data from the firm size distribution from Kwon, Ma, and Zimmermann (2024).

Finally, we show the visual fit of the Gompertz model to the log firm size data using the hazard-based approach. We report the data for 2018 in Figure 2.12.

We see that Gompertz fits the log hazard rates very well. Despite the estimated low value of γ, the Gompertz hazard still fits much better than the exponential hazard. The divergence happens beyond $w = 4$; the advantage of this dataset is that the sample size is sufficiently large such that w continues until 12 or so. The empirical failure of the epxonential distribution therefore becomes especially pronounced in Figure 2.12.

Figure 2.12: Cumulative Hazard Rates for the U.S. Firm Size Distribution



Notes: Figure plots the empirical cumulative hazard rate (black) against the fitted Gompertz hazard (blue, dotted) and an exponential hazard with $\alpha = 1$ (red). Data are for the 2018 top 1% of the U.S. firm size distribution.

2.8. Implications

Wealth, city size, and firm size all follow a Weibull rather than a Pareto distribution. This section explores some of the implications of this finding. We discuss these implications in a wider setting, since our results might carry over to other contexts where Pareto has commonly been assumed, such as income or trade. We first discuss some stochastic properties of Weibull and Pareto, and notions of convergence. Then, we discuss the implications for optimal taxation. Afterwards, we briefly discuss some other theoretical settings where Pareto is often used. Finally, we discuss the implications for the "granular hypothesis" (Gabaix 2011).

Convergence to Pareto and Weibull: The Pareto distribution is the quintessential example of a fat-tailed distribution. These distributions are defined by a concept from extreme-value theory called *regular variation*, which ensures that a function behaves like a Pareto distribution in the limit (cf. Benhabib and Bisin 2018; Bingham, Goldie, and Teugels 1989). Formally, a function $f(x)$ is called regularly varying if

$$\lim_{x\to\infty} \frac{f(xy)}{f(y)} = x^{-1/\alpha}, \quad \forall x > 0. \tag{2.17}$$

The interest in Pareto has therefore not only arisen because of its simple theoretical properties, but also because it seems to emerge as the stationary distribution for many stochastic processes. Gabaix (2009) gives many examples of diffusion processes which lead to Pareto in the limit. In general, these processes are variations on random growth models (e.g., geometric Brownian motions), with some frictions added to stabilize the distribution.

There is a tight link between convergence to Pareto and the decay of the distribution, which is governed by the hazard rate of its log-transform. Contrary to the Gompertz distribution (the log transform of Weibull) the hazard rate of the exponential distribution (the log transform of Pareto) is constant and *memoryless*: it has a constant hazard rate regardless of the choice of lower bound: $H(w) = H(x) = \alpha^{-1}, \forall \omega$. It can be shown that a distribution is fat-tailed if and only if the hazard rate of its log-transform converges to a constant (Beirlant, Goegebeur, Segers, and Teugels 2006). Our results, in contrast, show an exponentially increasing hazard.

Stochastic processes can converge to other distributions than Pareto. The exact process for convergence to truncated-Weibull can be modeled using a diffusion of the form $\mathrm{d}W_t = \mu(W_t)\mathrm{d}t + \sigma(W_t)\mathrm{d}\mathcal{B}(t)$, where $\mu(\cdot)$ and $\sigma(\cdot)$ are functions governing the drift and diffusion, respectively, and where $\mathcal{B}$ is a standard Brownian motion. We can reverse-engineer this process by inserting the truncated-Weibull distribution into the Kolmogorov Forward Equation, and solving for $\mu(\cdot)$ and $\sigma(\cdot)$. Results available on request indicate that this is possible for wealth both in levels and in logs, but that these functional forms are highly non-linear.

It has recently been proved that Gompertz emerges as the distribution of the length of *self-avoiding walks* (SAW) on stochastic Erdős–Rényi–Gilbert networks (Tishby, Biham, and Katzav 2016). These networks are formed by a graph $\mathrm{ER}[n, p]$ with n nodes, where each edge connecting two nodes has probability p of being generated. The connected nodes form a network. A SAW on such a network is defined as follows. Start at any node i that is part of the network. A SAW is a random walk from one node to another node that is connected to that node by the network until one cannot proceed to an new node not visited previously. Tishby, Biham, and Katzav (2016) prove that the distribution of SAW path lengths is Gompertz.

Stochastic networks are increasingly explored by economic theorists, see Goyal (2023). We can interpret the exponentially increasing hazard rate of the next node being the endpoint of a SAW as a “capacity constraint”: ultimately, a SAW can never be longer than the number of nodes minus one, $n - 1$. This could serve as a microfoundation for the emergence of Gompertz in our settings. For instance, cities can only grow on the available land; hence, their size is naturally bounded by the total land mass. Likewise, firm can only grow by "invading" new parts of the economy; hence, their size is bounded by the size of the economy at large. Similar analogues can be constructed for income and wealth.

Optimal taxation: A well-known result in the Mirrlees (1971) framework is that the marginal tax rate on the richest individual should be zero when the income distribution is bounded (Sadka 1976). A marginal tax rate just yields distortions at its level of incidence. However, it comes with the benefit of a higher tax on on higher income levels. For the highest income, there are no higher income levels, so the benefit of a positive marginal tax rate is zero and hence this marginal tax should be zero. This is not necessarily true when we allow the income distribution to be unbounded, see Saez (2001). When the distribution of top income is Pareto and the uncompensated elasticity of earned income with respect to the marginal tax rate (denoted η) is the same for all income levels, the Diamond (1998)–Saez (2001) formula for the optimal top income tax rate τ^* satisfies:[18]

$$\tau^* = \frac{\alpha}{\alpha + \eta}. \tag{2.18}$$

τ^* can be calculated from just two statistics: the tail index α^{-1} and the elasticity of earned income η. The parameter α appears in the formula is because the optimal tax rate equates the cost of a marginal tax at W (which is $\eta \times W \times \Pr\left[\underline{W} = W\right]$) and the additional tax revenue on an even higher income $\underline{W} > W$ enabled by this a higher marginal rate (which is $\Pr\left[\underline{W} > W\right]$). For Pareto, this ratio is constant, so that the top tax rate should converge to a constant. For Weibull, the corresponding expression reads:

$$\tau^* = \frac{\alpha e^{-\gamma w}}{\alpha e^{-\gamma w} + \eta},$$

so that the optimal tax rate converges to zero. However, this analysis maintains the assumption that the earned income elasticity for the top earners η converges to a constant. It might be the case that a formal model that generates a Weibull right tail of the income distribution also generates a declining elasticity of earned income with respect to the marginal tax rate since it becomes increasingly difficult to increase one's income or wealth due to "capacity constraints", as in our discussion

[18] The Diamond-Saez formula ignores externalities of taxing top earners. Jones (2022a) argues that top earners generate new ideas that increase productivity growth. In that case, optimal tax rates could be lower.

of networks with SAWs above. In that case, η would also converge to zero for the top earner and τ^* might again be constant.

Firm Productivity and Endogenous Growth: Kortum (1997) models endogenous growth arising from researchers that are heterogeneous in their productivity, where productivity is the highest draw ('best idea') from a distribution. He argues that exponential productivity growth can only be consistent with Pareto-distributed productivity. This result has been challenged by Jones (2023), who shows that if ideas are combinations of existing ideas, productivity need not be fat-tailed: combinatorial draws from a thin-tailed distribution will also result in exponential productivity growth. He arrives at a convenient closed-form expression for the growth rate of the best idea (g_Z) if productivity is Weibull-distributed (in our notation): $g_Z^{\text{Weibull}} = g_N/\gamma$, where g_N is the population growth rate. Jones calls γ the rate at which ideas are getting harder to find. This mirrors our interpretation of γ.

International trade: Chaney (2018) derives an explanation for the gravity equation based on Pareto-distributed firm productivity. Core to his argument is a linear relationship between log rank and log sizes. In general, Pareto provides closed-form expressions for trade models based on heterogeneous firms in the spirit of Melitz (2003); see also Helpman, Melitz, and Yeaple (2004) and Chaney (2008). Our results show that the actual firm size distribution is described better by Weibull than Pareto.

Granularity: Pareto is also popular because of the *granular hypothesis*, first formulated by Gabaix (2011). Gabaix argues that if firm size is Pareto, idiosyncratic shocks to firms do not wash out in the aggregate, and hence can partly account for aggregate fluctuations. Core to his argument is the fact that if α is sufficiently large, the firm size distribution does not have finite variance, and hence the law of large numbers does not apply as normal (which would imply that volatility decays on the order of $\sqrt{N}^{-1}$). Gabaix shows that if the firm size distribution has infinite variance, the decay is much slower, on the order of $1/\ln N$. Weibull, in contrast, has finite variance, and hence idiosyncratic shocks will decay faster than under Pareto. A Pareto distribution of firm size is no prerequisite for idiosyncratic shocks to generate aggregate fluctuations; as long as the distribution of firm connections is sufficiently right-skewed, shocks can propagate along the production network (Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi 2012).

2.9. Conclusion

We test the default assumption of Pareto for right-skewed distributions, using the *Forbes List of Billionaires*. We develop test statistics $\mathcal{R}_k$ based on normalizations of kth-order moments of the log-transformed data. We find strong evidence against a Pareto tail based on our statistics $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ in all cases. Our rejection of Pareto conforms to the arguments of Blanchet, Fournier, and Piketty (2022); however, where they argue for a non-parametric distribution, we show that a parametric alternative – (truncated-)Weibull – fits the data remarkably well. Our evidence for Weibull comes from the fit of within region/year maximum likelihood estimates, and from cross-equation restrictions on between region/year observations (the cross-equation restrictions in regressions of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$ on $\ln\widehat{\alpha}$ and $\ln\widehat{\gamma}$, and of $\widehat{\mathcal{R}_3}$ on $\widehat{\mathcal{R}_2}$). Our model yields sharp testable predictions, which are all accepted. Weibull-based predictions of mean wealth are almost perfect, whereas the predictions based on Pareto are nonsensically large or even infinite.

Our results cannot be driven by measurement error or small-sample bias. Our conclusions also apply to two other settings, city size and firm size. We conjecture our rejection of Pareto and embrace of Weibull to apply to even more settings where Pareto has commonly been used. Further empirical research will easily be able to test this hypothesis, since our statistics $\mathcal{R}_k$ are easily calculated and our Weibull alternative yields easily testable predictions.

The predictions of the truncated-Weibull distribution differ on two accounts from Pareto. First, whereas the Pareto distribution has a constant hazard rate, the truncated-Weibull distribution has an exponentially increasing hazard rate. Hence, the ratio of the density at the threshold and the fraction above the threshold is increasing in the threshold. However, while the Weibull hazard does not converge to Pareto, the moments of Weibull do converge to Pareto far in the upper tail; however, this only applies to the upper-most parts of the distribution. For the majority of the upper tail, Pareto has no defined moments for the empirically relevant values of α^{-1}, while the moments of the truncated-Weibull are always defined. These two differences make the truncated-Weibull distribution particularly suited for the study of upper tails.

The clarity of our results raise the question why they have not been found before. Existing research appears to hold a prior that the hazard rate will converge to a constant, perhaps implictly reasoning from the stochastic convergence of diffusion processes to Pareto discussed in Section 2.8. This prior is at odds with the data in all cases we study in this chapter. A second reason might be because the primary diagnostic to check Paretianity is the log rank regression. This tool has weak discriminatory power between Pareto and Weibull for small γ, introduces correlations in the error term because the data are ordered, and is sensitive to measurement error. Our test statistics $\mathcal{R}_k$ and estimation methods do not suffer from these problems.

Our results have important theoretical and empirical implications. If wealth is Weibull, log wealth is Gompertz. Gompertz emerges as the stationary distribution of self-avoiding walks on stochastic networks. This model structure has intriguing parallels with the environments we study, because we can microfound these SAWs as capacity constraints. Our results further have implications for optimal taxation. Mechanically, the optimal tax rate decreases because the hazard is now an increasing function of the threshold. However, the capacity constraints argument may mean that the behavioral elasticity is also no longer constant; hence, the full implications of our findings are ambiguous and need to be explored in a full model of optimal taxation. Finally, Weibull shows why the number of billionaires has increased since 2000; in effect, the threshold to be a billionaire, Ω, has decreased as countries have become richer and stock market capitalization has increased. Since Weibull is sensitive to the lower bound, these changes translate into a fattened upper tail. We explore these arguments further in Chapter 3.

2.A. Bias and variance of $\widehat{\mathcal{R}_2}$ and $\widehat{\mathcal{R}_3}$

Let $\underline{g} := w/\alpha$ be distributed exponentially with decay parameter 1 with moments $\mathrm{E}\left[\underline{g}^k\right] = k!$ and let g be a realisation of $\underline{g}$. $\mathcal{R}_k$ is defined in equation (2.6) as $\mathcal{R}_k = \mathrm{E}\left[\underline{g}^k\right] / \left(k!\mathrm{E}\left[\underline{g}\right]^k\right)$ for any $k \geq 1$. The estimator $\widehat{\mathcal{R}_k} = \overline{g^k} / \left(k!\overline{g}^k\right)$ of $\mathcal{R}_k$ uses the sample means $\overline{g^k}$ as the estimator for $\mathrm{E}\left[\underline{g}^k\right]$. However, since $\widehat{\mathcal{R}_k}$ is a non-linear transform of $\overline{g^k}$ and $\overline{g}$ this estimator is biased: $\mathrm{E}\left[\widehat{\mathcal{R}_k}\right] \neq \mathcal{R}_k$. This section asseses the variance and bias of $\widehat{\mathcal{R}_k}$.

2.A.1. Variance $\overline{g}^k$

The variance of $\overline{g}^k$ satisfies

$$\mathrm{Var}\left[\overline{g}^k\right] = \mathrm{E}\left[\overline{g}^{2k}\right] - \mathrm{E}\left[\overline{g}^k\right]^2 .$$

The variance of $\mathrm{E}\left[\overline{g}^k\right]$ satisfies:

$$\begin{aligned}
\mathrm{E}\left[\overline{g}^k\right] &= \mathrm{E}\left[\left(N^{-1}\sum_i \underline{g}_i\right)^k\right] \\
&= N^{-k}\left(\frac{N!}{(N-k)!}\mathrm{E}\left[\underline{g}\right]^k + \frac{N!}{(N-k+1)!}\binom{k}{2}\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2} + O\left(N^{k-2}\right)\right).
\end{aligned}$$

$\left(\sum_i \underline{g}_i\right)^k$ consists of N^k terms $\prod_{i\in K_n} \underline{g}_i$, where K_n is one of the $n = \{1, ..., N^k\}$ feasible combinations of k draws from the set of N draws. We retain only these terms where at most $\underline{g}_i$ are the same.

For $\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1}(N-m)$ terms, all $\underline{g}_i$ are different. The number of these terms is of order $O\left(N^k\right)$. Using $\mathrm{E}\left[\underline{g}_i\underline{g}_j\right] = \mathrm{E}\left[\underline{g}_i\right]\mathrm{E}\left[\underline{g}_j\right]$ the expectation of these terms is $\mathrm{E}\left[\underline{g}\right]^k$.

For $\frac{N!}{(N-k+1)!} = \prod_{m=0}^{k-2}(N-m)$ combinations of $\underline{g}_i$, all $\underline{g}_i$ but one pair are different. $\prod_{m=0}^{k-2}(N-m)$ is of order $O\left(N^{k-1}\right)$. For each combination, the pair of equal $\underline{q}_i'$s can be drawn from the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$ in $\binom{k}{2}$ different ways. The expectation of these terms is $\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{2k-2}$.

The number of terms where more than than two draws are the same is of order N^{k-m} with $m > 1$. These terms are therefore captured in the term $O\left(N^{k-2}\right)$.

Using $\mathrm{E}\left[\underline{g}^k\right] = k!$ and $\left(\frac{N!}{(N-k+1)!}\right)^2 = O\left(N^{2k-2}\right)$ these relations can be simplified

$$\mathrm{E}\left[\overline{g}^k\right] = N^{-k}\left(\frac{N!}{(N-k)!} + \frac{N!}{(N-k+1)!}k(k-1) + O\left(N^{k-2}\right)\right) \tag{2.19}$$
$$= 1 - \frac{k(k-1)}{2N} + \frac{k(k-1)}{N} + O\left(N^{-2}\right) = 1 + \frac{k(k-1)}{2N} + O\left(N^{-2}\right).$$

using

$$\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1}(N-m) = N^k - N^{k-1}\sum_{m=0}^{k-1} m + O\left(N^{k-2}\right)$$
$$= N^k - \frac{k(k-1)}{2}N^{k-1} + O\left(N^{k-2}\right),$$

which implies $\frac{N!}{(N-k+1)!} = N^{k-1} + O\left(N^{k-2}\right)$.

The same argument applies to $\mathrm{E}\left[\overline{g}^{2k}\right]$, replacing k by $2k$.

$$\mathrm{E}\left[\overline{g}^{2k}\right] = 1 + \frac{k(2k-1)}{N} + O\left(N^{-2}\right).$$

$\mathrm{E}\left[\overline{g}^k\right]^2$ satisfies

$$\mathrm{E}\left[\overline{g}^k\right]^2 = \left(1 + \frac{k(k-1)}{2N} + O\left(N^{-2}\right)\right)^2 = 1 + \frac{k(k-1)}{N} + O\left(N^{-2}\right) \tag{2.20}$$

Combining both terms yields

$$\mathrm{Var}\left[\overline{g}^k\right] = 1 + \frac{k(2k-1)}{N} - 1 - \frac{k(k-1)}{N} - O\left(N^{-2}\right) = \frac{k^2}{N} + O\left(N^{-2}\right). \tag{2.21}$$

2.A.2. Variance $\overline{g^k}$

$$\mathrm{Var}\left[\overline{g^k}\right] = N^{-1}\mathrm{Var}\left[\underline{g}^k\right] = N^{-1}\left(\mathrm{E}\left[\underline{g}^{2k}\right] - \mathrm{E}\left[\underline{g}^k\right]^2\right) = \frac{(2k)! - k!^2}{N}. \tag{2.22}$$

2.A.3. Covariance $\overline{g}^k$ and $\overline{g^k}$

$$\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right] = \mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right] - \mathrm{E}\left[\overline{g^k}\right]\mathrm{E}\left[\overline{g}^k\right]^2.$$

where we use the expression for $\mathrm{E}\left[\overline{g}^k\right]^2$ in equation (2.19). We have:

$\mathrm{E}\left[\overline{g^k}\right] = \mathrm{E}\left[\underline{g}^k\right]$, while $\mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right]$ satisfies

$$\begin{aligned}\mathrm{E}\left[\overline{g^k}\,\overline{g}^k\right] &= \mathrm{E}\left[\left(N^{-1}\sum_i \underline{g}_i^k\right)\left(N^{-1}\sum_i \underline{g}_i\right)^k\right] \\ &= N^{-k-1}\left(\begin{array}{c} \frac{N!}{(N-k-1)!}\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}\right]^k + \frac{N!}{(N-k)!}\binom{k}{2}\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2} \\ +\frac{N!}{(N-k)!}k\mathrm{E}\left[\underline{g}^{k+1}\right]\mathrm{E}\left[\underline{g}\right]^{k-1} + O\left(N^{k-2}\right)\end{array}\right).\end{aligned}$$

$\left(\sum_i \underline{g}_i^k\right)\left(\sum_i \underline{g}_i\right)^k$ consists of N^{k+1} terms $\prod_{i\in K_n} \underline{g}_i \underline{g}_j^k$, where K_n is one of the $n = \left\{1, \ldots, N^k\right\}$ feasible combinations of k draws from the set of N draws. We retain only these terms where at most one pair of these k factors with $\underline{g}_i$ are the same and where they are different from $\underline{g}_j$ in $\underline{g}_j^k$ or where one these k factors has the same $\underline{g}_i$ as $\underline{g}_j$ in $\underline{g}_j^k$.

For $\frac{N!}{(N-k-1)!} = \prod_{m=0}^k (N-m)$ terms, all $\underline{g}_i$ are different from each other and from the $\underline{g}_j$. The number of these terms is of order $O\left(N^{k+1}\right)$. Using $\mathrm{E}\left[\underline{g}_i\underline{g}_j\right] = \mathrm{E}\left[\underline{g}_i\right]\mathrm{E}\left[\underline{g}_j\right]$ the expectation of these terms is $\mathrm{E}\left[\underline{g}\right]^k\mathrm{E}\left[\underline{g}^k\right]$.

For $\frac{N!}{(N-k)!} = \prod_{m=0}^{k-1} (N-m)$ combinations of $\underline{g}_i$, either all $\underline{g}_i$ but one pair are different and they are all different from $\underline{g}_j$ in $\underline{g}_j^k$ or they are all different, but one of them is equal to $\underline{g}_j$ in $\underline{g}_j^k$. For the first group, the pair of equal $\underline{g}_i'$s can be drawn from the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$ in $\binom{k}{2}$ different ways. The expectation of these terms is $\mathrm{E}\left[\underline{g}^k\right]\mathrm{E}\left[\underline{g}^2\right]\mathrm{E}\left[\underline{g}\right]^{k-2}$. For the second group, the $\underline{g}_i$ that is equal to $\underline{g}_j$ in $\underline{g}_j^k$ can be drawn from each of the k factors of the product $\left(\sum_i \underline{g}_i\right)^k$. The expectation of these terms is $\mathrm{E}\left[\underline{g}^{k+1}\right]\mathrm{E}\left[\underline{g}\right]^{k-1}$.

The numbers of terms where more than than two observations are the same are of order N^{k-m} with $m > 1$. These terms are therefore captured in the term $O\left(N^{k-1}\right)$.

Using $\mathrm{E}\left[\underline{g}^k\right] = k!$ and previous relations, we obtain

$$
\begin{aligned}
\mathrm{E}\left[\overline{g^k}\overline{g}^k\right] &= N^{-k-1}\left(\frac{N!}{(N-k-1)!}k!\right. \\
&\quad \left.+\frac{N!}{(N-k)!}k\,(k-1)\,k! + \frac{N!}{(N-k)!}k\,(k+1)! + O\left(N^{k-2}\right)\right) \\
&= k! - \frac{k\,(k+1)}{2N}k! + \frac{k\,(k-1)\,k! + k\,(k+1)!}{N} + O\left(N^{-2}\right) \\
&= k!\left(1 + \frac{k\,(3k-1)}{2N}\right) + O\left(N^{-2}\right).
\end{aligned}
$$

Combining these results yields

$$
\begin{aligned}
\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right] &= k!\left(1+\frac{k\,(3k-1)}{2N}\right) - k!\left(1+\frac{k\,(k-1)}{2N}\right) + O\left(N^{-2}\right) \qquad (2.23)\\
&= \frac{k^2 k!}{N} + O\left(N^{-2}\right).
\end{aligned}
$$

2.A.4. The expectation and variance of $\widehat{\mathcal{R}_k}$

A Taylor expansion for the expectation and variance of a quotient reads (Mood, Graybill, and Boes 1974, pp. 180)

$$
\begin{aligned}
\mathrm{E}\left[\frac{\underline{z}}{\underline{y}}\right] &= \frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]} - \frac{\mathrm{Cov}\left[\underline{y},\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]^2} + \frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]^3}\mathrm{Var}\left[\underline{y}\right] + O\left(N^{-2}\right), \\
\mathrm{Var}\left[\frac{\underline{z}}{\underline{y}}\right] &= \left(\frac{\mathrm{E}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{y}\right]}\right)^2\left(\frac{\mathrm{Var}\left[\underline{z}\right]}{\mathrm{E}\left[\underline{z}\right]^2} - 2\frac{\mathrm{Cov}\left[\underline{y},\underline{z}\right]}{\mathrm{E}\left[\underline{z}\right]\mathrm{E}\left[\underline{y}\right]} + \frac{\mathrm{Var}\left[\underline{y}\right]}{\mathrm{E}\left[\underline{y}\right]^2}\right) + O\left(N^{-2}\right).
\end{aligned}
$$

Using equation (2.19), (2.21), (2.22) and (2.23) and

$$
\mathrm{E}\left[\overline{g}^k\right]^{-1} = \left(1+\frac{k\,(k-1)}{2N}\right)^{-1} + O\left(N^{-2}\right) = 1 - \frac{k\,(k-1)}{2N} + O\left(N^{-2}\right),
$$

we obtain

$$
\begin{aligned}
\mathrm{E}\left[\widehat{\mathcal{R}_k}\right] &= \frac{\mathrm{E}\left[\overline{g^k}\right]}{k!\mathrm{E}\left[\overline{g}^k\right]} - \frac{\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right]}{k!^2\mathrm{E}\left[\overline{g}^k\right]^2} + \frac{\mathrm{E}\left[\overline{g^k}\right]}{k!^3\mathrm{E}\left[\overline{g}^k\right]^3}\mathrm{Var}\left[\overline{g}^k\right] + O\left(N^{-2}\right) \\
&= 1 - \frac{k\left(k-1\right)}{2N} - \frac{k^2 k!}{k!N} + \frac{(2k)! - k!^2}{k!^2 N} + O\left(N^{-2}\right) \\
&= 1 + N^{-1}\left(\frac{(2k)!}{k!^2} - \frac{3k^2 - k + 2}{2}\right) + O\left(N^{-2}\right).
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left[\widehat{\mathcal{R}_k}\right] &= \left(\frac{\mathrm{E}\left[\overline{g^k}\right]}{k!\mathrm{E}\left[\overline{g}^k\right]}\right)^2 \left(\frac{\mathrm{Var}\left[\overline{g^k}\right]}{\mathrm{E}\left[\overline{g^k}\right]^2} - 2\frac{\mathrm{Cov}\left[\overline{g^k}, \overline{g}^k\right]}{\mathrm{E}\left[\overline{g^k}\right]\mathrm{E}\left[\overline{g}^k\right]} + \frac{\mathrm{Var}\left[\overline{g}^k\right]}{\mathrm{E}\left[\overline{g}^k\right]^2}\right) + O\left(N^{-2}\right) \\
&= \frac{(2k)! - k!^2}{k!^2 N} - \frac{2k^2 k!}{k!N} + \frac{k^2}{N} + O\left(N^{-2}\right) \\
&= N^{-1}\left(\frac{(2k)!}{k!^2} - k^2 - 1\right) + O\left(N^{-2}\right).
\end{aligned}
$$

2.B. Moments of the Gompertz and Weibull distribution

Define the parameter $\theta := \alpha\gamma$ and the normalized random variable $\underline{q} := \gamma\underline{w}$ for $\underline{q} \geq 0$. Its complement distribution function $S(q)$ reads:

$$\Pr\left[\underline{q} > q\right] =: S\left(q\right) = \exp\left(-\frac{\mathrm{e}^{q} - 1}{\theta}\right).$$

The moments of $\underline{q}$ for $k > 0$ satisfy:

$$\begin{aligned}
\mathrm{E}\left[\underline{q}^{k}\right] &= \int_{0}^{\infty} q^{k} s\left(q\right) \mathrm{d}q \\
&= -\overbrace{\left[q^{k} S\left(q\right)\right]_{q=0}^{\infty}}^{=0} + \overbrace{k \int_{0}^{\infty} q^{k-1} S\left(q\right) \mathrm{d}q}^{=h(\theta,k)}, \\
h\left(\theta, k\right) &:= k \int_{0}^{\infty} q^{k-1} \exp\left(-\frac{\mathrm{e}^{q} - 1}{\theta}\right) \mathrm{d}q = k\mathrm{e}^{1/\theta} \int_{0}^{\infty} q^{k-1} \exp\left(-\mathrm{e}^{q}/\theta\right) \mathrm{d}q, \\
h\left(\theta, 1\right) &= \mathrm{e}^{1/\theta} \int_{0}^{\infty} \exp\left(-\mathrm{e}^{q}/\theta\right) \mathrm{d}q = \mathrm{e}^{1/\theta} \int_{1/\theta}^{\infty} y^{-1}\mathrm{e}^{-y} \mathrm{d}y = \mathrm{e}^{1/\theta} \mathrm{Ei}\left(1/\theta\right), \\
y &:= \mathrm{e}^{q - \ln\theta} \Rightarrow q = \ln y + \ln\theta \Rightarrow \frac{\mathrm{d}q}{\mathrm{d}y} = \frac{1}{y},
\end{aligned}$$

where $s\left(q\right) := -S'\left(q\right)$ is the density function and where $\mathrm{Ei}(z)$ is the standard Exponential integral. The second line uses the definition of $S\left(q\right)$ and $h\left(\theta, k\right)$ and applies integration by parts. The third line gives the definition of $h(\theta, k)$. The fourth line analyses the special case $h\left(\theta, 1\right)$. The definitions of $\underline{q}$ and θ imply:

$$\mathrm{E}\left[\underline{w}^{k}\right] = \gamma^{-k}\mathrm{E}\left[\underline{q}^{k}\right] = \gamma^{-k} h\left(\alpha\gamma, k\right).$$

for the derivation of $\mathrm{E}\left[\underline{w}^{k}\right]$.

The expectation of $\underline{W} = \mathrm{e}^{\underline{w}}$ satisfies:

$$
\begin{aligned}
\mathrm{E}\left[\underline{W}\right] &= \frac{\gamma}{\theta}\mathrm{e}^{1/\theta}\int_0^\infty \exp\left(\left(1+\gamma\right)w - \theta^{-1}\mathrm{e}^{\gamma w}\right)\mathrm{d}w \\
&= \frac{\gamma}{\theta}\mathrm{e}^{1/\theta}\int_{\theta^{-1}}^\infty \frac{1}{\gamma y}\left(\theta y\right)^{\left(1+\gamma\right)/\gamma}\mathrm{e}^{-y}\mathrm{d}y \\
&= \theta^{1/\gamma}\mathrm{e}^{1/\theta}\int_{\theta^{-1}}^\infty y^{1/\gamma}\mathrm{e}^{-y}\mathrm{d}y = \theta^{1/\gamma}\mathrm{e}^{1/\theta}\Gamma\left(1+\gamma^{-1},\theta^{-1}\right) \\
y := \theta^{-1}\mathrm{e}^{\gamma w} &\Rightarrow w = \gamma^{-1}\ln\left(\theta y\right) \Rightarrow \frac{\mathrm{d}w}{\mathrm{d}y} = \frac{1}{\gamma y}
\end{aligned}
$$

compare the expressions for the moments of $\underline{W}$ for the truncated Weibull distribution in Wingo (1989).

2.C. Likelihood of the Gompertz Distribution

The density function of the Gompertz distribution reads:

$$f(w) = \alpha^{-1}\exp\left(\gamma w - \frac{e^{\gamma w} - 1}{\alpha\gamma}\right).$$

Hence, the log likelihood reads:

$$N^{-1}\log\mathcal{L}\left(\alpha,\gamma\right) = -\ln\alpha + \gamma\overline{w} - \left(\alpha\gamma\right)^{-1}\left(\overline{e^{\gamma w}} - 1\right). \tag{2.24}$$

The first order condition for α reads:

$$N^{-1}\frac{\partial\log\mathcal{L}\left(\alpha,\gamma\right)}{\partial\alpha} = \left(-1 + \frac{\overline{e^{\gamma w}} - 1}{\widehat{\alpha}\gamma}\right)\widehat{\alpha}^{-1} = 0 \Rightarrow$$

$$\widehat{\alpha} = \gamma^{-1}\left(\overline{e^{\gamma w}} - 1\right). \tag{2.25}$$

The second order condition for α reads:

$$N^{-1}\frac{d^2\log\mathcal{L}\left(\alpha,\gamma\right)}{(d\alpha)^2} = \frac{\alpha\gamma - 2\left(\overline{e^{\gamma w}} - 1\right)}{\alpha^3\gamma},$$

$$N^{-1}\left.\frac{d^2\log\mathcal{L}\left(\alpha,\gamma\right)}{(d\alpha)^2}\right|_{\alpha=\widehat{\alpha}} = -\widehat{\alpha}^{-2},$$

$$\mathrm{SE}(\widehat{\alpha}) = \widehat{\alpha}\sqrt{N}^{-1}.$$

In the second line we use equation (2.25) to simplify the first line.

Substitution of equation (2.25) in equation (2.24) yields the concentrated log likelihood (up to a constant):

$$N^{-1}\log\mathcal{L}\left(\gamma\right) = \ln\gamma + \gamma\overline{w} - \ln\left(\overline{e^{\gamma w}} - 1\right).$$

The first (FOC) and second order condition read:

$$
\begin{aligned}
N^{-1} \left.\frac{\mathrm{d}\log\mathcal{L}(\gamma)}{\mathrm{d}\gamma}\right|_{\gamma=\widehat{\gamma}} &= \widehat{\gamma}^{-1} + \overline{w} - \frac{\overline{w\mathrm{e}^{\widehat{\gamma}w}}}{\overline{\mathrm{e}^{\widehat{\gamma}w}} - 1} = 0, \\
N^{-1}\frac{\mathrm{d}^2\log\mathcal{L}(\gamma)}{(\mathrm{d}\gamma)^2} &= -\gamma^{-2} + \left(\frac{\overline{w\mathrm{e}^{\gamma w}}}{\overline{\mathrm{e}^{\gamma w}} - 1}\right)^2 - \frac{\overline{(w^2+1)\,\mathrm{e}^{\gamma w}}}{\overline{\mathrm{e}^{\gamma w}} - 1}, \\
N^{-1} \left.\frac{\mathrm{d}^2\log\mathcal{L}(\gamma)}{(\mathrm{d}\gamma)^2}\right|_{\gamma=\widehat{\gamma}} &= \left(2\widehat{\gamma}^{-1} + \overline{w}\right)\overline{w} - \frac{\overline{(w^2+1)\,\mathrm{e}^{\widehat{\gamma}w}}}{\overline{\mathrm{e}^{\widehat{\gamma}w}} - 1} \\
&= -\frac{\overline{(w^2+1)\,\mathrm{e}^{\widehat{\gamma}w}} - 2\overline{w}\,\overline{w\mathrm{e}^{\widehat{\gamma}w}}}{\overline{\mathrm{e}^{\widehat{\gamma}w}} - 1} - \overline{w}^2, \\
\mathrm{SE}(\widehat{\gamma}) &= \sqrt{N\left(\frac{\overline{(w^2+1)\,\mathrm{e}^{\widehat{\gamma}w}} - 2\overline{w}\,\overline{w\mathrm{e}^{\widehat{\gamma}w}}}{\overline{\mathrm{e}^{\widehat{\gamma}w}} - 1} + \overline{w}^2\right)}^{-1}.
\end{aligned}
$$

In the third line we use the first line to simplify the second line.

2.D. Gompertz Hazard Rate Estimation

The Gompertz hazard is characterized by

$$H(w) = \frac{1}{\alpha} \exp\left(\gamma w\right). \tag{2.26}$$

The hazard collapses to the constant Exponential hazard as $\gamma \to 0$. We use a parametric duration model[19] to estimate γ and α, with maximum likelihood. Specifically, we use the `eha` package in R. This package uses two parametrizations for the Gompertz hazard, which the author calls the "rate" and "canonical" parameterizations.[20] The former is closest to our notation, and parameterizes the hazard as

$$H(w) = p \exp\left(kw\right), \tag{2.27}$$

where p is called the `shape` parameter and k the `rate`; $p = \alpha^{-1}$ and $k = \gamma$. To calculate the standard error of α, we use standard Taylor approximation arguments (i.e., $\text{Var}[g(x)] \approx (g'(\text{E}[x]))^2 \cdot \text{Var}[x]$) to calculate $\text{SE}(\alpha) \approx \alpha^{-2} \cdot \text{SE}\left(\alpha^{-1}\right)$.

[19]Specifically, we use an accelerated failure time model, since this results in more stable estimations than proportional hazards. Since we do not use covariates, accelerated failure time models are equivalent to proportional hazards models.

[20]See: https://cran.r-project.org/web/packages/eha/vignettes/gompertz.html.

Chapter 3

Why Has the Number of Billionaires Increased So Much?

This chapter is based on Teulings and Toussaint (2024).

3.1. Introduction

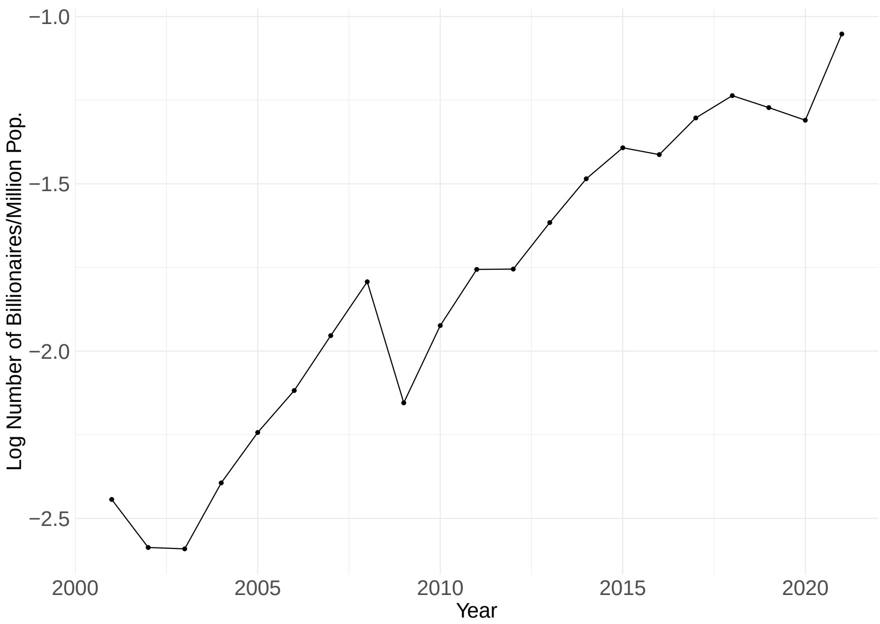
The global ratio of billionaires to population size has more than quadrupled since 2001, see Figure 3.1, panel (a). The regional variation in this trend is enormous, see panel (b). The ratio has increased in China and India by a factor 40 and 25 respectively, while in Europe it has increased only fourfold. What factors explain this general increase and its regional variation?

Given the often heated debate surrounding billionaire wealth (e.g., Saez and Zucman 2019), it is surprising that these questions have not been extensively studied. The purpose of this chapter is to fill this gap. We develop a minimalist model that can both capture salient features of the billionaire wealth distribution while remaining very tractable. Our point of departure is a robust finding, documented and extensively tested in Chapter 2, that billionaire wealth is not distributed Pareto but Weibull. If wealth is Weibull, log wealth is Gompertz, a distribution characterized by an exponentially increasing hazard rate. In our setting, this means that it becomes exponentially less likely to observe a billionaire with, say, 100 billion compared to one with 10 billion. While this finding is important in its own right, in this chapter we solely use it to motivate a *network model* of billionaire wealth.

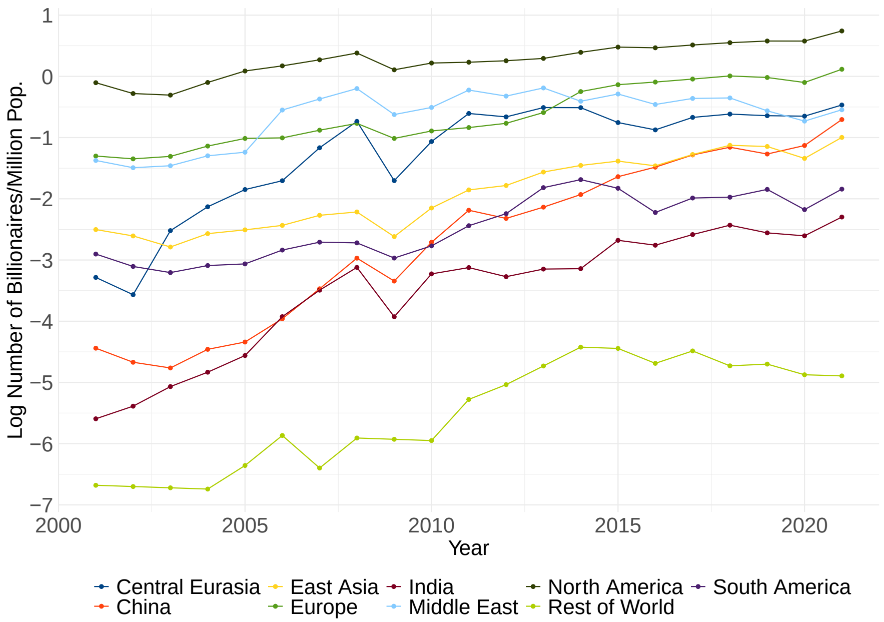
We model an economy which is represented by a random Erdős-Rényi graph. This network consists of L nodes, each of which has probability p of forming a link. We think of nodes in this setting as customers, with a link between two nodes signifying an expansion of a firm's customer base. Then, the extent of business growth (a longer uninterrupted path between nodes) is limited by the extent of the market, since a larger number of nodes ensures more connections are possible. Mathematically, we let a business owner's wealth be represented by a *Self-Avoiding Walk* (SAWs) on the graph, meaning a path that touches no nodes more than once. It has recently been proved that SAWs on random graphs are distributed Gompertz (Tishby, Biham, and Katzav 2016), completing our analogy. A longer SAW, in our context, means a firm which has successfully expanded its operations to serve many customers (many links between nodes).

The length of the longest SAW – equivalently, the wealth of the richest person – is governed by the hazard rate. In our model, the hazard rate is completely determined by the parameter p, which serves as the elasticity of the hazard rate with respect to wealth: a 1% increase in wealth leads to an increase of the hazard rate of p%. A high p means few super-rich individuals (a high hazard rate). As $p \to 0$, the length of the upper tail increases. Empirically, we observe widely varying values of p, ranging anywhere from just above 0 to almost 0.7. A direct implication of our network model is that p should solely be determined by the log number of nodes. In our interpretation of the network size as corresponding to market size, this means

Figure 3.1: Global and Regional Trends in Billionaire Numbers



(a) Global Log Number of Billionaires per Million Population, 2001–2021



(b) Regional Variation in Log Billionaire Numbers, 2001–2021

Notes: Panel (a) shows the evolution of the global log number of billionaires, normalized by population size (in millions); panel (b) shows the same per region, with regions defined in Table 3.1.

that p should solely be determined by log population size in a region, and that the relationship should be negative. This is a startling and strong prediction: More populous countries should have more tail inequality.

We also study entry into the network by analyzing parameters governing the lower bound. We view business owners as a set of individuals who can capitalize rental income into wealth, and model the threshold for that capitalization as a function of local and global market conditions. Specifically, we will allow the log lower bound to be a function of log nominal GDP per capita and the log of a global financial wealth-income ratio. Entry into billionaire ranks is governed by two parameters, h and d. h is the inverse hazard at the capitalization lower bound, conditional on p; the higher h, the lower the hazard rate at the beginning of our sample and hence the fatter the right tail of top wealth. Conditional on p and h, d measures the probability that somebody is a billionaire. For both these parameters, our model gives strong predictions: They should depend on log GDP per capita and the log wealth-income ratio, with coefficients equal to one, and on nothing else.

We use data from the *Forbes List of Billionaires* from 2001 to 2021 to test our hypotheses. We find that our predictions by and large hold. We robustly find that p solely depends on population, and not on the other covariates, with a negative coefficient as predicted. We test for the inclusion of fixed effects. Remarkably, our model does not need time fixed effects, explaining all time-series variation in p. We do, however, need region fixed effects. This indicates that there is more cross-sectional variation than our minimalist model can capture.

Similar patterns emerge for the other two parameters. The conditional baseline inverse hazard, h, should depend solely on log GDP per capita and our asset-market factor, with unit coefficients. The unit coefficients restriction holds quite well. We find again that time fixed effects are not needed, but region fixed effects are. Finally, the conditional billionaire probability, d, should also move one-for-one with log GDP per capita and the log market capitalization factor and nothing else. Our results are similar to those for h, with the restriction that these variables have unit coefficients holding reasonably well.

We conclude from our exercise that our minimal model explains the increase in billionaires remarkably well. Another way of evaluating this statement is by looking at the model's predictive capacity. To that end, we disaggregate our data to the country level, including countries with way fewer than 32 billionaires. We find that our model predicts both the fraction of billionaires as well as their mean (log) wealth extremely well. Moreover, we also track the global values for these moments very closely. We can decompose the global increase in billionaire numbers and mean (log) wealth into the contributions played by our variables (log population, log

GDP per capita, and log market capitalization). We find that the change in global GDP per capita played by far the largest role, explaining 70% of the model-implied increase in billionaire numbers and mean log wealth, and more than 60% of the increase in mean wealth. Similar conclusions follow if we focus solely on the United States. There, too, most of the billionaire increases and their increase in mean (log) wealth has been driven by the increases in GDP per capita. We conclude that our simple model has quite some power.

We interpret our model and results in light of heterogeneous-firm models. In this class of models[1], a country's market size determines average firm size and the number of entrants. Opening up to trade disproportionately benefits the largest firms. We see our results as the counterpart results of the owners of these firms. Large countries have larger firms and hence more wealthy firm-owners. Changes in local and global conditions that are more conducive to business dynamism benefit the entire market, but most disproportionately benefit the upper tail. Hence, the elasticity of the wealth hazard rate γ declines in population size while mean (log) wealth are increasing in GDP per capita and the global wealth-income ratio.

In addition, we can link our model to endogenous-growth models where the number of ideas is proportional to population size (Jones 2022b). A larger population means more ideas; moreover, more ideas mean more possibilities for connecting ideas, driving technological progress. It is an interesting implication of our particular Erdős-Rényi structure that the network features few clusters. While this is generally viewed as an unrealistic feature of these models (Goyal 2023), in our context a lack of clustering is natural: If the network were highly clustered, many SAWs would get "stuck" in their cluster, meaning few really successful firms could emerge.

Our model naturally leaves quite some detail unexplained. Our reliance on region fixed effects suggests that some countries have more billionaires than they "deserve" based on fundamentals. Inspecting the standardized fixed effects, it becomes clear that Scandinavian countries as well as small low-tax regions like Israel, Hong Kong, Singapore and Switzerland overperform. It is plausible that institutional differences make up the bulk of this variation in fixed effects, with high-fixed effect regions having favorable tax regimes and other institutional settings that attract billionaires to reside there.

Related Literature: We build on three literatures. First, we contribute to the literature on cross-country inequality differences. Much of the literature has focused on variation in income levels (e.g., Acemoglu and Ventura 2002; Teulings and van Rens 2008; Lakner and Milanovic 2016; Chancel and Piketty 2021), with data

[1]See Melitz (2003) for the seminal contribution.

limitations precluding extensive study of wealth inequality. The few papers which do study wealth from a global comparative perspective include Davies, Sandström, Shorrocks, and Wolff (2011) and Bauluz, Blanchet, Martínez-Toledano, and Sodano (2022). We contribute by studying a well-defined and interesting part of the wealth distribution, namely the very upper tail, with one of the few cross-country datasets fit for this purpose.

Second, we contribute to the literature which seeks to explain top wealth inequality and its increase. As surveyed in Benhabib and Bisin (2018), many models either take the incomplete-markets or random-growth form (see e.g. Benhabib, Bisin, and Zhu 2011; Achdou et al. 2022); both of which are not capable of explaining cross-country differences since all variation in wealth is due to different realizations of the stochastic processes underlying wealth accumulation (labor income risk in a standard Bewley-Aiyagari-Huggett model; stochastic rates of return in Benhabib, Bisin, and Zhu 2011). We relate more closely to models with entrepreneurs or firm owners (e.g, Quadrini 1999; Cagetti and De Nardi 2006; Aghion et al. 2019; Jones and Kim 2018). Novel to our approach is using the Gompertz distribution to show that the wealth distribution is more unequal for more populous countries.

Finally, we relate to the burgeoning literature on networks in (international) macroeconomics (Goyal 2023). Prominent work has focused on input-output linkages (e.g., Baqaee and Farhi 2019a; Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi 2012; Carvalho and Gabaix 2013; Liu and Tsyvinski 2024). While the use of networks in this chapter is more abstract than in these papers, we do have concrete economic interpretations in mind in the style of Melitz (2003)-style models where market size and access to global markets determines domestic firm size and dispersion.

Chapter Outline: Section 3.2 recapitulates the main arguments in Chapter 2. Section 3.3 presents our model, while Section 3.4 analyzes its identification and estimation. In Section 3.5, we discuss our data. Section 3.6 presents the results. We use our model for counterfactual exercises in Section 3.7. We discuss and interpret our results in Section 3.8.

3.2. Stylized Facts

We need a model to match the following facts:

Fact # 1: Billionaire Wealth is not Pareto, but Weibull. It might seem strange to model log wealth as Gompertz, given that the Pareto distribution for wealth is the default (and hence for log wealth, the exponential distribution). However, Pareto is typically taken for granted and not tested. We refer the reader to Chapter 2 for our full testing procedure, including all robustness checks and discussion of measurement error. Briefly, our testing procedure is based on a scaled ratio of log moments. If wealth is Pareto, log wealth is exponential. The exponential distribution has the attractive feature that all moments are well-defined (unlike for Pareto, where only moments up to α^{-1} are defined, where α is the inverse tail index coefficient). The exponential distribution has the moment function $\mathrm{E}\left[w^k\right] = \alpha^k k!$ for integer moment k. Our test of Pareto essentially divides the left-hand side of the moment function by the right-hand side, where we replace α by its maximum likelihood estimator, mean log wealth. Under the null that wealth is Pareto, this test statistic should therefore equal 1.

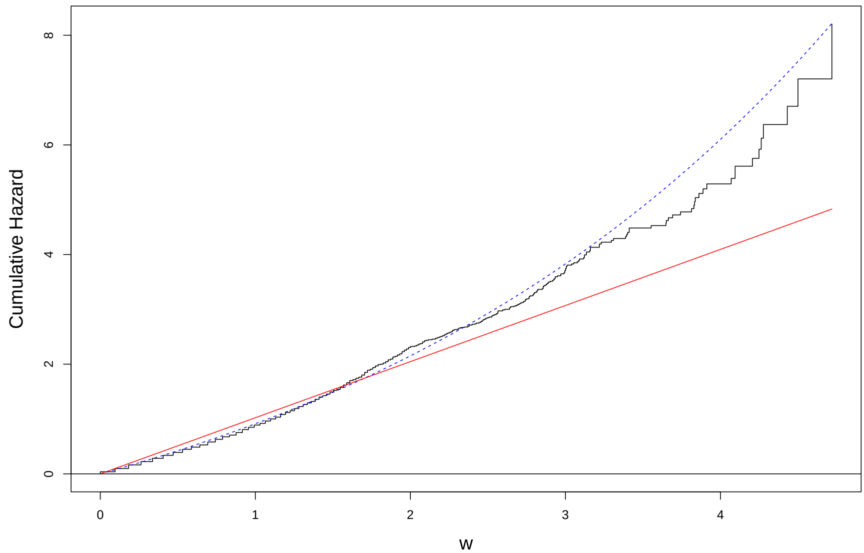
We find robustly, across all regions and years, that Pareto is rejected, i.e., the test statistic is significantly different from 1. Moreover, the rejection is systematic: the test based on the second moment (i.e., $k = 2$) delivers a range of statistics centered around 0.85, whereas the third-moment test is centered around 0.65. This suggests persistent deviations from Pareto, where higher-order moments die out relative to lower-order moments; in other words, there are "too few" super-rich billionaires relative to "ordinary" billionaires. This suggests a systematically increasing hazard of log wealth. Of all the distributions within the exponential family which nest exponential as a special case, the only distribution consistent with these fact is the *Gompertz* distribution.

The defining feature of Gompertz is an exponentially increasing hazard rate. We visualize this feature by plotting the empirical cumulative hazard rates of the global wealth distribution in 2018. As seen in Figure 3.2, the empirical cumulative hazard is very closely matched by the parametric Gompertz hazard (in blue). By contrast, if wealth would be Pareto, the hazards should be close to the parametric hazard of the exponential distribution (in red). This is clearly not the case.

If log wealth is Gompertz, wealth in levels is (truncated-)Weibull. In Chapter 2, we provide further evidence based on cross-equation restrictions that Weibull provides a better fit to the data than Pareto. We also find this rejection of Pareto and the better fit of Weibull for city size and, importantly for this chapter, firm size. The model in the present chapter is intended to rationalize the Gompertz distribution of log wealth.

Fact # 2: Large Variation in Billionaire Numbers and Mean Log Wealth. We can observe the trends for billionaire numbers in Figure 3.1. The ratio of billionaires to population has quadrupled, with large variation between regions. Even zooming in within these regions, there is substantial heterogeneity. For instance, within Europe, there are large differences between, say, Italy and Switzerland. We can

Figure 3.2: Empirical and Parametric Hazard Rates, Billionaire Log Wealth Distribution

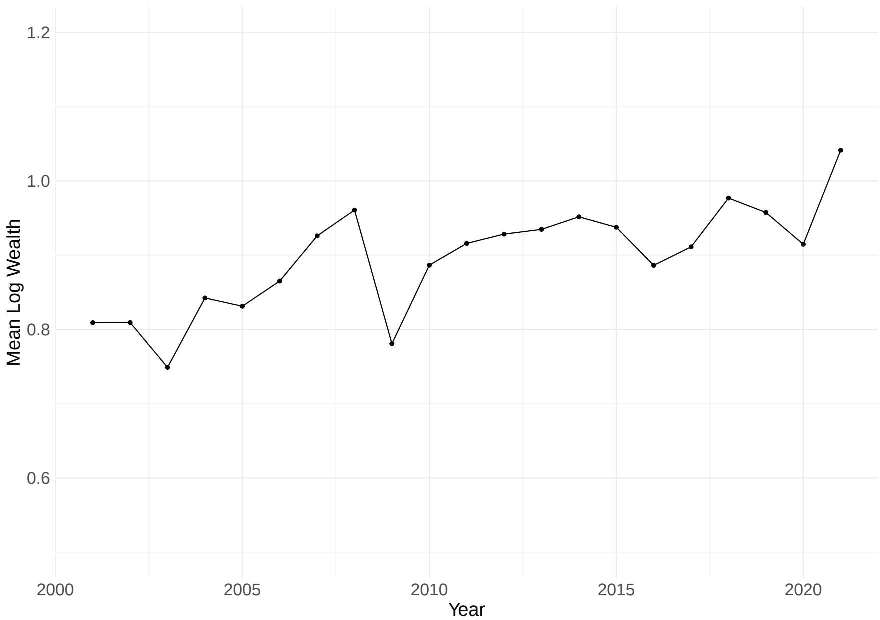


Notes: Figure plots the Kaplan-Meier estimator of the empirical cumulative hazard of the billionaire log wealth distribution, using the 2018 *Forbes List of Billionaires* and pooling all observations. The blue dotted line is a fitted Gompertz hazard, and solid red is an Exponential hazard.

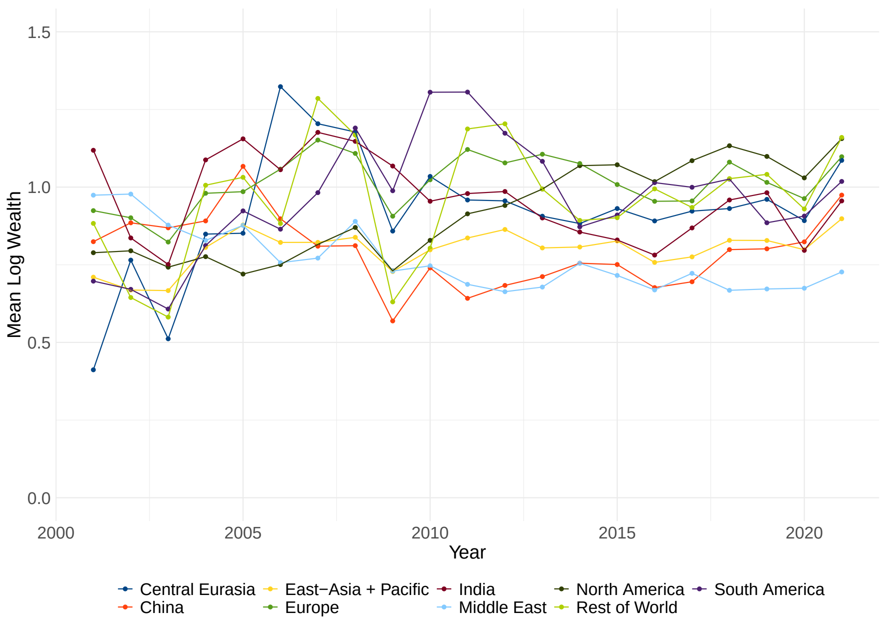
also investigate whether billionaires' wealth increased. Figure 3.3 shows global and regional trends in mean log wealth, analogous to Figure 3.1.

We again observe substantial increases over time. Pooling all observations, panel (a) shows that billionaires in 2001 owned on average $\exp(0.809) = 2.25$ billion USD; by 2021, this had increased to $\exp(1.04) = 2.8$ billion, almost a 25% increase. As panel (b) shows, these global trends mask substantial regional variation. There is a marked decline around the 2008 crisis, covered by a quick rebound. There is a sharp increase from 2020 to 2021; panel (b) shows that this increase is common to most regions (with the partial exception of the Middle East).

Figure 3.3: Global and Regional Trends in Billionaire Mean Log Wealth



(a) Global



(b) Regional Variation

3.3. The Model

Consider an economy in region j in year t with population $L_{jt} > 0$. We typically think of a region as a large country or a group of geographically close countries that are similar in economic characteristics, as in Chapter 2. However, we will also test the validity of our model at the country level. This economy consists of a number of Erdős-Rényi networks $\text{ER}\left[L_{jt}, p_{jt}\right]$ where L_{jt} is the number of nodes in this network and where $p_{jt} \in (0,1)$ is probability that there exists a link between two nodes. The expected number of links of a node to other nodes c_{jt} satisfies:

$$c_{jt} = p_{jt}\left(L_{jt} - 1\right) \cong p_{jt}L_{jt} \tag{3.1}$$

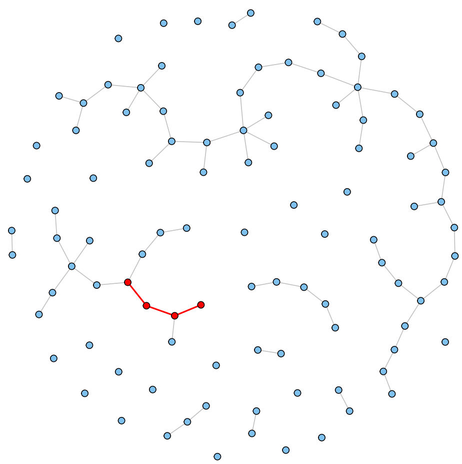
where the latter approximation applies for large L_{jt}, as in our application. We assume $c_{jt} > 1$, such that a giant cluster component arises in each network including a macroscopic fraction of the nodes; that is, a nonvanishing fraction of nodes are connected as L_{jt} gets large. One interpretation of these networks is that each network represents the customers of a particular industry. A link represents for example either a shared technology of two nodes which requires inputs from that industry, a flow of information between two nodes inducing them to buy the same product, or other factors that coordinate the consumption patterns of two nodes. Though there presumably are other interpretations of this network structure, we use this interpretation for explaining the mechanics of the model. In each industry — that is: on each $\text{ER}\left[L_{jt}, p_{jt}\right]$ network — there is a dominant supplier. The customer-base of this dominant supplier is represented by a *Self-Avoiding Walk* (SAW) on this network with length $\underline{z}$ (throughout the chapter random variables are underlined). A SAW is a random walk along the links of the network from one node to another, where the walker cannot revisit a node that she has visited before; the walk ends when the walker arrives at a node without links to nodes that she has not visited before, see Goyal (2023) for a discussion and Figure 3.4 for an illustration.[2]

If p_{jt} were independent of log population size $\ell_{jt} := \ln L_{jt}$, equation (3.1) implies that the expected number of links of a node c_{jt} would be proportional to L_{jt}. Throughout, lower cases denote the log of the corresponding upper case. This is unrealistic: one would not expect the number of links of people in large countries to rise proportionally to the size of the population. It would imply, for example, that people in China have more than a hundred times more links than people in Sweden. Hence, we assume that p_{jt} depends negatively on ℓ_{jt}:

$$\ln p_{jt} = \ln p\left(\ell_{jt}\right) = \gamma_0 - \gamma\ell_{jt}, \tag{3.2}$$

[2] Terminology differs slightly between fields, with some authors calling a SAW a *path*. We stick to SAWs to emphasize the implication that a node cannot be visited twice.

Figure 3.4: Self-Avoiding Walk on an Erdős-Rényi Network



Notes: Figure shows an ER[100, 0.015] network. There are many self-avoiding walks on this network; one particular SAW (starting from the left-most node) is shown in red.

where $\gamma_0 \in \mathbb{R}$ and $\gamma \in (0,1)$ are parameters (throughout the chapter, parameters are denoted by Greek letters). If γ were 0, then p_{jt} would be independent of ℓ_{jt} and c_{jt} proportional to L_{jt} indeed. If γ were 1, then p_{jt} would be inversely related to L_{jt} and the expected number of links c_{jt} independent of ℓ_{jt};[3] for all intermediate cases, c_{jt} increases sublinearly in L_{jt}.

Tishby, Biham, and Katzav (2016) show that the right tail of the distribution of $\underline{z}$ converges to Gompertz with complement distribution function:

$$\Pr\left[\underline{z} \geq z\right] = \exp\left(-\frac{e^{p(\ell_{jt})z} - e^{p(\ell_{jt})d(\ell_{jt})}}{p\left(\ell_{jt}\right)} e^{-p(\ell_{jt})h(\ell_{jt})}\right). \tag{3.4}$$

[3] As is well known, in this case, the degree distribution converges to Poisson

$$\Pr\left[\underline{c} = c\right] = \frac{e^{-\mu}\mu^c}{c!} \tag{3.3}$$

where c is a particular realisation of $\underline{c}$ and $\mu := \mathrm{E}[\underline{c}]$ is the mean degree. Clearly, the distribution of $\underline{c}$ is increasing in μ, and is therefore an increasing function of market size.

where $d\left(\ell_{jt}\right)$ and $h\left(\ell_{jt}\right)$ are functions of $\ell_{jt} \equiv \ln L_{jt}$. Let $d^*\left(\ell_{jt}\right) > d\left(\ell_{jt}\right)$ be the lower bound above which the distribution of $\underline{z}$ has converged to the Gompertz distribution. Hence, the probability that a random draw from the full distribution of $\underline{z}$ is drawn from this right tail is:

$$\Pr\left[\underline{z} \geq d^*\left(\ell_{jt}\right)\right] = \exp\left(-\frac{e^{p\left(\ell_{jt}\right)d^*\left(\ell_{jt}\right)} - e^{p\left(\ell_{jt}\right)d\left(\ell_{jt}\right)}}{p\left(\ell_{jt}\right)} e^{-p\left(\ell_{jt}\right)h\left(\ell_{jt}\right)}\right).$$

We shall assume that all billionaires have realisations of $\underline{z} \geq d^*\left(\ell_{jt}\right)$, such that the distribution of $\underline{z}$ follows Gompertz for this group. Hence, this lower bound is non-binding for our empirical application and we can ignore it. We present it here just as a warning that this model cannot be applied for the wealth distribution of the full population, since many people have realisations of $\underline{z} < d^*\left(\ell_{jt}\right)$, where Gompertz does not apply.

Each individual i in region j at time t "owns" an SAW with length z_{ijt}; z_{ijt} is her income relative to the average income Y_{jt} in region j at time t. Wealthy individuals can capitalize their income into wealth at a global wealth-income ratio K_t. Individual i's wealth W_{ijt} is therefore equal to:

$$W_{ijt} = e^{z_{ijt}} Y_{jt} K_t \Rightarrow z_{ijt} = w_{ijt} - y_{jt} - k_t. \tag{3.5}$$

Substitution of equation (3.5) in equation (3.4) yields:

$$\Pr\left[\underline{w} \geq w_{ijt}\right] = \exp\left(-\frac{e^{p_{jt}w_{ijt}} - e^{p_{jt}d_{jt}}}{p_{jt}} e^{-p_{jt}h_{jt}}\right) \tag{3.6}$$

where

$$h_{jt} := h\left(\ell_{jt}\right) + y_{jt} + k_t = \eta_0 + y_{jt} + k_t + \eta\ell_{jt}, \tag{3.7}$$

$$d_{jt} := d\left(\ell_{jt}\right) + y_{jt} + k_t = \delta_0 + y_{jt} + k_t + \delta\ell_{jt}. \tag{3.8}$$

In the second step, we use linear approximations of $h\left(\ell_{jt}\right)$ and $d\left(\ell_{jt}\right)$; $p_{jt} = p\left(\ell_{jt}\right)$ safisfies equation (3.2).

For our purpose, it is convenient to denote all variables in billions of nominal US dollars. A billionaire is therefore someone for whom $W_{ijt} \geq 1$ or $w_{ijt} \geq 0$. Equation (3.6) implies:

$$\frac{\Pr\left[\underline{w} = w_{ijt}\right]}{\Pr\left[\underline{w} \geq w_{ijt}\right]} = e^{p_{jt}\left(w_{ijt} - h_{jt}\right)}, \tag{3.9}$$

$$\Pr\left[\underline{w} \geq 0\right] = \exp\left(\frac{e^{p_{jt}d_{jt}} - 1}{p_{jt}} e^{-p_{jt}h_{jt}}\right). \tag{3.10}$$

Equation (3.9) is the hazard rate of the distribution of top wealth, equation (3.10) is the probability that somebody is a billionaire. These equations provide an appealing interpretation of the roles of p_{jt}, h_{jt}, and d_{jt} in our model: p_{jt} is the elasticity of the hazard rate with respect to W_{jt}; a higher value of p_{jt} corresponds to a faster increase in the hazard rate. Conditional on p_{jt}, h_{jt} measures the level of the inverse hazard rate for $w_{ijt} = 0$; the higher h_{jt}, the lower is the hazard rate and the fatter is the right tail of the distribution of top wealth and the higher expected (log) top-wealth of billionaires. Conditional on p_{jt} and h_{jt}, d_{jt} measures the probability that somebody is a billionaire.

The critical difference between the exponential distribution (the distribution of log wealth when wealth itself is distributed Pareto) and the Gompertz distribution is the shape of hazard rate. To see this, define $a_{jt} := \mathrm{e}^{-p_{jt}h_{jt}}$. The distribution function for the exponential distribution is obtained by taking the limit $p_{jt} \to 0$ while keeping $a_{jt} = \mathrm{e}^{-p_{jt}h_{jt}}$ constant:

$$\lim_{p_{jt}\to 0} \Pr\left[\underline{w} \geq w_{ijt}\right] = \lim_{p_{jt}\to 0} \exp\left(-\frac{\mathrm{e}^{p_{jt}w_{ijt}} - \mathrm{e}^{p_{jt}d_{jt}}}{p_{jt}} a_{jt}\right) = \mathrm{e}^{-a_{jt}\left(w_{ijt}-d_{jt}\right)}.$$

The hazard rate of Gompertz and the exponential distribution can therefore be written as:

$$\frac{\Pr\left[\underline{w} = w_{ijt}\right]}{\Pr\left[\underline{w} \geq w_{ijt}\right]} = \begin{cases} \text{exponential} & : \quad a_{jt} \quad = \quad \mathrm{e}^{-p_{jt}h_{jt}} \\ \text{Gompertz} & : \quad a_{jt}\mathrm{e}^{p_{jt}w_{ijt}} \quad = \quad \mathrm{e}^{p_{jt}\left(w_{ijt}-h_{jt}\right)} \end{cases}.$$

While the hazard rate is constant for the exponential distribution, it is increasing exponentially for the Gompertz distribution. The Gompertz distribution therefore has a thinner right tail than the exponential distribution. In fact, Gompertz has an even thinner tail than the Normal distribution, as can be seen from

$$\lim_{w\to\infty} w^{-1}\frac{\Pr\left[\underline{w} = w\right]}{\Pr\left[\underline{w} \geq w\right]} = \lim_{w\to\infty} w^{-1}\frac{\phi\left(w\right)}{\Phi\left(-w\right)} = 1,$$

where $\phi\left(\cdot\right)$ and $\Phi\left(\cdot\right)$ denote the density and distribution function respectively of the standard Normal distribution. The hazard rate of the Normal distribution increases linearly with w for large w, while it increases exponentially for Gompertz. The fact that our empirical results strongly support a Gompertz rather than an exponential distribution for log top-wealth — and hence a Weibull rather than a Pareto distribution for the level of top-wealth — therefore comes as a surprise, since the common wisdom holds that top wealth is fat tailed, while the empirical evidence in Chapter 2 suggests that log top-wealth is even thinner tailed than the Normal distribution.

The model presented above provides an interpretation of this phenomenon. In the extreme right tail, the size of the population imposes a capacity constraint on the potential of a dominant supplier to increase its customer base even further and hence on her ability to increase her wealth. The larger the size of the population, the higher the level of wealth at which this capacity constraint becomes binding and the more skewed to the right is therefore the distribution of top-wealth.

The moments for $k > 0$ of $\underline{W}$ and $\underline{w}$ read:

$$\mathrm{E}\left[\underline{W}^k\right] = G_{jt}^{k/p_{jt}} \mathrm{e}^{G_{jt}^{-1}} \Gamma\left(1 + k/p_{jt}, G_{jt}^{-1}\right), \tag{3.11}$$

$$\mathrm{E}\left[\underline{w}^k\right] = p_{jt}^{-k} h\left(G_{jt}, k\right), \tag{3.12}$$

$$G_{jt} := p_{jt} \mathrm{e}^{p_{jt} h_{jt}},$$

where $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function and where

$$h(G, k) := k \mathrm{e}^{G^{-1}} \int_0^\infty q^{k-1} \exp\left(-G^{-1} \mathrm{e}^{-q}\right) \mathrm{d}q,$$

$$h(G, 1) = \mathrm{e}^{G^{-1}} \mathrm{Ei}\left(G^{-1}\right),$$

where $\mathrm{Ei}(\cdot)$ is the exponential integral, see Chapter 2.

3.4. Estimation

3.4.1. Parameter Estimation

Our model of the top wealth distribution has three reduced-form parameters for each economy $\{j, t\}$: p_{jt}, h_{jt}, and d_{jt}. The variation in these reduced-form parameters between economies is determined by six structural parameters: the intercepts γ_0, η_0, and δ_0, and the coefficients γ, η, and δ measuring the effect of ℓ_{jt}. We first discuss the estimation of the reduced form parameters p_{jt}, h_{jt}, and d_{jt} for each economy and then discuss how we estimate the structural parameters from the variation in these reduced-form parameters.

The parameters p_{jt} and h_{jt} are estimated by maximum likelihood estimation from the distribution of w_{ijt} for each $\{j, t\}$, see Chapter 2 for a more detailed discussion. The density function of the conditional Gompertz distribution for billionaires reads:

$$f(w_{ijt}) = a_{jt} \exp\left(p_{jt} w_{ijt} - a_{jt} \frac{\mathrm{e}^{p_{jt} w_{ijt}} - 1}{p_{jt}}\right).$$

Hence, the log likelihood reads:

$$N_{jt}^{-1} \ln \mathcal{L}\left(a_{jt}, p_{jt}\right) = \ln a_{jt} + p_{jt} \overline{w_{jt}} - a_{jt} \frac{\overline{\mathrm{e}^{p_{jt} w_{jt}}} - 1}{p_{jt}}, \tag{3.13}$$

where a bar on top of a variable denotes its mean for region-year $\{j,t\}$: $\overline{x}_{jt} := N_{jt}^{-1}\sum_i x_{ijt}$, where N_{jt} is the number of billionaires in $\{j,t\}$. The first-order condition for a_{jt} reads:

$$N_{jt}^{-1}\frac{\partial \ln \mathcal{L}\left(a_{jt},p_{jt}\right)}{\partial a_{jt}} = a_{jt}^{-1} - p_{jt}^{-1}\left(\overline{e^{p_{jt}w_{jt}}} - 1\right) = 0$$

$$\Rightarrow \widehat{a}_{jt} = \widehat{p}_{jt}\left(\overline{e^{\widehat{p}_{jt}w_{jt}}} - 1\right)^{-1}, \tag{3.14}$$

where a hat on top of a parameter denotes its maximum likelihood estimator. Substitution of equation (3.14) in equation (3.13) yields the concentrated log likelihood (up to a constant)

$$N_{jt}^{-1}\ln \mathcal{L}\left(p_{jt}\right) = \ln p_{jt} - \ln\left(\overline{e^{p_{jt}w_{jt}}} - 1\right) + p_{jt}\overline{w_{jt}}.$$

The first-order condition reads

$$N_{jt}^{-1}\frac{\mathrm{d}\ln \mathcal{L}\left(p_{jt}\right)}{\mathrm{d}p_{jt}} = p_{jt}^{-1} - \frac{\overline{w_{jt}}}{\overline{e^{p_{jt}w_{jt}}} - 1} = 0 \Rightarrow \widehat{p}_{jt} = \frac{\overline{e^{\widehat{p}_{jt}w_{jt}}} - 1}{\overline{w_{jt}}}. \tag{3.15}$$

We first calculate $\widehat{p}_{jt}$ as the solution to equation (3.15), then use this solution to calculate $\widehat{a}_{jt}$ as the solution to equation (3.14) and to solve $\widehat{h}_{jt}$ from $\widehat{h}_{jt} = -\widehat{p}_{jt}^{-1}\ln\widehat{a}_{jt}$. Finally, we use the solutions for $\widehat{p}_{jt}$ and $\widehat{a}_{jt}$ to solve $\widehat{d}_{jt}$ from equation (3.10)

$$\widehat{d}_{jt} = \widehat{p}_{jt}^{-1}\ln\left(1 + \frac{\widehat{p}_{jt}}{\widehat{a}_{jt}}\ln\left(\frac{N_{jt}}{L_{jt}}\right)\right), \tag{3.16}$$

using the fraction of billionaires in the data N_{jt}/L_{jt} for region j at time t as an estimate for $\Pr\left[\underline{w} \geq 0\right]$. We use these estimates for p_{jt}, h_{jt}, and d_{jt} to the estimated structural parameters γ_0, γ, η_0, η, δ_0, and δ in equations (3.2), (3.7) and (3.8) by means of standard linear techniques like OLS or WLS.

3.4.2. Identification

The model has three reduced-form parameters, p, h, and d, which we allow to vary across regions j and time t. Stacking these parameters across economies, we obtain three vectors $\mathbf{p}$, $\mathbf{h}$, and $\mathbf{d}$. By equation (3.2), $\ln p_{jt}$ varies negatively with log population size, ℓ_{jt}. By equations (3.7) and (3.8), the baseline hazard h_{jt} and the conditional billionaire probability d_{jt} are directly affected by log GDP per capita y_{jt} and the log wealth-income ratio k_t. We gather these covariates into a matrix

$\mathbf{X} := \begin{bmatrix} \ell & \mathbf{y} & \mathbf{k} \end{bmatrix}$. We might need (time or region) fixed effects as well. Compactly, our empirical model can be written as

$$\theta = \mathbf{X}'\beta + \zeta + \varepsilon, \tag{3.17}$$

where $\theta := (\ln \mathbf{p}, \mathbf{h}, \mathbf{d})$. We potentially allow the set of covariates $\mathbf{X}$ to vary with the parameter. ζ is a vector of fixed effects, which can include region fixed effects, year fixed effects, or a constant intercept (if no fixed effects are needed). ε is an error term.

Our null hypothesis is that this model can be simplified to:

$$\begin{aligned} \ln p_{jt} &= \gamma_0 + \gamma \ell_{jt} + \varepsilon^p_{jt} \\ h_{jt} &= \eta_0 + y_{jt} + k_{jt} + \varepsilon^h_{jt} \\ d_{jt} &= \delta_0 + y_{jt} + k_{jt} + \varepsilon^d_{jt} \end{aligned}$$

where $\gamma < 0$. The motivation for assuming $\gamma < 0$ comes from equation (3.2). We use log population size ℓ as a proxy for log network size. The motivation for assuming y and k to have unit coefficients comes from the assumption of a common income distribution across j and t. Suppose there is neutral inflation raising both log GDP y_{jt} and log wealth w of all billionaires living in economy $\{j, t\}$ by some additive constant π. A unit coefficient implies that this affects neither the number of "real" (that is: corrected for inflation) billionaires nor the distribution of their log wealth. Alternatively, this assumption states that the distribution of wealth is independent of log nominal GDP per capita and the log wealth-income ratio. These assumptions will by and large hold, except that we will have to allow for region fixed effects, since our parsimonious model will not capture all cross-sectional variation in these reduced-form parameters.

3.5. Data

We use the *Forbes List of Billionaires* for the years 2001–2021. The dataset provides the names of billionaires, their net worth, their country of origin, their age and their citizenship. In the years 2011 to 2021, the dataset also provides information on the origin of the wealth of the listed billionaires. We classify billionaires according to their citizenship.

Forbes calculates net worth at the individual level, but aggregates family wealth under one person. Understandably, their methodology is not fully transparent, as the composition of their list requires idiosyncratic choices for each individual billionaire. Forbes splits family wealth if each family member has 1 billion USD or more after the split. As observed by Piketty (2014), this is likely to create an upward bias on individual fortunes. Moreover, they use available documentation and sometimes data provided by billionaires themselves to estimate their net worth. This implies that the number of billionaires is likely to be underestimated, especially in less developed countries or if wealth is derived from nefarious activities. Nevertheless, we follow the existing literature which uses rich lists like Forbes, as other sources are likely to underestimate the number of billionaires (Vermeulen 2016; Novokmet, Piketty, and Zucman 2018; Piketty, Yang, and Zucman 2019; Gomez 2023). For a lack of a more rigorous account of billionaire wealth, we take the Forbes data as given.

We cluster countries in regions. The guiding principles for this clustering is to merge countries that are geographically connected and close in terms of GDP per capita. With this in mind we define eighteen regions as geographical units consisting of one or a small number of countries. As a rough threshold for the minimum number of billionaires for a country or a group of countries to be considered as a region we use 40 billionaires in 2019. Countries that cannot easily be included in a region are excluded from the region classification. Hence, the region classification excludes countries with a small number of billionaires which cannot be easily grouped with other countries. Table 3.1 gives an overview of these regions.

We proxy mean income by GDP per capita, presuming that depreciation is an approximately fixed share of GDP. Since we apply $\ln Y_{jt}$ in all equations, a fixed share will only affect the intercept of the regressions. We construct population and GDP data for all regions and regions for the period 2001–2021 using the World Bank's population and GDP data. The global wealth-income ratio is based on the US because of this country has the largest and most developed capital market and it therefore is the centre of the world's financial system. The ratio is computed as the market capitalization of the US stock market in at the first of January relative to US GDP. This definition fits the assumptions of our model of a common income distribution. Hence, the income corresponding to the market capitalization at the US stock market is proportional to U.S. GDP. Clearly, this is a simplification, since part of the market capitalization at the U.S. stock market reflects income earned in other countries. An alternative for U.S. GDP as the measure of capitalized income would be US dividends. Dividends are more procyclical than GDP. The choice between both measures is therefore analogous to that between the standard price-to-dividend ratio and Shiller's Cyclically Adjusted Price Earnings (CAPE) ratio, see Campbell and Shiller (1988a, 1988b). Campbell and Shiller argue that the standard price-to-dividend ratio is less appropriate. Figure 3.5 plots the US market capitalisation to GDP ratio against CAPE. We plot both ratios relative to their value for 2019. Both indices move in parallel, showing tops during the dot-com

Table 3.1: Region Classification & Descriptive Statistics

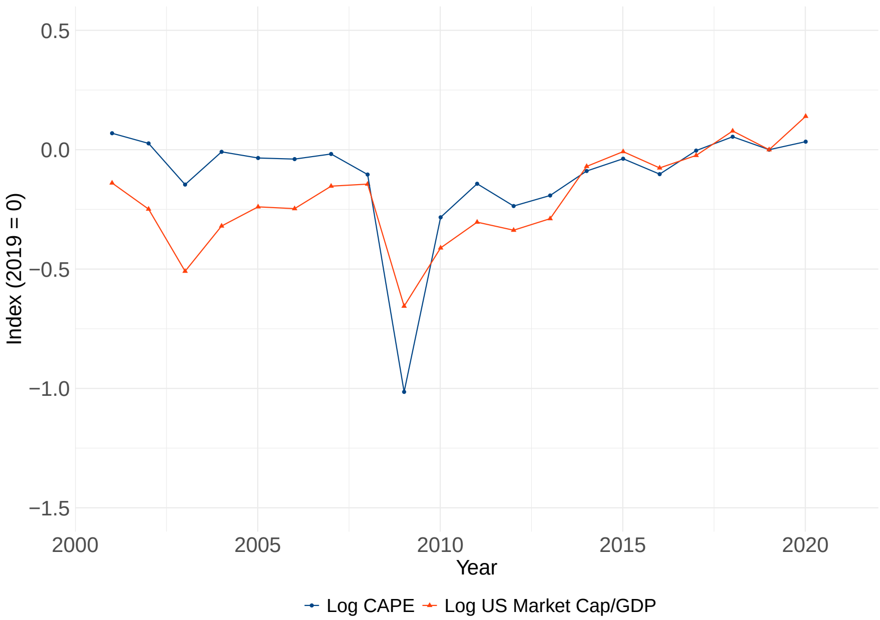
Region Classification		Statistics (2001–2021 Average)			
Region	Area	Y	$\overline{w}$	ℓ	N
North America					
U.S.		0.822	0.923	19.6	443
Canada		0.678	0.894	17.4	29.2
Europe					
Germany		0.651	1.12	18.2	74.1
British Islands	U.K. + Ireland	0.682	0.828	18	40.6
Scandinavia	Sweden + Denmark + Norway + Finland	0.878	1.17	17.1	30.6
France	incl. Monaco	0.605	1.27	18	26.5
Alps	Switzerland + Austria + Liechtenstein	0.939	1.05	16.6	25
Italy		0.526	1.01	17.9	24.1
(South)East Asia & Oceania					
China	excl. Taiwan, incl. Hong Kong	0.089	0.794	21	188
Southeast Asia	Thailand + Malaysia + Singapore	0.129	0.923	18.4	31.1
Asian Islands	Taiwan + Philippines + Indonesia	0.06	0.727	19.7	38.5
South Korea		0.389	0.673	17.7	18.7
Japan		0.627	0.875	18.7	26.6
Australia		0.754	0.736	16.9	19
India		0.02	0.964	20.9	56.1
Other					
Russia		0.151	0.956	18.8	68.8
Brazil		0.128	0.861	19.1	30
Israel+Turkey		0.18	0.585	18.2	36.2
World		**0.15**	**0.896**	**22.7**	**1325**

Notes: World totals include billionaires from countries not part of the regions. China and India count both as regions and regions. N = total number of billionaires; ℓ is log population; $\overline{w}$ = mean log wealth. GDP per capita Y is normalized by the GDP per capita of the U.S. in 2018.

bubble in 2001 and the second IT wave after 2017 and troughs after the bursting of the dot-com bubble in 2003 and after the 2008 financial crisis. This shows that our measure, like the CAPE, is a better measure of the underlying capitalization possibilities in financial markets than the raw price-dividend ratio.

Table 3.1 provides summary statistics for all regions and regions. We average all statistics over the 2001–2021 period. We observe wide heterogeneity. We report GDP per capita (Y) relative to the U.S. in 2018 for comparability. European regions have GDP per capita close to or above the U.S.; in contrast, India had on average only 2% of U.S. GDP per capita. Likewise, mean log wealth $\overline{w}$ varies from less than 0.6 in Israel+Turkey to almost 1.3 in France (Bernard Arnault again). That log population ℓ varies widely between, say, China and the Alps needs no further comment.

Figure 3.5: Trends in Stock Market Capitalization



Notes: Figure shows the log Cyclically Adjusted Price-Earnings index from Robert Shiller's website (http://www.econ.yale.edu/ shiller/data.htm), as well as log US Market Capitalization/US GDP. Both series have been normalized such that 2019 equals 0.

3.6. Results

We test the model by running weighted least squares (WLS) regressions for all region-year observations for which the number of billionaires, N_{jt}, is large enough. Concretely, we run our model on observations with $N_{jt} \geq 32$. In addition, our maximum likelihood estimates result in some estimates for p numerically identical to zero. Were we to take the log of these observations, they would become very negative indeed, distorting our results. We overcome this by trimming the top and bottom 5% of all observations.[4] We weight observations by the square root number of billionaires, $\sqrt{N_{jt}}$ to increase efficiency. To account for autocorrelation and potential cross-sectional correlation, we report Driscoll and Kraay (1998) standard

[4]The observations dropped by this trimming are: The U.S. in 2003 and 2005; the British Islands in 2007 and 2013; the Alps in 2015, 2017, 2018, 2019, and 2020; China in 2009, 2011, and 2012; South Korea in 2019 and 2021; Japan in 2021; India in 2010; Russia in 2006 and 2009; and Israel+Turkey in 2011.

errors, which generalize Newey and West (1987) standard errors to allow for arbitrary spatial correlation as well as autocorrelation. All variables in all specifications will be demeaned, such that the intercept can be interpreted as the average effect.

We begin with our regressions on $\ln p_{jt}$, reported in Table 3.2. Column (1) begins by testing all variables in $\mathbf{x}$. Recall our hypothesis: Only log population ℓ should matter for $\ln p_{jt}$, and the coefficient should be negative. This hypothesis is clearly supported by the data. All other variables are insignificant, and a formal F-test accepts the restriction that they are jointly zero. We therefore can restrict ourselves to our hypothesized model in column (2), with only log population. The coefficient is stable, going from -0.32 to -0.28.

We then sequentially test whether adding fixed effects significantly improves the model fit. Column (3) reports the results of adding time fixed effects. The restriction that these effects are jointly insignificant is upheld (F-test value of 1.782). The coefficient on log GDP per capita remains insignificant (the log wealth-income ratio drops out). Hence, our minimalist model does not need time fixed effects. This is rather different for region fixed effects, as shown in columns (5) and (6). The F-statistic of 8.4 is clearly significant. We can also directly observe from the significant reduction in the RMSE and the improved R^2 that region fixed effects significantly improve the model fit. The coefficients on log GDP per capita and the log wealth-income ratio remain insignificant. More concerning is the fact that log population, while significant, has flipped sign and increased by an order of magnitude. This is not easy to interpret; it indicates that within-region variation in the level of log population has a positive association with $\ln p_{jt}$. Conceptually, if we only would use within-region variation, it would make more sense to evaluate the effect of a change in log population. This is what we do in column (8), where we time-average all variables, and then run a cross-sectional regression. Since there is no cross-sectional variation in k, this variable drops out. We observe that the cross-sectional relationship between log population and $\ln p_{jt}$ are again as predicted, given by the coefficient on the time-averaged $\overline{\ell}_j$. Moreover, the deviations from the cross-sectional averages, $\widetilde{\ell}_{jt}$, show up positively and significantly, and are quantitatively similar to the results found in columns (5) and (6) with region fixed effects. We conclude that the main predictions of our model hold.

Next, we study the results for the conditional inverse baseline hazard, h. In calculating h, using the procedure described in Section 3.4, we use the fitted values of $\widehat{p}_{jt}$ from Table 3.2, column (7).

Table 3.3 reports the results of regressing our fitted $\widehat{h}_{jt}$ on $\mathbf{x}$. Our hypothesis was that log GDP per capita and log market capitalisation solely determine $\widehat{h}_{jt}$, both with unit coefficients. Column (1) shows that this does not quite hold. The

Table 3.2: Regressions on $\ln \widehat{p}_{jt}$

Dependent Variable:	$\ln \widehat{p}_{jt}$						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-1.2***	-1.2***					3.8***
	(0.06)	(0.05)					(0.77)
ℓ_{jt}	-0.32***	-0.28***	-0.29***	-0.28***	6.8*	7.2**	
	(0.07)	(0.04)	(0.05)	(0.04)	(2.7)	(2.1)	
y_{jt}	-0.08		-0.03		0.14		
	(0.06)		(0.03)		(0.27)		
k_t	0.57				-0.006		
	(0.32)				(0.50)		
$\overline{\ell}_j$							-0.28***
							(0.04)
$\widetilde{\ell}_{jt}$							5.9***
							(1.5)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fixed-effects							
Year			✓	✓			
Region					✓	✓	
Fit statistics							
Observations	173	173	173	173	173	173	173
R^2	0.221	0.189	0.365	0.363	0.613	0.612	0.307
Within R^2			0.224	0.222	0.208	0.206	
RMSE	2.05	2.09	1.85	1.85	1.44	1.45	1.93
F-test (only $\ell \neq 0$)	1.953		1.175		0.246		
F-test (FE)			1.782		8.408***		

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Driscoll-Kraay standard errors in parentheses. ℓ_{jt} = log population, y_{jt} = log GDP per capita, k_t = log global wealth-income ratio. Column (7) splits log population ℓ_{jt} in a region-specific average $\overline{\ell}_j := T_j^{-1} \sum_t^{T_j} \ell_{jt}$, and deviations from that mean $\widetilde{\ell}_{jt} := \ell_{jt} - \overline{\ell}_j$.

Table 3.3: Regressions on $\widehat{h}_{jt}$

Dependent Variable:	$\widehat{h}_{jt}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.36^{***}	0.31^{***}				
	(0.07)	(0.06)				
y_{jt}	0.33^{*}	0.41^{**}	0.36^{*}	0.44^{***}	0.56	0.66
	(0.14)	(0.11)	(0.13)	(0.10)	(0.73)	(0.71)
k_t	1.1^{**}	1.1^{**}			1.2^{*}	1.3^{**}
	(0.33)	(0.31)			(0.48)	(0.44)
ℓ_{jt}	-0.13		-0.12		1.5	
	(0.07)		(0.07)		(2.2)	
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fixed-effects						
Year			✓	✓		
Region					✓	✓
Fit statistics						
Observations	173	173	173	173	173	173
R^2	0.253	0.238	0.324	0.312	0.711	0.710
Within R^2			0.252	0.238	0.245	0.242
RMSE	2.67	2.69	2.53	2.56	1.66	1.66
F-test ($\ell = 0$)	2.993		2.983		0.481	
F-test ($y = k = 1$)	11.759^{***}	13.68^{***}	25.122^{***}	29.429^{***}	0.176	0.244
F-test (FE)			0.903		26.918^{***}	

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Driscoll-Kraay standard errors in parentheses. ℓ_{jt} = log population, y_{jt} = log GDP per capita, k_t = log global wealth-income ratio.

coefficient on market capitalisation is indeed close to 1, but the coefficient on log GDP per capita is much lower, at 0.33. Log population does enter insignificantly, as predicted. In column (2), we therefore drop log population. The coefficient on log GDP per capita now increases to 0.41, which still far from 1. The F-test therefore does reject the restriction that they are jointly equal to 1. We proceed by imposing the unit coefficient restriction. Column (3) reveals that the RMSE does go up substantially, highlighting that this restriction does not help our model fit.

We test the inclusion of fixed effects. Column (4) shows that time fixed effects again are not jointly significant (F-stat of 0.9). Log population remains insignificant. As before, log market cap drops out of columns (4) and (5). The remaining covariate, log GDP per capita, has similar coefficients to columns (1) and (2), and hence is far from 1. Column (6) shows that region fixed effects are highly significant. Again, log population is insignificant. Interestingly, the coefficient in log GDP per capita has now increased further (although insignificant), and cannot statistically be distinguished from 1. Hence, our F-test on our restriction is now accepted. In columns (8) and (9), we again time-average all variables to create a purely cross-sectional regression. Here again, log population is insignificant, but log GDP per capita is far from one. Therefore, our model predictions for h only hold if we rely on region fixed effects.

Finally, we test our hypotheses for the conditional billionaire probability d_{jt}. Like h_{jt}, this parameter should solely be determined by log GDP per capita and log market capitalisation, with unit coefficients. Unlike our results for $\widehat{h}_{jt}$, this restriction is upheld much more easily for $\widehat{d}_{jt}$. We use the fitted values for $\widehat{h}_{jt}$ from Table 3.3, column (9) to calculate $\widehat{d}_{jt}$. The results are reported in Table 3.4. Column (1) shows that both variables of interest have highly significant coefficients that are close to 1. A formal F-test rejects the restriction that they equal 1. This is unsurprising, since these coefficietns are highly precisely estimated, since the variation in billionaire probabilities is very large across regions and years. This is in contrast to variation in mean (log) wealth, which essentially is the variation used to identify h_{jt}. Hence, we cannot formally accept our restriction, but in magnitudes it almost holds. Log population shows up insignificantly, as hypothesized. We proceed as before. In column (2), we drop log population. The resulting coefficients for log GDP per capita and log market capitalisation are barely changed.

We sequentially add fixed effects as before. Time fixed effects are again not needed, see column (3). Column (5) shows, as before, that region fixed effects are needed. However, note that although the RMSE drops significantly, the R^2 only marginally increases, having been very high even in the non-saturated specifications. This suggests that our simple model without region fixed effects does have strong

Table 3.4: Regressions on $\widehat{d}_{jt}$

Dependent Variable:	$\widehat{d}_{jt}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Constant	1.6***	1.6***				
	(0.009)	(0.009)				
y_{jt}	0.89***	0.88***	0.89***	0.89***	1.2***	1.1***
	(0.01)	(0.008)	(0.009)	(0.007)	(0.08)	(0.07)
k_t	0.74***	0.74***			0.88***	0.82***
	(0.07)	(0.07)			(0.06)	(0.05)
ℓ_{jt}	0.01		0.01		-0.60	
	(0.009)		(0.008)		(0.57)	
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fixed-effects						
Year			✓	✓		
Region					✓	✓
Fit statistics						
Observations	173	173	173	173	173	173
R^2	0.927	0.927	0.930	0.930	0.987	0.987
Within R^2			0.926	0.926	0.879	0.877
RMSE	0.831	0.832	0.814	0.815	0.345	0.347
F-test ($\ell = 0$)	1.965		1.561		1.080	
F-test ($y = k = 1$)	52.548***	108.59***	125.91***	236.76***	2.739	5.589*
F-test (FE)			0.415		69.594***	

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Driscoll-Kraay standard errors in parentheses. ℓ_{jt} = log population, y_{jt} = log GDP per capita, k_t = log global wealth-income ratio.

explanatory power. In columns (5) and (6), we revisit our restrictions, like we did for $\widehat{h}_{jt}$. We again find that region fixed effects suffice to make all restrictions hold in column (5). In column (6), the F-test is marginally significant at the 5% level. Finally, the cross-sectional regressions confirm these findings. Log GDP per capita cannot be statistically distinguished from 1, and log population remains insignificant throughout. The R^2 in the cross-sectional regressions is strikingly large, reflecting that variations in the billionaire probability are sufficiently large to make our model precisely estimated.

Our hypotheses for p_{jt} are easily upheld, while those for h_{jt} and d_{jt} only hold with region fixed effects. We do see that our hypothesis more closely holds for d than for h. This suggests that our model works well for the conditional billionaire probability, but less so for the conditional inverse baseline hazard.

3.7. Accounting for Billionaire Growth

In this section, we test the implications of our model and regression results, both on the level of individual countries and on a global level. First, we test our model on individual countries. We begin by re-estimating all equations at the country level, using the coefficients found before. Specifically, we use a country-year-level $\widehat{p}_{jt} = \exp\left(-1.2 - 0.28\ell_{jt}\right)$, the fitted value of Table 3.2, column (2). Then, we use this $\widehat{p}$ to calculate an observation-specific $\widehat{d}$, which we decompose into a fixed effect (d_j) and the contributions of log GDP per capita y and log market capitalisation k, both with unit coefficients. With p and d, we calculate h as outlined in Section 3.4, which we decompose like d into a fixed effect h_j and the contributions of y and k with unit coefficients.

Once we have country-specific estimates for our parameters, we can test the model implications and compare them to the data. We will look at three key moments: the fraction of the population which is a billionaire (expressed in millions of population), mean log wealth, and mean wealth. To calculate these moments, we use equations (3.10), (3.11) and (3.12). Since we re-estimate all parameters at the country level, these moments should fit very well. The reason is that the billionaire probability is used to estimate h, and transformed mean log wealth is part of the estimate of p, see equation (3.15). However, mean wealth in levels is untargeted, so our model need not necessarily fit this moment well.

Table 3.5 shows our model predictions, where we take the 2001–2021 average of our model predictions and compare them to the averaged empirical data. The model in general fits really well. The billionaire probability is a function of p, d, and h. Since the latter two parameters contain fixed effects (re-estimated at the country

Table 3.5: Predictions versus Realizations, 2001–2021 Average

	$\Pr[\underline{w} \geq 0]$		$\mathrm{E}[\underline{w}]$		$\mathrm{E}[\underline{W}]$		Fixed Effects	
Country	Model	Data	Model	Data	Model	Data	h	d
Australia	0.673	0.756	0.708	0.722	2.61	2.55	0.323	0.752
Brazil	0.122	0.14	0.83	0.858	3.23	3.24	0.984	-0.349
Canada	0.769	0.787	0.871	0.893	3.43	3.33	0.996	0.968
China	0.052	0.096	0.55	0.496	2.06	2.04	-0.267	-0.517
France	0.341	0.36	1.26	1.33	6.87	6.32	2.62	0.925
Germany	0.831	0.865	1.07	1.12	4.86	4.28	1.88	1.41
Hong Kong	4.87	5.2	1.04	1.07	4.62	4.41	1.87	2.57
India	0.034	0.040	0.89	0.964	3.62	3.9	1.38	-1.97
Italy	0.366	0.383	0.97	1.02	4.06	3.93	1.44	0.557
Japan	0.199	0.2	0.863	0.876	3.36	3.18	0.877	-0.16
Mexico	0.107	0.108	1.18	1.2	5.98	7.21	2.29	-0.719
Russia	0.405	0.462	0.927	0.949	3.86	3.78	1.47	0.832
South Korea	0.283	0.345	0.619	0.672	2.26	2.36	-0.21	0.144
Spain	0.356	0.369	0.83	0.81	3.19	3.95	0.793	0.357
Sweden	1.51	1.58	1.26	1.45	6.9	6.54	2.66	2.28
Switzerland	2.08	2.18	0.969	1.08	4.08	3.47	1.52	1.87
Taiwan	0.772	0.867	0.706	0.775	2.6	2.53	0.256	0.798
Turkey	0.257	0.309	0.419	0.463	1.67	1.75	-1.87	-0.25
United Kingdom	0.519	0.535	0.798	0.801	3.02	2.81	0.62	0.562
United States	1.35	1.37	0.917	0.911	3.72	3.91	1.2	1.4

Notes: Model predictions are made with a p_{jt} fitted using Table 3.2, column (2); values for h_{jt} using this fitted $\widehat{p}_{jt}$, decomposed into fixed effects and y and k with unit coefficients; and d_{jt} using these fitted $\widehat{p}_{jt}$ and $\widehat{h}_{jt}$, decomposed into fixed effects and y and k with unit coefficients. The probability of billionaires is expressed per millions of population. The fixed-effect components of parameters h and d are demeaned.

level), the close fit should not be a great surprise. Divergence between predicted and fitted values can be seen most clearly in countries where the probability increased strongly over time, like Hong Kong. In those cases, our model tends to underperform somewhat.

Mean log wealth is a function of p and d. Since d is estimated with the empirical likelihood (and later decomposed into fixed effects and covariates), the extremely close fit should again not be a surprise. Nevertheless, it is worthwhile to note that rival models fail to predict mean log wealth with the same accuracy, see Teulings and Toussaint (2023) for a comparison with the Pareto distribution.

The final moment is unmatched, mean wealth. Here, we are heartened by a strong fit for most cases. The model overpredicts for some countries (i.e., France), and underpredicts for others (Mexico). An interpretation here is that French billionaires are less wealthy than they "should be" given fundamentals, while Mexican billionaires are wealthier than they should. The presence of Carlos Slim may matter here, but it is interesting that Bernard Arnault skews the data less than the model.

We should note that we have grouped billionaires here by citizenship, not by country of birth. This may matter in some cases, in particular for countries which have institutional features making them attractive to high-net worth individuals. We can for instance interpret the large positive Switzerland fixed effect in this regard.

Finally, we turn to the global change in these three moments. We first re-estimate all parameters at a global level, analogous to the country-specific exercises above. This gives us predictions for each moment. In Table 3.6, Panel A, we report the predicted value for our first year (2001), and the last (2021). We see that our model fits the global data very well. It slightly overpredicts the billionaire fraction in 2001 (0.11 billionaires per million people instead of 0.09), but underpredicts the fraction in 2021 (0.3 vs 0.35). It fits mean log wealth extremely well (0.79 vs 0.8 in 2001, and 1.05 vs 1.04 in 2021). The initial value for mean wealth is underestimated (2.9 vs 3.2); however, the final value matches very closely (4.7 vs 4.75).

Table 3.6: Decomposition of the Change in Billionaire Numbers and Mean (Log) Wealth, 2001–2021

	Value 2001		Value 2021		Δ		$\Delta\ell$	Δy	Δk	ε
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: World										
$\Pr\left[\underline{w} \geq 0\right]$	0.087	0.113	0.349	0.294	0.262	0.181	-0.007	0.121	0.048	0.018
$\mathrm{E}\left[\underline{w}\right]$	0.809	0.789	1.04	1.05	0.232	0.259	0.01	0.174	0.076	-0.000
$\mathrm{E}\left[\underline{W}\right]$	3.21	2.94	4.75	4.68	1.54	1.74	0.076	1.01	0.399	0.25
Panel B: United States										
$\Pr\left[\underline{w} \geq 0\right]$	0.944	1.1	1.85	2.09	0.908	0.989	-0.024	0.623	0.306	0.085
$\mathrm{E}\left[\underline{w}\right]$	0.793	0.863	1.04	1.04	0.25	0.182	0.004	0.117	0.06	0.001
$\mathrm{E}\left[\underline{W}\right]$	3.37	3.34	4.8	4.64	1.43	1.3	0.047	0.75	0.364	0.139

Notes: Columns (1) and (3) give the first and last values of the empirical moments, and columns (2) and (4) do the same for the model-implied moments. Columns (5)–(6) give the absolute change (Δ). The model-implied change (6) is decomposed into a component due to log population ($\Delta\ell$), log GDP per capita (Δy) and log market capitalisation (Δk), in columns (7)–(9). Since the moments are non-linear, the decomposition is not exact, with the unexplained residual denoted ε. Panel A does this exercise for all billionaires worldwide, while Panel B concentrates on the United States.

We do an accounting exercise, where we decompose the model-implied change between in 2001 and 2021 into components due to population change, a change in global GDP per capita, and a change in the global stock market, holding the other variables fixed. Consider for instance the billionaire fraction. The data show an increase from 0.09 to 0.35; our model implies an increase from 0.11 to 0.3, or 0.19 points. We can decompose this increase of 0.19 into the three variables. For instance, a change in population alters p, with negative coefficient -0.28 per log point increase. A change in p subsequently alters d and therefore also h. We can therefore calculate a counterfactual value for the billionaire fraction in where we use the 2021 value for population in these calculations but keep all other variables fixed at their 2001 values. Call this counterfactual moment $\mathrm{Pr}^{\sharp}_{\ell}$. The contribution of population size to the increase in billionaire numbers is then given as $\Delta\ell = \mathrm{Pr}^{\sharp}_{\ell} - \mathrm{Pr}^{\natural}$, where the natural sign ♮ indicates the 2001 model prediction. The contributions of y and k follow analogously. Since the moments are highly non-linear, these variables interact in the model predictions. This means that our additive decomposition is not exact; changing one variable at a time while holding the others fixed does not add up to the full model-implied value. We denote the residual by ε; we can also think of this as an interaction term between the three variables.

Table 3.6 reports the results. The change in predicted moments is almost entirely driven by the increase in global GDP per capita, explaining 0.121/0.181 = 66.8% of the increase in the billionaire fraction, 67.2% of the increase in mean log wealth and 58% of the increase in mean wealth. The log market capitalisation factor k also contributed positively, and accounts for a respective 26.5%, 29.2% and 22.9% of the change in the moments. In contrast, the contribution of $\Delta\ell$ is very small and even slightly negative for the billionaire fraction. This is because in our model, population has two offsetting effects: a direct effect (a decrease in p) and the indirect effects (the resulting increases in d and h).[5] The decomposition shows that these effects work against each other, resulting in small net effects. In contrast, GDP per capita and market capitalization have unambiguously signed effects (one-for-one increases in d and h). Of the two, GDP per capita mattered much more in our sample.

We also investigate the dynamics of billionaire increases in the United States. This is reported in panel B of Table 3.6. Here, we do not re-estimate the equations and fixed effects, instead taking the values from Table 3.5. Nevertheless, columns (1)–(4) reveal that the predicted values remain very close to the data. The billionaire

[5] There is an additional channel which works through the denominator, i.e., total population in millions. In results available on request, we have also analyzed the number of billionaires directly, eliminating this denominator effect. The results are qualitatively identical to those reported in Table 3.6, suggesting that the denominator effect is second-order.

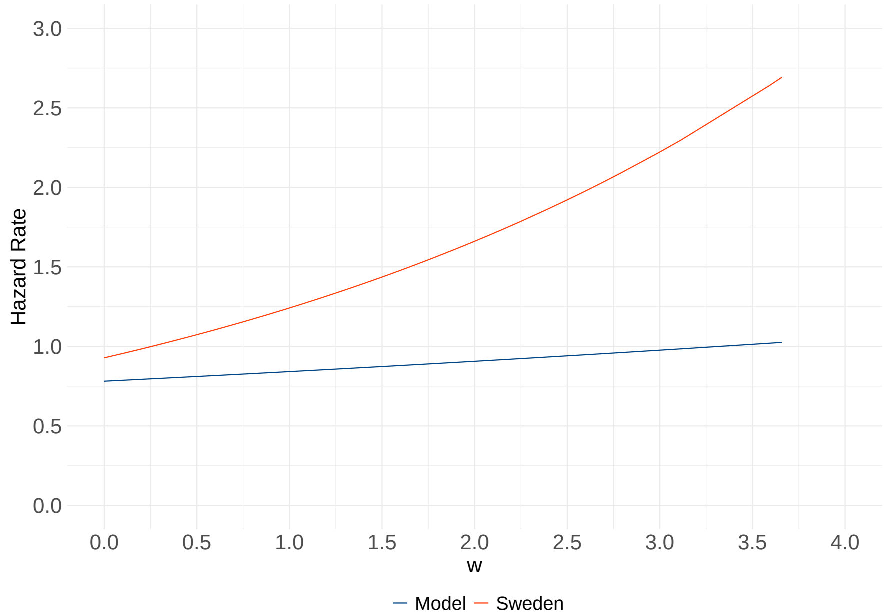
probability is slightly overestimated both at the beginning and end of our sample period, but the resultant predicted increase is therefore close to the actual increase. The billionaire probability doubled between 2001 and 2021, both in the data and in our model. The model attributes almost two thirds of this increase to the change in U.S. GDP per capita. Around 30% comes from the change in the wealth-income ratio. Population has a net negative effect, like in Panel A. The residual or interaction term is comparable to that in Panel A, and makes up almost 10% of the increase.

The predictions for mean log wealth and mean wealth again show a dominant role for GDP per capita, a secondary role for the wealth-income ratio, and a negligible role for population. The residual term for mean wealth is relatively sizable, like in Panel A, suggesting that interactions matter much more for wealth in levels than in logs.

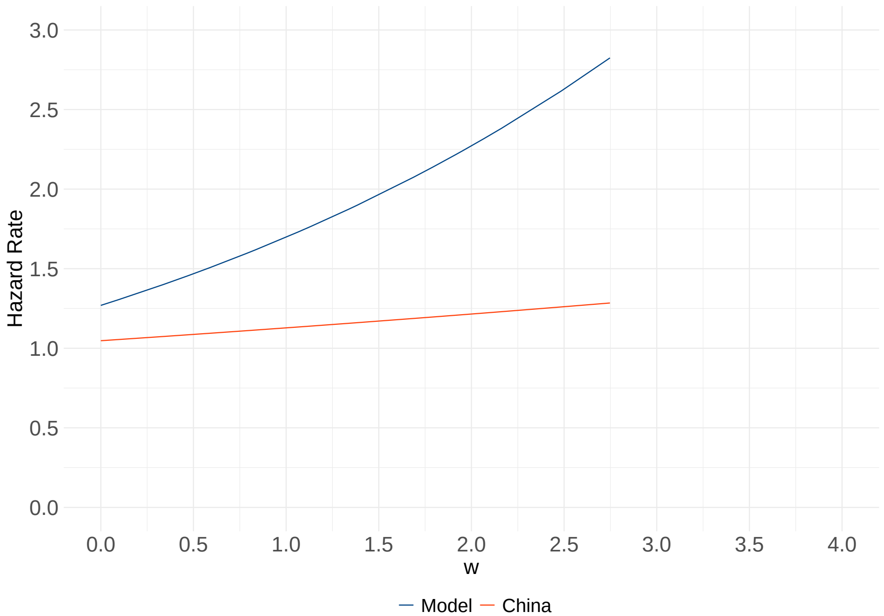
As a final exercise, we investigate the impact of the scale effect of population. To do so, we take China and Sweden in 2019, countries with vastly different populations. We calculate the mean p_{jt} among observations in 2019, and calculate the country-specific values for p_{jt} (and hence h_{jt}) in deviations from that mean. This results in a China-specific value of $p_{\text{China}} = 0.07$ and a Sweden-specific value of $p_{\text{Sweden}} = 0.29$. Then, we investigate what would happen to each country's hazard rate if they had the other country's values for p and h. Since h depends on w, this swapping will not result in exactly identical hazard rates.

Figure 3.6 plots the results. The scale effects in the hazard rates are strong. China, with Sweden's parameters, would have a much higher hazard rate, implying that it would have far fewer super-wealthy billionaires relative to "ordinary" billionaires. For Sweden, the picture is reversed: China's parameters imply a much longer and fatter upper right tail. Swapping the parameters does not result in identical hazard rates; the baseline hazards start at different values (around 0.75 for model-implied China in panel (a) and slightly above 1 for counterfactual Sweden in panel (b)), and the counterfactual hazard rates are steeper. This implies that the interaction with log wealth is also important in determining the fit of the model.

Figure 3.6: Population Scale Effects in Hazard Rates, China and Sweden



(a) China



(b) Sweden

Notes: Figures show the model-implied hazard rates for China and Sweden in 2019, contrasted with hazard rates where the values for p_{jt} and h_{jt} are taken from the other country.

3.8. Interpretation

Our most robust finding is that the elasticity of the wealth hazard rate, p_{jt}, is declining in log population. In other words, more populous countries have longer upper tails of the wealth distribution. We can interpret this result using two kinds of models. First, in heterogeneous-firm models with fixed entry costs à la Melitz (2003), firm profits increase in population size. However, a population increase does not benefit all firms equally; most of the gains go to the upper tail of the firm productivity distribution. Hence, an increase in ℓ would cause an increase in mean productivity but also an increase in inequality. Translated to firm owners, the implications are clear.

An even stronger association is found in endogenous growth models. As reviewed in Jones (2022b), essentially all models in this literature predict population to have a positive effect on the number of ideas and hence technological change.[6] If we interpret the number of nodes L as the number of ideas that can be successfully embodied into output, the link with endogenous growth models follow naturally. Our setup, however, is slightly artificial in that it is static: the network is formed at the dawn of time and never changes. This implies, among other things, that once an individual is rich (has accumulated many ideas), he never drops out. This seems to sharply contrast with Schumpeterian models in the vein of Aghion and Howitt (1992), where incumbent firms (and hence their owners' wealth) are constantly cannibalized by more successful upstarts.

We think that a dynamic extension of our model could resolve this tension, where each period links can also disappear with some probability (and new links can also be formed).[7] We conjecture that such a model would feature many of the dynamics that are congruent to Schumpeterian models; yet in each snapshot $\{j, t\}$ the distribution of Self-Avoiding Walks would still be Gompertz. Naturally, individual entrepreneurs might drop out (their links get broken) and others take their place. Hence, our model is a description of the cross-sectional distribution in a region in a point in time, and not a model of the trajectory of individual billionaires. We view an extension of our model along these lines as a fruitful avenue for future research.

[6] There is some difference between fully endogenous growth models such as Romer (1990) or Aghion and Howitt (1992) and semi-endogenous growth models such as Jones (1995) and Kortum (1997). In the latter category, population growth has level effects, while in the former, population growth increases the growth *rate* of technological change.

[7] For instance, agents might enter and leave the network at some Poisson rate λ, as in Akbarpour, Li, and Gharan (2020).

Chapter 4

Robust Estimation of Private Business Wealth

This chapter is based on Toussaint (2024).

4.1. Introduction

Private businesses are key drivers of economic activity. In the United States, private businesses account for almost half of aggregate sales and profits (Campbell and Robbins 2023). Moreover, their ownership is highly concentrated among the top of the income and wealth distribution (Kopczuk and Zwick 2020). Therefore, for many pressing debates in economics, good data on private businesses are indispensable. In macroeconomics, the decline in productivity growth has been attributed to numerous causes; yet, this measured decline might to a significant part be an artifact of mismeasured capital stocks (Crouzet and Eberly 2021). Related, the debate on the apparent decline in the labor share also relies on correct measures of the capital share (Karabarbounis and Neiman 2014; Barkai 2020). For distributional dynamics, it is likewise key to get the magnitudes of private business wealth right (Saez and Zucman 2016; Smith, Zidar, and Zwick 2023).

The key challenge is that private businesses, by definition, are not listed. Hence, their market values are unobserved. Firm datasets do typically include balance sheet components such as a firm's book value and capital stock. However, these components are unlikely to capture the true market values of a firm. Conceptually, firm book value misses intangibles such as goodwill, brand reputation and other things that are capitalized in a market price but not necessarily recorded accurately in a firm balance sheet. Moreover, since these balance sheet components are based on accounting measures, they are inherently backward-looking. In contrast, a firm's market value is forward-looking: it is the expected present value of future income streams.

There are four common responses to this problem. The first is to leave the accounting values as is, without further adjustment. This is common in most of the wealth inequality literature, but risks seriously understating top wealth shares. A second response is to use industry-level market-to-book ratios, perhaps with some illiquidity discount, and apply these to private firms of similar industries (Bach, Calvet, and Sodini 2020; Damodaran 2012). This strategy introduces mechanical correlations between firms at the industry level. Moreover, it is debatable whether listed firms are representative of their industry. A third response is to capitalize cashflows using some estimated discount factor. This is a forward-looking measure and in principle results in correct values; however, this introduces mechanical correlations between firm values and cashflows. This makes it impossible to estimate, return heterogeneity, for instance (Fagereng, Guiso, Malacrino, and Pistaferri 2020). Moreover, it requires the estimation of a (stochastic) discount factor, which opens the door for further measurement errors. A final strategy is to look at resales of private firms, and regress the updated values on firm characteristics (Campbell and

Robbins 2023). This approach relies on resales being representative, which is again debatable.

This chapter introduces a new approach. I treat the missing market value problem as an econometric problem. From standard neoclassical investment theory, we expect a firm's value to be a linear function of its capital stock; under the neoclassical assumptions, a regression of value on capital has a coefficient equal to Tobin's q. I show that even in the more general model of Crouzet and Eberly (2023), where firms potentially charge markups and/or face decreasing returns to scale, approximate linearity still holds. Hence, we would expect a firm's market value to be structurally related to its capital stock, and this structural relation should be identifiable using a linear regression.

Since we do not observe the firm's value, nor its true capital stock, we cannot directly run this regression. Instead, I use the following three-step approach. First, I construct estimates of firms' market value, using the capitalized cashflows approach. I assume that these estimates are contaminated with measurement error, as is the firm's capital stock. The measurement errors in value and capital are highly correlated; hence, a naive regression would be biased. I then regress the estimated market value on the capital stock. To account for the measurement error in the left-hand and right-hand side of this regression, I use an instrumental variables strategy to filter out the errors. In an instrumental variables or generalized method of moments (GMM) framework, these instruments then deliver fitted values of capital (from the first stage) and fitted values of market value (from the second stage). Under the identifying assumptions, these fitted values are free from measurement errors. Therefore, I treat these fitted values as the "true" values of market value and capital.

Clearly, my approach relies on valid instruments. I develop an approach using time-series restrictions, exploiting the panel nature of my dataset. As noted by Griliches and Hausman (1986), we can use a variety of estimators (within, first-difference, second-difference, etc.) to estimate panel data in the presence of unobserved fixed effects. Under measurement error, these estimators are all biased, but in *different ways*; hence, we can use differences in these estimators as instruments to identify the parameters of the model. This setup yields more instruments than parameters to be estimated. Therefore, we can test over-identifying restrictions, verifying whether the model holds in the data. Specifically, I use an instrument based on the deviations between the within and the first-difference estimator, and instrument based on on the deviations between the first-difference and the second-difference estimator, and an instrument based on the deviations between the within and the second-difference estimator. Under the identifying assumption that the measurement errors are stationary and uncorrelated over time, these are valid instruments. In extensions, I weaken these assumptions to allow the measurement error to follow a MA(1) process, and derive valid instruments under this assumption.

I use administrative data from the Netherlands 2008–2020. I link the universe of incorporated firms to their owners, which enables me to investigate dynamics both at the firm and the household level. For my econometric methods, I look at the firm data only. I use the univese of corporate income tax returns, which have rich information on firms' balance sheet components and income statements. I construct the capital stock as the sum of a firm's physical and intangible capital. I construct my raw estimate of a firm's market value by taking a three-year moving average of its profits as my forecast of its future profits. I use three different estimates for the firm discount rate. The first two are from Gormsen and Huber (2023, 2024), who use firms' earnings calls to extract their actual discount rates. Moreover, they manually reconstruct firms' weighted-average cost of capital (WACC). I use the average of these two measures for the Dutch firms in their sample each year. A third measure, following Barkai (2020) and others, simply takes the risk-free interest rate and adds a constant risk premium of 5%. Obviously, these discount rates are all imperfect and subject to error. This is why my GMM approach is important: it filters out the error in these estimations by retaining only the components that are structurally related to a firm's true capital stock.

Across the three different discount rate measures, we can apply the three instruments in all possible combinations, resulting in a total of 21 different specifications. Across specifications, the estimated coefficient on the capital stock, β, is significant and large. The availability of multiple instruments permit overidentification tests. The J-statistics are uniformly insignificant and quantitatively almost indistinguishable from zero, meaning that the restrictions easily hold. The coefficients with the best test statistics cluster together in two groups. The high group hovers around 4.5, and the low group around 3.5. I term these two coefficients β^{+} and β^{-}, and use these to update private business wealth estimates.

Once we have estimated values for firm and capital stocks, we can revisit many of the aggregate and distributional debates mentioned above. The updated firm values matter for top wealth shares, since private firms are concentrated at the top of the wealth distribution: it accounts for almost 80% of the top 0.01%'s portfolio. I use the fitted values from the GMM specifications, using β^{+} and β^{-}. This results in two different vectors of fitted values. I aggregate these to the firm-owner level using a comprehensive Shareholder Registry, which allows me to link corporate tax returns to Dutch households for the universe of firms incorporated in the Netherlands.

On the aggregate level, my adjustments increase the total value of private business wealth markedly, in particular in the early years of my sample. The unadjusted microdata show a steep increase in this wealth component's aggregate value from 200 billion € in 2008 to around 450 billion € a decade later. In contrast, the updated values are almost 150 billion € are higher in early years, but stay more stable over the period. Hence, at the end of the sample, the microdata and the updated series almost coincide. This shows that while the procedure does generally increase the measured values, it does not do so mechanically. This also addresses concerns voiced by Cochrane (2020), Greenwald, Leombroni, Lustig, and Van Nieuwerburgh (2021), and Fagereng et al. (2024) and others whether the increase in wealth inequality is merely a valuation effect ("paper gains"). My results show that it is the combination of discount rates and profits that drive the changes in private business wealth, not merely declining discount rates.

I analyze why the book values show different trends and levels than my updated series and conclude that this is likely driven by portfolio reallocations of investors as interest rates declined. The Dutch tax system taxes savings and directly-held financial assets quite heavily, effectively using a wealth tax. The decline in deposit interest rates since 2013 coincided with a major reallocation of capital towards private firms, which are are not subject to the wealth tax.[1] This suggests that book values of Dutch firms were quite heavily undervalued at the beginning of my period, and mechanically increased in value as capital started flowing in. In contrast, the firms' capital stock remained quite stable, which means that my updated measures also remained stable. I conclude that while book-value measures of private firm wealth are vulnerable to fiscal manipulation, my alternative series is more robust.

I also use the aggregate data to investigate why the coefficients β^+ and β^- are so sizable. Intuition based on Tobin's q suggests that these values should be lower, on the order of 1.5.[2] However, the model developed in Section 4.2 suggests two answers to this question. First, if either market power and/or decreasing returns to scale are present, firms have positive rents in equilibrium. The net present value of these rents shows up in the reduced-form parameter which governs the curvature of the profit function, μ, and through μ also shows up in the regression coefficient β. I estimate the structural parameters of the model and conclude that rents are indeed present. μ is larger than 1, which indicates positive rents. In fact, its value is substantially larger than the estimates in Crouzet and Eberly (2023) for U.S. listed corporations. This suggests that applying the model to the universe of firms reveals larger rents than just among listed firms. However, applying the structural model also reveals puzzlingly low estimated capital costs as a share of total firm costs. This suggests that the capital stock is underrecorded in firms' accounts. An upward adjustment of the firm's capital stock would result in a lower μ and hence also a

[1] Dividends and realized capital gains are subject to income taxation. See Section 4.4 for an extensive discussion of the Dutch capital income taxation system.

[2] According to the United States Flow of Funds, Tobin's average Q averages around 1.3 since 2008 in the United States.

lower β. Since these two adjustments cancel out, the fitted values would remain the same.

My results have major implications for wealth inequality. Based on the low coefficient β^-, the top 1% wealth share increases by 3 percentage points on average relative to the headline series, to about 36%. The top 0.1% share increases by a similar amount, to about 17%. These are the lower-range results from my specifications. If I use the higher coefficient β^+, the top 1% share increases further to almost 39% at its peak, and the top 0.1% share almost equals 20%. Clearly, these are estimates, and the exact quantitative upward adjustment is open to reasonable debate. Nevertheless, my results indicate unambiguously that inequality is underestimated in the microdata. Based on these adjustments, the Netherlands would rank among the most unequal developed countries in the world on wealth inequality. However, series from other countries likewise suffer from underestimating private business wealth; hence it seems plausible that my procedure would result in higher top wealth shares in other countries as well.

A final implication of my results concerns the debate on return heterogeneity. As observed by Gabaix, Lasry, Lions, and Moll (2016), existing models of wealth inequality have difficulty accounting for the speed of inequality increases. Models with heterogeneous returns, however, are able to account for these dynamics. Several papers have documented return heterogeneity (Fagereng, Guiso, Malacrino, and Pistaferri 2020; Bach, Calvet, and Sodini 2020; Xavier 2020), which seems to confirm the hypothesis. However, I show that these results are highly sensitive to measurement error. Since returns have wealth (or assets) in the denominator, a regression of returns on wealth yields mechanical correlations between left- and right-hand side if there is measurement error in wealth. The sign of this bias depends on whether the numerator is also measured with error, as seems likely in the case of private business wealth. These theoretical results underscore the need to take measurement seriously in accounting for wealth dynamics.

With my fitted values of firm value, I can revisit the return heterogeneity literature. Importantly, since I use the fitted values from the GMM estimation and not the ‘raw’ estimates based on discounted cashflows, these regressions are no longer mechanically correlated. In effect, I have retained only the component of capitalized cashflows that is structurally related to a firm’s (true) capital stock. I find that adjusting firms’ values makes a major difference. Returns based on unadjusted returns are essentially flat or weakly increasing across the firm size distribution. Adjusted returns, on the other hand, strongly increase with firm size, with the gradient becoming the steepest in the top decile. These conclusions also hold when we aggregate to the household level. Therefore, heterogeneity in returns (either at the firm or firm-owner level) is substantial, with most of the heterogeneity coming from the top. Unadjusted returns, due to measurement error or other noise, understate this effect.

Related Literature: This chapter integrates several strands of literature. Methodologically, I build on a long literature seeking to identify models in the presence of measurement error, as reviewed by Schennach (2016, 2022). Within this literature, my approach builds on Griliches and Hausman (1986), but is also inspired by the influential stream of papers which seek to robustly test the neoclassical investment model, as developed by Hayashi (1982) and Abel and Eberly (1994) and many others. These papers typically use higher-order moment restrictions in a GMM procedure to identify investment-q regressions (Erickson and Whited 2000, 2012; Erickson, Jiang, and Whited 2014), which goes back to insights by Reiersøl (1950), Kapteyn and Wansbeek (1983) and Lewbel, Schennach, and Zhang (2023) and many others. Relative to these papers, I seek solutions to the inverse problem: fitted values that can serve as estimates of the market value of private firms.

Substantively, this chapter relates to many papers in finance that seek to account for the correct valuation of private businesses (Damodaran 2012; Kaplan and Schoar 2005; Korteweg 2019; Gupta and Van Nieuwerburgh 2021). More broadly, my methods and contributions fit within the literature in empirical corporate finance seeking to robustly estimate key firm objects, such as intangibles (Crouzet and Eberly 2021, 2023), labor and capital shares (Barkai 2020; Karabarbounis and Neiman 2014), and firm-level discount rates (Gormsen and Huber 2023, 2024). Central to all these literatures is a concern with correctly measuring firm capital and risk (and hence value), which is the central goal of this chapter.

My results relate to the wealth inequality literature (Saez and Zucman 2016; Smith, Zidar, and Zwick 2023; Saez and Zucman 2020). Many authors have noted concerns with the reliable measurement of top wealth shares in the presence of unlisted private businesses (Smith, Zidar, and Zwick 2023; Kopczuk and Zwick 2020; Toussaint, van Bavel, Salverda, and Teulings 2020; Toussaint, de Vicq, Moatsos, and van der Valk 2022). As yet, there is no agreed-upon method to address these concerns. My method is straightforward to implement for researchers with access to firm balance sheet data, and thus provides a theoretically grounded, econometrically robust first step toward measuring the true extent of top wealth inequality.

Finally, my results relate to the debate on return heterogeneity (Gabaix, Lasry, Lions, and Moll 2016; Fagereng, Guiso, Malacrino, and Pistaferri 2020; Bach, Calvet, and Sodini 2020; Xavier 2020). Measurements from Norway and Sweden seem to establish correlations between wealth and returns, but it is still unclear what the underlying mechanism is underlying these features. Explanations include Jones and Kim (2018), Kacperczyk, Nosal, and Stevens (2019), Gerritsen, Jacobs, Spiritus, and Rusu (2024), and Guvenen et al. (2023). My results first clarify that indeed, returns are heterogeneous, and that most of the action is concentrated at the top of the firm and firm-owner distributions. Future theoretical and empirical work should therefore focus on the interplay between high-return firms and high-return owners.

Chapter Outline: The rest of the chapter is organized as follows. Section 4.2 discusses the structural model that underpins the use of regressions of firm value on capital. In Section 4.3, I discuss the two estimation strategies I employ to overcome measurement error. Section 4.4 introduces the data used and provides summary statistics. Section 4.5 shows the results of the econometric strategies. The next three sections discuss several implications. Aggregate levels of private business wealth are discussed in Section 4.6, distributional implications in Section 4.7, and the consequences for heterogeneous returns are covered in Section 4.8. Section 4.9 concludes.

4.2. Framework

In neoclassical models of investment, market value is related to capital via Tobin's marginal q, the shadow value of an additional unit of investment. Hayashi (1982) shows that if firms are price takers and face constant returns to scale, marginal q and average Q (the firm's ratio of market value to capital stock) are the same: Market value is a linear function of capital, or

$$V_t = qK_t. \tag{4.1}$$

This result rests on strong assumptions. Besides the Walrasian assumptions already mentioned, the firm is assumed to be able to adjust capital subject to a convex adjustment cost function $\Phi(K_t)$. The convexity ensures that the firm smoothly and continually adjusts its investment; irreversibilities are ruled out. However, Abel and Eberly (1994) show that mild forms of fixed costs in investment can also be accommodated in this framework. As long as the fixed costs are proportional to the profit function Π_t, firm value can still be written as a linear function of capital.

Under these assumptions, my strategy is straightforward: compute some measure of market value (which may be ridden with measurement error), regress it on capital (instrumented by some instrument to correct for endogeneity), and use the fitted values as the true measure of market value. However, we need to verify whether this argument also holds in more general models where firms have market power and face decreasing returns to scale. I now construct a model along those lines and show that, approximately, linearity holds.

The Model: The model is based on Crouzet and Eberly (2023). Time is discrete. I use capital letters for aggregate variables, and lower-case letters for firm-level or input-level variables. Consider a representative firm that uses G distinct variable inputs $\{m_{gt}\}_{g=1}^{G}$, as well as capital K, to produce and sell output to consumers, whose demand function is given by:

$$C_t = P_t^{-\frac{\lambda}{\lambda-1}} D_t,$$

where C_t is consumption of the final good, P_t is the price, D_t indexes aggregate demand, and $\lambda \geq 1$ is the firm's markup over marginal cost. The price of variable input g is given by p_t^g. The total input of capital K_t (which is made up of tangible and intangible capital), is quasi-fixed: it is chosen dynamically but cannot be modified immediately. The firm has a production function

$$F(A_t, K_t, M_t) = A_t \left(K_t^{\delta} M_t^{1-\delta}\right)^{\eta}, \tag{4.2}$$

where A_t is an aggregate productivity shifter, and M_t is a Cobb-Douglas aggregator of variable inputs, including labor:

$$M_t := \prod_{g=1}^{G} m_{gt}^{\nu_g}, \quad \sum_{g=1}^{G} \nu_g = 1. \tag{4.3}$$

The firm can have rents both from market power on the product power, given by λ, and quasi-rents from decreasing returns to scale of degree η in capital and variable inputs. To obtain an interior solution, I follow Crouzet and Eberly (2023) and assume $\eta \leq \lambda$. Interesting cases are typically those where $\eta \leq 1$, so that the firm has decreasing returns to scale; but constant or increasing returns are also covered in this model as long as the markup λ is sufficiently high.

The firm chooses variable inputs and an output price to maximize its profits:

$$\Pi_t = \max_{\{m_{gt}\}_{g=1}^{G}, P_t} P_t^{-\frac{\lambda}{\lambda-1}} D_t - \sum_{g=1}^{G} p_{gt} m_{gt}, \tag{4.4}$$

$$\text{s.t.} \quad A_t \left(K_t^{\delta} M_t^{1-\delta}\right)^{\eta} \geq P_t^{-\frac{\lambda}{\lambda-1}} D_t. \tag{4.5}$$

An attractive property of this model is that the optimization problem is static; this is done by not explicitly modeling capital adjustment. Doing so, for instance by including a convex adjustment cost function $\Phi(K_t)$, would complicate the algebra without altering the substantive results. A second advantage of the model is that it admits a closed-form representation of the profit function Π_t that is highly tractable, as derived by Crouzet and Eberly (2023) and reproduced in Appendix 4.A for completeness:

$$\Pi_t = H_t^{1-\frac{1}{\mu}} K_t^{\frac{1}{\mu}} \tag{4.6}$$

where:

$$\mu := 1 + \frac{\varphi - 1}{\delta} \geq 1, \tag{4.7}$$

$$\varphi := \frac{\lambda}{\eta} \geq 1, \tag{4.8}$$

and H_t is a combination of parameters and variables that does not depend on K_t:

$$H_t := \left(\frac{\varphi}{1-\delta}\right)^{-\frac{\varphi}{\varphi-1}} \left(\frac{\varphi}{1-\delta} - 1\right)^{\frac{\varphi-(1-\delta)}{\varphi-1}} D_t^{\frac{\varphi-\eta}{\varphi-1}} \mathcal{P}_t^{-\frac{1-\delta}{\varphi-1}} A_t^{\frac{1}{\eta(\varphi-1)}}$$

where $\mathcal{P}_t$ is a weighted aggregate of input prices p_{gt}:

$$\mathcal{P}_t := \prod_{g=1}^{G} \left(\frac{p_{gt}}{\nu_g}\right)^{\nu_g}.$$

This result has several implications. First, the impact of market power and returns to scale is pinned down by the reduced-form parameter $\varphi = \lambda/\eta$ (although η and λ are generally not separately identified). If the firm has constant returns to scale ($\eta = 1$) and is a price taker ($\lambda = 1$), $\mu = 1$ and profits are a linear function of capital.[3] In the general case where $\varphi > 1$, equation (4.7) shows that this impacts the profit curvature parameter μ, which will then be larger than 1. I now show that even when $\mu > 1$, profits are still approximately linear in capital:

[3] In fact, in this static framework, profits are identically equal to capital when $\mu = 1$. In a more general formulation with convex adjustment costs to capital, this coefficient (which equals marginal q) would be larger than one (Hayashi 1982).

$$\Pi_t \approx \sum_{n=0}^{\infty}(K_t-1)^n \binom{\frac{1}{\mu}}{n} H_t^{\frac{\mu-1}{\mu}}$$

$$\approx \underbrace{H_t^{\frac{\mu-1}{\mu}}\left(1-\frac{1}{\mu}\right)}_{=:\alpha} + \underbrace{\frac{1}{\mu}H_t^{\frac{\mu-1}{\mu}}}_{=:\beta} K_t + \frac{\frac{1}{\mu}-1}{2\mu}(K_t-1)^2 H_t^{\frac{\mu-1}{\mu}} + O\left(K^3\right). \tag{4.9}$$

Equation (4.9) shows that, to first order, profits are linear in capital. The second-order term vanishes for minor deviations from $\mu = 1$. Crouzet and Eberly (2023) find in their application of American listed corporations a μ ranging from 0.984 to 1.290.[4] These values imply that the fraction premultiplying the second-order term ranges from 0.01 to −0.09, which are one to two orders of magnitude smaller than β. Hence, I conclude that even with market power and returns to scale, approximate linearity holds. As a result, firm value, which is the discounted present value of profits, is also a linear function of capital:

$$V_t = \sum_{t=0}^{\infty}\left(\frac{1}{1+\varrho}\right)^t \Pi_t = \sum_{t=0}^{\infty}\left(\frac{1}{1+\varrho}\right)^t \left(\alpha+\beta K_t\right). \tag{4.10}$$

Since the firm has rents, its value will include the capitalized value of expected future rents. Through Equation (4.6), the profit function includes the capitalized value of rents through H_t, and the importance of rents is governed by μ. Under the frictionless benchmark ($\mu = 1$), rents do not show up in the profit function and hence also not in firm value.

So far, the model has been set up without uncertainty. Incorporating this into the model is straightforward, by letting $\{\Pi_t\}$ follow a stochastic process. In particular, to facilitate the operationalization of my model, I make an important assumption:

Assumption 1. *The profit function* $\{\Pi_t\}$ *follows a martingale process.*

Assumption 1 drastically simplifies the computation of the expected present value (equation (4.10)), since it implies that we can use the current value of a firm's profits as the best linear forecast of future profits. Since profits tend to be highly volatile, I operationalize this assumption by taking a three-year moving average

[4] Values smaller than 1 are strictly speaking not possible under the model written above; however, it is clear that this value would be statistically hard to tell apart from 1, although Crouzet and Eberly do not report standard errors.

of firm i's profits, $\overline{\pi}_{it}$ (lower-case letters denoting firm-level variables). Then, we estimate v_{it} as

$$v_{it} = \sum_{t=0}^{\infty} \left(\frac{1}{1+\varrho} \right)^t \overline{\pi}_{it} = \frac{1+\varrho}{\varrho} \overline{\pi}_{it}. \tag{4.11}$$

Estimating firm value in this way is grounded in the model presented above (and holds in most economic models of firm investment, see Crouzet and Eberly (2023)). However, this measure is highly sensitive to the stochastic discount rate ϱ, which is an object to be estimated. Even absent measurement error, the fact that we use estimated values for ϱ will introduce estimation uncertainty into v_{it}, biasing the results. Moreover, it is quite plausible that the discount rate and profits are measured with error, further attenuating the results.

My estimation strategy seeks to overcome this measurement error issue. It is based on the observation, from equation (4.9), that firm value is an approximately linear function of the capital stock. However, the capital stock is also likely to be mismeasured, since intangibles are poorly recorded in a firm's accounts. Importantly, mismeasured intangibles and other mismeasurements of the capital stock would affect both the capital stock and firm profits, hence my estimate of firm value. As I show in Section 4.3, this introduces endogeneity; moreover, left- and right-hand side are mechanically correlated. To overcome this endogeneity, we need some valid instrument that is correlated with the true capital stock, but uncorrelated with the measurement errors. I detail in the next section how I construct this instrument. Summarizing, my procedure to estimate firms' market values is as follows:

1. Construct an estimate of v_{it}, using some discount rate ϱ and profits $\overline{\pi}_{it}$. Call this estimate y_{it}. Take a measure of the firm's capital stock from its accounts; call this measure x_{it}.

2. Construct a valid instrument z_{it} that is correlated with the true capital stock k_{it} but uncorrelated with the measurement error.

3. Regress y_{it} on x_{it}, instrumented with z_{it}; obtain fitted values $\widehat{y}$, which are free from measurement error.

In the next section, I more elaborately formulate the econometric problem, and discuss my identification strategy.

4.3. Estimation Strategy

4.3.1. Measurement Error

I write the measurement error problem as follows. We are interested in the model

$$v_{it} = \alpha + k_{it}\beta + \varepsilon_{it} \tag{4.12}$$

where v_{it} is the market value of firm i in year t, and k is the firm's capital stock. ε_{it} is an error term capturing all other factors that might influence a firm's market value. Under the frictionless conditions of Hayashi (1982), $\beta = q$, i.e., β equals Tobin's marginal q, the shadow value of an additional unit of investment. For my purposes, it is irrelevant whether $\beta = q$, as long as a linear and positive structural relation between capital and value holds. Equation (4.9) shows that this approximately holds even in models with markups and decreasing returns to scale.

Instead of observing a firm's true market value and capital stock, however, we observe the accounting values of these variables, which we treat as measurement-error-contaminated proxies for the true variables. For simplicity, I assume that firm value and capital are contaminated with the same measurement error ψ_{it}, which contains a fixed firm-specific component χ_i and idiosyncratic innovations ξ_{it}. In the context of this chapter, we can think of χ_i as the initial or average error when first writing down the firm's capital stock in the accounts, while ξ_{it} are the noise in the annual investments (which cumulate into the capital stock). Note that all results go through if we allow for different measurement erors in v and k, as long as the identifying assumptions apply to them symmetrically.[5] Inserting $\psi_{it} := \chi_i + \xi_{it}$ into the structural model, we have:

$$x_{it} = k_{it} + \chi_i + \xi_{it}, \tag{4.13}$$

$$y_{it} = v_{it} + \chi_i + \xi_{it}. \tag{4.14}$$

We can rewrite this system to obtain the reduced form

$$y_{it} = x_{it}\beta + \underbrace{\alpha + (1-\beta)\chi_i}_{=:\gamma_i} + \underbrace{\varepsilon_{it} + (1-\beta)\xi_{it}}_{=:u_{it}}. \tag{4.15}$$

[5]Specifically, measurement error in v that is uncorrelated with k would just increase the standard errors but not affect point estimates. Hence, we can think of ψ as the component of measurement error that is common to v and k. Of course, k may have additional measurement error which would cause further attenuation bias; but as long as the time-series restrictions on ξ also apply to this k-specific measurement error, the instrumental variable strategy developed below will also eliminate this measurement error.

A regression of y_{it} on x_{it} is biased for two reasons: first, because of the fixed effect γ_i, which absorbs the intercept α, and second, because the error term u_{it} is correlated with the regressor x_{it}. Stacking across observations, we have

$$\mathbf{y} = \mathbf{x}\beta + \gamma + \mathbf{u} = \mathbf{x}\beta + (1-\beta)\chi + \varepsilon + (1-\beta)\xi \tag{4.16}$$

with $\text{Var}[\varepsilon] = \Sigma$, and $\text{Var}[\xi] = \Xi$. The goal is to obtain consistent estimators for β using instrumental variables or the Generalized Method of Moments (GMM). Once we have found consistent estimators, we can obtain fitted values for $\mathbf{k}$ and $\mathbf{v}$, which I will treat as the true values of k_{it} and v_{it}. I now discuss my identification strategy, based on time-series restrictions.

4.3.2. Time-Series Identification

We seek to obtain consistent estimators for β using restrictions on the unobservable variables k and ξ. Since my application is a panel data setting, we can tap into the large literature on internal instruments in dynamic settings (e.g., Anderson and Hsiao 1982; Arellano and Bond 1991; Blundell and Bond 1998). Specifically, I use the framework laid out in Griliches and Hausman (1986), who develop a class of estimators based on restrictions on the time-series properties of k and ξ.

To build intuition, recall that the regression of $\mathbf{y}$ on $\mathbf{x}$ is biased by the presence of the fixed effect γ and the endogeneity of $\mathbf{x}$. Assume that ξ_{it} is stationary and uncorrelated over time. Consider two common estimators developed to remove the fixed effect γ, the within or fixed-effects estimator and the first-difference estimator:

$$\begin{aligned}\widehat{\beta}_{FE} &= (\ddot{\mathbf{x}}'\ddot{\mathbf{x}})^{-1}\ddot{\mathbf{x}}'\ddot{\mathbf{y}},\\ \widehat{\beta}_{FD} &= (\mathbf{\Delta x}'\mathbf{\Delta x})^{-1}\mathbf{\Delta x}'\mathbf{\Delta y},\end{aligned}$$

where $\ddot{x}_{it} := x_{it} - T^{-1}\sum_t x_{it}$, $\Delta x_{it} := x_{it} - x_{i,t-1}$ and so on. Both estimators address the fixed effect γ, but since the regressor is endogenous with the error term, both are biased. It is not difficult to work out that as $N \to \infty$,

$$\begin{aligned}\text{plim}\,\widehat{\beta}_{FD} &= \beta\left(1 - \frac{2\sigma_\xi^2}{\sigma_{\Delta x}^2}\right),\\ \text{plim}\,\widehat{\beta}_{FE} &= \beta\left(1 - \frac{T-1}{T}\frac{\sigma_\xi^2}{\sigma_{\ddot{x}}^2}\right).\end{aligned}$$

The insight of Griliches and Hausman (1986) is to recognize that this is a system of two equations in two unknowns, β and σ_ξ^2; all other variances and objects are known to the econometrician once she's run the two regressions. Hence, we can solve this system to obtain a consistent estimator for β:

$$\beta = \frac{\frac{2\widehat{\beta}_{FE}}{\sigma^2_{\Delta x}} - \frac{(T-1)\widehat{\beta}_{FD}}{T\sigma^2_{\ddot{x}}}}{\frac{2}{\sigma^2_{\Delta x}} - \frac{T-1}{T}\sigma^2_{\ddot{x}}} \tag{4.17}$$

$$\sigma^2_{\xi} = \frac{\left(\beta - \widehat{\beta}_{FD}\right)\sigma^2_{\Delta x}}{2\beta}. \tag{4.18}$$

This solution is a special case of a general estimation strategy, where the deviations between various panel-data estimators are used as instrumental variables. Formally, we seek instrumental variables estimators of the form $\widehat{\beta}_{IV} = (\mathbf{z}'\mathbf{x})^{-1}\mathbf{z}'\mathbf{y}$, where $\mathbf{z} = (\mathbf{I}_N \otimes \mathbf{P})\mathbf{x}$ for some $T \times T$ matrix $\mathbf{P}$. I write the following conditions for $\mathbf{z}$ to be a valid instrument:

Assumption 2 (Time-Series)**.** *Let ι be a vector of all ones. We require*

$$\iota'\mathbf{P} = 0, \tag{4.19}$$

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{P}\mathbf{u}_i = 0, \tag{4.20}$$

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i'\mathbf{P}\mathbf{x}_i \neq 0. \tag{4.21}$$

The first condition ensures that the fixed effect γ is eliminated. The other two conditions are the usual exogeneity and relevance conditions. The instrument matrix $\mathbf{P}$ plays a crucial role. In the example above, we can find $\mathbf{P}$ by noting that equation (4.17) is the solution to an equation of the form for which we can rewrite the numerator as

$$\mathbf{x}'\mathbf{R}_{\text{FD}}{}'\mathbf{R}_{\text{FD}}\mathbf{y} - \mathbf{x}'\mathbf{R}_{\text{FE}}{}'\mathbf{R}_{\text{FE}}\mathbf{y} = \mathbf{x}'\left[\mathbf{R}_{\text{FD}}'\mathbf{R}_{\text{FD}} - \mathbf{R}_{\text{FE}}'\mathbf{R}_{\text{FE}}\right]\mathbf{y} =: \mathbf{x}'\mathbf{P}\mathbf{y} =: \mathbf{z}'\mathbf{y}.$$

In the equation above, $\mathbf{R}_{\text{FD}}$ is a differencing matrix (a bi-diagonal matrix with -1 on the diagonal and $+1$ on the superdiagonal), and $\mathbf{R}_{\text{FE}}$ is the within-transforming matrix, $\mathbf{R}_{\text{FD}} = \mathbf{I} - \mathbf{J}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{J}$ is $1/T$ times a matrix of all ones. In other words, the matrices $\mathbf{R}$ appropriately transform $\mathbf{x}$ to eliminate the fixed effect, and the instrument matrix $\mathbf{P}$ gathers the quadratic deviations of these different transformations.

I use three instruments based on this setup. My first instrument, $\mathbf{z_1}$, is based on the deviations between the first-difference and the within-transformation, as in the examples above. For overidentification, I also construct an instrument $\mathbf{z_2}$, based on the deviations between the second and first difference. The second difference is given by

$$\Delta_2 x_{it} := x_{it} - x_{i,t-2}.$$

Finally, I construct an instrument $\mathbf{z_3}$, based on the difference between the within-estimator and the second-difference estimator. I derive in Appendix 4.A that the vectors of instruments $\mathbf{z}$ take the form

$$\mathbf{z_1} = \begin{pmatrix} \bar{x} - x_2 \\ \bar{x} - x_1 + x_2 - x_3 \\ \bar{x} - x_2 + x_3 - x_4 \\ \vdots \\ \bar{x} - x_{T-1} \end{pmatrix}, \tag{4.22}$$

$$\mathbf{z_2} = \begin{pmatrix} x_1 - 2x_2 + x_3 \\ -2x_1 + 4x_2 - 3x_2 + x_4 \\ x_1 - 3x_2 + 4x_3 - 3x_4 + x_5 \\ \vdots \\ x_{T-3} - 3x_{T-2} + 4x_{T-1} - 2x_T \\ x_{T-2} - 2x_{T-1} + x_T \end{pmatrix}, \tag{4.23}$$

$$\mathbf{z_3} = \begin{pmatrix} \bar{x} + x_1 - 3x_2 + x_3 \\ \bar{x} - 3x_1 + 5x_2 - 4x_3 \\ \bar{x} - 3x_2 + 5x_3 - 4x_4 \\ \vdots \\ \bar{x} - 3x_{T-1} + x_T \end{pmatrix}, \tag{4.24}$$

where $\bar{x} := T^{-1} \sum_t x_{it}$ is the time-series average of x for firm i. This setup is easily extended to unbalanced panels by making the period length observation-specific, i.e., T_i instead of T.

In principle, I could use more instruments to improve efficiency. The optimal number depends on the nature of the restrictions I impose on $\mathbf{k}$ and ξ. The previous instruments were derived under the assumption that ξ is stationary and uncorrelated over time. This may be a strong assumption, but given the overidentification, it is testable. Griliches and Hausman (1986) derive the optimal number of instruments as the largest number which all contain independent information about V and K. The maximal number of instruments is T^2. The optimal number of instruments is therefore T^2 minus the number of unique linear restrictions provided by equations (4.19) and (4.20) (since (4.21) is typically not binding). Requirement (4.19) imposes

T restrictions. The optimal number of instruments further depends on the time-series properties of the measurement errors. To see this, rewrite (4.20) as

$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{k}_i+\xi_i)'\mathbf{P}(\varepsilon_i-\xi_i(1+\beta))=0\Rightarrow$$
$$\text{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\xi_{it}\xi_{i\tau}P_{i\tau}=0.$$

The implied restrictions on $\mathbf{P}$ depend on Ξ, the covariance matrix of ξ_i. If ξ_{it} are not stationary nor serially correlated, the exogeneity condition requires that $p_{t\tau}$ equal zero whenever $t=\tau$. This imposes T additional restrictions on $\mathbf{P}$. If ξ_{it} are stationary, then only $\text{tr}(\mathbf{P})$ must equal 0. This is one additional restriction. If the measurement errors follow a MA(m) process, this imposes restrictions of either $m+1$ (if $m<T-2$) or $T-1$ (if $m\geq T-2$).

In my main application, I will assume that ξ_{it} is stationary and serially uncorrelated; under these conditions, the instrument vectors $\mathbf{z_1}$, $\mathbf{z_2}$ and $\mathbf{z_3}$ found above are consistent. In extensions, I relax the assumptions on the covariance structure of ξ by allowing for MA(1) dependence and hence autocorrelation. I maintain stationarity throughout.

4.4. Data

I use administrative data from Statistics Netherlands. I merge several large datasets. I first describe the data I use to estimate firm discount rates. Then, I detail the firm data I use, followed by the household data.

4.4.1. Discount Rates

To construct my estimate of market value, y_{it}, I need discount rates. Since firm equity is a risky asset, the discount rate must be risky. Constructing a discount rate for non-listed assets is challenging, however, since both the costs for equity and for debt are unobserved (Damodaran 2012). For listed assets, one could use one of the various asset pricing models (e.g., Fama and French 1993). Bach, Calvet, and Sodini (2020), for instance, obtain expected returns in private equity by estimating such an asset pricing model on listed stocks, regressing the obtained risk factors on listed firm characteristics, and then assuming that these coefficients are the same for unlisted businesses. These are strong assumptions that strongly depend on the chosen asset-pricing model.

I use three different measures of the discount rate, each following different strands of the literature. The first measure is based on the conceptually correct discount rate, the weighted-average cost of capital:

$$\varrho_t^{\text{wacc}} = \omega_t \times (1 - \tau) \times r_t^d + (1 - \omega_t) \times r_t^e, \tag{4.25}$$

where ω is the leverage ratio (i.e, the share of debt in total firm liabilities), τ is the firm's tax rate, and r^d and r^e are the costs of debt and equity. Computing the WACC for private firms is again challenging. I opt for an indirect approach, using data by Gormsen and Huber (2023, 2024), who calculate the WACC for a large sample of firms based on rich asset-pricing models. The firms in their sample are listed firms who make conference calls, in which they also mention the firms' self-reported discount rates. Gormsen and Huber find that these internal discount rates persistently differ from the WACC. For my first two measures, I take the Dutch firms from their sample, and simply use the average of their data per year. This results in a series ϱ^{wacc}, based on the weighted-average cost of capital, and a series ϱ^{gh}, based on firms' self-reported discount rates.

As a final measure, I follow a large stream of literature, including Barkai (2020), and estimate ϱ^{b} as the ten-year Dutch government bond yield plus a constant risk premium of 5%. Since bond yields have declined strongly, this measure of ϱ declines as well, which will blow up valuations.

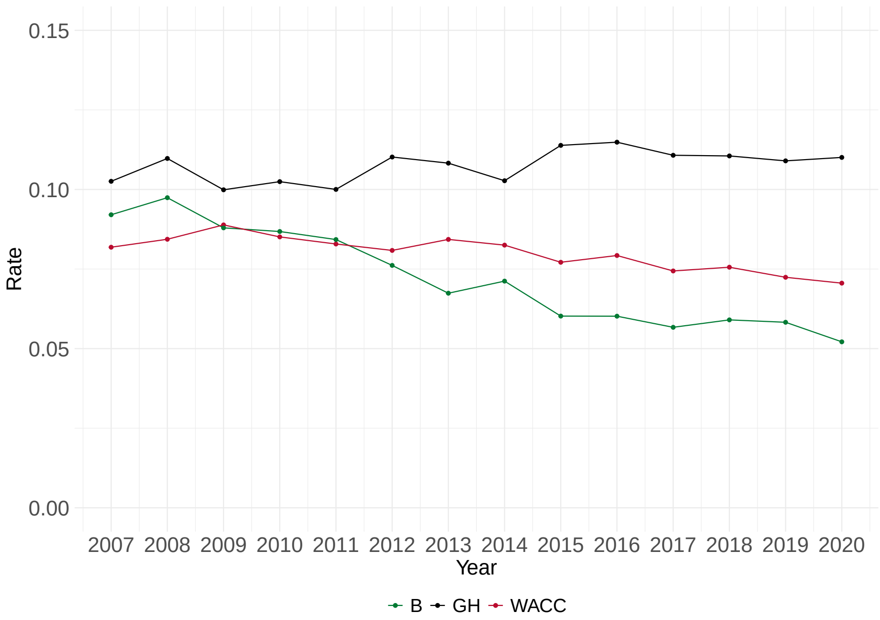
Figure 4.1 plots the different measures. The three discount rates comove, but also display important differences. The series for ϱ^{b} declines strongly following the financial crisis in 2008, whereas the other two series remain much more stable. The self-reported discount rates, ϱ^{gh}, are persistently the highest and stable around 11%. The weighted-average cost of capital, which is conceptually the closest to the true discount rate, is in between these two series for most of the period.

Obviously, all three discount rate measures are highly imperfect; this is why my measurement-error framework is important. Under the identifying assumptions of the model, my GMM estimates will retain only those parts of these series that are structurally related to firms' capital stock. Ideally, the instruments should lead to similar point estimates across the series, since that gives confidence that our initial choice of discount rate does not drive the estimation results.

4.4.2. Firm Data

My main data source is the *Bedrijfsgegevensstatistiek* (BG), which covers the universe of Dutch corporate income tax returns 2007–2019, where dates are December 31st

Figure 4.1: Discount Rates



of a given year. Hence, all my results apply to incorporated firms.[6] The Netherlands levies a progressive corporate income tax with two brackets. For most of my sample period, the first bracket applied to fiscal profits up to 200,000 € (with a rate of 20%), and the second bracket applied to all profits in excess of 200,000 € (taxed at 25%). From the corporate income tax returns, I obtain the full balance sheet and profit and loss statement of each firm. In addition, I observe firm characteristics such as industry, corporate form, and firm age, although coverage varies for these background variables.

The predominant corporate form in my data is the non-listed limited-liability company (*besloten vennootschap*, BV). This is a widely used form that covers a wide variety of firms. In particular, it is not uncommon to observe nests of firms, where firm A owns firm B which owns firms C and so on. The BG data shows balance sheet and income statement variables both at the consolidated and unconsolidated level. Since my ultimate interest is to link firms to owners, I use consolidated data throughout.

---

[6]Note that partnerships and other unincorporated business forms are less prevalent in the Netherlands among the top of the distribution than in other countries, such as the United States (Kopczuk and Zwick 2020). This is because unincorporated businesses are subject to the personal income tax, whereas incorporated firms have distinct tax advantages, as I detail in the next section.

For firm's capital **x**, I use the sum of the firm's fixed capital stock and intangible capital. This choice of variables is done for substantive reasons. The neoclassical investment model emphasizes the role of productive capital in firms' investment decisions. Hence, while balance sheet items such as property contribute to a firm's value, they are not part of the firm's productive capital stock.

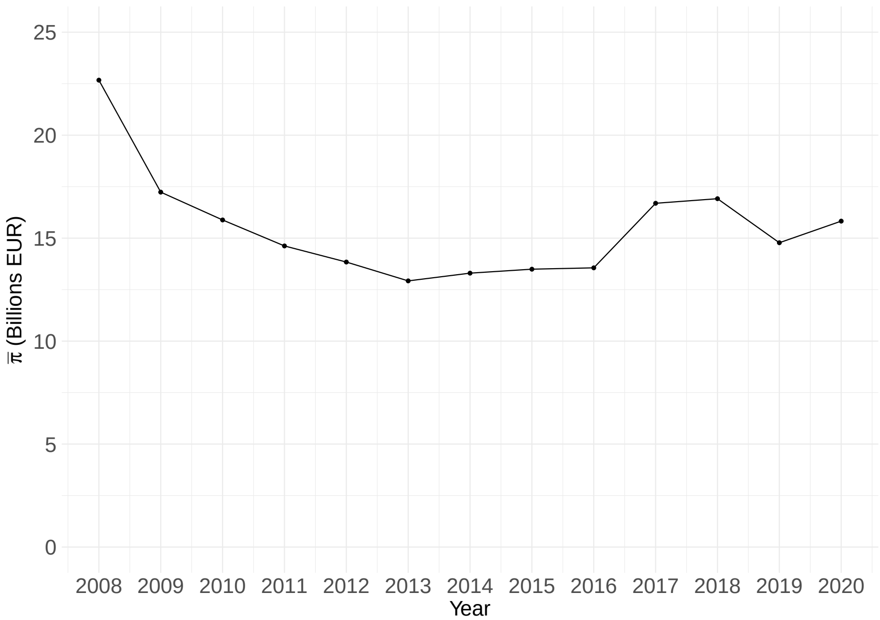
In Dutch accounting rules, intangible assets include goodwill, development (not research), software, and so on. There are two criteria to include intangible assets on the balance sheet. First, it must be likely that the asset generates future economic benefits. Second, the costs of the asset can be reliably assessed. Some costs do not meet these criteria and are instead expensed on the income statement, such as costs for training, advertisement, startup costs for a new product or activity. Research is expensed, not an asset; development can be included as an asset if the to be developed intangible asset is (a) technically feasible; (b) the intention exists to use or sell the asset; (c) the capacity to produce/use the asset must be there; (d) it must be likely to generate future economic benefits; (e) the capacity is there to complete the asset; (f) the costs can be reliably assessed.

To construct my measure of firms' market value, **y**, I combine the discount rates from the previous section with a moving average of firm profits. I use the three-year moving average as a smoothed forecast of future profits.[7] Figure 4.2 plots the aggregate trends of this measure, $\overline{\Pi}_t$. The series is mostly stable, hovering between 12.5 billion € and 17.5 billion € for most of my sample period. In the first year, averaged profits were higher, with the 2008 value peaking at around 22.5 billion €.

My estimation methods rely on existing values for **x**; hence, I drop all observations where this variable is missing. This results in a dataset with about 8 million firm-year observations. Note that not all these observations can be linked to Dutch households, to which I turn next.

[7] In Crouzet and Eberly (2023)'s model, Π_t refers to a firm's operating surplus, i.e., total revenue minus variable costs. The difference between operating surplus and profits in the data is driven by net capital income accruing to the firm, for instance dividends from shareholdings the firm has. Since many firm-owners use their firms (at least partly) as investment funds, these flows should be capitalized into the firm value; excluding it would make little difference to my inequality results, but might affect the interpretation of the structural parameters. I return to this issue when estimating the model's structural parameters in Section 4.6.

Figure 4.2: Aggregate Trends in $\overline{\Pi}_t$



4.4.3. Household Data

I use administrative data on Dutch households, which are available from 2006–2020, measured on January 1st.[8] This dataset covers households' total wealth and its composition. We can distinguish between deposits, financial assets (which are stocks, bonds, and other securities), owner-occupied housing, other real estate, non-incorporated business wealth, incorporated business wealth, other assets, mortgages, student loans, and other liabilities. Since capital-funded pensions are not taxed, they are excluded from all wealth measures.[9]

[8]Hence, when linking firm and household data, I assign firm data from December 31st in year t to year $t+1$.

[9]It is debatable whether this omission is justified. From a permanent-income perspective, funded pension claims enter the net present value of future consumption and hence should be included in wealth. However, the Dutch pension system is different from comparable systems in countries such as the United States in that individuals cannot consume, trade, or otherwise claim their pension assets before retirement; in other words, households hold no property rights over their pension claims. This makes the Dutch capital-funded system more akin to Social Security, which are typically not included in household wealth. In incomplete financial markets, there is a meaningful distinction between present-value-based and property-rights-based definitions of wealth. See Martínez-Toledano, Sodano, and Toussaint (2023) for further discussion of these points and their implications.

The Netherlands levies different types of taxes on different types of assets under its income tax system, which is organized in three 'boxes'. The imputed rent from owner-occupied housing is taxed together with labor earnings (including earnings of the self-employed) in Box 1. This imputed rent increases in the cadastral value of the property, making this Box nominally progressive in capital income; however, since the interest on mortgage debt is deductible, net taxable income from owner-occupied housing is typically negative.[10] Box 2 taxes dividends and realized capital gains from corporations in which a household has more than 5% of the shares (*aanmerkelijk belang*, lit. "significant stake") This tax is progressive and is paid on top of the corporate income tax described in the previous section.

All other assets (except pension claims) are taxed in Box 3, which taxes a presumptive or fictitious rate of return.[11] For most of my sample period, Box 3 charged a flat tax of 30% on a presumptive return of 4% (implying an effective wealth tax of 1.2%), above some threshold. From 2017, progressive rates were introduced in Box 3, based on a presumed portfolio composition at different wealth levels, where the presumptive return in each bracket was the weighted average of the historical return of these assets (deposits, stocks, bonds, etc.).[12]

This wide variation in tax rates according to asset class leads to wide variation in effective tax rates on households' assets, and therefore also provides fiscal incentives to move assets to low-tax Boxes. In particular, Box 2 provides incentives to defer taxation by deferring dividend payouts (although the corporate income tax continues to apply). Firm-owners can borrow against the value of their firm and use these loans to finance their consumption; since these loans depress the value of their assets, this might reduce additionally reduce their tax burden in Box 3 (IBO Vermogensverdeling 2022). Indeed, in their analysis of Dutch tax progressivity, Bruil et al. (2022) find that the Dutch overall tax system is regressive, which is largely due to the fiscally advantageous role of Box 2.

Since wealth is typically owned jointly by household members, all data are on the household level. We link this dataset to firm data using the Shareholder

[10] The Netherlands has uniquely large mortgage debt, in the range of 100% of national income in the period I cover (Toussaint, de Vicq, Moatsos, and van der Valk 2022). This has been driven by government stimulus of owner-occupied housing, as exemplified by interest-only mortgages (Bernstein and Koudijs 2024).

[11] Since legally the tax is on a return to assets, it is called a capital income tax. Since the rate of return is not the realized return but a presumptive return, Box 3 is a *de facto* wealth tax (Jacobs 2015).

[12] This differentiation was challenged in court and declared illegal by the Supreme Court in 2022, who held that it discriminated against individuals who did not hold this exact portfolio composition. As a result of this ruling, the current Box 3 system based on presumptive taxation has been declared illegal and the government is obligated to move towards a system of taxing realized returns.

Registry (*Aandeelhoudersregister* (AR)), see Appendix 4.B. This dataset provides information on the ownership structure of each incorporated firm at quarterly frequency. Linking the datasets results in about 7 million firm-owner-year observations overall. I merge this linked set with the full dataset on the Dutch wealth distribution, ending up with more than 120 million observations at the household level, about 7 million of which therefore are firm-owner observations.

4.4.4. Summary Statistics

Table 4.1 provides summary statistics for all firms which are matched to Dutch household owners. It provides selected items from the balance sheets and income statements, as well as ownership concentration.

It is clear that firm values are widely dispersed. Some firms are worth in excess of 45 billion €, while the median firm is worth only 89,000 €. Similarly, there is wide variation in the structure of firms' balance sheets. Half of all firms have negative or negligible values of capital stocks, and similarly for most other components. The income statement also shows wide variation. Personnel costs and depreciation are less dispersed than financial variables such as the firm's capital income flows. Panel C reveals that most firms are fully owned by a single owner. Given the large values of firm wealth, this suggests that firms play an outsize role in wealth concentration.

This suggestion is confirmed by Table 4.2, which dissects the wealth distribution for the year 2019. Note that the summary statistics in this Table are all at the household level.

Table 4.1: Summary Statistics, Firms & Firm Owners

Item	Min	P25	Median	Mean	P75	Max
Panel A: Balance Sheet						
Intangible Capital	0	0	0	0.031	0	6,217
Property	-0.368	0	0	0.228	0	1,824.529
Fixed Capital	-0.685	0	0.001	0.321	0.048	3,394.027
Shares	-1,564	0	0	0.834	0.018	44,670
Inventory	-296.648	0	0	0.750	0	3,831.089
Claims	-8.58	0.008	0.043	0.475	0.159	235,900
Securities	0	0	0	0.233	0	9,654
Total Assets	-1,552	0.102	0.310	2.509	0.902	243,000
Long-Run Liabilities	-0.179	0	0	0.294	0	21,970
Short-Run Liabilities	-2,475	0.005	0.025	0.557	0.109	16,680
Net Worth	-853.8	0.008	0.089	1.367	0.437	45,650
Panel B: Income Statement						
Net Revenue	-7.265	0	0.045	0.959	0.198	227,800
Personnel Costs	-16.559	0	0.03	0.174	0.096	508.922
Depreciation	-9.651	0	0	0.008	0	43.216
Total Costs	-91.1	0.004	0.057	0.841	0.174	78,530
Net Financial Result	-250.342	-0.003	0	-0.002	0.005	2,015.167
Shareholdings Result	-7,173	0	0	0.098	0	5,265
Fiscal Profit	-884.602	-0.004	0.003	0.065	0.035	4,341.089
Panel C: Ownership						
Shares (%)	0.128	50	100	76.563	100	100

Notes: All observations pooled (6,319,593 observations); values in millions of nominal EUR. Not all items on the balance sheet and profit & loss statement are summarized above.

Table 4.2: Wealth Distribution, 2019

					Composition (Share of Assets)				
Bracket	# Households	Threshold	Mean	Share	Deposits	Financial	Real Estate	Business	Debt
Population	8,045,662		212,638	1.000	0.125	0.071	0.622	0.182	−0.334
Bottom 50%	4,022,825		−1,907	−0.004	0.152	0.012	0.830	0.006	−1.033
Middle 40%	3,218,268	48,838	197,988	0.372	0.134	0.023	0.815	0.029	−0.404
Top 10%	804,569	466,418	1,343,958	0.632	0.113	0.124	0.420	0.344	−0.145
Top 1%	80,457	2,197,062	5,935,055	0.279	0.056	0.148	0.222	0.574	−0.123
Top 0.1%	8,046	9,749,227	24,404,472	0.115	0.033	0.139	0.127	0.700	−0.099
Top 0.01%	805	40,537,908	93,433,157	0.044	0.023	0.116	0.067	0.794	−0.064

Notes: Values for the threshold and mean in nominal EUR. Financial assets include stocks, bonds, and other securities. Real estate is the sum of owner-occupied housing and other real estate. Business is the sum of incorporated and non-incorporated business wealth. Debt is the sum of mortgages, student loans and other liabilities.

We first note that wealth is highly concentrated. The top 1% of households owns almost a third of all wealth. The top 0.01% – 805 households – own 5%. Another observation is that the composition of wealth varies dramatically over the wealth distribution. The bottom 90% have most of their wealth in housing and deposits, differing only in the relative value of their house vis-a-vis their mortgage. Only in the top 10% do financial assets start to play a meaningful role. In fact, most of the top 10% looks similar in wealth composition to the bottom 90%, and only from the top 1% or so upwards do we see financial and business wealth become prominent. Of the two, the latter is by far the most important. The top 0.01% hold almost 80% of their assets in private business wealth. This confirms that the correct valuation of these assets is of key importance for studying wealth concentration at the top.

4.5. Estimation Results

In this section, I report the results from my econometric procedure. I begin with simple OLS estimates to give a benchmark. Then, I move to the main results of the chapter. I finish the section by exploring some extensions.

4.5.1. OLS

We begin our investigation by simple OLS regressions. Recall that I use three distinct discount rates: ϱ^{wacc}, ϱ^{gh}, and ϱ^{b}. I use these discount rates to capitalize smoothed profits, $\overline{\pi}_{it}$, to create three distinct raw series of firm value, y_{it}. In Table 4.3, I regress these series on the firms' capital stock. I report Driscoll-Kraay standard errors, which are robust to arbitrary serial correlation and heteroskedasticity.

I first investigate a simple regression, in columns (1), (4), and (7). Then, in succeeding columns, I add fixed effects for year and firm sector (at the two-digit level). Across specifications, coefficients remain stable and are always highly significant. Adding fixed effects makes no difference. To some extent, this is to be expected: our measures of y_{it} are based on the same discount rate for all firms within a year; hence, all variation must be within-year variation. Nevertheless, it is surprising that adding industry fixed effects does not alter any of the results. Based on these specifications, we would expect a coefficient of β in the region of 3.5–4.5. However, the measurement error concerns noted earlier imply that the true value could well be outside this interval too. Next, I investigate whether using instruments in a GMM procedure improves identification.

Table 4.3: OLS Estimates

	e^{wacc}			e^{gh}			e^{b}		
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β	4.23***	4.23***	4.24***	3.37***	3.37***	3.37***	4.54***	4.55***	4.55***
	(0.664)	(0.665)	(0.665)	(0.576)	(0.576)	(0.577)	(0.533)	(0.533)	(0.533)
Fixed-effects									
Year		✓	✓		✓	✓		✓	✓
Industry			✓			✓			✓
Fit statistics									
N	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564
R^2	0.335	0.335	0.335	0.335	0.335	0.335	0.340	0.340	0.340
Within R^2		0.335	0.335		0.335	0.335		0.340	0.340

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*
Notes: Driscoll-Kraay standard-errors in parentheses.

4.5.2. Time-Series Identification

Next, we look at the time-series results. Table 4.4 shows the manual identification of β using equations (4.17) and (4.18). Column (1) shows the regression of the within estimator, and (2) that of the first-difference estimator. Note that the identification of Griliches and Hausman (1986) relies on both estimators having the same numbers of observations. Therefore, whenever a lagged value is not available, I set that value to 0 instead of missing. This is consistent with the econometric framework developed in Appendix 4.A.2.

Table 4.4: Within vs. First-Difference Transformation

Discount Rate:	ϱ^{wacc}		ϱ^{gh}		ϱ^{b}	
Dependent Variable:	$\ddot{y}_{it}$	Δy_{it}	$\ddot{y}_{it}$	Δy_{it}	$\ddot{y}_{it}$	Δy_{it}
Model:	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{\beta}$	2.92	1.81	2.06	1.43	4.02	1.84
	(2.46)	(1.59)	(1.78)	(1.28)	(3.03)	(1.53)
Fit statistics						
Observations	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564	6,713,564
R^2	0.788	0.062	0.799	0.069	0.773	0.052
Within R^2	0.093		0.078		0.138	
Separate Parameter Estimates:						
$\sigma^2_{\ddot{x}}$	2.23×10^{15}		1.35×10^{15}		2.72×10^{15}	
$\sigma^2_{\Delta x}$		6.65×10^{14}		3.70×10^{14}		8.33×10^{14}
Joint Parameter Estimates:						
β	3.09		2.15		4.38	
σ^2_{ξ}	1.38×10^{14}		6.21×10^{13}		2.41×10^{14}	

Driscoll-Kraay standard-errors in parentheses
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Both values are positive, but again insignificant. Note that $\widehat{\beta}_{FE} > \widehat{\beta}_{FD}$, which is to be expected; Griliches and Hausman (1986) argue that this indicates a panel structure with a declining autocorrelogram in the variables. In the bottom panel, we identify $\beta \in (2.1, 4.4)$, using the estimated variances of the regressions and point estimates from the three measures. This is a wide interval, but again, given the insignificance of the point estimates, it is unclear where in this interval the true value lies. The manual calculations underlying this Table do not allow to weight the data, which results in a loss of efficiency.

I now systematically employ the instrumental variables strategy based on the time-series restrictions. I use two-step GMM (Hansen 1982), which is asymptotically consistent. Moreover, since the weighting matrix converges to the optimal weighting matrix, the results should be more efficient than either the OLS or manual calculations reported above. I use heteroskedasticity and autocorrelation-robust standard errors, based on Andrews (1991).[13] The results are in Table 4.5.

Table 4.5 consists of three panels, one for each measure for y_{it}. Each panel has seven columns, which each report estimation results based on all possible combinations of my three instrumental variables. I report robust first-stage F-statistics based on Montiel Olea and Pflueger (2013), which are robust to serial correlation, clustering, and heteroskedasticity. The endogenous right-hand side variable is x_{it}, which is identical across all 21 specifications; hence, it should be no surprise that the first-stage F-statistic does not vary between panels in the same column.

Start with the just-identified columns (1)–(3). Column (1) is insignificant across panels. Columns (2) and (3) are significant and the coefficients are sizable.The robust first-stage F-statistics are quite sizable for those columns, although strictly below the thresholds where the asymptotic bias of the GMM estimator is no more than 10% of the OLS bias (which is the criterion used by Montiel Olea and Pflueger 2013). Column (2) is significant at the 5% level in all panels, whereas column (3) is significant at the 0.1% level. Moreover, coefficients in column (3) are generally larger than in column (2), with the exception of Panel C.

Next, we move to the specifications with more instruments than endogenous variables, which permit overidentification tests. Columns (4)–(7) show J-tests which are uniformly insignificant. In columns (4) and (6), these test statistics are actually almost equal to 0. The statistics in column (7) are larger in magnitude, but given the extra degree of freedom these statistics are actually very insignificant. These results indicate that the identifying assumptions underpinning my GMM procedure are easily upheld. Coefficients in these columns tend to cluster closer together within each panel, although there is quite some variation across panels. The coefficients for ϱ^{wacc} are in the 3.9–4.5 range, while those for ϱ^{gh} cluster between

[13] I use the Bartlett kernel, with a bandwidth of three.

Table 4.5: GMM Estimation, Time-Series Identification

Dependent Variable:	y_{it}						
Instruments	z_1	z_2	z_3	z_1, z_2	z_2, z_3	z_1, z_3	z_1, z_2, z_3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: ϱ^{wacc}							
β	4.610	4.089*	4.543***	3.882*	4.530***	4.529***	4.347***
	(3.077)	(2.001)	(1.238)	(1.865)	(1.237)	(1.141)	(1.031)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				0.093	0.078	0.001	0.159
J-test p-value				0.761	0.780	0.977	0.923
First Stage F-statistic	1.934	6.205	7.785	2.932	6.507	8.658	7.066
Panel B: ϱ^{gh}							
β	3.242	2.909*	3.580***	2.782*	3.507***	3.640***	3.423***
	(2.221)	(1.433)	(0.977)	(1.337)	(0.962)	(0.929)	(0.828)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				0.070	0.303	0.038	0.338
J-test p-value				0.792	0.582	0.845	0.845
First Stage F-statistic	1.934	6.205	7.785	2.932	6.507	8.658	7.066
Panel C: ϱ^{b}							
β	6.347	5.289*	5.052***	4.845*	5.031***	4.743***	4.615***
	(3.940)	(2.590)	(1.393)	(2.383)	(1.380)	(1.177)	(1.059)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				0.255	0.014	0.176	0.263
J-test p-value				0.614	0.907	0.675	0.877
First Stage F-statistic	1.934	6.205	7.785	2.932	6.507	8.658	7.066

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Two-step GMM with heteroskedasticity and autocorrelation-robust standard errors in parentheses. Robust first-stage F-statistic by Montiel Olea and Pflueger (2013). Each panel shows GMM results with as independent variable the **y** computed using the respective discount rate.

2.7–3.6 and those for ϱ^{b} range between 4.6 and 5.0. Coefficients in column (4) are significant at the 5% level; those in (5)–(7) are significant at the 0.1% level.

The first-stage F-statistic in columns (4)–(5) is reduced compared to columns (3). This indicates that these combinations of instruments does not add additional information that can be used to significantly identify the first-stage coefficient. Column (6), on the other hand, has higher F-statistics than (3), and column (7) is comparable in magnitude. These coefficients also do not meet the Montiel Olea and Pflueger (2013) threshold above which an instrumental-variables estimate is asymptotically less biased than the comparable OLS estimate; although it should be noted that the standard Kleibergen-Paap F-statistic is much higher for column (7), on the order of 43. Thus, formally my coefficients are estimated with weak instruments (Andrews, Stock, and Sun 2019); nevertheless, they appear reasonably informative.

Column (6) combines $\mathbf{z_1}$ and $\mathbf{z_3}$. Interestingly, the coefficients for ϱ^{wacc} and ϱ^{b} in Panels A and C are now almost identical around 4.6, where they diverged quite substantially in specifications (1)–(6). The coefficient in Panel B remains lower at 3.6. Column (7) combines all three instruments. The resulting coefficients are again very similar in panels A and C, and consistent with the coefficients in column (6). The coefficient in Panel B remains lower, but is very close to that in column (6) as well.

I conclude that my GMM procedure holds up. The framework results in coefficients which are reasonable in magnitude, highly significant, with reasonably strong first stages and insignificant overidentification tests. The fact that coefficients for ϱ^{wacc} and ϱ^{b} converge, despite being based on very different discount rate trends and despite having very different OLS coefficients (see Table 4.3), gives confidence that these coefficients are close to the true value. On the other hand, from a data-driven perspective there is no reason to reject the coefficients in Panel B. Obviously, the lower coefficients in Panel B will result in lower values for private business wealth and hence inequality.

In the end, I use two coefficients. The first, dubbed $\beta^{+} \approx 4.56$, is the average of columns (6) and (7), Panels A and C. The second, $\beta^{-} \approx 3.53$, is the average of columns (6) and (7) in Panel B. With this high and low coefficient, I will estimate top wealth shares and heterogeneous returns to wealth. In Section 4.6, I investigate whether the magnitudes of these coefficients make economic sense at the aggregate level.

4.5.3. Extensions

So far, the time-series identification rested on the identifying assumption that the measurement errors ξ_{it} were stationary and serially uncorrelated. Griliches and Hausman (1986) show that it is possible to relax these assumptions and still obtain valid instruments. I now assume that the measurement errors are stationary but follow an MA(1) process. The strategy is, as before, to find an estimating framework that gets rid of the fixed effect γ, and finding valid instruments to that effect. Following Griliches and Hausman (1986), we end up with the following system of three equations:

$$\Delta y_{it} = \Delta x_{it} + \Delta u_{it} \tag{4.26}$$

$$\Delta_2 y_{it} = \Delta_2 x_{it} + \Delta_2 u_{it} \tag{4.27}$$

$$\Delta_3 y_{it} = \Delta_3 x_{it} + \Delta_3 u_{it} \tag{4.28}$$

where $\Delta_\tau y_{it} := y_{it} - y_{i,t-\tau}$, i.e., the long difference of order τ. Under the assumption that ξ is stationary and follows an MA(1) process, valid instruments for the equations are, respectively:

$$\mathbf{z_4} = \left(x_{i,t-1} + x_{it}, \quad x_{i,t+2}, \quad x_{i,t-2}\right)' \tag{4.29}$$

$$\mathbf{z_5} = \left(x_{i,t-1}, \quad x_{i,t+1}\right)' \tag{4.30}$$

$$\mathbf{z_6} = \left(x_{i,t-2}, \quad x_{i,t-1}, \quad x_{i,t+1}\right)' \tag{4.31}$$

We simultaneously estimate this system of equations using system-GMM. Table 4.6 reports the results.

Table 4.6: System GMM Estimation, Time-Series Identification

	ϱ^{wacc}			ϱ^{gh}			ϱ^{b}		
y	Δy_{it}	$\Delta_2 y_{it}$	$\Delta_3 y_{it}$	Δy_{it}	$\Delta_2 y_{it}$	$\Delta_3 y_{it}$	Δy_{it}	$\Delta_2 y_{it}$	$\Delta_3 y_{it}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β	1.937	1.895	1.583	1.509	1.488	1.236	2.206	1.996	1.653
	(1.841)	(1.175)	(0.997)	(1.411)	(0.934)	(0.802)	(2.033)	(1.183)	(0.981)
x	Δx_{it}	$\Delta_2 x_{it}$	$\Delta_3 x_{it}$	Δx_{it}	$\Delta_2 x_{it}$	$\Delta_3 x_{it}$	Δx_{it}	$\Delta_2 x_{it}$	$\Delta_3 x_{it}$
z	z_4	z_5	z_6	z_4	z_5	z_6	z_4	z_5	z_6
N	6,713,561	6,713,561	6,713,561	6,713,561	6,713,561	6,713,561	6,713,561	6,713,561	6,713,561
F-test	2.621	5.205	14.862	2.621	5.205	14.862	2.621	5.205	14.862
J-test		2.613			2.766			2.201	
p-value		0.759			0.736			0.821	

The table shows a value for β that is positive and mostly uniform across the equations, but lower in magnitude than the results obtained previously. Moreover, the coefficients are all insignificant, indicating that there is insufficient within-firm variation in long differences to identify β (standard errors are clustered at the firm level). The robust first-stage F-statistics of these models are insignificant except for the third equation, which is highly significant. The joint J-test is insignificant, indicating that the over-identifying assumptions fail to be rejected.

Since the coefficients are insignificant, it is unclear whether the true coefficient is lower than those found in Table 4.5, since many of the coefficients from that Table are within a standard error from the coefficients reported here. Moreover, these specifications are in first and longer differences, whereas the earlier results are for variables in levels. Given the lack of unambiguous differences relative to Table 4.5, I stick with the coefficients from that table.

4.6. Aggregate Implications

In this Section, I investigate the aggregate implications from my empirical results. First, I contrast my alternative measures of firm value to the recorded book values. Second, I investigate whether the data are consistent with the model laid out in Section 4.2. Both exercises allow me to check whether the results obtained in the previous Section make economic sense.

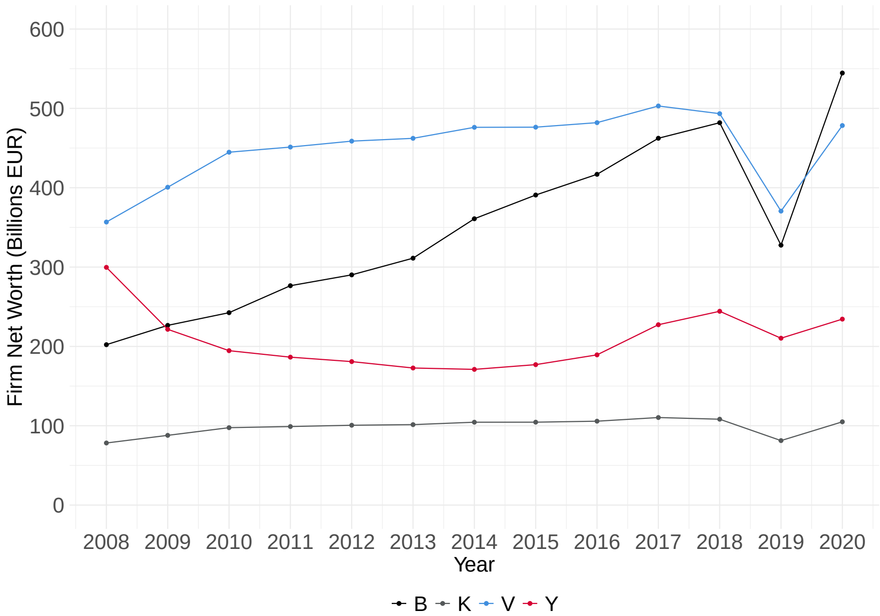
4.6.1. Aggregate Values

With the results from Section 4.5 in hand, we can estimate the changes in private firm value. I first present estimates at the aggregate level. To do so, I aggregate three series: the microdata (based on book values, denoted b_{it}), the raw estimated market values based on a discount rate (y_{it}), and the corrected market values based on the GMM procedure ($\widehat{y}_{it} =: v_{it}$). I then aggregate these as

$$B_t = \sum_i b_{it},$$
$$Y_t = \sum_i y_{it},$$
$$V_t = \sum_i v_{it}.$$

In Figure 4.3, I report the results of this aggregation. I use y^{wacc} and I use the fitted values using β^{+}. When moving to wealth shares in the next section, I more systematically compare different fitted values. I also show values for the aggregate capital stock $K_t = \sum_i k_{it}$.

Figure 4.3: Adjusted Values for Total Private Business Wealth

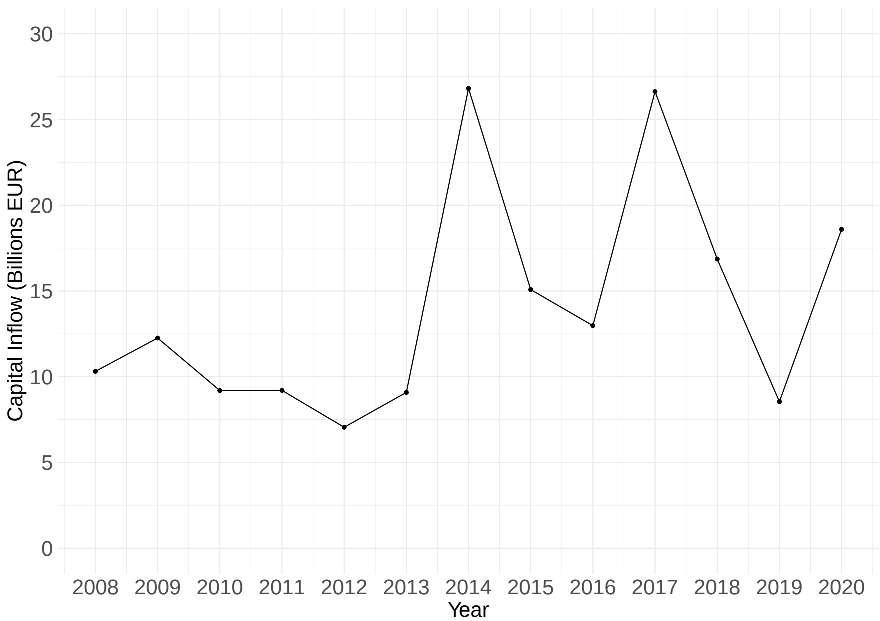


Notes: All values are aggregated across firms. B = book value (recorded private business wealth); K is the capital stock, Y is the raw alternative measure of firm value (based on $\overline{\Pi}_t$ and using the ϱ^{wacc} discount rate); and V is the fitted value from Table 4.5 (using β^{+}).

The microdata, based on book values B_t, show a steep upward trend. In contrast, the raw alternative measure Y_t is relatively stable, only increasing somewhat after 2016. It is also generally lower than B_t. The relative stability is a function of the relative stability of both $\overline{\Pi}_t$ and ϱ^{wacc}. Figure 4.3 also shows the importance of adjusting the raw Y_t using the GMM procedure. Effectively, V_t is the component of Y_t that is structurally related to the true capital stock K_t and is not attenuated by measurement error. As a result, the aggregate updated series for private business market value, V_t, is generally larger than both the book value and Y_t. It is also increasing over time, but at a slower rate than the book value, reflecting that part of the steep increase in B_t is an accounting artefact rather than a true increase in economic value. After 2016, the two series more ore less coincide. In the final two years, the book value outpaces my market-value measure. This is reassuring, since it means that my adjustments do not mechanically increase market values. Instead, the updated V_t seems to be a different series than B_t (although obviously correlated), lending confidence to our results.

The steep increase in B_t is puzzling, because many of the fundamentals seem not to have changed. The reported capital stock K_t hovers around 100 billion €, and Figure 4.2 shows that aggregate profits are also stable. Of course, profits accumulate into book value, so stable values for $\overline{\Pi}_t$ would result in a constant increase in B_t. The steepness of the increase, however, cannot solely be explained by profit accumulation. Inspection of the balance sheets suggests that a substantial part of the increase might be driven by fiscal motives. As explained in Section 4.4, financial assets other than private businesses are subject to a wealth tax in Box 3, based on a presumptive rate of return. As interest rates declined, this presumptive rate of return became increasingly difficult to achieve, especially for deposits and other safe assets. Reallocating toward private businesses allowed investors to escape the increasing tax burden in Box 3.

Figure 4.4: Capital Inflows, 2008–2020



Notes: Figure shows aggregate values for gross capital inflows into firms.

The granular data of the corporate income tax returns allows me to partially investigate this hypothesis. Figure 4.4 shows the time series of aggregate gross capital inflows into firms, i.e., capital reallocated towards the firm by its owners. This variable is distinct from profits and is purely a financial reallocation. I use gross inflows because I am interested in an allocation towards private business firms, which are taxed in the fiscally advantageous Box 2. Using net inflows (where I correct for outflows) makes little difference to these results.

Figure 4.4 shows that inflows were substantial in this period. In particular, after 2013 inflows increased substantially. The series is quite volatile, but the post-2013 average is clearly higher than the pre-2013 average. This coincides with the increase in the gradient of the B_t series in Figure 4.3. The two peaks in 2014 and 2017, moreover, coincide with the steepest increases in B_t. This suggests that capital inflows are a significant driver of the increase in recorded book values. The post-2013 period also saw a sustained decline in global safe interest rates, as can be gleaned from the ϱ^{b} series in Figure 4.1. Of course, safe interest rates were already declining before 2013 as well. However, bank deposit interest rates declined steeply after 2013, as can be seen in Figure 4.5. This Figure plots monthly deposit interest rates for Dutch account-holders since 2003. It is clear that deposit interest rates collapsed after 2013, precisely the year when capital inflows pick up in Figure 4.4 and book values start increasing markedly in Figure 4.3.

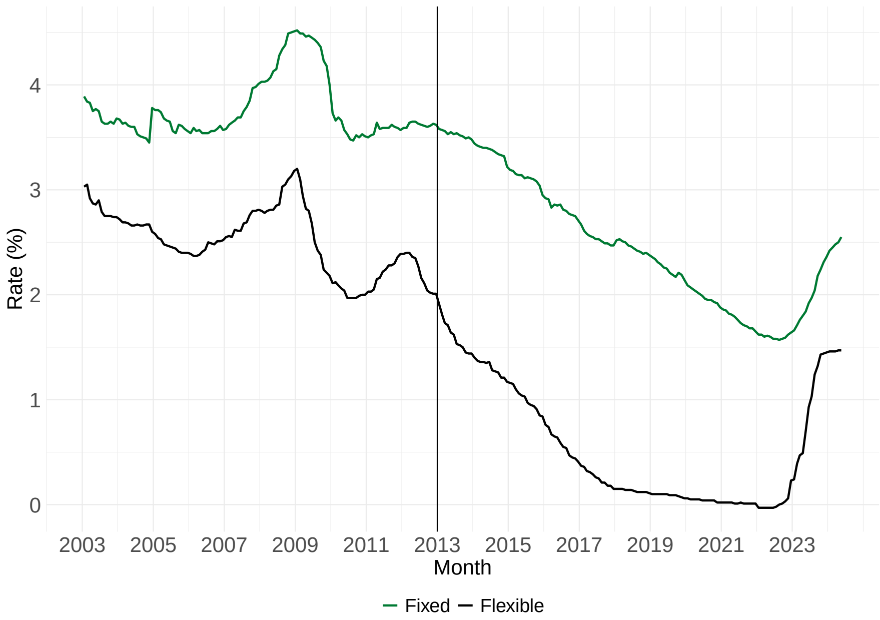
2013 is also the year in which some exemptions in Box 3 were phased out by the Rutte II government. These exemptions concerned social investments, cultural investments, and investments in venture capital. In particular the latter category may be relevant for the wealthiest investors. As those categories were no longer exempt from the wealth tax in Box 3 from 2013 onward, there were strengthened incentives to reallocate capital toward private firms in Box 2.

I conclude from the preceding analysis that the book values as recorded in the data are very sensitive to fiscal considerations. While precise causality cannot be established with my data, the timing of interest rate declines together with the ending of exemptions in Box 3 suggests that the increase in book values is to a significant extent artificial. In contrast, my alternative measures for firm value are based on firm capital and profits. While profits are in principle susceptible to fiscal manipulations, the corporate income tax environment remained quite stable in this period. Moreover, my use of three-year moving averages in $\overline{\Pi}_t$ should alleviate the worst of profit manipulations. Taken together, the fitted values V_t are more a more robust and plausible measure of firm value than the book values.

4.6.2. Parameter Implications

Another way of investigating the results from Section 4.5 is by pondering their implications for the model in Section 4.2. In standard neoclassical frameworks, a

Figure 4.5: Deposit Interest Rates, 2003–2024



Notes: Figure shows monthly deposit interest rates for Dutch households. Red line shows flexible deposits, blue shows long-term fixed deposits. Data from De Nederlandsche Bank.

value of $\beta = 4.6$ would imply that on the margin, a euro of investment yields 4.6 euros in firm value. This is an implausibly high number; indeed, Tobin's average Q for U.S. nonfinancial corporations – which should equal marginal q under Hayashi (1982)'s assumptions – has not exceeded 1.75 since World War II, according to the U.S. Flow of Funds. In my model, there are no capital adjustment costs for simplicity; hence, absent market imperfections, Tobin's q should identically equal 1.

The point of the model is of course that market imperfections do exist and do matter. Equation (4.9) shows that β equals the product of two terms. The first term, $1/\mu$, is the inverse of the curvature of the profit function. μ depends on the capital output elasticity δ and the ratio of market power to returns to scale $\varphi = \lambda/\eta$. The second term, H_t, is a product of aggregate demand, the price level, and aggregate productivity, and also depends on δ and φ. The variables that enter H_t are unobservable to the econometrician, making it impossible to exactly disentangle the contributions of each variable to β. However, we can use the model to estimate all parameters separately, and investigate whether they are mutually consistent and make economic sense.

To identify the parameters, I follow Crouzet and Eberly (2023). In the model, Π_t is a firm's profits. Define υ as the ratio of profits to net revenue, and ϕ as the ratio of capital costs to operating surplus:

$$\upsilon := \frac{\Pi_t}{P_t C_t}, \tag{4.32}$$

$$\phi := \frac{\varrho K_t}{\Pi_t}. \tag{4.33}$$

Then, using equations (4.4)–(4.6), all reduced-form parameters can be identified, using

$$\varphi := \frac{\lambda}{\eta} = \frac{1}{\phi + (1-\phi)(1-\upsilon)}, \tag{4.34}$$

$$\delta = 1 - \frac{1-\upsilon}{\phi + (1-\phi)(1-\upsilon)}, \tag{4.35}$$

$$\mu = \frac{\varphi - (1-\delta)}{\delta} = \frac{1}{\phi}. \tag{4.36}$$

To stay close to the model, I measure operating surplus as net revenue minus variable costs, where variable costs are the sum of total labor costs, material inputs, and bought services. The resulting measure for Π_t is very comparable to the series using net profits, but is generally a bit lower since profits also include the net extraordinary revenue (such as capital income from shareholdings). Net revenue is directly measured in the data, and is quite stable, with one exception. In 2017, measured net revenue jumps to about 600 billion €, more than doubling relative to surrounding years. Inspection of the data reveals that there are two firms in this year that each report revenue in excess of 200 billion €, while at the same time reporting extraordinary costs of also more than 200 billion €. Since all other firm-level variables for these firms are at much smaller magnitudes, this suggests that these firms engaged in fiscal manipulation and are dropped as outliers.[14] Capital costs are approximated using ϱ^{wacc} found above, which should give the weighted-average cost of capital. I multiply ϱ^{wacc} with the total value of the capital stock in a given year to approximate total capital costs.

Table 4.7 reports the results. The ratio of capital costs to operating surplus, ϕ, is 0.73 on average. Over the years, this ratio fluctuates between 0.46 to 1.02

[14]None of the other results in this chapter are affected by these outliers, since these firms' capital stocks and profits are not outliers.

Table 4.7: Parameters

Year	ϕ	υ	φ	δ	μ
2008	0.46	0.04	1.02	0.02	2.18
2009	0.50	0.05	1.03	0.03	2.02
2010	0.53	0.05	1.03	0.03	1.90
2011	0.92	0.03	1.00	0.02	1.09
2012	0.94	0.02	1.00	0.02	1.06
2013	1.02	0.03	1.00	0.03	0.98
2014	1.01	0.04	1.00	0.04	0.99
2015	0.69	0.05	1.02	0.04	1.46
2016	0.79	0.04	1.01	0.03	1.27
2017	0.82	0.04	1.01	0.04	1.22
2018	0.70	0.05	1.01	0.03	1.43
2019	0.59	0.06	1.02	0.03	1.70
2020	0.59	0.06	1.02	0.03	1.70
Average	**0.73**	**0.04**	**1.01**	**0.03**	**1.46**

Notes: ϕ is the ratio of capital costs to operating surplus; υ is the ratio of operating surplus to net revenue. φ is the reduced-form parameter governing returns to scale and markups; δ is the output elasticity of capital with respect to variable inputs; and μ gives the curvature of the profit function. All variables are aggregated each year to compute the parameters.

in 2013. Operating surplus is on average about 4% of total net revenue, which is quite stable. These two parameters combine to identify quite stable values for the markup/returns-to-scale ratio φ and the capital output elasticity δ. These values average at 1.01 and 0.03, respectively. Finally, the parameter we are most interested in, the profit curvature parameter μ averages at around 1.5.

The results imply that the second-order term in equation (4.9) is small but not negligible, since a $\mu \approx 1.5$ implies a coefficient premultiplying the second-order term of around -0.11. This suggests that omitting the quadratic term slightly biases the estimation results. The direction of this omitted-variable bias, however, is downward rather than upward, since the coefficient premultiplying the second-order term is negative, and the correlation between the first and second order terms is positive. My results should therefore be read as a lower bound.

These parameter estimates also help explain the large values found for β. Recall that β only equals Tobin's marginal q under neoclassical assumptions, which require no market power ($\lambda = 1$) and constant returns to scale ($\eta = 1$). As can be verified from Table 4.7, this condition is violated in almost every year, since the ratio of λ to η, $\varphi \neq 1$. The source of this violation cannot be unpacked; Crouzet and Eberly (2023) show that λ and η are generally not separately identifiable. Nevertheless, the fact that $\varphi > 1$ indicates a positive net present value of rents. This positive value shows up in H_t and hence in β. This explains why, even absent any adjustment costs in capital, $\beta > 1$, and rationalizes why the coefficient is of a large magnitude (both β^- and β^+ are substantially larger than estimates of Tobin's q common in the literature).

The capital elasticity, δ, is surprisingly small. We are used to aggregate values of δ on the order of 0.3, not 0.03 (Barkai 2020). There can be two explanations for this divergence. First, in the model δ represents the elasticity with respect to all variable inputs. These inputs include labor but also materials and other variable costs. On the aggregate level, the production function is typically written using only labor and capital as inputs. There is in general no reason why micro and macro elasticities would have to be the same (Baqaee and Farhi 2019b). A second explanation might be that the capital stock as recorded in firm balance sheets underestimates the capital stock at the aggregate level. Indeed, in Figure 4.3, the capital stock hovers around a value of 100 billion €. From the National Accounts, we know that the aggregate capital stock (excluding housing) is much larger, hovering around 1 trillion € in this period. Of course, not all of this tenfold increase can be allocated to firms' capital stocks; it is unclear, however, whether firms' capital stocks should be part of the corporate sector or the household sector. The household sector's capital stock is very close to the numbers reported in Figure 4.3 (Toussaint, de Vicq, Moatsos, and van der Valk 2022), suggesting that the capital stocks from firm balance sheets conform most easily to the household capital stock in the National Accounts.

Nevertheless, as discussed throughout the chapter, capital stocks might be underrecorded both in the National Accounts and the microdata. Intangibles, in particular, are hard to measure and value. The upshot of mismeasured capital stocks for the parameters in Table 4.7 is the following. A larger K_t would increase the estimated costs of capital $\varrho^{\text{wacc}} K_t$; hence, ϕ would increase one-for-one. The other parameters are affected by less than one-for-one. Moreover, not all the changes wrought by a larger K_t go in the "right" direction. Increasing K_t by a factor 10 – in effect assuming that firms own the entire capital stock – does result in a $\delta \approx 0.25$. However, the markups/returns-to-scale parameter φ declines to 0.8, which is less than 1 and hence not possible. Likewise, μ declines to 0.21, another impossibility. The reason for these countervailing changes is that while the capital stock might be increased, firm revenues and profits remain unchanged. Hence, to keep the model consistent, the other parameters have to decline.

I conclude that the model is useful for understanding aggregate dynamics for Dutch firms. I find higher curvature of the profit function, as measured by μ, than Crouzet and Eberly (2023). Even allowing for some mismeasurement in capital, this larger value is consistent with the structure of the firm accounts and hence is likely to be a robust finding. This suggests that private firms might have a larger flow value of rents than the public firms investigated in Crouzet and Eberly (2023). This rationalizes the large values for β found in the previous Section.[15]

4.7. Distributional Implications

4.7.1. Top Wealth Shares

Recall that the results in Table 4.5 converge to two coefficients: a high coefficient, β^{+}, and a low one, β^{-}. With these two values for β, I compute fitted values $\widehat{y}_{it}$, which under the identifying assumptions are free from measurement error. Hence, I will treat these fitted values as the true market values v_{it}. With these values of v in hand, we can update wealth shares. In Appendix 4.B, I detail how I compute household's firm wealth. Throughout, I stay close to the official Statistics Netherlands procedure (Menger 2021). The brief procedure is as follows. We can link firms to their owners using the Shareholder Registry. I retain values for firms which can be linked to Dutch households. The Registry also provides information on ownership shares; hence, we can assign the updated values for firm wealth to households in proportion to their shareholdings. I deviate in one important aspect from Statistics Netherlands: Statistics Netherlands assigns ownership shares based on the share reported in the fourth quarter of the previous fiscal year, except for the first year in the dataset, where they use the first quarter of the current fiscal year. I adapt this procedure, by making the starting year firm-specific. If, for instance, firm A enters in 2008, that is the first year in my procedure, and I use the first-quarter ownership share for that year. Under the Statistics Netherlands procedure, this firm would not be included until 2009. As detailed in the Appendix, the Statistics Netherlands procedure assigns values for private business wealth to individuals who have no observable link to a firm, if they do declare firm dividend income. Since I have no

[15] Crouzet and Eberly also investigate the properties of their model using National Accounts data from the United States. As discussed above, there is a general mismatch between aggregate statistics for the capital stock and microdata. Hence, if I were to use their model using Dutch National Accounts, I might well find different parameter values. For the purpose of this chapter, however, the parameters reported in Table 4.7 confirm that the model is broadly consistent with my data.

meaningful improvement to those imputations, I keep those values, only replacing firm wealth for those households I can link to firms.

Once I have adjusted values of firm wealth at the household level, I recompute their position in the wealth distribution, where I adjust wealth totals as well. For instance, the series based on β^+ results in a new estimate of private business wealth. To compute top wealth shares for this series, I replace private business wealth for all matched households with this new measure, resum all new wealth totals, and recompute top wealth shares. This is important, since the adjusted values of private businesses could potentially accrue to households in the middle of the 'old' wealth distribution, who therefore by rights should be further up in the distribution than the microdata suggests.

This procedure results in three series for wealth shares: a raw series from the microdata, where the official aggregated business wealth is taken as given; and the fitted series based on β^+ and β^-. We begin by analyzing the top 1% wealth share. Figure 4.6 shows the unadjusted top 1% share in blue, as well as the two adjusted series. The high series, based on β^+, is in red; the low series based on β^- is in yellow.

Figure 4.6: Updated Top 1% Wealth Shares, 2008–2020

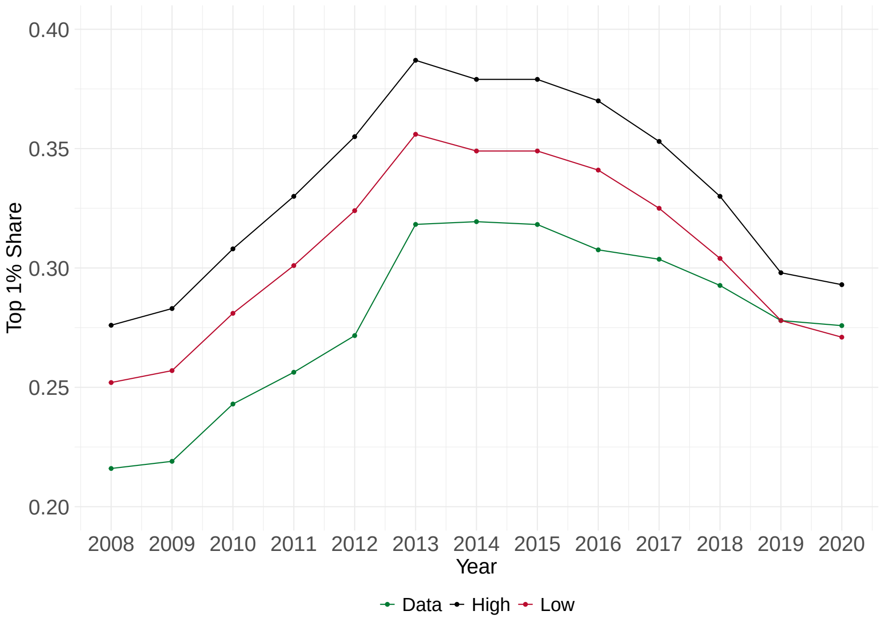


Figure 4.6 reports the results. In this timeframe, the microdata show a steep increase in wealth inequality, following the 2008 financial crisis, with the top 1% share rising from about 22% to almost one-third. After 2014 or so, inequality declined again, stabilizing at around 27% in 2020. The reasons for these dynamics

are well-understood, and are due to differential portfolio composition of the richest households relative to the middle class (Toussaint, de Vicq, Moatsos, and van der Valk 2022; Toussaint 2022). The majority of households predominantly owns housing, whereas the rich hold most of their portfolio in private businesses and other financial assets. After the crisis, housing prices collapsed, while the stock market was faster to recover, resulting in a gain in wealth of the top 1% relative to most of the distribution. After 2014, housing prices started to increase rapidly, leading to a reversal of this trend. These dynamics mirror the "race between the housing market and the stock market" documented by Kuhn, Schularick, and Steins (2020).

The adjusted wealth shares also display these trends, but start on higher levels. Start with the high series, based on β^{+}. This series results in the highest wealth shares, which peak at 38.7% in 2013. The low series shows similar trends to the high series, but peaks at a lower level of 35.6% in 2013. This is still a substantial top wealth share, comparable to the United States (Saez and Zucman 2016; Smith, Zidar, and Zwick 2023).[16] The low series overlaps with the unadjusted top 1% share at the end of the period, stabilizing around 27.5%. The high series remains close to 30% even in those final years.

Both updated series rise and fall more or less similarly, showing that while the level of these series is determined by differences in the construction of the discount rate, their trends are driven more by time-series variation in firm profits. This nuances concerns by some authors that the trends in wealth inequality are primarily driven by interest rates (Greenwald, Leombroni, Lustig, and Van Nieuwerburgh 2021; Cochrane 2020). My results show that it is the interaction between discount rates and profits that matter; both prices and quantities, not just prices.

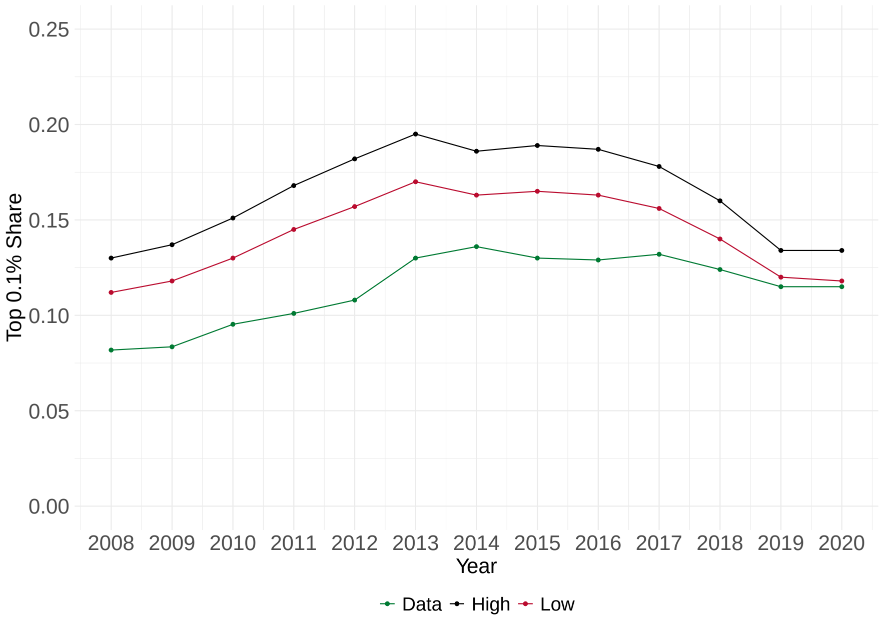
In conclusion, wealth shares are higher than recorded in the microdata. The extent of the upward adjustment differs based on specification. If we take the low coefficients as truth, the data should be adjusted by about 2.8 percentage points on average. If we take the high series as a reasonable average of the statistically strongest specifications, the adjustment should be in the order of 5.5 percentage points on average.

Next, we move to the top 0.1% share. Since private business wealth is even more concentrated among this bracket than among the broader top 1%, as revealed by

[16] For international comparisons, two complications arise. First, the unit of analysis differs between studies. In Saez and Zucman (2016) and Smith, Zidar, and Zwick (2023), the unit is a tax unit, which is not quite the same as a household. Second, the large values of Dutch private pensions may distort comparisons. As explained in Section 4.4, this wealth component is not included in Dutch wealth statistics, including this chapter. Excluding it, however, ignores the large redistributive effect of the welfare state, which is less extensive in the United States. See Martínez-Toledano, Sodano, and Toussaint (2023) for further discussion.

Table 4.2, the resulting adjustments for the top 0.1% share should be even stronger. As before, Figure 4.7 plots the microdata together with the high and low series.

Figure 4.7: Updated Top 0.1% Wealth Shares, 2008–2020



Interestingly, the microdata show a far less marked increase in top wealth shares through my sample, rising from about 7% to stabilize in the 11–12% range. Given the clear dynamics in the overall 1% share, this relative stability is puzzling. In contrast, both adjusted series show much stronger dynamics, which are fully in line with the dynamics observed for the top 1% share. The high series peaks at 19.5% in 2013, is stable around 19% until 2017, and declines in the final years to stabilize around 13.4% in 2020. The low series remains closer to the high series in relative terms than was the case for the top 1% share in Figure 4.6. The low series peaks at 17%, remains stable around 16.5%, then declines to 11.8%. For the high series, the upward adjustment clocks in at about five percentage points on average, which is similar in absolute terms to the upward adjustment in Figure 4.6, and therefore much more significant in relative terms. The low series shows an average upward adjustment of 2.8 percentage points, which is identical quantitatively to the adjustment in Figure 4.6. These results show that the extent of wealth concentration at the top is much larger than can be gleaned from the microdata, with the wealthiest households capturing most of the increase in firm value.

4.7.2. Wealth Composition

A second implication of the adjusted series is that the portfolio composition of the wealthiest may have changed. Figure 4.8 plots the old and adjusted shares of private business wealth in total wealth for the top 1%.

Figure 4.8: Adjusted Private Business Share of Total Wealth, Top 1%

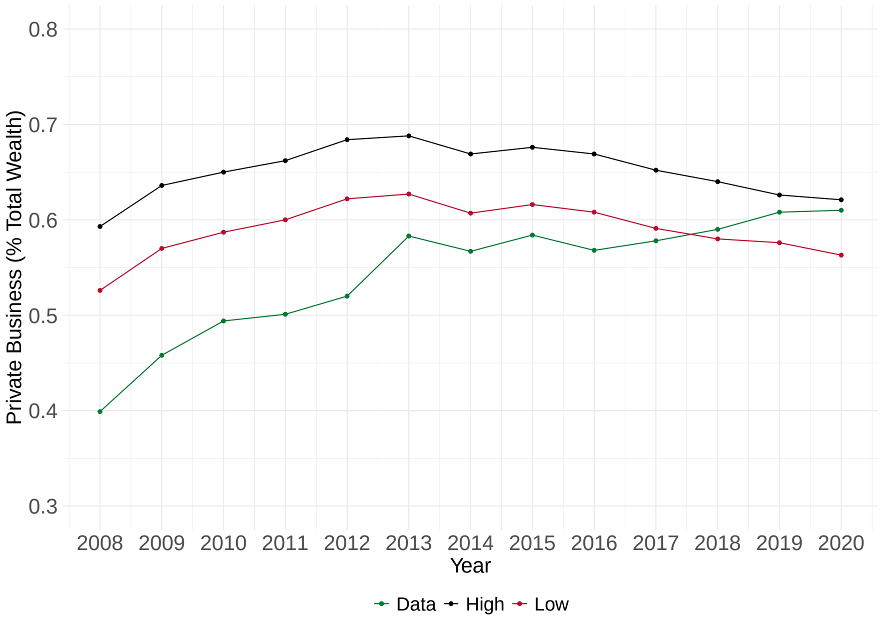


Figure 4.8 shows that private business wealth became increasingly dominant in the microdata. This could be both due to changes in prices and quantities. A decline in interest rates may have increased profitability and asset values of these corporations, boosting their value. At the same time, the richest may have increasingly reallocated their wealth towards private businesses. Fiscal considerations likely played a substantial role. As discussed in Section 4.4 and 4.6, households rebalanced financial assets toward private firms. Since firm-owners are concentrated at the top of the distribution, this portfolio reallocation results in a significant upward trend in the private firm share for the top 1%.

In contrast, the adjusted measures show a much more stable portfolio share. In these values, we also observe an increase between 2008 and 2013, but a relative stabilization afterward; and if anything, the private business share seems to decline somewhat after 2016. Since my alternative measures are based on firms' capital stocks and their profits, they are less sensitive to capital flows resulting from fiscal considerations. Thus, while the accounting importance of private business wealth for top portfolios has increased steeply in this period, the economic importance has remained stable.

4.8. Return Heterogeneity

In a final application, I turn to return heterogeneity. As is well known, standard incomplete-markets models have trouble generating the heavy tails of the wealth distribution seen in the data (Benhabib and Bisin 2018). Modifications that do match the levels of inequality require some multiplicative stochastic process, with some friction added to prevent wealth from becoming unboundedly negative after a succession of bad draws; see Gabaix (2009) for examples of such processes.[17]

In an important paper, Gabaix, Lasry, Lions, and Moll (2016) show that while such random-growth models can generate the levels of inequality, they cannot capture the speed of changes in inequality. To account for those dynamics, models need to be enriched such that returns correlate with wealth. The exact mechanism that generates this wealth-returns correlation is still unsettled (e.g., Jones and Kim 2018; Kacperczyk, Nosal, and Stevens 2019). Nevertheless, the centrality of return heterogeneity to accounting for inequality makes it a crucial concept to better understand.

Empirically, return heterogeneity has been documented in several instances. Piketty (2014) noted that U.S. universities with larger endowments earn higher returns; Saez and Zucman (2016) noted a similar scale effect for foundations. The most systematic evidence comes from Norway (Fagereng, Guiso, Malacrino, and Pistaferri 2020) and Sweden (Bach, Calvet, and Sodini 2020). These papers document wide heterogeneity in (realized or expected) returns across the wealth distribution. To a large extent, this is unsurprising, since portfolios vary dramatically and hence so do risk exposures (Cioffi 2021). However, Fagereng, Guiso, Malacrino, and Pistaferri (2020) show that even within narrow asset classes, returns increase with wealth.

Besides potentially accounting for inequality dynamics, this finding has major normative implications, since it provides a rationale for positive optimal capital (income) taxation (Gerritsen, Jacobs, Spiritus, and Rusu 2024; Guvenen et al. 2023; Guvenen, Kambourov, Kuruscu, and Ocampo-Diaz 2024). Hence, properly accounting for return heterogeneity is of first-order importance.

[17] Formally, these processes converge to a stationary Pareto distribution. Teulings and Toussaint (2023) show that top wealth is actually distributed Weibull, not Pareto. Teulings and Toussaint (2024) provide a network-based model with capacity constraints that can rationalize the Weibull distribution. Note that the test statistics developed in Teulings and Toussaint (2023) also show a clear rejection of Pareto in the dataset in this chapter; results available upon request.

Unfortunately, these results are highly sensitive to even slight measurement error. This is particularly pressing for private business wealth, since most of the return heterogeneity documented in the literature is driven by the top of the distribution. As I have argued throughout the chapter, private business wealth is subject to major measurement error. The problem with returns-wealth correlations is that it divides a cashflow (which potentially also has error) by wealth, which has measurement error, and then relates this ratio to wealth, which has the *same* measurement error. Hence, even absent true heterogeneity, left-hand and right-hand side would be mechanically correlated.

I formulate these claims in the following two propositions. Let $r^*_{it} := \pi_{it}/w^*_{it}$ be the (firm or firm-owner level) return, where π is the (correctly measured, for now) total cash-flow and capital gain out of true wealth w^*.[18] Now assume that wealth w^* is measured with error, so that we only observe $w = w^* + \xi$, with $\mathrm{E}[\pi\xi] = \mathrm{E}[w^*\xi] = 0$. Assume a finite variance of w^* equal to $\sigma^2_{w^*}$. Then, both returns and wealth will be mismeasured, and the common measurement error component will induce spurious correlation.

Proposition 4.1. *Consider the return $r^*_{it} := \pi_{it}/w^*_{it}$ where π is the correctly measured total cash-flow and capital gain out of wealth w^*. Then, when we consider the regression of returns on wealth $r = \beta w + \varepsilon$, the estimate $\widehat{\beta}$ is biased by*

$$\mathrm{E}\left[\widehat{\beta}\right] = \beta\left(1 - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*} + \sigma^2_{\xi}}\right) - \mathrm{E}\left[r^*\xi^2\right]\mathrm{E}\left[w^*\right].$$

Proof. See Appendix 4.A.3. □

The result is driven by the fact that a return is a ratio. The expectation of a ratio does not equal the ratio of expectations; instead, there are second-order terms which do not disappear. As a result, a returns regression is biased by two terms. The first term, $\left(1 - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*}+\sigma^2_{\xi}}\right)$ is the usual attenuation formula, which biases the coefficient toward zero. The second term, $-\mathrm{E}\left[r^*\xi^2\right]\mathrm{E}\left[w^*\right]$ is the result of the nonadditivity of the measurement error. The signs of these coefficients indicate that β will be attenuated toward zero.

Now include measurement error in cash-flows: $\pi_t := \pi^* + \xi$. This is plausible if capital gains are not properly measured, which is highly likely since market values are unobserved. This results in even worse bias in regressions, as formalized in the next proposition.

[18] To emphasize that these regressions can run at the firm or household level, I use w^* instead of v to denote wealth.

Proposition 4.2. *Include measurement error in cash-flows:* $\pi_t := \pi^* + \xi$*. Now, we have*

$$\mathrm{E}\left[\widehat{\beta}\right] = \frac{\mathrm{Cov}\,[r, w]}{\mathrm{Var}\,[w]} = \beta\left(1 - \frac{\sigma_\xi^2}{\sigma_{w^*}^2 + \sigma_\xi^2}\right) + \left(\sigma_\xi^2 - \mathrm{E}\left[r^* \xi^2\right]\right)\mathrm{E}\,[w^*]\,. \qquad (4.37)$$

Proof. See Appendix 4.A.3. □

Relative to Proposition 4.1, the second term is modified to $\left(\sigma_\xi^2 - \mathrm{E}\left[r^* \xi^2\right]\right)\mathrm{E}\,[w^*]$. The sign of this term is ambiguous and depends on the relative size of the variance of measurement error, σ_ξ^2, to the covariance between returns and squared measurement error. Hence, the bias in a regression where both numerator and denominator of returns are mismeasured is sizable but ambiguous in sign.

Note that if the measurement error in cashflows is not identical with that in wealth, but correlated, (i.e., $\pi = \pi^* + \zeta$, $\mathrm{E}\,[\xi\zeta] \neq 0$), the previous result goes through entirely, except that the σ_ξ^2 in the second term is replaced by $\mathrm{E}[\xi\zeta]$.

These two propositions underscore the importance of properly measuring private business wealth and returns. My GMM procedure is a first step toward addressing this issue. Since we have corrected values of firm value v and capital gains, bias will disappear from these regressions, and we can systematically investigate whether returns and wealth are correlated.

4.8.1. Measurement

I follow Fagereng, Guiso, Malacrino, and Pistaferri (2020) in constructing my returns measures. Details can be found in Appendix 4.B. I consider returns both at the level of firm i and household j. Household j's return to firm i in year t is given by

$$r_{ijt} := \frac{\pi_{it} + \kappa_{it}}{v_{ijt-1} + \frac{1}{2} f_{it}} \cdot s_{ijt}. \qquad (4.38)$$

Here, π are total profits, κ are capital gains, and s is the ownership share. As in Fagereng, Guiso, Malacrino, and Pistaferri (2020), the denominator not only includes beginning-of-period assets but also accounts for the fact that there are net inflows during the year, which might capitalize into profits and/or capital gains. Hence, the second factor corrects for the net inflows f, assuming that these occur about halfway during the years on average.

I aggregate this to the household-level return using

$$r_{jt} := \sum_i r_{ijt} \cdot \underbrace{\frac{v_{ijt}s_{ijt}}{\sum_i v_{ijt}s_{ijt}}}_{=:\theta_{jt}}. \tag{4.39}$$

Hence, firm-level returns are aggregated to the household level using value weights θ_{jt}.

I calculate returns based on microdata for all firms and firm-owners in my sample, and do likewise with the adjusted firm-value series. For the microdata, I use total assets as my measure of firm value; this stays close to Fagereng, Guiso, Malacrino, and Pistaferri (2020) and ensures that there are no interpretation difficulties when both numerator and denominator are negative, as is common at the bottom of the distribution. For my adjusted measures, I only have total value, not gross assets. However, since these values are mostly positive (since they are positively related to firms' capital stock, which is generally positive), this issue is less prevalent. I approximate κ in the microdata as the annual change in firm value net of in- and outflows of capital, which are separately recorded in my data and can thus be cleanly removed. For the adjusted series, I simply take the first difference of the adjusted values: $\kappa_{it} = \Delta v_{it}$. Since this corresponds purely to a change in value (since changes in profits are separately included in π_{it} and v_{it} is based on a smoothed three-year moving average of profits), this is a reasonable measure of capital gains.

For the results reported next, I drop all observations where returns do not exist (which happens if all components in Equation (4.38) do not exist or the denominator equals zero). Moreover, since small denominators can result in very large returns, I trim the top and bottom 0.5% of the sample in each year for all return measures. This results in a dataset of about 3.3 million firm-year observations.

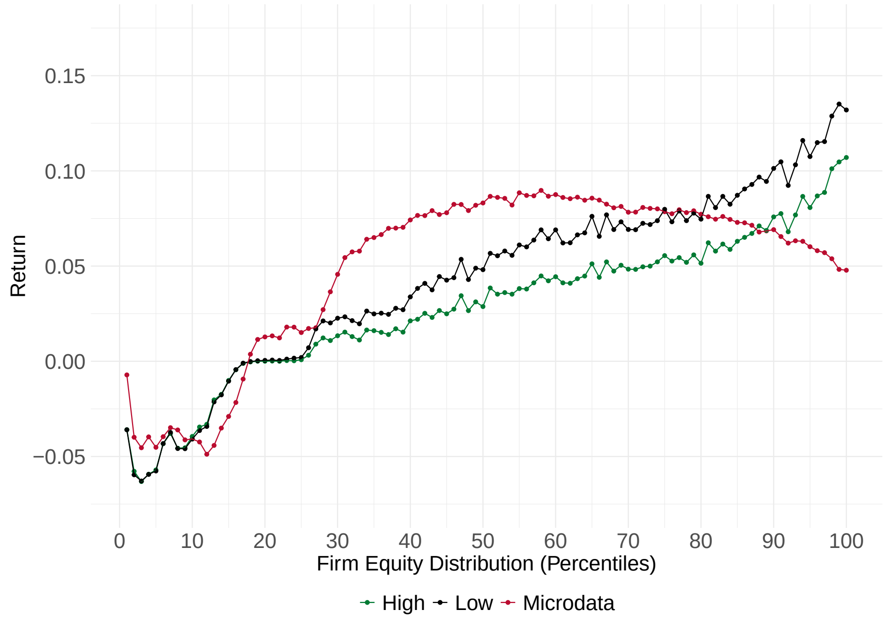
4.8.2. Results

I report visual evidence for return heterogeneity in Figure 4.9. Panel (a) shows heterogeneity on the firm-level, with firms ranked by their position in the firm equity distribution. For each percentile of this distribution in each year, I collect all returns. For robustness, I then report the median of these returns for each percentile.[19] Panel (b) does the same for returns aggregated to the household level.

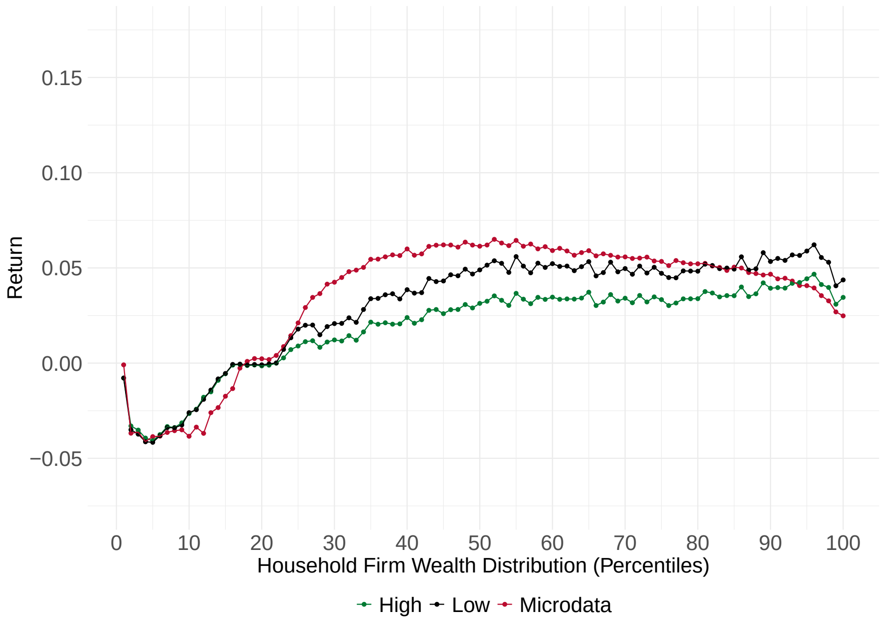
We observe clear return heterogeneity in all series. However, for the microdata, most heterogeneity is driven by a difference between the bottom 30% and the upper

[19] Results based on the mean for each percentile are broadly similar but with stronger volatility in the microdata series in the bottom 30%. This is presumably because even after the data cleaning, outliers remain due to denominator effects. Median estimates are robust to these outliers.

Figure 4.9: Return Heterogeneity over the Firm and Firm-Owner Distribution



(a) Firms



(b) Firm-Owners

Notes: Figures show heterogeneous returns along the distribution. Panel (a) shows the firm-size distribution, panel (b) the firm-owner distribution. Along each distribution in each year, firms are grouped in percentiles; per percentile (across years), the median return is calculated for the microdata series and the two adjusted series based on the WACC.

70%. For the majority of the distribution, median returns are remarkably stable at around 8%. In fact, returns taper off at the top of the distribution, suggesting decreasing returns to scale in returns (as in Boar, Gorea, and Midrigan (2023)). These results also hold at the firm-owner level. At first glance, this seems to contradict results in Fagereng, Guiso, Malacrino, and Pistaferri (2020). Moreover, if high-value firms (or their owners) have *lower* returns than average, this would contradict most mechanisms proposed in the literature to tackle the Gabaix, Lasry, Lions, and Moll (2016) challenge.

Happily, the adjusted returns come to the rescue. Both series show a clear and steep gradient, suggesting that wealth is systematically correlated with returns. Moreover, the gradient increases above the 90th percentile, suggesting that the top of the distribution plays an outsized role in return heterogeneity. Moving from the median to the 100th percentile would cause a firm to see an increase in returns of almost 10 percentage points. For households, the gradient is less steep, but still sizable, with an increase in returns of more than 7 percentage points moving from the median to the top. The median returns using the adjusted values are now under the median returns using unadjusted data for most of the distribution. However, the gradient is much clearer than for the unadjusted data.

Figure 4.9 highlights the importance of properly measuring firm market values for return heterogeneity. However, median returns at each percentile might not be representative of overall return heterogeneity. To investigate these results more systematically, I run regressions of the form

$$r_{ijt} = \beta D_{ijt} + \varepsilon_{ijt}, \tag{4.40}$$

$$r_{jt} = \beta D_{jt} + \varepsilon_{jt}, \tag{4.41}$$

where D_{ijt} is a dummy for the decile of the distribution firm i is in. To make sure that the heterogeneity we capture is not driven by systematic differences between firms, I also estimate these regressions with a battery of fixed effects. I successively investigate whether firm, year, and firm × year fixed effects affect the estimations. I run these regressions both at the firm level (equation (4.40) and the firm-owner level (4.41).[20]

[20] Since a firm can be owned by multiple shareholders, there is a risk of double-counting in these regressions. However, Table 4.1 shows that the vast majority of firms are owned by a single owner, making this a minor concern.

Table 4.8: Return Heterogeneity Along the Firm Distribution

	r_{ijt}				r^+_{ijt}				r^-_{ijt}			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	2.63				-0.060				-0.062			
	(1.82)				(0.110)				(0.133)			
Mean FE		8.536***	2.579***	1.042***		2.034***	-0.046	0.337***		2.635***	-0.051	0.438***
		(1.359)	(0.282)	(0.285)		(0.013)	(0.208)	(0.006)		(0.017)	(0.241)	(0.008)
Decile 2	-1.38	-2.96	-1.38	-0.020	0.120*	-0.270*	0.120*	0.283**	0.128	-0.422**	0.128	0.366**
	(1.33)	(2.67)	(1.33)	(0.012)	(0.052)	(0.111)	(0.052)	(0.083)	(0.063)	(0.131)	(0.063)	(0.091)
Decile 3	-0.677	-6.35	-0.677	-0.005	0.752***	-0.137	0.752***	0.696***	0.930***	-0.287	0.930***	0.888***
	(2.32)	(3.35)	(2.32)	(0.019)	(0.100)	(0.185)	(0.100)	(0.103)	(0.134)	(0.220)	(0.134)	(0.133)
Decile 4	0.451	-6.18	0.451	-0.022	1.28***	-0.031	1.28***	1.02***	1.61***	-0.133	1.61***	1.29***
	(2.59)	(3.42)	(2.59)	(0.015)	(0.112)	(0.237)	(0.112)	(0.153)	(0.150)	(0.284)	(0.150)	(0.201)
Decile 5	-2.12	-7.73*	-2.12	-0.016	1.59***	-0.008	1.59***	1.10***	2.01***	-0.090	2.01***	1.39***
	(2.27)	(3.49)	(2.27)	(0.014)	(0.133)	(0.264)	(0.133)	(0.188)	(0.176)	(0.320)	(0.176)	(0.243)
Decile 6	-2.23	-6.57	-2.23	-0.008	1.80***	-0.066	1.80***	1.32***	2.28***	-0.151	2.28***	1.68***
	(1.86)	(3.15)	(1.86)	(0.015)	(0.159)	(0.280)	(0.159)	(0.209)	(0.208)	(0.343)	(0.208)	(0.264)
Decile 7	-2.43	-6.73*	-2.43	-0.024	1.95***	-0.074	1.95***	1.55***	2.47***	-0.147	2.47***	1.97***
	(1.85)	(3.00)	(1.85)	(0.016)	(0.158)	(0.239)	(0.158)	(0.231)	(0.206)	(0.295)	(0.206)	(0.290)
Decile 8	-3.06	-6.93*	-3.06	-0.002	2.12***	-0.095	2.12***	1.71***	2.69***	-0.160	2.69***	2.17***
	(1.90)	(2.99)	(1.90)	(0.019)	(0.157)	(0.216)	(0.157)	(0.277)	(0.204)	(0.270)	(0.204)	(0.338)
Decile 9	-2.51	-6.58*	-2.51	-0.001	2.38***	-0.075	2.38***	2.02***	3.03***	-0.117	3.03***	2.56***
	(1.82)	(2.79)	(1.82)	(0.017)	(0.144)	(0.171)	(0.144)	(0.291)	(0.185)	(0.217)	(0.185)	(0.353)
Decile 10	-2.24	-4.90	-2.24	-0.002	2.76***	-0.145	2.76***	2.67***	3.51***	-0.191	3.51***	3.39***
	(1.92)	(2.92)	(1.92)	(0.021)	(0.171)	(0.153)	(0.171)	(0.365)	(0.219)	(0.190)	(0.219)	(0.436)
Fixed-effects												
Firm		✓				✓				✓		
Year			✓				✓				✓	
Firm × Year				✓				✓				✓
Fit statistics												
Observations	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312
R^2	0.000	0.561	0.000	1.00	0.009	0.363	0.014	0.991	0.009	0.363	0.013	0.991
Within R^2		0.000	0.000	0.000		0.000	0.009	0.012		0.000	0.009	0.012
RMSE	430.3	285.2	430.3	0.365	9.31	7.46	9.28	0.901	12.0	9.62	12.0	1.16

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Driscoll-Kraay standard errors in parentheses. r_{ijt} is the firm-level return based on unadjusted data, r^+_{ijt} is the firm-level return using β^+; r^-_{ijt} is the corresponding return using β^-.

Table 4.8 shows the results for the firm-level regressions. Because of multicollinearity, the first decile dummy drops out. I report the intercept (in the regressions without fixed effects) and the mean of the fixed effects, to indicate a benchmark against which the coefficients can be compared. I use Driscoll-Kraay standard errors to account for heteroskedasticity and autocorrelation.

The microdata-based returns (r_{ijt}) are insignificant in the first specification (column (1)), with none of the deciles having a systematically differential return. These conclusions are unchanged if we add fixed effects. Some decile dummies in column (2) are significant at the 5% level, indicating that using within-firm variation only, increasing firm size would *decrease* the firm return to wealth. Adding year or firm × year fixed effects again results in insignificant results. This indicates that accounting values cannot reveal systematic return heterogeneity using only within-firm variation. If anything, returns would seem to decline with scale, contradicting the return heterogeneity literature.

Contrast this with the results based on the adjusted returns. Across columns (5)–(12), results are highly significant and remain so after the inclusion of the fixed effects. Moreover, the gradient is sizable and increases in steepness at the top of the distribution. These results survive after the inclusion of the fixed effects. The results are qualitatively identical whether using r^{+}_{ijt} or r^{-}_{ijt}, although magnitudes differ somewhat. Results using only within-firm variation are generally insignificant (columns (6) and (10)). This is unsurprising: the adjusted returns are constructed at the firm level, leaving no room for within-firm variation. Within-year variation is significant, however. Moreover, using firm × year fixed effects does result in highly significant and economically sizable coefficients. This indicates that allowing firms to have separate time trends does aid identification.

My results indicate that return heterogeneity is sizable, mostly driven by the top, and is systematically related to firm wealth, even after allowing firms to have differential time trends and focusing on within-firm variation.

Table 4.9: Return Heterogeneity Along the Firm-Owner Distribution

	r_{jt}				r^{+}_{jt}				r^{-}_{jt}			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	2.54 (1.36)				-0.001 (0.096)				0.015 (0.118)			
Mean FE		6.115*** (0.798)	2.510*** (0.192)	0.634*** (0.206)		1.914*** (0.013)	0.010 (0.191)	0.827*** (0.006)		2.481*** (0.017)	0.022 (0.223)	1.068*** (0.008)
Decile 2	-1.61 (1.05)	-3.59 (2.09)	-1.61 (1.05)	0.046** (0.011)	0.160* (0.059)	-0.176 (0.105)	0.160* (0.059)	0.205* (0.068)	0.182* (0.073)	-0.291* (0.123)	0.182* (0.073)	0.255** (0.083)
Decile 3	-1.28 (1.57)	-4.75* (2.05)	-1.28 (1.57)	0.080*** (0.018)	0.950*** (0.119)	-0.032 (0.189)	0.950*** (0.119)	0.445*** (0.091)	1.18*** (0.157)	-0.145 (0.222)	1.18*** (0.157)	0.563*** (0.117)
Decile 4	-1.33 (1.86)	-5.54* (2.40)	-1.33 (1.86)	0.090*** (0.019)	1.40*** (0.131)	0.032 (0.224)	1.40*** (0.131)	0.575*** (0.076)	1.77*** (0.170)	-0.041 (0.268)	1.77*** (0.170)	0.725*** (0.100)
Decile 5	-2.28 (1.38)	-4.88* (2.06)	-2.28 (1.38)	0.093*** (0.015)	1.61*** (0.141)	0.006 (0.237)	1.61*** (0.141)	0.642*** (0.099)	2.03*** (0.182)	-0.060 (0.288)	2.03*** (0.182)	0.809*** (0.129)
Decile 6	-2.32 (1.37)	-4.74* (1.94)	-2.32 (1.37)	0.097*** (0.020)	1.75*** (0.153)	-0.012 (0.219)	1.75*** (0.153)	0.767*** (0.105)	2.22*** (0.198)	-0.069 (0.268)	2.22*** (0.198)	0.968*** (0.133)
Decile 7	-3.16 (1.57)	-5.22* (2.03)	-3.16 (1.57)	0.101*** (0.019)	1.78*** (0.144)	-0.045 (0.186)	1.78*** (0.144)	0.892*** (0.133)	2.26*** (0.186)	-0.100 (0.230)	2.26*** (0.186)	1.13*** (0.167)
Decile 8	-2.44 (1.37)	-4.76* (1.81)	-2.44 (1.37)	0.103*** (0.018)	1.86*** (0.123)	-0.019 (0.133)	1.86*** (0.123)	1.03*** (0.147)	2.36*** (0.158)	-0.055 (0.167)	2.36*** (0.158)	1.31*** (0.183)
Decile 9	-2.45 (1.37)	-4.39* (1.72)	-2.45 (1.37)	0.104*** (0.019)	1.98*** (0.122)	-0.001 (0.109)	1.98*** (0.122)	1.14*** (0.165)	2.52*** (0.157)	-0.024 (0.138)	2.52*** (0.157)	1.44*** (0.203)
Decile 10	-2.47 (1.37)	-2.96 (1.98)	-2.47 (1.37)	0.106*** (0.022)	1.97*** (0.122)	-0.040 (0.089)	1.97*** (0.122)	1.54*** (0.209)	2.50*** (0.158)	-0.067 (0.109)	2.50*** (0.158)	1.96*** (0.257)
Fixed-effects												
Firm		✓				✓				✓		
Year			✓				✓				✓	
Firm × Year				✓				✓				✓
Fit statistics												
Observations	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312	2,898,312
R^2	0.000	0.481	0.000	1.00	0.006	0.364	0.011	0.987	0.006	0.364	0.010	0.987
Within R^2		0.000	0.000	0.001		0.000	0.006	0.005		0.000	0.006	0.005
RMSE	295.5	213.0	295.5	0.198	9.01	7.21	8.99	1.04	11.6	9.29	11.6	1.34

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*

Notes: Driscoll-Kraay standard errors in parentheses. r_{jt} is the firm-owner-level return based on unadjusted data, r^{+}_{jt} is the firm-owner-level return using β^{+}; r^{-}_{jt} is the corresponding return using β^{-}.

Table 4.9 does the same exercise for household-level returns. Results are mostly in line with those from Table 4.8, with some important differences. First, after including firm fixed effects, the accounting returns in column (2) are significantly negative for most of the deciles. Second, firm × year fixed effects now cause significant coefficients for the accounting returns in column (4), which are all positive. This indicates that allowing firms to have separate time trends results in household-level returns that have a positive gradient over the wealth distribution. In Table 4.9 as in Table 4.8, the adjusted return measures show a steeper gradient over the distribution than the accounting returns. The overall message is clear: return heterogeneity matters, mostly at the top, and particularly in the adjusted series.

These findings are not only informative for the return heterogeneity debate, they also help reconcile the firm-owner-level results of Fagereng, Guiso, Malacrino, and Pistaferri (2020) and others with the firm-level results by Boar, Gorea, and Midrigan (2023). These authors show that accounting returns exhibit decreasing returns to scale, and build to model to argue that financial constraints might drive a wedge between accounting returns and economic returns. My updated measures aim to capture the economic return to a firm, which is increasing in firm size. When aggregated to the household level, my adjusted returns continue to show a positive gradient. Accounting returns, on the other hand, are insignificant or even decreasing over the firm distribution (Table 4.8) but increasing when aggregated to the firm-owner level (Table 4.9). Hence, the results of Fagereng, Guiso, Malacrino, and Pistaferri (2020) and Boar, Gorea, and Midrigan (2023) are mutually consistent.

At the same time, my results also indicate that focusing on accounting returns can be misleading. When mapping theory to data, we are interested in the *economic* return: the compensation for risk-taking and management of the firm. My adjusted measures, while doubtlessly imperfect, are therefore closer to the corresponding variables in our economic models.

4.9. Conclusion

Private business wealth is a key determinant of distributional and aggregate dynamics. In this chapter, I have developed a robust econometric procedure to identify the market value of private firms, based on time-series restrictions. The resulting series are consistent regardless of the choice of discount rate, result in more stable aggregate values of private business wealth, higher top wealth shares, and a steeper wealth-returns gradient.

My econometric procedure is straightforward to implement in settings where researchers can link firms to their owners. Even without this link, my results could be useful to adjust wealth shares. With data on firm balance sheets alone (such as from Compustat or Orbis), my GMM procedure could be implemented to create fitted values of private business wealth. Then, researchers could allocate the additional wealth along the wealth distribution in proportion to the weight of private business wealth in households' portfolios across the distribution. While imperfect, this adjustment is simple to do and would improve upon a status quo where accounting values are used as-is.[21]

My results have additional implications for the growing literature on return heterogeneity. Beyond its theoretical relevance, return heterogeneity has first-order effects for optimal taxation (Guvenen et al. 2023; Guvenen, Kambourov, Kuruscu, and Ocampo-Diaz 2024; Gerritsen, Jacobs, Spiritus, and Rusu 2024). My results show a steep returns-wealth gradient. This suggests a positive and sizable tax on wealth or capital income. Intriguingly, Guvenen, Kambourov, Kuruscu, and Ocampo-Diaz (2024) argue for a book-value tax as the optimal capital tax that minimizes economic distortions in the presence of return heterogeneity due to heterogeneous ability. Implementation of the book-value tax is robust to concerns of misspecification in my GMM procedure, but is theoretically justified by the return heterogeneity found using my procedure.

My results are a first step towards a systematic understanding of heterogeneous returns at the top of the distribution. Future work should consider mechanisms that generate return heterogeneity and wealth inequality and are consistent with the evidence found in this chapter. Increasing availability of linked data, like in my setting, will hopefully permit econometric evaluation of competing theories to explain the dynamics at the top of the wealth distribution.

[21] Toussaint, van Bavel, Salverda, and Teulings (2020) implement such a procedure in the Dutch context; the top wealth shares found in that study are broadly similar to those found in this chapter, suggesting that the simplified procedure is a reasonable quick fix in the absence of linked data.

4.A. Mathematical Appendix

4.A.1. Closed-Form Expression of Profit Function

The derivation in this section follows Crouzet and Eberly (2023, Online Appendix IA.B.3.3) and is presented here for completeness. To economize on notation, define $\vartheta := (1-\delta)\eta$ and $\Psi_t := A_t K_t^{\delta\eta}$. Now the firm problem becomes

$$
\begin{aligned}
\Pi_t = &\max_{\{m_{gt}\}_{g=1}^{G}, P_t} P_t^{-\frac{\lambda}{\lambda-1}} D_t - \sum_{g=1}^{G} p_{gt} m_{gt}, \\
\text{s.t.} \quad & \Psi_t M_t^{\vartheta} \geq C_t,
\end{aligned}
$$

where $M_t := \prod_{g=1}^{G} m_{gt}^{\nu_g}$ and $C_t := P_t^{-\frac{\lambda}{\lambda-1}} D_t$. The dual to this maximization problem is to minimize variable costs:

$$
\begin{aligned}
\text{VC}_t = &\min_{\{m_{gt}\}_{g=1}^{G}} \sum_{g=1}^{G} p_{gt} m_{gt}, \\
\text{s.t.} \quad & \Psi_t M_t^{\vartheta} \geq C_t.
\end{aligned}
$$

The Cobb-Douglas structure of the problem ensures that the cost input share is equal to the output elasticity ν_g:

$$
\frac{p_{gt} m_{gt}}{\mathcal{P}_t M_t} = \nu_g.
$$

As a result, we have

$$
\begin{aligned}
\text{VC}_t &= \mathcal{P}_t M_t = \mathcal{P}_t \left(\frac{C_t}{\Psi_t} \right)^{\frac{1}{\vartheta}} \\
M_t &= \left(\frac{C_t}{\Psi_t} \right)^{\frac{1}{\vartheta}}.
\end{aligned}
$$

Substitution of these expressions into the original firm problem yields

$$\Pi_t = \max_{\{m_{gt}\}_{g=1}^{G}, P_t} P_t^{-\frac{\lambda}{\lambda-1}} D_t - \mathcal{P}_t \left(\frac{C_t}{\Psi_t}\right)^{\frac{1}{\vartheta}},$$

$$\text{s.t.} \quad C_t \geq P_t^{-\frac{\lambda}{\lambda-1}} D_t,$$

which has first-order conditions

$$P_t = \lambda \text{MC}_t,$$

$$\text{MC}_t = \frac{\mathcal{P}_t}{\vartheta C_t} \left(\frac{C_t}{\Psi_t}\right)^{\frac{1}{\vartheta}}$$

so that:

$$C_t = \left(\frac{\lambda}{\vartheta}\right)^{-\frac{\vartheta\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} D_t^{\frac{\vartheta(\lambda-1)}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \mathcal{P}_t^{-\frac{\vartheta\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \Psi_t^{\frac{\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}}.$$

Inserting this into the solution for the cost-minimization problem, we get

$$\text{VC}_t = \left(\frac{\lambda}{\vartheta}\right)^{-\frac{\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} D_t^{\frac{\lambda-1}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \mathcal{P}_t^{-\frac{\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \Psi_t^{\frac{1}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}}.$$

Profits are given by

$$\Pi_t = \left(\frac{\lambda}{\vartheta} - 1\right) \text{VC}_t,$$

so that

$$\Pi_t = \left(\frac{\lambda}{\vartheta} - 1\right) \left(\frac{\lambda}{\vartheta}\right)^{-\frac{\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} D_t^{\frac{\lambda-1}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \mathcal{P}_t^{-\frac{\lambda}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}} \Psi_t^{\frac{1}{\vartheta(\lambda-1)+(1-\vartheta)\lambda}}.$$

Using the definitions of ϑ and Ψ_t, these expressions can be rewritten as

$$
\begin{aligned}
\Pi_t &= H_t^{1-\frac{1}{\mu}} K_t^{\frac{1}{\mu}}, \\
\mu &\coloneqq 1 + \frac{\varphi - 1}{\delta}, \\
\varphi &\coloneqq \frac{\lambda}{\eta}, \\
H_t &\coloneqq \left(\frac{\varphi}{1-\delta}\right)^{-\frac{\varphi}{\varphi-1}} \left(\frac{\varphi}{1-\delta} - 1\right)^{\frac{\varphi-(1-\delta)}{\varphi-1}} D_t^{\frac{\varphi-\eta}{\varphi-1}} \mathcal{P}_t^{-\frac{1-\delta}{\varphi-1}} A_t^{\frac{1}{\eta(\varphi-1)}}.
\end{aligned}
$$

4.A.2. Derivation of Instrument Matrices

Let $\mathbf{R}_{\mathrm{FE}}$ be the within-transforming matrix, i.e.

$$\mathbf{R}_{\mathrm{FE}} = \mathbf{I} - \mathbf{J} = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 1 \end{pmatrix} - \frac{1}{T} \begin{pmatrix} 1 & 1 & \dots & \dots & 1 \\ 1 & 1 & \dots & \dots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & \dots & \dots & \dots & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{T-1}{T} & -\frac{1}{T} & \dots & \dots & -\frac{1}{T} \\ -\frac{1}{T} & \frac{T-1}{T} & -\frac{1}{T} & \dots & -\frac{1}{T} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -\frac{1}{T} & \dots & \dots & \dots & \frac{T-1}{T} \end{pmatrix}.$$

It is easy to see that $\mathbf{R}_{\mathrm{FE}}'\mathbf{R}_{\mathrm{FE}} = \mathbf{R}_{\mathrm{FE}}$. Let $\mathbf{R}_{\mathrm{FD}}$ be the first-differencing matrix, i.e.,

$$\mathbf{R}_{\mathrm{FD}} = \begin{pmatrix} -1 & 1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & -1 & 1 \\ 0 & \dots & \dots & \dots & \dots & -1 \end{pmatrix}.$$

Hence,

$$\mathbf{R}_{\mathrm{FD}}'\mathbf{R}_{\mathrm{FD}} = \begin{pmatrix} 1 & -1 & 0 & \dots & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & -1 & 2 & -1 \\ 0 & \dots & \dots & \dots & \dots & -1 & 1 \end{pmatrix}.$$

The second-difference matrix $\mathbf{R}_{\mathrm{SD}}$ is equal to $\mathbf{R}_{\mathrm{FD}}'\mathbf{R}_{\mathrm{FD}}$, hence:

$$
\mathbf{R}_{\mathrm{SD}}{}'\mathbf{R}_{\mathrm{SD}} = \begin{pmatrix}
2 & -3 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\
-3 & 6 & -4 & 1 & 0 & \dots & \dots & \dots & 0 \\
1 & -4 & 6 & -4 & 1 & 0 & \dots & \dots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \dots & \dots & \dots & 1 & -4 & 6 & -4 & 1 \\
0 & \dots & \dots & \dots & \dots & 1 & -4 & 6 & -3 \\
0 & \dots & \dots & \dots & \dots & \dots & 1 & -3 & 2
\end{pmatrix}.
$$

The respective instrument matrices $\mathbf{P_1}$, $\mathbf{P_2}$ and $\mathbf{P_3}$ can be found as

$$
\begin{aligned}
\mathbf{P_1} &= \mathbf{R}_{\mathrm{FD}}{}'\mathbf{R}_{\mathrm{FD}} - \mathbf{R}_{\mathrm{FE}}{}'\mathbf{R}_{\mathrm{FE}}, && (4.42)\\
\mathbf{P_2} &= \mathbf{R}_{\mathrm{SD}}{}'\mathbf{R}_{\mathrm{SD}} - \mathbf{R}_{\mathrm{FD}}{}'\mathbf{R}_{\mathrm{FD}}, && (4.43)\\
\mathbf{P_3} &= \mathbf{R}_{\mathrm{SD}}{}'\mathbf{R}_{\mathrm{SD}} - \mathbf{R}_{\mathrm{FE}}{}'\mathbf{R}_{\mathrm{FE}}. && (4.44)
\end{aligned}
$$

$\mathbf{P_1}$ has the form:

$$
\mathbf{P_1} = \begin{pmatrix}
\frac{1}{T} & -\frac{T-1}{T} & \frac{1}{T} & \dots & \dots & \dots & \frac{1}{T} \\
-\frac{T-1}{T} & \frac{T+1}{T} & -\frac{T-1}{T} & \frac{1}{T} & \dots & \dots & \frac{1}{T} \\
\frac{1}{T} & -\frac{T-1}{T} & \frac{T+1}{T} & -\frac{T-1}{T} & \frac{1}{T} & \dots & \frac{1}{T} \\
\vdots & \frac{1}{T} & \ddots & \ddots & \ddots & \ddots & \vdots \\
\frac{1}{T} & \dots & \dots & \dots & \dots & -\frac{T-1}{T} & \frac{1}{T}
\end{pmatrix}
$$

and hence the vector of instruments $\mathbf{z_1}$ takes the form

$$
\mathbf{z_1} := \mathbf{Px} = \begin{pmatrix}
\overline{x} - x_2 \\
\overline{x} - x_1 + x_2 - x_3 \\
\overline{x} - x_2 + x_3 - x_4 \\
\vdots \\
\overline{x} - x_{T-1}
\end{pmatrix} \tag{4.45}
$$

where $\overline{x} := T^{-1} \sum_t x_{it}$ is the time-series average of x for firm i. This setup is easily extended to unbalanced panels by making the period length observation-specific, i.e., T_i instead of T.

$\mathbf{P_2}$ is of the form

$$\mathbf{P_2} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ -2 & 4 & -3 & 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & -3 & 4 & -3 & 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & -3 & 4 & -3 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & 1 & -3 & 4 & -2 \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 & -2 & 1 \end{pmatrix}$$

and hence we get an instrument vector $\mathbf{z_2}$ of the form

$$\mathbf{z_2} = \begin{pmatrix} x_1 - 2x_2 + x_3 \\ -2x_1 + 4x_2 - 3x_2 + x_4 \\ x_1 - 3x_2 + 4x_3 - 3x_4 + x_5 \\ \vdots \\ x_{T-3} - 3x_{T-2} + 4x_{T-1} - 2x_T \\ x_{T-2} - 2x_{T-1} + x_T \end{pmatrix}. \tag{4.46}$$

Finally, I construct an instrument $\mathbf{z_3}$, based on the difference between the within-estimator and the second-difference estimator. We get a matrix

$$\mathbf{P_3} = \begin{pmatrix} \frac{T+1}{T} & -\frac{3T-1}{T} & \frac{T+1}{T} & \frac{1}{T} & \dots & \dots & \frac{1}{T} \\ -\frac{3T-1}{T} & \frac{5T+1}{T} & -\frac{4T-1}{T} & \frac{1}{T} & \dots & \dots & \frac{1}{T} \\ \frac{1}{T} & -\frac{3T-1}{T} & \frac{5T+1}{T} & -\frac{4T-1}{T} & \frac{1}{T} & \dots & \frac{1}{T} \\ \vdots & \frac{1}{T} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \frac{1}{T} & \dots & \dots & \dots & \dots & -\frac{3T-1}{T} & \frac{T+1}{T} \end{pmatrix}$$

and hence an instrument vector

$$\mathbf{z_3} = \begin{pmatrix} \bar{x} + x_1 - 3x_2 + x_3 \\ \bar{x} - 3x_1 + 5x_2 - 4x_3 \\ \bar{x} - 3x_2 + 5x_3 - 4x_4 \\ \vdots \\ \bar{x} - 3x_{T-1} + x_T \end{pmatrix}. \tag{4.47}$$

4.A.3. Proofs of Propositions

Proof of Proposition 4.1. We have:

$$
\begin{aligned}
r = \frac{\pi}{w} = \frac{\pi}{w^* + \xi} &= r^* \left(1 - \frac{\xi}{w^* + \xi}\right), \\
\mathrm{E}\left[w\right] &= \mathrm{E}\left[w^* + \xi\right] = \mathrm{E}\left[w^*\right], \\
\mathrm{E}\left[r\right] &= \mathrm{E}\left[r^*\right] - \mathrm{E}\left[\frac{r^*\xi}{w^* + \xi}\right] \\
&\approx \mathrm{E}\left[r^*\right] - \frac{\mathrm{E}\left[r^*\xi\right]}{\mathrm{E}\left[w^* + \xi\right]} + \frac{\mathrm{Cov}\left[r^*\xi, w^* + \xi\right]}{\mathrm{E}\left[w^* + \xi\right]^2} - \frac{\mathrm{E}\left[r^*\xi\right]}{\mathrm{E}\left[w^* + \xi\right]^3}\mathrm{Var}\left[w^* + \xi\right] \\
&= \mathrm{E}\left[r^*\right] + \frac{\mathrm{E}\left[r^*\xi^2\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}}, \\
\mathrm{E}\left[rw\right] &= \mathrm{E}\left[r^* w^*\right], \\
\mathrm{Cov}\left[r, w\right] &= \mathrm{E}\left[rw\right] - \mathrm{E}\left[r\right]\mathrm{E}\left[w\right] = \mathrm{E}\left[r^* w^*\right] - \left(\mathrm{E}\left[r^*\right] + \frac{\mathrm{E}\left[r^*\xi^2\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}}\right)\mathrm{E}\left[w^*\right] \\
&= \mathrm{Cov}\left[r^*, w^*\right] - \frac{\mathrm{E}\left[r^*\xi^2\right]\mathrm{E}\left[w^*\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}}, \\
\mathrm{Var}\left[w\right] &= \sigma^2_{w^*} + \sigma^2_{\xi},
\end{aligned}
$$

where we have repeatedly made use of the orthogonality of ξ with r^* and w^*, as well as the mean-zero expectation of ξ, and we have used a second-order approximation to the expectation of a ratio (Mood, Graybill, and Boes 1974). Collecting terms, we have

$$
\begin{aligned}
\mathrm{E}\left[\widehat{\beta}\right] = \frac{\mathrm{Cov}\left[r, w\right]}{\mathrm{Var}\left[w\right]} &= \frac{\mathrm{Cov}\left[r^*, w^*\right] - \frac{\mathrm{E}\left[r^*\xi^2\right]\mathrm{E}\left[w^*\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}}}{\sigma^2_{w^*} + \sigma^2_{\xi}} \\
&= \beta\left(1 - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*} + \sigma^2_{\xi}}\right) - \mathrm{E}\left[r^*\xi^2\right]\mathrm{E}\left[w^*\right],
\end{aligned}
$$

which is the result in the main text. □

Proof of Proposition 4.2. We follow the same steps as in Proposition 4.1. We have

$$
\begin{aligned}
r &= \frac{y}{w} = \frac{\pi^* + \xi}{w^* + \xi} = r^* \left(1 - \frac{\xi}{w^* + \xi}\right) + \frac{\xi}{w^* + \xi}, \\
\mathrm{E}\left[w\right] &= \mathrm{E}\left[w^* + \xi\right] = \mathrm{E}\left[w^*\right], \\
\mathrm{E}\left[r\right] &= \mathrm{E}\left[r^*\right] - \mathrm{E}\left[\frac{r^*\xi}{w^* + \xi}\right] + \mathrm{E}\left[\frac{\xi}{w^* + \xi}\right] \\
&\approx \mathrm{E}\left[r^*\right] - \frac{\mathrm{E}\left[r^*\xi\right]}{\mathrm{E}\left[w^* + \xi\right]} + \frac{\mathrm{Cov}\left[r^*\xi, w^* + \xi\right]}{\mathrm{E}\left[w^* + \xi\right]^2} - \frac{\mathrm{E}\left[r^*\xi\right]}{\mathrm{E}\left[w^* + \xi\right]^3}\mathrm{Var}\left[w^* + \xi\right] \\
&+ \frac{\mathrm{E}\left[\xi\right]}{\mathrm{E}\left[w^* + \xi\right]} - \frac{\mathrm{Cov}\left[\xi, w^* + \xi\right]}{\mathrm{E}\left[w^* + \xi\right]^2} + \frac{\mathrm{E}\left[\xi\right]}{\mathrm{E}\left[w^* + \xi\right]^3}\mathrm{Var}\left[w^* + \xi\right] \\
&= \mathrm{E}\left[r^*\right] + \frac{\mathrm{E}\left[r^*\xi^2\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}} - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*} + \sigma^2_{\xi}}, \\
\mathrm{E}\left[rw\right] &= \mathrm{E}\left[r^* w^*\right], \\
\mathrm{Cov}\left[r, w\right] &= \mathrm{E}\left[rw\right] - \mathrm{E}\left[r\right]\mathrm{E}\left[w\right] \\
&= \mathrm{E}\left[r^* w^*\right] - \left(\mathrm{E}\left[r^*\right] + \frac{\mathrm{E}\left[r^*\xi^2\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}} - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*} + \sigma^2_{\xi}}\right)\mathrm{E}\left[w^*\right] \\
&= \mathrm{Cov}\left[r^*, w^*\right] - \frac{\left(\mathrm{E}\left[r^*\xi^2\right] + \sigma^2_{\xi}\right)\mathrm{E}\left[w^*\right]}{\sigma^2_{w^*} + \sigma^2_{\xi}}, \\
\mathrm{Var}\left[w\right] &= \sigma^2_{w^*} + \sigma^2_{\xi},
\end{aligned}
$$

hence, collecting terms, we have

$$
\mathrm{E}\left[\widehat{\beta}\right] = \frac{\mathrm{Cov}\left[r, w\right]}{\mathrm{Var}\left[w\right]} = \beta\left(1 - \frac{\sigma^2_{\xi}}{\sigma^2_{w^*} + \sigma^2_{\xi}}\right) + \left(\sigma^2_{\xi} - \mathrm{E}\left[r^*\xi^2\right]\right)\mathrm{E}\left[w^*\right].
$$

□

4.B. Data Construction

4.B.1. Overview

I use the following administrative datasets:

- VEHTAB: administrative records on the Dutch wealth distribution. All variables are dated January 1st of a year. VEHTAB contains anonymized household classifiers (`RINPERSOON`), as well as values for total wealth and all its components: total assets (consisting of deposits, other financial assets, private firm wealth, pass-through business equity, housing, other real estate, and other) and total liabilities (the sum of mortgages, student loans, and other loans).
- SZO AB+: administrative records linking firms to their owners, dated December 31st of a given year. SZO AB+ consists of two sources: Bedrijfsgegevens (BG), which contains the universe of corporate income tax records (merged with records from the Value-Added Tax, although these are less complete), and the Shareholder Registry (Aandeelhoudersregister, AR), which contains ownership information for all incorporated firms in each quarter of a fiscal year.
- BG contains, for each firm in each year, a full balance sheet and profit & loss statement. In addition, it includes information on firm age, number of employees, industry (at the five-digit level), and legal form. Note that not all variables are equally well observed. In particular, firm characteristics are not always observed. BG contains both consolidated and unconsolidated versions of balance-sheet and profit & loss variables.
- There are three kinds of identifiers: one household identifier (the `RINPERSOON`), and two types of firm identifiers:
  - BG has two variables for firm IDs per tax return: one for the firm filing the return (`RsinAangever_crypt`), and one for the firm for which the return is filed (`RsinAangegevene_crypt`). These two IDs may be the same, if the firm consists of one layer. For holdings and other multilayer structures, the uppermost layer (the "mother") will file for multiple firms, including itself and its "daughters".
  - AR has one variable for firm ID and one for ownership ID (which may be a firm or individual ID, i.e., a `RINPERSOON`). Hence, per firm-year observation, we have the firm ID and the ID of its owner. We also know if the owner is a corporation or natural person.

4.B.2. Linking Firms to Owners

I merge the BG and AR datasets for all years I have access to. I merge on the firm for which the return is filed, (`RsinAangegevene_crypt`). This creates a dataset which has all layers of all firms, and their ownership structure. We retain firms that are owned by natural persons, i.e., whose owner in the AR dataset is a `RINPERSOON`.

The firm in question may consist of a single layer, in which case there are no complications. Often, however, a firm has a complex structure, with a mother owning multiple daughters, which may in turn own firms. Mother firms report both consolidated and unconsolidated balance sheets and flow variables. We retain consolidated variables only, at the holding level. It may be that a household owns shares in a holding as well as direct shares in one or more of the daughters. If so, we retain the daughters as well, since by definition the holding will not own 100% of the daughter firms; hence, the mother balance sheets do not double-count.

This procedure will capture most of the holding structures that are linkable to individuals. However, complex ownership structures may still be undercounted. These structures are most often encountered at the top of the distribution (IBO Vermogensverdeling 2022). My results should therefore be seen as a lower bound on the true extent of inequality and heterogeneity of returns to wealth.

4.B.3. Estimating Private Business Wealth

Total wealth held by household j in firm i is simply obtained by multiplying j's share in the firm with i's wealth at beginning of the calendar year. Household variables are at January 1 and firm balance sheets at December 31, so I use lagged values of firm variables to obtain the relevant wealth. The data also has variables on beginning-of-year firm equity. Inspection reveals, however, that this variable is highly noisy, often not corresponding at all to the end-of-year values of one year earlier, where these should be the same. Hence, I opt to use lagged values of end-of-year firm equity throughout. The only exception is for the initial year 2007, where this is not possible. I follow the approach by Statistics Netherlands here (Menger 2021), which is to take the end-of-year balance sheet of 2007 but multiply it by ownership share in the first quarter.

My approach is similar to the one in Menger (2021), but there are some minor differences in sources used and aggregation rules. A major difference is that I impute missing values of ownership shares based on surrounding years: if at time t an ownership share is missing, but is present in either $t-1$ or $t+1$, that value is imputed (with the earlier date given preference). This results in more observations retained than in Menger (2021). Another difference is that sometimes the individual ID classifier is missing, which can similarly be imputed.[22]

The approach in Menger (2021) is as follows:

1. Determine net equity per firm, correcting for goodwill and other intangibles

2. Select ownership structures that generate wealth (excluding cooperations and so on)

3. Correct ownership shares:

    a) Ownership shares are taken to be lagged shares on December 31 in year $t-1$, except for the first year 2007, where the share at the end of the first quarter of 2007 is used.
    b) If ownership shares sum to more than 100%, they are proportionately scaled back.
    c) If ownership shares are missing, there are two scenarios:
        i. The corporation is either listed, or large (>100 million in equity); in both cases, a conservative share of 5% is presumed
        ii. For all other scenarios, the capital is split equally among the observed shareholders
    d) Remove outliers, which are defined as firms with at least a billion in equity but with no economic activity, which also have negative liquid assets (a potential sign of cooking the books).

4. Obtain indirect shares: this is important for a holding structure known as STAK (Stichting Administratiekantoor), which allows family offices to distribute shares among family members.

5. Impute missing observations: if a shareholder owns shares in a firm in $t-1$ and $t+1$ but ostensibly not in t, he is presumed to also own the firm in t.

6. Aggregate shares per person. At this stage, no correction is made for negative equity, even though these firms should be limited-liability. There are many fiscal constructs through which an owner can borrow from a firm, and as such the firm could have negative net worth.

[22] The anonymized individual IDs consist of a classifier followed by a unique string of digits. I match on the string of digits throughout. The classifier is either "R" (real person), or "F" (fictitious person, i.e., a corporation). If this classifier is missing altogether, CBS doesn't compute wealth totals, even though the digit-string and all other information is present, and even in cases where the classifier does appear in other years.

7. If individuals receive firm dividends for at least three years, but have no observable claim to a firm, they are assigned firm wealth equal to their last received dividend multiplied by 10.

8. Aggregate to the household level. Here, negative equity is transformed to a symbolic value of 1 EUR.

My approach differs in step 3(a) and 3(c). Step 3(a) is problematic for firms whose first recorded year is later than 2007, since their entire first year is ignored. From a conceptual point of view, there is no good reason to treat those first years differently from the ones whose first year is observed in 2007. Hence, I apply the 2007 procedure for all first-year observations. Instead of fixed assignment rules like in 3(c), I impute ownership shares based on surrounding observations.

My data do not allow me to apply step 7, since I do not observe dividend payouts. In general, the private business wealth recorded in the wealth statistics VEHTAB have more observations than I am able to link to firms. For households which I cannot link to firms, I retain the private firm value recorded in VEHTAB. For households which I can link to firms, I replace recorded private business wealth with my measures estimated from my GMM procedure.

4.B.4. Constructing Returns to Wealth

Main variables the main variables I need for the main exercises are returns and gross assets. The return on assets of household i for firm j at time t is defined as

$$r_{ijt} := \frac{\pi_{jt} + \kappa_{jt}}{a_{ijt-1} + \frac{1}{2} f_{jt}} \cdot s_{ijt}. \tag{4.48}$$

here, π are total profits, κ are capital gains, and s is the ownership share. As in Fagereng, Guiso, Malacrino, and Pistaferri (2020), the denominator not only includes beginning-of-period assets but also accounts for the fact that there are net inflows during the year, which might capitalize into profits and/or capital gains. Hence, the second factor corrects for the net inflows f, assuming that these occur about halfway during the years on average. In what follows, t denotes a calendar year, where all flow variables are occurring during the year, and all stock variables are dated at December 31 of a given year. I match all of these objects to my data as follows:

- π_{jt} = `OND1_0789_BELASTBARE_WINST`, which is the total taxable profit of firm j at time t

- κ_{jt} = OND1_0793_TOTAAL_VERMOGENSVERSCHIL, which is the net difference in total wealth between $t-1$ and t, net of inflows and outflows of capital.

- $a_{ijt-1} = a_{jt-1} \cdot s_{ijt-1}$ = OND1_1376_TOTAAL_ACTIVA_EB * BetrokkenheidUltKw4, total assets at the end of calendar year $t-1$, multiplied by the share at the end of the fourth quarter of period $t-1$.

- f_{jt} = OND1_0803_STORTINGEN_VAN_KAPITAAL_IN_HET_BOEKJAAR − OND1_0798_TERUGBETALING_KAPITAAL_PRIVE_ONTTREKKING, i.e., the difference between inflows of capital during the year and outflows (either repayments or withdrawals for private use). The difference is therefore the net inflow of capital during period t.

- s_{ijt} = BetrokkenheidUltKw4, the ownership share at the end of the fourth quarter of period t.

As discussed in the main text, for the updated returns measures, I replace assets a_{jt} with updated firm wealth v_{jt} and measure capital gains κ as the first-difference of updated firm wealth, $\kappa = \Delta v_{jt}$.

4.C. Additional Figures and Tables

The GMM estimations in the main text use heteroskedasticity-robust standard errors. If we instead cluster standard errors at the firm level, we get the estimations in Table 4.10.

Table 4.10: GMM Estimation, Time-Series Identification, Clustered Standard Errors

Dependent Variable:	y_{it}						
Instruments	z_1	z_2	z_3	z_1, z_2	z_2, z_3	z_1, z_3	z_1, z_2, z_3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: ϱ^{wacc}			
β	4.610**	4.089*	4.543	4.040*	4.224**	4.597**	2.859***
	(1.704)	(1.819)	(2.415)	(1.831)	(1.553)	(1.665)	(0.711)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				0.870	0.027	0.001	2.611
J-test p-value				0.351	0.870	0.975	0.271
First Stage F-statistic	10.889	10.095	10.486	9.131	7.367	11.392	9.050
				Panel B: ϱ^{gh}			
β	3.242**	2.909*	3.580	3.061*	3.061**	3.237**	2.831**
	(1.169)	(1.298)	(1.959)	(1.223)	(1.117)	(1.164)	(0.874)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				0.505	0.086	0.037	1.887
J-test p-value				0.478	0.769	0.848	0.389
First Stage F-statistic	10.889	10.095	10.486	9.131	8.951	11.392	9.050
				Panel C: ϱ^{b}			
β	6.347**	5.289*	5.052*	4.706*	5.179**	5.570*	2.645***
	(2.252)	(2.350)	(2.256)	(2.325)	(1.965)	(2.243)	(0.469)
Observations	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563	6,713,563
J-test Statistic				1.668	0.006	0.307	4.417
J-test p-value				0.196	0.936	0.579	0.110
First Stage F-statistic	10.889	10.095	10.486	9.131	8.951	11.392	9.050

*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*
Notes: Two-step GMM with standard errors clustered at the firm level in parentheses. Robust first-stage F-statistic by Montiel Olea and Pflueger (2013). Each panel shows GMM results with as independent variable the $\mathbf{y}$ computed using the respective discount rate.

The results are broadly comparable, with the exception of column (7), which is uniformly lower across specifications (and is almost identical across panels). Since heteroskedasticity is a major concern in my setting (see Table 4.1), I stick with the specifications in the main text.

4.D. Alternative Method: Higher-Order Moments

The results in the main text are derived using time-series instruments. I have also investigated an alternative method, based on higher-order moment restrictions. Although theoretically valid, the GMM estimates did not converge for these estimators. For completeness, I include the framework here.

We can purge error in **x** and **y** by using higher-order moments as instruments, following Lewbel, Schennach, and Zhang (2023) and Erickson and Whited (2000). Observe that equations (4.13) and (4.14) form a "triangular" system; that is, the measurement error ψ shows up in both equations. Lewbel et al. show that under assumptions on the distribution of ξ and k, *all* coefficients and distributions of variables can be point identified. To start, we demean y and x using the sample-wide mean:

$$\widetilde{y}_{it} := y_{it} - \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it},$$
$$\widetilde{x}_{it} := x_{it} - \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}.$$

This demeaning ensures that the fixed effect γ drops out and also has the advantage of simplifying the cumulant equations used in constructing the moment conditions. The identifying assumptions for the higher-order moment approach are summarized in Assumption 3.

Assumption 3 (Higher-Order Moments)**.** $\widetilde{\mathbf{x}}$ *and* ξ *are mutually independent random variables with zero mean, neither of which is normally distributed.*

The assumption has three substantive parts. The first part, mutual independence, is restrictive. Note, however, that the model can be weakened to hold conditional on covariates and fixed effects. For instance, if we are concerned that measurement errors might increase with firm size, adding a (correctly-measured) proxy for firm size and performing the estimation on the residualized variables would make this part of Assumption 3 hold.[23] The second part, zero-mean values, holds by construction. The third part is important for using higher-order moments as instruments. Since products of Normal variables are also Normal, higher-order

[23]In the empirical application, I have experimented with adding the firm's wage bill and/or tax bill as controls, since these should proxy for firm size; the tax bill moreover addresses concerns that firms might strategically underreport their accounts for fiscal reasons. These controls were insignificant in every specification I investigated; hence, for simplicity I proceed without controls.

(products of) moments would be unable to disentangle k and ξ (Reiersøl 1950; Kapteyn and Wansbeek 1983).

Under Assumption 3, Lewbel, Schennach, and Zhang (2023) obtain the following valid moment conditions:

$$\mathrm{E}\left[\widetilde{xy} - \mu_{\widetilde{xy}}\right] = 0 \tag{4.49}$$

$$\mathrm{E}\left[\widetilde{x}^2 - \mu_{\widetilde{xx}}\right] = 0 \tag{4.50}$$

$$\mathrm{E}\left[(\widetilde{y} - \beta\widetilde{x})(\widetilde{y} - (\beta+1)\widetilde{x})\widetilde{x}\right] = 0 \tag{4.51}$$

$$\begin{aligned}\mathrm{E}\Big[(\widetilde{y} - \beta\widetilde{x})(\widetilde{y} - (\beta+1)\widetilde{x})\left(\widetilde{x}^2 - \mu_{\widetilde{xx}}\right) - \\ 2\left(\mu_{\widetilde{xy}} - \beta\mathrm{E}\left[\widetilde{x}^2\right]\right)(\widetilde{y} - (\beta+1)\widetilde{x})\,\widetilde{x}\Big] = 0\end{aligned} \tag{4.52}$$

where $\mu_{\widetilde{xy}} := \mathrm{E}\left[\widetilde{xy}\right]$ and $\mu_{\widetilde{xx}} := \mathrm{E}\left[\widetilde{x}^2\right]$, which are estimated in the GMM routine as nuisance parameters. This is a system of four equations in three unknowns, allowing for overidentification tests.

To build intuition for these moment conditions, consider the first one, equation (4.49). Under Assumption 3, k and ξ are mutually independent, and hence

$$\mathrm{E}\left[\widetilde{xy} - \mathrm{E}[\widetilde{xy}]\right] = \mathrm{E}\left[(k+\xi)(v+\xi) - \mathrm{E}[(k+\xi)(v+\xi)]\right] = \mathrm{E}[vk - \mathrm{E}[vk]].$$

In words, since the measurement errors and the true values are uncorrelated, the covariance between measured x and y is a valid instrument for the unobserved k, since any correlation between xy and x must be due to k. Similar results can be obtained for the other moment conditions. Lewbel et al. rely on the following equation linking higher-order cumulant equations:[24]

$$M_p(1,\beta) := \Upsilon_{x,y}^{1+p,2} - \Upsilon_x^{3+p} - (\beta+1)\left(\Upsilon_{x,y}^{2+p,1} - \Upsilon_x^{3+p}\right). \tag{4.53}$$

Here, Υ_x^p is the cumulant equation for variable x of order p, and $\Upsilon_{x,y}^{p,p'}$ is the product cumulant equation for variables x and y or order p and p', respectively. Lewbel et al. show that for any integer $p' > p$, the two moment constraints point-identify the parameters of the model

$$M_p(1,\beta) = 0$$
$$M_{p'}(1,\beta) = 0$$

[24] Note that their formula is slightly more general than mine since they allow the measurement error to have a coefficient different from 1 in equation (4.13).

under the condition that $\Upsilon_k^{3+p'}\Upsilon_\xi^{3+p} \neq \Upsilon_\xi^{3+p'}\Upsilon_k^{3+p}$, as well as that the moments $\mathrm{E}\left[|x|^{p'}\right]$, $\mathrm{E}\left[|\xi|^{p'}\right]$, and $\mathrm{E}\left[|\varepsilon|^{p'}\right]$ exist. The moment conditions above were derived by setting $p = 0$ and $p' = 1$. The equations become:

$$M_0(1,\beta) = \Upsilon_{x,y}^{1,2} - \Upsilon_x^3 - (\beta+1)\left(\Upsilon_{x,y}^{2,1} - \Upsilon_x^3\right),$$
$$M_1(1,\beta) = \Upsilon_{x,y}^{2,2} - \Upsilon_K^4 - (\beta+1)\left(\Upsilon_{x,y}^{3,1} - \Upsilon_x^4\right).$$

Since the variables are mean-zero, we can write the cumulant equations as products of lower-order moments, using:

$$\begin{aligned}
\Upsilon_{x,y}^{1,2} &= \mathrm{E}[xy^2]\\
\Upsilon_{x,y}^{2,1} &= \mathrm{E}[y^2x]\\
\Upsilon_x^3 &= \mathrm{E}[y^3]\\
\Upsilon_{x,y}^{2,2} &= \mathrm{E}[x^2y^2] - \mathrm{E}[y^2]\mathrm{E}[x^2] - 2\mathrm{E}[xy]\mathrm{E}[xy]\\
\Upsilon_{x,y}^{3,1} &= \mathrm{E}[xy^3] - 3\mathrm{E}[xy]\mathrm{E}[x^2]\\
\Upsilon_x^4 &= \mathrm{E}[x^4] - 3\mathrm{E}[x^2]\mathrm{E}[x^2]
\end{aligned}$$

To make the cumulant equations amenable for GMM estimation, some of the products of moments need to be linearized; this is why we need the nuisance parameters introduced above. Inserting the definitions of the cumulants into equation (4.53), together with the nuisance equations $\mu_{xx} = \mathrm{E}[x^2]$ and $\mu_{xy} = \mathrm{E}[xy]$, yields the moment conditions used above.

Since these conditions were obtained using cumulants of order 0 and 1, the restriction on the moments is that $\Upsilon_k^4\Upsilon_\xi^3 \neq \Upsilon_\xi^4\Upsilon_k^3$. This restriction might be rejected; this would be the case if the distribution of k and ξ were symmetric, if either of their distributions were Normal, or if they were distributed exactly the same; which are all cases ruled out by Assumption 3. Since we have overidentification, this restriction is testable. If we are willing to use even higher moments (namely by setting $p'' = 2$), we obtain the additional moment conditions

$$\mathrm{E}\left[y^2 - \mu_{yy}\right] = 0, \tag{4.54}$$

$$\begin{aligned}\mathrm{E}\big[&-3xy^2\mu_{xx} - 6\mu_{xy}x^2y - \mu_{yy}x^3 + x^3y^3 - x^5 + 10x^3\mu_{xx} - \\ &(\beta+1)\left(-6\mu_{xx}x^2y - 4\mu_{xy}x^3 + x^4y - x^5 + 10x^3\mu_{xx}\right)\big] = 0.\end{aligned} \tag{4.55}$$

Equations (4.54) and (4.55) add two additional moment conditions and one additional nuisance parameter, $\mu_{yy} := \mathrm{E}\left[y^2\right]$, which provide further opportunities to test the over-identifying restrictions. These additional conditions come under the additional restriction that at least one of the inequalities $\Upsilon_\xi^4 \Upsilon_k^3 \neq \Upsilon_k^4 \Upsilon_\xi^3$, $\Upsilon_\xi^5 \Upsilon_k^3 \neq \Upsilon_k^5 \Upsilon_\xi^3$, or $\Upsilon_\xi^5 \Upsilon_k^4 \neq \Upsilon_k^5 \Upsilon_\xi^4$ hold.

Bibliography

’t Hart, Marjolein, Joost Jonker, and Jan Luiten van Zanden. 1997. *A Financial History of the Netherlands.* Cambridge University Press.

Abel, Andrew B., and Janice C. Eberly. 1994. “A Unified Model of Investment under Uncertainty.” *American Economic Review* 84 (5): 1369.

Acciari, Paolo, Facundo Alvaredo, and Salvatore Morelli. 2024. “The Concentration of Personal Wealth in Italy 1995–2016.” *Journal of the European Economic Association* 22 (3): 1228–1274.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. “The Network Origins of Aggregate Fluctuations.” *Econometrica* 80 (5): 1977–2016.

Acemoglu, Daron, and Jaume Ventura. 2002. “The World Income Distribution.” *Quarterly Journal of Economics* 117 (2): 659–694.

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2022. “Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach.” *Review of Economic Studies* 89 (1): 45–86.

Aghion, Philippe, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous. 2019. “Innovation and Top Income Inequality.” *Review of Economic Studies* 86 (1): 1–45.

Aghion, Philippe, and Peter Howitt. 1992. “A Model of Growth Through Creative Destruction.” *Econometrica* 60 (2): 323–351.

Ahuja, J.C., and Stanley W. Nash. 1967. “The Generalized Gompertz-Verhulst Family of Distributions.” *Sankhyā: The Indian Journal of Statistics, Series A,* 141–156.

Aiyagari, S. Rao. 1994. “Uninsured Idiosyncratic Risk and Aggregate Saving.” *Quarterly Journal of Economics* 109 (3): 659–684.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan. 2020. “Thickness and Information in Dynamic Matching Markets.” *Journal of Political Economy* 128 (3): 783–815.

Albers, Thilo N., Charlotte Bartels, and Moritz Schularick. 2022. "Wealth and its Distribution in Germany, 1895-2018." *CEPR Discussion Paper* DP17269.

Alvaredo, Facundo, Anthony B. Atkinson, and Salvatore Morelli. 2018. "Top Wealth Shares in the UK Over More than a Century." *Journal of Public Economics* 162:26–47.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. *World Inequality Report 2018.* Harvard University Press.

Amsterdamsch Effectenblad. 1942. "Statistiek der naamlooze vennootschappen: Interessante gegevens van het C.B.S." (July 21, 1942).

Anderson, T.W., and Cheng Hsiao. 1982. "Formulation and Estimation of Dynamic Models using Panel Data." *Journal of Econometrics* 18 (1): 47–82.

Andolfatto, David. 2002. "A Theory of Inalienable Property Rights." *Journal of Political Economy* 110 (2): 382–393.

Andrews, Donald W.K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 59 (3): 817–858.

Andrews, Isaiah, James H. Stock, and Liyang Sun. 2019. "Weak Instruments in Instrumental Variables Regression: Theory and Practice." *Annual Review of Economics* 11:727–753.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58 (2).

Artola Blanco, Miguel, Luis Bauluz, and Clara Martínez-Toledano. 2020. "Wealth in Spain, 1900–2017: A Country of Two Lands." *Economic Journal* 131 (633): 129–155.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49 (1): 3–71.

Atkinson, Antony B., and Alan J. Harrison. 1978. *The Distribution of Personal Wealth in Britain.* 89–108.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135 (2): 645–709.

Axtell, Robert L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science* 293 (5536): 1818–1820.

Bach, Laurent, Laurent E. Calvet, and Paolo Sodini. 2020. "Rich Pickings? Risk, Return, and Skill in Household Wealth." *American Economic Review* 110 (9): 2703–47.

Bakker, Gert den. 2019. "The Dutch Interwar Economy Revisited: Reconstruction and Analysis of the National Accounts 1921–1939." PhD diss., Utrecht University.

Balkema, August A., and Laurens de Haan. 1974. "Residual Life Time at Great Age." *The Annals of Probability,* no. 5 (2): 792–804.

Baqaee, David R., and Emmanuel Farhi. 2019a. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica* 87 (4): 1155–1203.

———. 2019b. "The Microeconomic Foundations of Aggregate Production Functions." *Journal of the European Economic Association* 17 (5): 1337–1392.

Barkai, Simcha. 2020. "Declining Labor and Capital Shares." *Journal of Finance* 75 (5): 2421–2463.

Baselgia, Enea, and Isabel Z. Martínez. 2023. "Wealth-Income Ratios in Free Market Capitalism: Switzerland, 1900–2020." *Review of Economics and Statistics,* Forthcoming.

Bauluz, Luis, Thomas Blanchet, Clara Martínez-Toledano, and Alice Sodano. 2022. "Global Wealth Dynamics: Understanding the Determinants." *Conference Draft.*

Bauluz, Luis, Filip Novokmet, and Moritz Schularick. 2022. "The Anatomy of the Global Saving Glut." *CEPR Discussion Paper* No. DP17215.

Bavel, Bas J.P. van. 2016. *The Invisible Hand?: How Market Economies Have Emerged and Declined Since AD 500.* Oxford University Press.

Bavel, Bas J.P. van, and Ewout H.P. Frankema. 2017. "Wealth Inequality in the Netherlands, c. 1950–2015.: The Paradox of a Northern European Welfare State." *Tijdschrift voor Sociale en Economische Geschiedenis* 14 (2): 29.

Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef L. Teugels. 2006. *Statistics of Extremes: Theory and Applications.* John Wiley & Sons.

Bengtsson, Erik, and Daniel Waldenström. 2018. “Capital Shares and Income Inequality: Evidence from the Long Run.” *Journal of Economic History* 78 (3): 712–743.

Benhabib, Jess, and Alberto Bisin. 2018. “Skewed Wealth Distributions: Theory and Empirics.” *Journal of Economic Literature* 56 (4): 1261–91.

Benhabib, Jess, Alberto Bisin, and Shenghao Zhu. 2011. “The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents.” *Econometrica* 79 (1): 123–157.

Bernstein, Asaf, and Peter Koudijs. 2024. “The Mortgage Piggy Bank: Building Wealth through Amortization.” *Quarterly Journal of Economics* 139 (3): 1767–1825.

Beurden, Tijn van, and Joost Jonker. 2021. “A Perfect Symbiosis: Curaçao, the Netherlands and Financial Offshore Services, 1951—2013.” *Financial History Review* 28 (1): 67–95.

Bewley, Truman. 1987. “Stationary Monetary Equilibrium With a Continuum of Independently Fluctuating Consumers.” Chap. 5 in *Contributions to Mathematical Economics in Honor of Gérard Debreu,* edited by Werner Hildenbrand and Andreu Mas-Colell, vol. 79. North-Holland.

Bie, Ronald van der. 2001. *Tweehonderd Jaar Statistiek in Tijdreeksen, 1800–1999.* Centraal Bureau voor de Statistiek.

Bingham, Nicholas H., Charles M. Goldie, and Jozef L. Teugels. 1989. *Regular Variation.* Cambridge University Press.

Blanchet, Thomas. 2022. “Uncovering the Dynamics of the Wealth Distribution.” *arXiv preprint arXiv:2211.15509.*

Blanchet, Thomas, Juliette Fournier, and Thomas Piketty. 2022. “Generalized Pareto Curves: Theory and Applications.” *Review of Income and Wealth* 68 (1): 263–288.

Blundell, Richard, and Stephen Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *Journal of Econometrics* 87 (1): 115–143.

Boar, Corina, Denis Gorea, and Virgiliu Midrigan. 2023. “Why Are Returns to Private Business Wealth So Dispersed?” *NBER Working Paper* 29705.

Boar, Corina, and Virgiliu Midrigan. 2023. “Should We Tax Capital Income or Wealth?” *American Economic Review: Insights* 5 (2): 259–74.

Boissevain, G.M. 1883. “De Omvang en de Verdeeling van het Stoffelijke Vermogen in Nederland.” *De Economist* 32 (1): 692–700.

———. 1891. “Het Particulier Vermogen in Nederland.” *De Economist* 40 (1): 643–657.

———. 1909. “Kapitaalsbezit en Kapitaals-vernietging.” *De Economist* 58 (1): 227–246.

———. 1910. “Successiebelasting en Staatsuitgaven.” *De Economist* 59 (1): 705–724.

Bolt, Jutta, and Jan Luiten van Zanden. 2020. “Maddison Style Estimates of the Evolution of the World Economy. A New 2020 Update.” *Maddison-Project Working Paper* WP-15.

Bonger, W.A. 1923. *Vermogen en Inkomen in Nederland en Gedurende den Oorlogstijd (1913-1920).* Amsterdam: Boekhandel en Uitgeversmaatschappij "Ontwikkeling".

Bos, N.J.P.M. 1990. “Vermogensbezitters en Bevoorrechte Belastingbetalers in de Negentiende Eeuw.” *Bijdragen en Mededelingen Betreffende de Geschiedenis der Nederlanden* 105 (4): 553–577.

Bosch, Kornelis Douwe. 1948. “De Nederlandse Beleggingen in de Verenigde Staten.” PhD diss., Nederlandsche Economische Hogeschool Rotterdam.

Bosch Kemper, M.J. de. 1950. *De Tegenwoordige Staat van Nederland : Samengesteld ter Gelegenheid van het 50-jarig Bestaan van het Centraal Bureau voor de Statistiek.* W. de Haan.

Bruil, Arjan, Céline van Essen, Wouter Leenders, Arjan Lejour, Jan Möhlmann, and Simon Rabaté. 2022. “Inequality and Redistribution in the Netherlands.” *CPB Discussion Paper* 436.

Buelens, Frans, and Ewout Frankema. 2016. “Colonial Adventures in Tropical Agriculture: New Estimates of Returns to Investment in the Netherlands Indies, 1919–1938.” *Cliometrica* 10:197–224.

Cagetti, Marco, and Mariacristina De Nardi. 2006. “Entrepreneurship, Frictions, and Wealth.” *Journal of Political Economy* 114 (5): 835–870.

Caminada, Coen L.J., Kees P. Goudswaard, and Marieke G. Knoef. 2014. "Vermogen in Nederland Gelijker Verdeeld Sinds Negentiende Eeuw." *Me Judice.*

Campbell, Cole, and Jacob Robbins. 2023. "The Value of Private Business in the United States." *Available at SSRN 4635369.*

Campbell, John Y., and Robert J. Shiller. 1988a. "Stock Prices, Earnings, and Expected Dividends." *Journal of Finance* 43 (3): 661–676.

———. 1988b. "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors." *Review of Financial Studies* 1 (3): 195–228.

Carvalho, Vasco, and Xavier Gabaix. 2013. "The Great Diversification and its Undoing." *American Economic Review* 103 (5): 1697–1727.

Catherine, Sylvain, Max Miller, and Natasha Sarin. 2023. "Social Security and Trends in Wealth Inequality." *Journal of Finance,* Forthcoming.

Centraal Bureau voor de Statistiek. 1947. "Uitkomsten en Enige Berekeningen Betreffende het Nationale Vermogen in Nederland in 1938." *Statistische en Econometrische Onderzoekingen* 3.

Chamley, Christophe. 1986. "Optimal Taxation of Capital Income in General Equilibrium With Infinite Lives." *Econometrica,* 607–622.

Champernowne, David G. 1953. "A Model of Income Distribution." *Economic Journal* 63 (250): 318–351.

Chancel, Lucas, and Thomas Piketty. 2021. "Global Income Inequality, 1820–2020: The Persistence and Mutation of Extreme Inequality." *Journal of the European Economic Association* 19 (6): 3025–3062.

Chaney, Thomas. 2008. "Distorted Gravity: the Intensive and Extensive Margins of International Trade." *American Economic Review* 98 (4): 1707–1721.

———. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Cioffi, Riccardo A. 2021. "Heterogeneous Risk Exposure and the Dynamics of Wealth Inequality." *Working paper.*

Cochrane, John. 2020. *Wealth and Taxes, part II.* Blog post.

Coenen, Ann. 2017. "Charting the Development of Wealth Inequality in the Netherlands since 1950: an On-going Quest." *The Low Countries Journal of Social and Economic History* 14 (2): 11–28.

Crouzet, Nicolas, and Janice C. Eberly. 2021. "Intangibles, Markups, and the Measurement of Productivity Growth." *Journal of Monetary Economics* 124:S92–S109.

———. 2023. "Rents and Intangible Capital: A *Q*+ framework." *Journal of Finance* 78 (4): 1873–1916.

Damodaran, Aswath. 2012. *Investment Valuation: Tools and Techniques for Determining the Value of Any Asset.* Vol. 666. John Wiley & Sons.

Davies, James B., Susanna Sandström, Anthony Shorrocks, and Edward N. Wolff. 2011. "The Level and Distribution of Global Household Wealth." *Economic Journal* 121 (551): 223–254.

De Nederlandsche Bank. 2000. *Nederlandse financiële instellingen in de twintigste eeuw: Balansreeksen en naamlijst van handelsbanken.* Amsterdam: De Nederlandsche Bank.

Dekkers, Arnold L.M., John H.J. Einmahl, and Laurens de Haan. 1989. "A Moment Estimator for the Index of an Extreme-Value Distribution." *The Annals of Statistics* 17 (4): 1833–1855.

Dell, Melissa, and Benjamin A. Olken. 2020. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java." *Review of Economic Studies* 87 (1): 164–203.

Derksen, J.B.D., and Jan Tinbergen. 1945. "Berekeningen over de Economische Betekenis van Nederlands-Indië voor Nederland." *Maandschrift van het Centraal Bureau voor de Statistiek* 40 (10/12): 210–216.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1): 83–95.

Driscoll, John C., and Aart C. Kraay. 1998. "Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (4): 549–560.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94 (5): 1429–1451.

———. 2009. "Gibrat's Law for (All) Cities: Reply." *American Economic Review* 99 (4): 1676–1683.

Eng, Pierre van der. 1998. “Exploring Exploitation: The Netherlands and Colonial Indonesia 1870–1940.” *Revista de Historia Economica - Journal of Iberian and Latin American Economic History* 16 (1): 291–321.

Erickson, Timothy, Colin Huan Jiang, and Toni M. Whited. 2014. “Minimum Distance Estimation of the Errors-in-Variables Model Using Linear Cumulant Equations.” *Journal of Econometrics* 183 (2): 211–221.

Erickson, Timothy, and Toni M. Whited. 2000. “Measurement Error and the Relationship between Investment and *q*.” *Journal of Political Economy* 108 (5): 1027–1057.

———. 2012. “Treating Measurement Error in Tobin’s *q*.” *Review of Financial Studies* 25 (4): 1286–1329.

Fagereng, Andreas, Matthieu Gomez, Emilien Gouin-Bonenfant, Martin Holm, Benjamin Moll, and Gisle Natvik. 2024. “Asset-Price Redistribution.” *World Inequality Lab Working Paper* 2024/14.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. “Heterogeneity and Persistence in Returns to Wealth.” *Econometrica* 88 (1): 115–170.

Fama, Eugene F., and Kenneth R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33:3–56.

Fedotenkov, Igor. 2020. “A Review of More than One Hundred Pareto-Tail Index Estimators.” *Statistica* 80 (3): 245–299.

Feldstein, Martin. 1976. “Social Security and the Distribution of Wealth.” *Journal of the American Statistical Association* 71 (356): 800–807.

Friedman, Milton. 1957. *A Theory of the Consumption Function.* National Bureau of Economic Research.

Gabaix, Xavier. 1999. “Zipf’s Law for Cities: An Explanation.” *Quarterly Journal of Economics* 114 (3): 739–767.

———. 2009. “Power Laws in Economics and Finance.” *Annual Review of Economics* 1 (1): 255–294.

———. 2011. “The Granular Origins of Aggregate Fluctuations.” *Econometrica* 79 (3): 733–772.

———. 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.

Gabaix, Xavier, and Rustam Ibragimov. 2011. "Rank −1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business & Economic Statistics* 29 (1): 24–39.

Gabaix, Xavier, and Ralph S.J. Koijen. 2024. "Granular Instrumental Variables." *Journal of Political Economy* 132 (7): 2274–2303.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123 (1): 49–100.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll. 2016. "The Dynamics of Inequality." *Econometrica* 84 (6): 2071–2111.

Garbinti, Bertrand, Jonathan Goupille-Lebret, and Thomas Piketty. 2020. "Accounting for Wealth-Inequality Dynamics: Methods, Estimates, and Simulations for France." *Journal of the European Economic Association* 19 (1): 620–663.

Gelderblom, Oscar, Joost Jonker, Ruben Peeters, and Amaury de Vicq. 2023. "Exploring Modern Bank Penetration: Evidence from Early Twentieth-Century Netherlands." *Economic History Review* 76 (3).

Gerritsen, Aart, Bas Jacobs, Kevin Spiritus, and Alexandra V. Rusu. 2024. "Optimal Taxation of Capital Income with Heterogeneous Rates of Return." *Economic Journal,* Forthcoming.

Goldin, Claudia, and Robert A. Margo. 1992. "The Great Compression: The Wage Structure in the United States at Mid- Century." *The Quarterly Journal of Economics* 14 (2): 11–28.

Gomez, Matthieu. 2023. "Decomposing the Growth of Top Wealth Shares." *Econometrica* 91 (3): 979–1024.

Gordon, Alec. 2010. "Netherlands East Indies: The Large Colonial Surplus of Indonesia, 1878-1939." *Journal of Contemporary Asia* 40 (3): 425–443.

Gormsen, Niels Joachim, and Kilian Huber. 2023. "Corporate Discount Rates." *NBER Working Paper,* no. w31329.

———. 2024. "Firms' Perceived Cost of Capital." *NBER Working Paper* 32611.

Goyal, Sanjeev. 2023. *Networks: An Economics Approach.* MIT Press.

Greenwald, Daniel L., Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh. 2021. "Financial and Total Wealth Inequality with Declining Interest Rates." *NBER Working Paper* 28613.

Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (1): 93–118.

Groote, Peter, Ronald Albers, and Herman De Jong. 1996. "A Standardised Time Series of the Stock of Fixed Capital in the Netherlands, 1900–1995." *Working Paper.*

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691–719.

Gumbel, Emil Julius. 1958. *Statistics of Extremes.* Columbia University Press.

Gupta, Arpit, and Stijn Van Nieuwerburgh. 2021. "Valuing Private Equity Investments Strip By Strip." *Journal of Finance* 76 (6): 3255–3307.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, Sergio Ocampo, and Daphne Chen. 2023. "Use It or Lose It: Efficiency and Redistributional Effects of Wealth Taxation." *Quarterly Journal of Economics* 138 (2): 835–894.

Guvenen, Fatih, Gueorgui Kambourov, Burhan Kuruscu, and Sergio Ocampo-Diaz. 2024. "Book-Value Wealth Taxation, Capital Income Taxation, and Innovation." *NBER Working Paper* 32585.

Haffner, Marietta, Joris Hoekstra, Michael Oxley, and Harry van der Heijden. 2009. *Bridging the Gap Between Social and Market Rented Housing in Six European Countries?* Housing and Urban Policy Studies 33. Amsterdam: IOS Press.

Hall, Robert E. 1967. "Essays on the Theory of Wealth." PhD diss., MIT.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (4): 1029–1054.

Hart, Oliver D., and John Moore. 1994. "A Theory of Debt Based on the Inalienability of Human Capital." *Quarterly Journal of Economics* 109 (4): 841–879.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica,* 213–224.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300–316.

Hill, Bruce M. 1975. "A Simple General Approach to Inference About the Tail of a Distribution." *The Annals of Statistics* 3 (5): 1163–1174.

Huggett, Mark. 1993. "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies." *Journal of Economic Dynamics and Control* 17 (5-6): 953–969.

Huisman, Ronald, Kees G. Koedijk, Clemens J.M. Kool, and Franz Palm. 2001. "Tail-Index Estimates in Small Samples." *Journal of Business & Economic Statistics* 19 (2): 208–216.

IBO Vermogensverdeling. 2022. *Licht Uit, Spot Aan: De Vermogensverdeling.* Rijksoverheid.

Jacobs, Bas. 2015. "Belastingen op Kapitaalinkomen in Nederland." *Tijdschrift voor Openbare Financiën* 47 (1): 24–48.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–784.

———. 2015. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality." *Journal of Economic Perspectives* 29 (1): 29–46.

———. 2022a. "Taxing Top Incomes in a World of Ideas." *Journal of Political Economy* 130 (9): 2227–2274.

———. 2022b. "The End of Economic Growth? Unintended Consequences of a Declining Population." *American Economic Review* 112 (11): 3489–3527.

———. 2023. "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail." *Journal of Political Economy* 131 (8): 1994–2031.

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy* 126 (5): 1785–1826.

Jonker, Joost. 1996. *Merchants, Bankers, Middlemen: The Amsterdam Money Market During the First Half of the 19th Century.* NEHA.

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. "The Rate of Return on Everything, 1870–2015." *Quarterly Journal of Economics* 134 (3): 1225–1298.

Judd, Kenneth L. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics* 28 (1): 59–83.

Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens. 2019. "Investor Sophistication and Capital Income Inequality." *Journal of Monetary Economics* 107:18–31.

Kaplan, Steven N., and Antoinette Schoar. 2005. "Private Equity Performance: Returns, Persistence, and Capital Flows." *Journal of Finance* 60 (4): 1791–1823.

Kapteyn, Arie, and Tom Wansbeek. 1983. "Identification in the Linear Errors in Variables Model." *Econometrica* 51 (6): 1847–1849.

Karabarbounis, Loukas, and Brent Neiman. 2014. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1).

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (1): 35–78.

Kleiber, Christian, and Samuel Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences.* John Wiley & Sons.

Knibbe, Merijn. 1993. *Agriculture in the Netherlands 1851–1950: Production and Institutional Change.* Neha Amsterdam.

———. 2014. "The Growth of Capital: Piketty, Harrod-Domar, Solow and the Long Run Development of the Rate of Investment." *Real-World Economics Review* 69:100–21.

Kondo, Illenin O., Logan T. Lewis, and Andrea Stella. 2023. "Heavy Tailed but not Zipf: Firm and Establishment Size in the United States." *Journal of Applied Econometrics* 38 (5): 767–785.

Kopczuk, Wojciech. 2015. "What Do We Know about the Evolution of Top Wealth Shares in the United States?" *Journal of Economic Perspectives* 29 (1): 47–66.

Kopczuk, Wojciech, and Emmanuel Saez. 2004. "Top Wealth Shares in the United States, 1916–2000: Evidence from Estate Tax Returns." *National Tax Journal* 57 (2): 445–487.

Kopczuk, Wojciech, and Eric Zwick. 2020. "Business Incomes at the Top." *Journal of Economic Perspectives* 34 (4): 27–51.

Korevaar, Matthijs, Marc Francke, and Piet Eichholtz. 2021. "Dure Huizen Maar Geen Zeepbel in Amsterdam." *Economisch Statistische Berichten* 106 (4793): 32–34.

Korteweg, Arthur. 2019. "Risk Adjustment in Private Equity Returns." *Annual Review of Financial Economics* 11 (1): 131–152.

Korthals Altes, Willem Lodewijk. 1986. "De betalingsbalans van Nederlandsch-Indië 1822–1939." PhD diss., Erasmus Universiteit Rotterdam.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–1419.

Koudijs, Peter, Abe de Jong, and Tim Kooijmans. 2022. "Plantation Mortgage-Backed Securities: Evidence from Surinam in the 18th Century." *Journal of Economic History* 83 (3): 874–911.

Krugman, Paul. 1996. "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and International Economies* 10 (4): 399–418.

Krusell, Per, and Anthony A. Jr. Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5): 867–896.

Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins. 2020. "Income and Wealth Inequality in America, 1949–2016." *Journal of Political Economy* 128 (9): 3469–3519.

Kumar, Rishabh. 2024. "Economic Growth and Indian Wealth-Income Ratios in the Long Run: 1860—2018." *Metroeconomica,* Forthcoming.

Kuvshinov, Dmitry, and Kaspar Zimmermann. 2022. "The Big Bang: Stock Market Capitalization in the Long Run." *Journal of Financial Economics* 145 (2, Part B): 527–552.

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2024. "100 Years of Rising Corporate Concentration." *American Economic Review* 114 (7): 2111–2140.

Kymmel, J. 1992. *Geschiedenis van de Algemene Banken in Nederland 1860–1914.* Vol. 1. Amsterdam: NIBE.

Lakner, Christoph, and Branko Milanovic. 2016. "Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession." *World Bank Economic Review* 30 (2): 203–232.

Lampman, Robert J. 1962. *The Share of Top Wealth-Holders in National Wealth, 1922–56.* 27–56. Princeton University Press.

Lewbel, Arthur, Susanne M. Schennach, and Linqi Zhang. 2023. "Identification of a Triangular Two Equation system Without Instruments." *Journal of Business & Economic Statistics,* 1–12.

Liu, Ernest, and Aleh Tsyvinski. 2024. "A Dynamic Model of Input-Output Networks." *Review of Economic Studies,* Forthcoming.

Luijt, Jan, and Martien Voskuilen. 2009. "Langetermijnontwikkeling van de agrarische grondprijs." *LEI Wageningen Nota* 09-014.

Luttmer, Erzo G. J. 2011. "On the Mechanics of Firm Growth." *Review of Economic Studies* 78 (3): 1042–1068.

Madsen, Jakob B. 2019. "Wealth and Inequality Over Eight Centuries of British Capitalism." *Journal of Development Economics* 138:246–260.

Martínez-Toledano, Clara, Alice Sodano, and Simon J. Toussaint. 2023. "Wealth Inequality in the Netherlands, 1894–2019: Updated Series With and Without Pension Wealth." *WID. World Technical Note,* nos. 2023/13.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Menger, Jasper. 2021. "Herziening Statistieken Aanmerkelijk Belang vanaf 2006." *CBS Publicatie.*

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2): 175–208.

Mogstad, Magne, Joseph P. Romano, Azeem M. Shaikh, and Daniel Wilhelm. 2024. "Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries." *Review of Economic Studies* 91 (1): 476–518.

Moll, Benjamin, Lukasz Rachel, and Pascual Restrepo. 2022. "Uneven Growth: Automation's Impact on Income and Wealth Inequality." *Econometrica* 90 (6): 2645–2683.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics* 31 (3): 358–369.

Mood, Alexander McFarlane, Franklin A. Graybill, and Duane C. Boes. 1974. *Introduction to the Theory of Statistics.* McGraw-Hill Kogakusha.

Moore, Lyndon. 2012. "World Financial Markets, 1900–1925." *Unpublished Manuscript.*

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–708.

Novokmet, Filip, Thomas Piketty, and Gabriel Zucman. 2018. "From Soviets to Oligarchs: Inequality and Property in Russia 1905–2016." *Journal of Economic Inequality* 16:189–223.

Orthofer, Anna, Stan Du Plessis, and Monique Reid. 2019. "Private Wealth in a Developing Country: Evidence from South Africa." *Review of Income and Wealth* 65 (3): 632–656.

Pareau, M. 1864. "Gaat de Welvaart in Nederland Achteruit?" *De Economist* 13 (1): 21–51.

Pareto, Vilfredo. 1896. *Cours d'Économie Politique: Professé á l'Université de Lausanne.* Vol. 1. F. Rouge.

Petersen, William. 1960. "The Demographic Transition in the Netherlands." *American Sociological Review* 25 (3): 334–347.

Philips, Robin. 2020. "Continuity or Change? The Spatial Evolution of Industry in the Netherlands and Belgium (1820–2010)." PhD diss., Utrecht University.

Pickands, James. 1975. "Statistical Inference using Extreme Order Statistics." *The Annals of Statistics* 3 (1): 119–131.

Piketty, Thomas. 2003. "Income Inequality in France, 1901–1998." *Journal of Political Economy* 111 (5): 1004–1042.

———. 2014. *Capital in the Twenty-First Century.* Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (1): 1–41.

Piketty, Thomas, Li Yang, and Gabriel Zucman. 2019. "Capital Accumulation, Private Property, and Rising Inequality in China, 1978–2015." *American Economic Review* 109 (7): 2469–2496.

Piketty, Thomas, and Gabriel Zucman. 2014. "Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010." *Quarterly Journal of Economics* 129, no. 3 (May): 1255–1310.

———. 2015. "Wealth and Inheritance in the Long Run." In *Handbook of Income Distribution,* edited by Anthony B. Atkinson and François Bourguignon, 2:1303–1368.

Potharst, Rob. 2022. "Estimating the Parameters of a Lognormal Distribution Using Grouped Censored Data." *Econometric Institute Report* EI 2022-01.

Quadrini, Vincenzo. 1999. "The Importance of Entrepreneurship for Wealth Concentration and Mobility." *Review of Income and Wealth* 45 (1): 1–19.

———. 2000. "Entrepreneurship, Saving, and Social Mobility." *Review of Economic Dynamics* 3 (1): 1–40.

Reiersøl, Olav. 1950. "Identifiability of a Linear Relation between Variables Which Are Subject to Error." *Econometrica* 18 (4): 375.

Roine, Jesper, and Daniel Waldenström. 2009. "Wealth Concentration over the Path of Development: Sweden, 1873–2006." *Scandinavian Journal of Economics* 111 (1): 151–187.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

Rosen, Kenneth T., and Mitchel Resnick. 1980. "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy." *Journal of Urban Economics* 8 (2): 165–186.

Rossi-Hansberg, Esteban, and Mark L.J. Wright. 2007. "Urban Structure and Growth." *Review of Economic Studies* 74 (2): 597–624.

Sadka, Efraim. 1976. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies* 43 (2): 261–267.

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68 (1): 205–229.

Saez, Emmanuel, and Gabriel Zucman. 2016. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131 (2): 519–578.

———. 2019. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay.* WW Norton & Company.

———. 2020. “Trends in US Income and Wealth Inequality: Revising After the Revisionists.” *NBER Working Paper* 27921.

Salverda, Wiemer. 2019. “Top Incomes, Income and Wealth Inequality in the Netherlands: The First 100 Years 1914–2014 – What’s Next?” *WID.world Working Paper* 2019/02.

Schennach, Susanne M. 2016. “Recent Advances in the Measurement Error Literature.” *Annual Review of Economics* 8 (1): 341–377.

———. 2022. “Measurement Systems.” *Journal of Economic Literature* 60 (4): 1223–1263.

Smeets, Mathieu. 1932. “Inkomen en Vermogen - En hun Betekenis voor de Welvaart.” PhD diss., University of Amsterdam.

Smith, Matthew, Owen Zidar, and Eric Zwick. 2023. “Top Wealth in America: New Estimates Under Heterogeneous Returns.” *Quarterly Journal of Economics* 138 (1): 515–573.

Smits, Jan Pieter Herman, Edwin Horlings, and Jan Luiten van Zanden. 2000. *Dutch GNP and its components, 1800–1913.* Groningen Growth / Development Centre.

Solow, Robert M. 1956. “A Contribution to the Theory of Economic Growth.” *Quarterly Journal of Economics* 70 (1): 65–94.

Stiglitz, Joseph E. 1969. “Distribution of Income and Wealth Among Individuals.” *Econometrica* 37 (3): 382–397.

Stuart, Verrijn. 1888. “Ons Maatschappelijk Vermogen voor 30 Jaren en Thans.” *Bijdrage van het Statistisch Instituut,* 267–298.

Teulings, Coen N., and Thijs van Rens. 2008. “Education, Growth, and Income Inequality.” *Review of Economics and Statistics* 90 (1): 89–104.

Teulings, Coen N., and Simon J. Toussaint. 2023. “Top Wealth Is Distributed Weibull, Not Pareto.” *CEPR Discussion Paper* 18364.

———. 2024. “Why Has the Number of Billionaires Increased So Much?” *CEPR Discussion Paper* 19191.

Thee, Kian Wie. 2012. *Indonesianization: Economic Aspects of Decolonization in the 1950s,* 3–23. ISEAS–Yusof Ishak Institute.

Tishby, Ido, Ofer Biham, and Eytan Katzav. 2016. "The Distribution of Path Lengths of Self Avoiding Walks on Erdős–Rényi Networks." *Journal of Physics A: Mathematical and Theoretical* 49 (28): 285002.

Toda, Alexis Akira. 2019. "Wealth Distribution with Random Discount Factors." *Journal of Monetary Economics* 104:101–113.

Toussaint, Simon J. 2022. "Technisch Rapport Vermogensverdeling." *Interdepartmentaal Beleidsonderzoek Vermogensonderdeling,* background document.

———. 2024. "Robust Estimation of Private Business Wealth." *Working paper.*

Toussaint, Simon J., Bas J.P. van Bavel, Wiemer Salverda, and Coen N. Teulings. 2020. "Nederlandse Vermogens Schever Verdeeld dan Gedacht." *Economisch Statistische Berichten* 105 (4789): 438–441.

Toussaint, Simon J., Amaury de Vicq, Michail Moatsos, and Tim van der Valk. 2022. "Household Wealth and its Distribution in the Netherlands, 1854–2019." *World Inequality Lab Working Paper* 2022/19.

Vaart, Aad W. van der. 2000. *Asymptotic Statistics.* Cambridge University Press.

Valk, Tim van der. 2019. "Household Finance in France and the Netherlands 1960–2000: An Evolutionary Approach." PhD diss., Utrecht University.

Vermeulen, Philip. 2016. "Estimating the Top Tail of the Wealth Distribution." *American Economic Review* 106 (5): 646–50.

———. 2018. "How Fat is the Top Tail of the Wealth Distribution?" *Review of Income and Wealth* 64 (2): 357–387.

Verstegen, Wybren. 1996. "National wealth and income from capital in the Netherlands, c. 1805-1910." *Economic and Social History in the Netherlands,* 73–108.

Vicq, Amaury de. 2021. "Exploring the Limits of the Limited Partnership: The Case of the Bank of Twente, 1860s–1920s." *Enterprise & Society* 23 (4): 1122–1147.

———. 2022. "Caught between Outreach and Sustainability: The Rise and Decline of Dutch Credit Unions." *Business History,* Forthcoming.

Vicq, Amaury de, and Christiaan van Bochove. 2023. "Historical Diversity in Credit Intermediation: Cosignatory Lending Institutions in Europe and North America, 1700s-–1960." *Social Science History* 47 (1): 95–119.

———. 2024. "Lending a Hand: Help Banks in the Netherlands, 1848–1898." *European Review of Economic History* 28 (2): 163–192.

Vicq, Amaury de, and Ruben Peeters. 2020. "Introduction to the Tafel v-bis Dataset: Death Duty Summary Information for The Netherlands, 1921: Social and Economic History." *Research Data Journal for the Humanities and Social Sciences* (Leiden, The Netherlands) 5 (1): 1–19.

Vicq, Amaury de, Simon Toussaint, Tim van der Valk, and Michail Moatsos. 2023a. "De vermogensongelijkheid stijgt sinds de jaren tachtig." *Economisch Statistische Berichten* 108 (4820): 156–157.

———. 2023b. "Kolonisatie Nederlands-Indië droeg sterk bij aan privaat vermogen." *Economisch Statistische Berichten* 108 (4820): 152–155.

Vries, Johannes de. 1976. *Een Eeuw vol Effecten: Historische schets van de Vereniging voor de Effectenhandel en de Amsterdamse Effectenbeurs 1876-1976.* Amsterdam: Vereniging voor Effectenhandel.

Waldenström, Daniel. 2017. "Wealth-Income Ratios in a Small, Developing Economy: Sweden, 1810–2014." *Journal of Economic History* 77 (1): 285–313.

———. 2021. "Wealth and History: An Update." *CEPR Discussion Paper* 16631.

Westrate, C. 1948. *Gedenkboek uitgegeven ter gelegenheid van het vijftigjarig bestaan der Coöperatieve Centrale Raiffeisen-Bank te Utrecht, 1898–1948.* Utrecht: Coöperatieve Centrale Raiffeisen-Bank.

Wijk, J. van der. 1939. *Inkomens- en Vermogensverdeling.* 26. Netherlands Economic Institute.

Wilterdink, Nico. 1984. *Vermogensverhoudingen in Nederland: Ontwikkelingen sinds de Negentiende Eeuw.* Amsterdam: Synopsis.

———. 2015. *Vermogensongelijkheid in Nederland: Ontwikkelingen sinds 1850.* Amsterdam: Van Gennep.

Wingo, Dallas R. 1989. "The Left-Truncated Weibull Distribution: Theory and Computation." *Statistical Papers* 30 (1): 39–48.

Wold, Herman O.A., and Peter Whittle. 1957. "A Model Explaining the Pareto Distribution of Wealth." *Econometrica,* 591–595.

Wolff, Edward N. 1998. "Recent Trends in the Size Distribution of Household Wealth." *Journal of Economic Perspectives* 12 (3): 131–150.

———. 2017. *A Century of Wealth in America.* Harvard University Press.

Xavier, Inês. 2020. "Wealth Inequality in the US: The Role of Heterogeneous Returns." *Working paper.*

Zanden, Jan Luiten van, and Daan Marks. 2013. *An Economic History of Indonesia: 1800-2010.* Routledge.

Zanden, Jan Luiten van, and Arthur van Riel. 2000. *Nederland 1780–1914: Staat, Instituties en Economische Ontwikkeling.* Uitgeverij Balans.

Zucman, Gabriel. 2019. "Global Wealth Inequality." *Annual Review of Economics* 11 (1): 109–138.

Nederlandstalige Samenvatting

Vermogen is de som van (de marktwaarde van) iemands bezit min diens schulden. We spreken van vermogens*concentratie* als het vermogen van de rijksten in de samenleving aanzienlijk toeneemt ten opzichte van de rest. Met vermogensconcentratie kunnen we ook de toegenomen dominantie van vermogen in de gehele economie beschrijven, ten opzichte van andere productiefactoren zoals arbeid. Dit proefschrift bestudeert de empirie en interpretatie van vermogensconcentratie, in beide opvattingen van het woord.

Vermogen is een essentieel raamwerk om de economie te begrijpen, en duikt op in tal van ogenschijnlijk ongerelateerde vraagstukken. Huishoudens kiezen hoeveel zij consumeren en hoeveel zij sparen voor later, en ze kiezen tussen het kopen en het huren van een huis. Bedrijven zoeken kapitaal om hun ideeën te financieren. Overheden financieren hun uitgaven door obligaties uit te geven. Al deze en nog vele andere voorbeelden gaan over de opbouw en verdeling van vermogen, en een integrale benadering is daarom noodzakelijk.

Ondanks het belang van het onderwerp is de vermogensverdeling tot voor kort nauwelijks bestudeerd. Vermogen is een moeilijk concept om scherp te definiëren en nog moeilijker om accuraat te meten; bovendien is de verdeling van vermogen vaak slecht in beeld. De empirische bijdragen van dit proefschrift omvatten de conceptualisatie van vermogen, het meten van privaat bedrijfsvermogen – een essentieel vermogensbestanddeel aan de top van de verdeling – en de bouw van betrouwbare langetermijndata over vermogensconcentratie. Behalve de empirische uitdagingen is in de literatuur ook het modelmatige begrip van de snelle fluctuaties in vermogensconcentratie onvolledig. De theoretische onderdelen van dit proefschrift dragen bij aan ons begrip hierover, door de canonieke kansverdeling die vermogens beschrijft te herzien, ons begrip van de verschillen in vermogensconcentratie tussen landen te verbeteren, en door nieuw bewijs te bieden voor een belangrijk mechanisme dat de snelle stijging in vermogensconcentratie kan verklaren, namelijk heterogene vermogensrendementen.

Dit proefschrift bestaat uit vier hoofdstukken, die elk zowel empirische als theoretische open vragen bestuderen. Hoofdstuk 1 onderzoekt langetermijntrends in vermogensconcentratie in Nederland, en verbindt veranderingen in indicatoren voor vermogensconcentratie aan sleutelmomenten in de Nederlandse economische geschiedenis zoals kolonisatie en de opbouw van de welvaartsstaat. Hoofdstuk 2 en 3 herzien de kansverdeling waarvan doorgaans wordt aangenomen dat die de vermogensverdeling goed beschrijft – de Paretoverdeling – en ontwikkelen modellen die consistenter zijn met de mondiale miljardairsverdeling. Hoofdstuk 4 richt zich op het meten van privaat bedrijfsvermogen, het dominante vermogensbestanddeel van de rijkste huishoudens. In de rest van deze Nederlandstalige samenvatting ga ik nader in op deze bijdragen. Ik begin door stil te staan bij de conceptualisatie van vermogen, wat met name in de Nederlandse beleidscontext geen sinecure is. Vervolgens bespreek ik de empirie van vermogensconcentratie en de bijdragen van dit proefschrift. Daarna introduceer ik de belangrijkste modelmatige inzichten die in dit proefschrift worden gebruikt, en hoe dit proefschrift die inzichten bekritiseert en verfijnt. Als laatste bespreek ik beleidsimplicaties van mijn onderzoek.

Wat is Vermogen?

Zoals gezegd is vermogen een nettobegrip: de waarde van iemands bezit min diens schulden. De "iemand" in deze definitie is bewust vaag gehouden, want de literatuur is hierover niet eenduidig. Standaard is om te kijken naar het vermogen van een *huishouden*, zeker aangezien de meeste huishoudens hun vermogen gezamenlijk beheren (denk aan het kopen van een huis). Andere keuzen zijn ook mogelijk en verdedigbaar, bijvoorbeeld om vermogen per volwassen individu te bekijken. In dit proefschrift beperk ik me in het algemeen tot vermogen op huishoudniveau. Een gedeeltelijke uitzondering hierop vormt het werk in hoofdstukken 2 en 3, waar ik miljardairslijsten gebruik. In theorie zijn die data op persoonsniveau, hoewel in de praktijk het vermogen van een miljardair zowat het hele vermogen van diens huishouden zal beslaan.

Wat is dan het bezit van een huishouden? Ook hier staat de onderzoeker voor keuzes. In de standaarddefinitie, zoals die bijvoorbeeld door het Centraal Bureau voor de Statistiek (CBS) wordt gebruikt, is bezit een object dat economische waarde genereert en waarover de bezitter eigendomsrechten kan uitoefenen. Voorbeelden zijn koopwoningen, spaartegoeden, aandelen, obligaties, en andere financiële producten. Merk echter op dat pensioenrechten zoals die in Nederland worden opgebouwd hier *niet* onder vallen: pensioenrechten zijn immers niet verhandelbaar, vererfbaar, of anderszins te consumeren voor pensionering.

Er valt echter ook veel voor te zeggen om pensioenrechten wel tot vermogen te rekenen, zelfs als het huishouden er geen directe eigenaar over is. Het zijn immers wel consumptiestromen in de toekomst waarvan redelijkerwijs verwacht kan worden dat het huishouden ze gaat ontvangen. Die verwachting kan je verdisconteren[25] en omrekenen tot een netto contante waarde; die netto contante waarde zou je pensioenvermogen kunnen noemen. Logischerwijs zouden dan ook andere aspecten van de welvaartsstaat, zoals de AOW, gekapitaliseerd kunnen en moeten worden. Ook verwacht toekomstig arbeidsinkomen, zogeheten menselijk kapitaal, zou dan onder deze ruimere definitie van vermogen moeten worden meegenomen. Het moge duidelijk zijn dat deze alternatieve visie op vermogen op veel aannames rust, bijvoorbeeld over de keuze van de discontovoet en over het berekenen van de verwachte waarde van toekomstige stromen. Niettemin is deze alternatieve definitie dicht bij het concept van "permanent inkomen" zoals Milton Friedman (1957) dat ooit formuleerde, en is als zodanig waarschijnlijk geschikter voor vraagstukken die de gehele levensloop aangaan.

In dit proefschrift doe ik geen expliciete aanbeveling voor een van beide definities. Gezien de databronnen die ik gebruik (impliciet) de eigendomsdefinitie volgen, doe ik dat ook. De enige plek waar ik hier enigszins van afwijk is in hoofdstuk 1. Daar reconstrueer ik het totale huishoudensvermogen in Nederland vanaf 1853, in lijn met het huidige Stelsel van Nationale Rekeningen. In dat Stelsel worden pensioenrechten als huishoudensvermogen gezien, dus doe ik dat ook voor het totale vermogen. In hoofdstuk 1 kijk ik ook naar de verdeling van vermogen. Omdat de bronnen die ik tot mijn beschikking heb geen verdeling van pensioenrechten bevatten, volg ik voor de ongelijkheidsstatistieken de standaard-eigendomsdefinitie. Dit is vooral relevant voor de meest recente decennia, aangezien de balans van Nederlandse pensioenfondsen sinds de jaren '90 dramatisch is gegroeid.

Een laatste keuze waar de onderzoeker voor staat is de bepaling van marktwaarde. Voor veel bestanddelen is dat eenvoudig: kijk naar de beurswaarde van aandelen, de waarde van spaartegoeden, enzovoorts. Voor woningen is dit al lastiger; vaak wordt de kadastrale waarde (WOZ in Nederland) gebruikt i.p.v. de marktwaarde. Maar het is vooral een probleem bij niet-beursgenoteerde bedrijven, omdat

[25] De discontovoet is het gewicht op toekomstige consumptie t.o.v. consumptie nu; bij een discontovoet van 5%, bijvoorbeeld, is consumptie morgen 0.95 keer zo waardevol als consumptie nu, en consumptie overmorgen $0.95^2 \approx 0.9$ keer zo waardevol. Een rekenvoorbeeld om verdisconteren te illustreren. Stel het toekomstige pensioeninkomen op 1000 € per jaar en de discontovoet op $r = 5\%$. Een persoon die 62 is verwacht pensioen op haar 67e en verwacht daarna nog tien jaar te leven. Zij moet vijf jaar wachten op haar eerste uitkering, die is haar nu dan $1000 \times 0.95^5 \approx 774$ euro waard. Elk volgend jaar wordt verder verdisconteerd, dus haar totale pensioenvermogen op haar 62e is

$$V_t = 0.95^5 \times 1000 + 0.95^6 \times 1000 + \ldots + 0.95^{15} \times 1000 \approx 6673.$$

Het rekenvoorbeeld maakt ook duidelijk dat zulke sommen erg fragiel zijn: als de discontovoet niet 5% maar 6% is, daalt het pensioenvermogen naar 6038 euro. Ook de aannames over de levensverwachting spelen vanzelfsprekend een grote rol.

daar de marktwaarde per definitie niet te zien is. Aangezien dit het voornaamste vermogensbestanddeel is van de allerrijkste huishoudens, is het voor een goed inzicht in vermogensconcentratie essentieel om de marktwaarde van private bedrijven goed in beeld te krijgen. Dit vraagstuk pak ik op in hoofdstuk 4.

Vermogensconcentratie: Empirie

Nu we uiteen hebben gezet wat ik onder vermogen versta, kunnen we kijken naar vermogens*concentratie*. Zoals eerder besproken kun je concentratie op twee manieren definiëren. Als we kijken naar de dominantie van vermogen in de algehele economie, gebruikt de literatuur sinds Piketty en Zucman (2014) de zogeheten "vermogens-inkomensratio." Deze maat deelt het totale (huishoudens-)vermogen, W_t, door het netto nationale inkomen Y_t.[26] Het resultaat is een getal, bijvoorbeeld 6, dat aangeeft hoeveel het totale vermogen waard is aan jaren nationaal inkomen. Als de ratio 6 is betekent dat dus dat het totale vermogen gelijk is aan 6 jaar nationaal inkomen. Dit is een eenvoudig te interpreteren maat; bovendien kun je vermogens-inkomensratio's tussen verschillende landen en perioden vergelijken, omdat je geen rekening met inflatie en andere prijsverschillen hoeft te houden.[27]

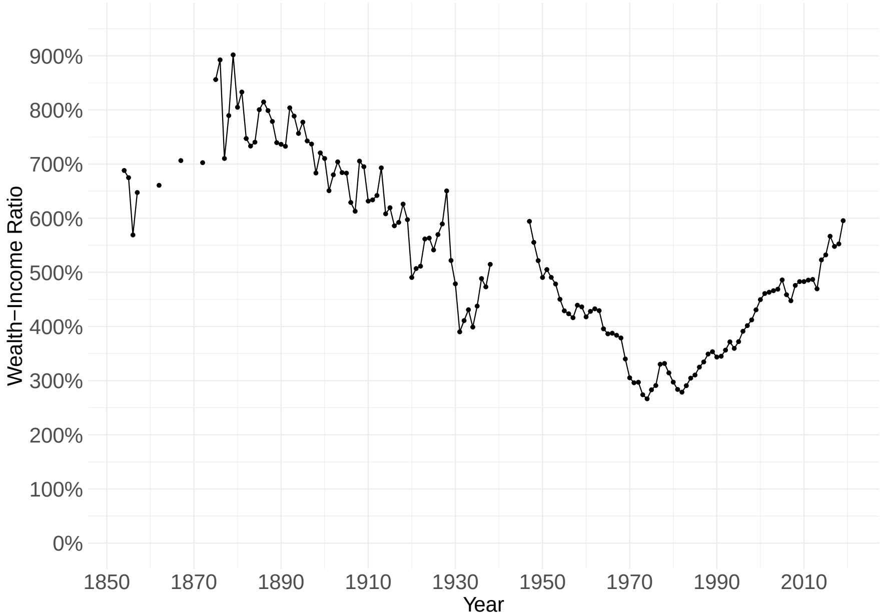
Onderstaande Figuur, ontleend aan hoofdstuk 1, toont de vermogens-inkomensratio in Nederland vanaf 1853. Het moge duidelijk zijn dat het niet eenvoudig is om het totale vermogen over een periode van 170 jaar te reconstrueren: er gaan vele verschillende databronnen, methoden en aannamen aan vooraf voordat we bij Figuur NL.1 uitkomen.

De Figuur toont grofweg een U-vorm over de loop van de 20e eeuw. Dit geeft aan dat vermogensconcentratie hoog was rond 1900 (waarden van 800%), sterk daalde tussen het interbellum en met name na de Tweede Wereldoorlog, en vanaf de jaren '80 weer stijgt. Dit is ook het beeld dat naar voren komt uit andere studies (cf. Piketty en Zucman 2014). Er vallen echter ook een paar andere dingen op. Ten eerste lijkt de vermogens-inkomensratio te stijgen tussen 1850 en 1900. In deze periode begon de Industriële Revolutie in Nederland, met alle transformaties vandien. Ook is dit de periode dat private investeringen in het buitenland flink beginnen te stijgen. Met name de investeringen in voormalig Nederlands-Indië stijgen flink vanaf 1880, en pieken in de jaren '20. Dit is duidelijk te zien in de Figuur: de neerwaartse trend

[26] Het netto nationale inkomen is vergelijkbaar met het bekendere bruto binnenlands product (BBP), behalve dat het netto inkomen verdiend in het buitenland wordt toegevoegd, en dat afschrijvingen op kapitaal worden afgetrokken (daarom is het netto nationaal inkomen en niet bruto).

[27] Dit is omdat W_t en Y_t in dezelfde nominale grootheid genoteerd staan, bijvoorbeeld guldens in 1938.

Figuur NL.1: Vermogens-Inkomensratio's in Nederland, 1853–2019



Noot: Deze reeks is o.b.v. vermogen inclusief pensioenrechten. Zie hoofdstuk 1 voor nadere details.

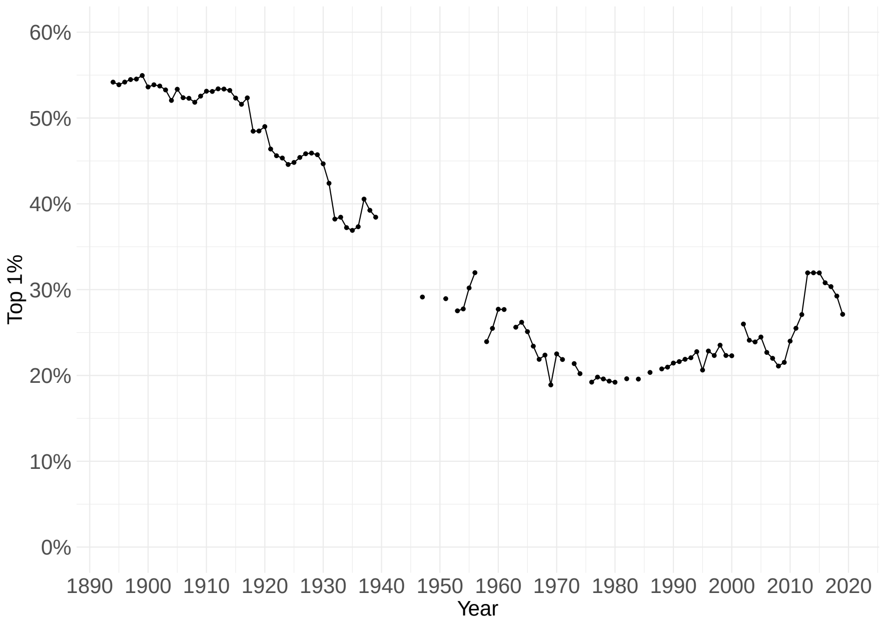
aan het eind van de Eerste Wereldoorlog wordt volledig ongedaan gemaakt door een enorme stijging die piekt vlak voor de Beurskrach in 1929. In de jaren '30 herstelt de ratio zich weer enigszins. Na de Tweede Wereldoorlog is het eerste datapunt, 1947, merkwaardig genoeg hoger dan de laatste voorloorlogse observatie, 1938.

De naoorlogse ontwikkelingen worden grotendeels verklaard door de dynamiek op de woningmarkt en de grote toename van het pensioenstelsel (zoals gezegd tellen pensioenrechten mee in deze maat). Eigenwoningbezit steeg flink vanaf de jaren '80, door een combinatie van overheidsbeleid (de ruime hypotheekrenteaftrek) en nieuwe financiële producten zoals de aflossingsvrije hypotheek. De bijkomstige grote toename van hypotheekschulden zorgden ervoor dat huizenvermogen netto niet veel bijdroeg aan de totale vermogensopbouw in de Figuur. De voornaamste bijdrage werd geleverd door de flinke toename van pensioenrechten vanaf de jaren '90.

Figuur NL.2 toont de voornaamste maatstaf van vermogensconcentratie binnen de vermogensverdeling, namelijk het aandeel van het totale vermogen dat in bezit is van de rijkste 1% huishoudens.

We zien grotendeels hetzelfde beeld als bij de vermogens-inkomensratio: een

Figuur NL.2: Top 1%-Aandeel, 1894–2019



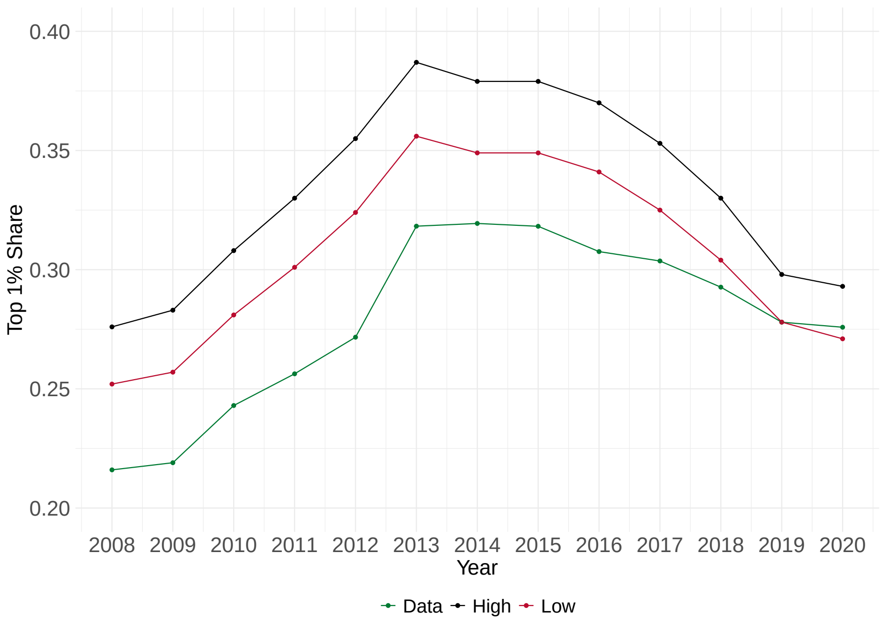
Noot: Deze reeks is o.b.v. vermogen exclusief pensioenrechten. Zie hoofdstuk 1 voor nadere details.

U-vorm, met een kleine heropleving in de jaren '20. De grote val in vermogensongelijkheid in de jaren '50 valt ook op: bijna 7 procentpunt in twee jaar. De timing correspondeert met de nationalisatie van de nog aanwezige Nederlandse bedrijven in Indonesië (geschat op 1,5 miljard gulden), hoewel verder onderzoek nodig is om dit verband oorzakelijk aan te tonen. Wat verder opvalt is de grote stijging in ongelijkheid na 2008. Dit correspondeert met een sterke daling in de huizenprijzen, waardoor het vermogen van de middenklasse drastisch afnam. Omdat de vermogendste huishoudens een andere samenstelling van hun vermogen hebben (hoofdzakelijk bedrijven en aandelen), steeg hun vermogen ten opzichte van de middenklasse. Hierdoor nam het 1%-aandeel toe. Deze trend werd omgekeerd toen de huizenprijzen weer gingen stijgen, omstreeks 2015.

In hoofstuk 1 neem ik de onderliggende CBS-data voor waar aan. In hoofdstuk 4 ga ik echter verder, en ontwikkel ik een methode om de marktwaarde van nietbeursgenoteerde bedrijven te schatten, die in officiële statistieken op boekwaarde worden genoteerd. Ik bouw een theoretische benadering die standhoudt in een breed scala aan modellen, en die stelt dat de marktwaarde van een bedrijf bij benadering lineair moet zijn in zijn kapitaalgoederenvoorraad. De methode is dan: pak een initiële schatting van marktwaarde, regresseer die op kapitaal, en houd de gefitte waarden.[28] Omdat zowel de initiële schatting als kapitaal met meetfouten worden waargenomen, ontwikkel ik een econometrische methode die deze fouten uit de regressie zeeft.

Het eindresultaat valt te bewonderen in onderstaande Figuur. In blauw zien we het top 1%-aandeel zoals die wordt gerapporteerd door het CBS. In rood en geel zien we de aangepaste waarden, die corresponderen met de lage en hoge schattingsresultaten van mijn regressies.

Figuur NL.3: Aangepaste Top-1%-aandelen



Noot: Deze reeks is o.b.v. vermogen exclusief pensioenrechten. Zie hoofdstuk 4 voor nadere details.

De eerder beschreven dynamiek door de huizenprijzen na de crisis is hier goed zichtbaar. De nieuwe resultaten tonen deze trends ook, maar hebben een hoger

[28]Een regressie vindt het best passende lineaire verband in de data. In formulevorm is de regressie

$$V_t = \beta K_t + \varepsilon_t,$$

waarbij V_t de initiële schatting van marktwaarde is en K de kapitaalgoederenvoorraad, beiden in jaar t; ε_t is het verschil tussen het geschatte lineaire verband en de data. Veronderstel een geschatte $\widehat{\beta} = 4.5$ (zoals uit mijn resultaten blijkt). Dan zou een bedrijf met een kapitaalgoederenvoorraad van 1 miljoen € een geschatte marktwaarde hebben van 4,5 miljoen € (de gefitte waarde).

niveau in het begin van de reeks. De piek van de vermogensongelijkheid, in 2013, lag tussen de 35% en 39%, wat in internationaal opzicht zeer hoge cijfers zijn. De belangrijkste conclusie van deze exercitie is dat het enorm belangrijk is om vermogen goed te meten, en dat dit een uitdagend vraagstuk is. De resultaten in hoofdstukken 1 en 4 zijn een eerste poging om dit vraagstuk te beantwoorden.

Vermogensconcentratie: Theorie

Het canonieke model om vermogensconcentratie te begrijpen is de Pareto-verdeling, dat beter bekend is als het "80-20 principe." Pareto (1896) toonde aan dat de rijkste 20% van de verdeling 80% van het vermogen bezat. Sterker nog: ook binnen de rijkste 20% was 80% van het vermogen in het bezit van de rijkste 20% (van de 20%, dus de rijkste 4% van de gehele verdeling). Deze "fractaliteit" is wat een Paretoverdeling kenmerkt. Wiskundig gezien vertaalt dat zich in de volgende formule: boven een bepaalde ondergrens Ω is de kans dat er iemand met vermogen $\underline{W}$ groter is dan een bepaalde realisatie W gelijk aan de ratio van die realisatie tot de ondergrens, tot een macht $-1/\alpha$:

$$\Pr\left[\underline{W} \geq W \,\middle|\, W \geq \Omega\right] = (W/\Omega)^{-1/\alpha}, \quad W, \Omega > 0, \alpha \in (0, 1).$$

Stel $\alpha = 1$ en $\Omega = 1$ miljoen €. Dan stelt de formule dat de kans om vermogens van meer dan 10 miljoen € te zien gelijk is aan (10 miljoen/1 miljoen)$^{-1}$ = 0.1, oftewel, er zijn tien keer zo weinig waarnemingen met 10 miljoen als met 1 miljoen. Het eigenaardige aan de Paretoverdeling is dat de verhoudingen constant blijven: de kans om vermogens van meer dan 100 miljoen te zien is 10 keer zo klein als die om vermogens van meer dan 10 miljoen te zien, enzovoorts. Dit is de fractaliteit die onder het 80-20 principe ligt.[29]

De Paretoverdeling is elegant en eenvoudig om mee te werken, en lijkt de top van de vermogensverdeling redelijk te omschrijven. Daarom wordt de verdeling vaak voor waar aangenomen, en zijn er weinig onderzoekers die de juistheid van deze verdeling bevragen. Dat is onfortuinlijk omdat Pareto ook belangrijke nadelen kent: voor te grote waarden van α bestaan belangrijke maten van de verdeling, zoals de variantie, niet meer. Sterker nog, voor $\alpha = 1$, zoals in het rekenvoorbeeld, zou het gemiddelde vermogen theoretisch gezien oneindig groot moeten zijn! In hoofstuk 2 ontwikkel ik een test die op eenvoudige wijze kan toetsen of een verdeling (bij

[29]Voor de verhouding 80-20 heb je een α nodig die niet gelijk is aan 1 maar aan $\ln 4/\ln 5 \approx 0.86$. Dit valt als volgt te zien: In een Paretoverdeling is het vermogensaandeel van de top p% gelijk aan $(100/p)^{\alpha-1}$. Zet $p = 20$ en stel de formule gelijk aan 0.8, neem links en rechts de logaritme en herschrijf voor het gewenste resultaat.

benadering) Pareto is. Ik gebruik de miljardairslijst van het zakenblad *Forbes* vanaf 2001, om de wereldwijde miljardairsverdeling aan deze test te onderwerpen. De test verwerpt de nulhypothese van Pareto overtuigend, over alle jaren en regio's.

Geen model is perfect, daarom is het op zich niet heel informatief dat Pareto niet (perfect) klopt. De tweede bijdrage van hoofstuk 2 is dan ook om een alternatief te ontwikkelen dat simpel is maar de data veel beter benadert. Dit alternatief is de Weibullverdeling. Weibull is een veralgemenisering van Pareto met een extra parameter, γ. Deze parameter bepaalt hoe snel topvermogens uitdunnen ten opzichte van de onderkant van de verdeling. Hoe verder in de verdeling, hoe (exponentieel) sneller de uitdunning. Als $\gamma = 0$ zijn we terug bij Pareto, die een constante uitdunning kent.

Hoofdstuk 2 toont aan dat de verwerping van Pareto ten faveure van Weibull niet verklaard kan worden door meetfouten in de miljardairslijst of te kleine steekproeven; het is een robuuste conclusie. Deze conclusie noopt tot het ontwikkelen van modellen van vermogensaanwas die een Weibullverdeling kunnen genereren. Het is welbekend hoe je modellen bouwt die een Paretoverdeling opleveren: begin met een multiplicatief stochastisch proces (technisch gezien een geometrische Brownse beweging), en voeg een frictie toe om de verdeling te stabiliseren.

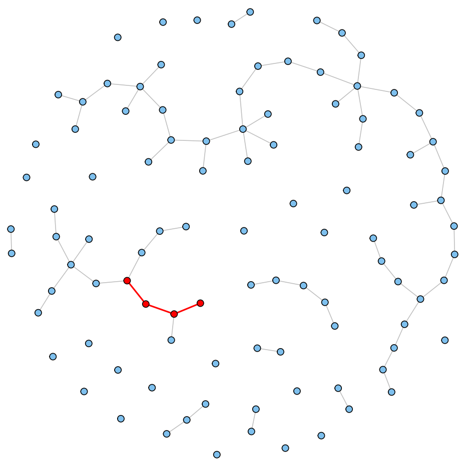
Dat werkt ruwweg als volgt. Gooi een muntje op: bij kop verdubbelt je vermogen, bij munt halveert het. Gooi nog een munt op: wederom verdubbelt of halveert je vermogen. In de limiet, als je oneindig veel muntjes opgooit (of beter gezegd, als de tijdsspanne tussen muntjes werpen naar nul gaat), heb je een geometrische Brownse beweging.[30] Om te voorkomen dat vermogen te klein wordt (in algemene diffusieprocessen wordt vermogen uiteindelijk vrijwel zeker negatief) heb je een 'ondergrens' nodig. Een voorbeeld van zo'n ondergrens in economische modellen is een leenrestrictie: een huishouden kan niet oneindig veel lenen om zijn investeringen te financieren. Zo'n ondergrens zorgt ervoor dat als een huishouden te lang achter elkaar pech heeft (te vaak munt), het 'terugstuitert.' Dit stabiliseert de verdeling, en je kan aantonen dat de stabiele langetermijnverdeling Pareto is. Topvermogens in zo'n model zijn huishoudens die ongewoon vaak geluk hebben gehad (heel vaak achter elkaar kop).

Je kunt het stochastische proces aanpassen om Weibull te genereren, maar in hoofdstuk 3 laat ik zien dat een alternatieve theoretische benadering ook werkt, waarbij we de economie niet als een massa atomen (huishoudens) zien, maar als een *netwerk*. Een succesvol bedrijf verbindt vele klanten in zijn industrie, maar de grootte van het netwerk beperkt hoe succesvol bedrijven kunnen worden. Een alternatieve interpretatie van het netwerk is dat het ideeën zijn die een bedrijf kan incorporeren

[30] Verdubbelen of halveren is niet strikt noodzakelijk; een geometrische Brownse beweging wordt gekenmerkt door een vaste procentuele stijging dan wel daling.

in zijn productieproces, waar wederom de hoeveelheid ideeën begrensd wordt door de totale bevolking. Onderstaande Figuur laat een netwerk zien: de punten zijn klanten (in de eerste interpretatie) of ideeën (in de tweede). Verbindingen tussen punten representeren bedrijven die meerdere klanten/ideeën verbinden. We zien een bedrijf als de langste wandeling langs verbonden punten zodat geen enkel punt tweemaal wordt aangedaan; dit noemen we een *zelfvermijdende wandeling*.

Figuur NL.4: Zelfvermijdende Wandeling



Figuur NL.4 toont een netwerk. We zien dat er een groot cluster is waar alle punten verbonden zijn (met dus grote bedrijven), en veel 'eilanden' (kleine bedrijven). Er zijn vele zelfvermijdende wandelingen op dit netwerk; een voorbeeld is in rood weergegeven (van het meest linkse punt naar het meest rechtse). Het blijkt dat de verdeling van de lengtes van zulke wandelingen (oftewel: van bedrijfsgroottes) overeenkomt met de verdeling die we in hoofdstuk 2 vonden.[31] Als we miljardairs zien als bedrijfseigenaren die vermogend zijn in proportie tot de grootte van hun bedrijven, verklaart dit de miljardairsverdeling. In hoofdstuk 3 gebruiken we het netwerkmodel zoals net geschetst.

Het model in hoofdstuk 3 doet drie boude voorspellingen. Ten eerste voorspelt het dat landen met een grotere bevolking meer supermiljardairs hebben ten opzichte

[31] Om precies te zijn is de lengteverdeling Gompertz; dit is de verdeling van de logaritme van vermogen als vermogen zelf een Weibullverdeling volgt.

van ‘gewone’ miljardairs. In termen van hoofdstuk 2: de parameter γ is lager, dus de uitdunning aan de top gaat langzamer. Verder voorspelt het model dat regio’s met een hoger BBP per capita meer miljardairs hebben (en dat deze relatie een-op-een is), en dat jaren met betere mondiale investeringscondities (zoals gemeten door de Amerikaanse vermogens-inkomensratio) ook leiden tot meer miljardairs, wederom met een een-op-eenrelatie. Alle drie de voorspellingen houden stand in de data van *Forbes*. Hoewel het model kaal is opgezet, verklaart het alle tijdreeksvariatie in de data. Niet alle variatie tussen regio’s wordt ondervangen: daarvoor hebben we regionale fixed effects nodig. Dit is logisch: Zwitserland heeft meer miljardairs dan waar ze ‘recht op hebben’ op basis van hun macroeconomische kernvariabelen.

In mijn proefschrift zit een laatste theoretische bijdrage besloten in hoofdstuk 4. Er is namelijk een ander probleem met de modellen gebaseerd op geometrische Brownse bewegingen (naast dat ze naar Pareto convergeren i.p.v. Weibull): de convergentie gaat te langzaam. Dit is logisch: iemand kan alleen miljardair worden in zo’n model door heel vaak achter elkaar kop te gooien, en dat komt heel zelden voor. In de data zien we echter zeer snelle fluctuaties in vermogen, en deze modellen kunnen die dynamiek niet verklaren. In een baanbrekend artikel formaliseren Gabaix, Lasry, Lions en Moll (2016) deze observatie en laten zien wat ervoor nodig is om deze dynamiek wel te vangen. Het missende ingrediënt is *rendementsheterogeniteit*. In de modellen hierboven is het rendement op vermogen onafhankelijk van het vermogen van een individu: bij kop verdubbelt het vermogen, ongeacht of dat vermogen 1000 € of 1 miljard € is. Als het zo is dat een miljardair een hoger rendement op haar vermogen heeft dan een miljonair, dan kan ongelijkheid veel sneller stijgen.

Een belangrijke reden waarom rendementen heterogeen zijn, wordt al lang goed begrepen: vermogende huishoudens hebben veel riskantere bezittingen dan minder vermogende huishoudens. Aandelen zijn riskanter dan spaartegoeden, en hebben daarom een hoger vereist rendement; dit is de basis van de financieringsleer. Maar recent bewijs laat zien dat zelfs binnen bezittingsklassen rendementen stijgen met het vermogen: iemand met 1 miljoen € aan aandelen haalt daar een hoger rendement op dan iemand met 1000 € aan aandelen, ceteris paribus (Fagereng, Guiso, Malacrino en Pistaferri 2020). Het is theoretisch nog een open vraag hoe dit kan, maar het is duidelijk dat dit potentie heeft om de dynamiek van de vermogensconcentratie te verklaren.

In hoofdstuk 4 laat ik zien dat de vermogendste huishoudens voor het overgrote deel hun vermogen in niet-beursgenoteerde bedrijven beleggen (tot wel 80% van hun portefeuille voor de top-0,01%). Rendementsheterogeniteit is dus met name voor dit bestanddeel belangrijk om in kaart te brengen. Hier speelt echter een probleem: als er meetfouten zitten in vermogen, tast dit rendementsregressies ernstig aan. Het idee is als volgt. In een rendementsregressie toetst de econometrist de relatie tussen een rendement op vermogen (dividenden en vermogenswinsten gedeeld door vermogen) en het vermogen zelf. Stel dat er een meetfout in vermogen zit, dan zit die zowel aan de linkerkant als de rechterkant van de regressie. Dit zorgt voor mechanische correlaties, zelfs als die er niet zouden zijn tussen het echte rendement en het echte vermogen.[32] De richting van de correlatie is ambigu: de relatie kan overschat worden, maar ook juist onderschat.

De econometrische methode die ik ontwikkel in hoofdstuk 4 om de marktwaarde van bedrijfsvermogen te schatten, zorgt ook voor foutenvrije schattingen van rendementen. Ik laat zien dat de correcties inderdaad flink uitmaken. De onaangepaste rendementsregressies laten nauwelijks een correlatie zien, maar de aangepaste rendementen zijn juist sterk positief gecorreleerd met bedrijfsvermogen. Mijn resultaten bevestigen dus het bestaan van rendementsheterogeniteit in bedrijfsvermogen.

Beleidsimplicaties

Mijn onderzoek heeft verscheidene beleidsimplicaties.[33] De eerdere beschouwing over wat vermogen is toont aan dat de definitie van vermogen tot op grote hoogte een beleidskeuze is: in Nederland hebben huishoudens geen eigendomsrechten over hun pensioenrechten, maar in andere landen wel. Beleid bepaalt dus niet alleen de verdeling van vermogen, maar zelfs wat telt als vermogen. Hoofdstuk 1 toont aan dat beleidskeuzes in grote mate bepalend zijn voor de aanwas, opbouw en verdeling van vermogen. Het is huishoudens door beleid aangemoedigd om hun vermogen goeddeels in woningen vast te zetten. Deze ontwikkeling heeft serieuze consequenties voor de spreiding van risico's en daarmee de stabiliteit van de economie, en ook voor het vermogen van huishoudens om met tegenslagen om te gaan.

Een tweede conclusie valt te trekken uit hoofdstukken 2 en 3. In deze hoofdstukken toon ik aan dat vermogensconcentratie aan de top van de (mondiale) vermogensverdeling in hoge mate wordt bepaald door macroeconomische kernvariabelen. De bevolkingsomvang vergroot staartongelijkheid, net zoals stijgingen in locale en mondiale investeringsmogelijkheden. Vermogensongelijkheid is dus in een niet onbelangrijke mate een endogene uitkomst in een markteconomie (van Bavel 2016).

[32] De technische reden voor dit resultaat is dat meetfout in de noemer van een breuk niet verdwijnt, omdat de verwachting van een breuk niet gelijk is aan de breuk van verwachtingen, door Jensens ongelijkheid.

[33] Zie hiervoor ook Toussaint, van Bavel, Salverda en Teulings (2020), de Vicq, Toussaint, van der Valk en Moatsos (2023a, 2023b), en Toussaint (2022).

Tegelijkertijd laat de grote mate van onverklaarde regionale heterogeniteit ook zien dat beleid wel degelijk de hoogte van vermogensconcentratie kan beïnvloeden.

Hoofdstuk 4 bevat verschillende beleidsconclusies. Ten eerste is het accuraat meten van niet-beursgenoteerd bedrijfsvermogen essentieel voor ons begrip van vermogensconcentratie. Een belangrijkere conclusie betreft rendementsheterogeniteit. Standaardmodellen, waarin het rendement op vermogen gelijk is voor iedereen en gelijk aan r, zien een belasting op vermogen van τ% als gelijk aan een belasting op kapitaalinkomen van τ/r%. Met andere woorden: een vermogensbelasting van 1% is, bij een rendement van 5%, gelijk aan een kapitaalinkomensbelasting van 20%, en vice versa. Als rendementen structureel verschillen tussen huishoudens, zelfs binnen het zelfde vermogensbestanddeel, valt deze equivalentie weg. Zodoende zou het optimaal kunnen zijn om vermogen te belasten i.p.v. kapitaalinkomen, hoewel de literatuur hierover nog geen eenduidig oordeel heeft geveld (Guvenen en andere 2023; Boar en Midrigan 2023; Gerritsen, Jacobs, Spiritus en Rusu 2024).[34] Rendementsheterogeniteit is dus van direct belang voor de structuur van kapitaalbelastingen.

In de Nederlandse context is dit punt van extra belang sinds het Kerstarrest van de Hoge Raad in 2022. Voor dit arrest werd financieel vermogen in box 3[35] belast met een *de facto* vermogensbelasting (Jacobs 2015): vermogen werd een fictief of forfaitair rendement van 4% toegerekend, wat met 30% werd belast. Dit komt neer op een vermogensbelasting van 1,2%. Deze 'vermogensrendementsheffing' werd in 2016 progressief gemaakt door voor verschillende schijven een verschillende vermogenssamenstelling te veronderstellen. Met de historische rendementen van deze vermogenbestanddelen werd zo een forfaitair rendement per schijf vastgesteld. Deze constructie werd succesvol aangevochten, wat ertoe heeft geleid dat een majeur gedeelte van de rijksinkomsten nu buiten werking zijn gesteld tot nader orde; op moment van schrijven worden zelfs 'herstelbetalingen' aan individuen gedaan die te veel belasting hebben betaald. Er wordt gewerkt aan een nieuw stelsel waar kapitaalinkomen op basis van feitelijk rendement wordt belast. Mijn resultaten laten zien hoe rendementen inderdaad verschillen over de vermogensverdeling, zelfs binnen een bestanddeel, en zijn dus relevant voor het ontwerp van het nieuwe belastingstelsel. Een succesvolle invoering van het nieuwe stelsel vereist een gedegen

[34] In simpele modellen is een vermogensbelasting een vorm van dubbele belasting en daarom is een kapitaalinkomensbelasting superieur; bij heterogene rendementen is dit niet noodzakelijkerwijs het geval. Guvenen et al. beargumenteren zelfs dat een vermogensbelasting doelmatigheidswinsten kan opleveren, omdat het ondernemers prikkelt om hogere rendementen te halen.

[35] Spaar- en banktegoeden, aandelenbelangen van minder dan 5%, obligaties en andere financiële producten.

vermogenskadaster, waarin alle bezittingen van huishoudens en hun rendementen secuur worden bijgehouden.

Wel is mijn onderzoek in hoofdstuk 4 gebaseerd op box 2 (aanmerkelijk belang), en niet box 3. Ook hier dringt zich een beleidsconclusie op. Ik analyseer in hoofdstuk 4 waarom mijn nieuwe reeks van bedijfsvermogen andere dynamieken vertoont dan de officiële data op basis van boekwaarde. Het blijkt dat er na de daling van depositorentes in 2013 er een grote instroom van kapitaal in private bedrijven te zien valt. Door de daling van de rente werd de (forfaitaire) belastingdruk in box 3 steeds hoger, met name voor huishoudens met veel spaartegoeden. Door hun vermogen in een bv te stallen in box 2, verlaagden zij hun relatieve belastingdruk. Dit stuwt de boekwaarde van niet-beursgenoteerde bedrijven kunstmatig op, iets waar mijn verbeterde reeks geen last van heeft. Deze resultaten bevestigen nogmaals dat arbitrage tussen verschillende boxen dermate groot is dat het op macroniveau te zien is. Dit is nog een voorbeeld van mijn stelling dat beleid de vermogenssamenstelling van huishoudens in verregaande mate beïnvloedt. Een uitgewogen belastingsysteem zorgt er dus voor dat deze arbitrage beperkt blijft, idealiter door een uniform tarief op alle vermogensbestanddelen (Jacobs 2015).

Curriculum Vitae

Simon Toussaint was born in 1997 in Utrecht. After finishing high school at the Utrechts Stedelijk Gymnasium (*summa cum laude*), he obtained a BSc. in Governance, Economics, and Development at Leiden University College (*cum laude*). Subsequently, he enrolled in the Research Master's programme at the Utrecht School of Economics. He graduated in 2020 (*cum laude*), and started his PhD at the U.S.E. the same year. He is a fellow of the *World Inequality Database*. In addition, he is an external researcher at Statistics Netherlands, working on Distributional Household Accounts. His research has been covered in various Dutch news outlets, such as *de Volkskrant*, *EenVandaag*, *de Jonge Economenpodcast*, and *NU.nl*. In 2022, he conducted research on behalf of a Dutch governmental report on the wealth distribution, published as Toussaint (2022). In 2023, he was one of six young economists presenting their research to Ministers Sigrid Kaag (Finance), Micky Adriaansens (Economic Affairs & Climate) and Karien van Gennip (Social Affairs & Employment), in the scenic Catshuis. He enjoys playing the cello, classical singing, and hiking in nature.

U.S.E. Dissertation Series

USE 001	**Bastian Westbrock**	(2010)	Inter-firm networks: economic and sociological perspectives.
USE 002	**Yi Zhang**	(2011)	Institutions and International Investments: Evidence from China and Other Emerging Markets.
USE 003	**Ryan van Lamoen**	(2011)	The Relationship between Competition and Innovation: Measuring Innovation and Causality.
USE 004	**Martijn Dröes**	(2011)	House Price Uncertainty in the Dutch Owner-Occupied Housing Market.
USE 005	**Thomas van Huizen**	(2012)	Behavioural Assumptions in Labour Economics: Analysing Social Security Reforms and Labour Market Transitions.
USE 006	**Martijn Boermans**	(2012)	International Entrepreneurship and Enterprise Development.
USE 007	**Joras Ferwerda**	(2012)	The Multidisciplinary Economics of Money Laundering.
USE 008	**Federico D'Onofrio**	(2013)	Observing the country: a history of Italian agricultural economics, 1900-1930.
USE 009	**Saraï Sapulete**	(2013)	Works Council Effectiveness: Determinants and Outcomes.
USE 010	**Britta Hoyer**	(2013)	Network Formation under the Threat of Disruption.
USE 011	**Coen Rigtering**	(2013)	Entrepreneurial Orientation: Multilevel Analysis and Consequences.
USE 012	**Beate Cesinger**	(2013)	Context and Complexity of International Entrepreneurship as a Field of Research.
USE 013	**Jan de Dreu**	(2013)	Empirical essays on the governance of financial institutions.

USE 014	**Lu Zhang**	(2013)	Industrial Specialization: Determinants, Processes and Consequences.
USE 015	**Matthias Filser**	(2013)	Strategic Issues in Entrepreneurship and Family Business Research.
USE 016	**Mikko Pohjola**	(2013)	A Compilation of Studies on Innovation in Firms: Capabilities, Strategies, and Performance.
USE 017	**Han-Hsin Chang**	(2013)	Heterogeneity in Development.
USE 018	**Suzanne Heijnen**	(2014)	Analyses of sickness absence.
USE 019	**Mark Kattenberg**	(2014)	The Economics of Social Housing: Implications for Welfare, Consumption, and Labor Market Composition.
USE 020	**Daniel Possenriede**	(2014)	The Economics of Temporal and Locational Flexibility of Work.
USE 021	**Dirk Gerritsen**	(2014)	The Relevance of Security Analyst Opinions for Investment Decisions.
USE 022	**Shiwei Hu**	(2014)	Development in China and Africa.
USE 023	**Saara Tamminen**	(2014)	Heterogeneous Firms, Mark-Ups, and Income Inequality.
USE 024	**Marcel van den Berg**	(2014)	Does Internationalization Foster Firm Performance?
USE 025	**Emre Akgündüz**	(2014)	Analyzing maternal employment and child care quality.
USE 026	**Jasper Lukkezen**	(2014)	From Debt Crisis to Sovereign Risk.
USE 027	**Vesile Kutlu**	(2015)	Essays on Subjective Survival Probabilities, Consumption, and Retirement Decisions.
USE 028	**Brigitte Crooijmans**	(2015)	Leiden fusies tot efficiëntere woningcorporaties? Een exploratieve studie naar schaalvoordelen in de sociale huisvesting.
USE 029	**Andrej Svorenčík**	(2015)	The Experimental Turn in Economics: a History of Experimental Economics.
USE 030	**Secil Danakol**	(2015)	Foreign Direct Investment, Foreign Aid and Domestic Entrepreneurship.
USE 031	**Ioana Deleanu**	(2015)	Anti-Money Laundering Efforts: Failures, Fixes and the Future.

USE 032	**Jaap Oude Mulders**	(2016)	Organizations, managers, and the employment of older workers after retirement.
USE 033	**Malka de Castro Campos**	(2016)	Private Consumption-Savings Behavior and Macroeconomic Imbalances.
USE 034	**Tahereh Rezai Khavas**	(2016)	Fairness concerns and cooperation in context.
USE 035	**Joyce Delnoy**	(2016)	Auctions with Competing Sellers and Behavioral Bidders.
USE 036	**Krista Bruns**	(2017)	Emergence and Diffusion of Institutions and their Effect on Economic Growth.
USE 037	**Daan van der Linde**	(2017)	Democracies under Rising Inequality: New Tests of the Redistributive Thesis.
USE 038	**Swantje Falcke**	(2017)	On the move: Analyzing immigration determinants and immigrant outcomes.
USE 039	**Joep Steegmans**	(2017)	House Prices and Household Mobility in The Netherlands: Empirical Analyses of Financial Characteristics of the Household.
USE 040	**Najmeh Rezaei Khavas**	(2017)	Essays in Information Economics.
USE 041	**Maryam Imanpour**	(2017)	The Role of Social Networks for Combating Money Laundering.
USE 042	**Ye Li**	(2018)	Hydrogen Infrastructure Decisions through a Real Option Lens.
USE 043	**Li Lin**	(2018)	Leadership across cultural contexts.
USE 044	**Werner Liebregts**	(2018)	Hidden entrepreneurship: Multilevel analyses of the determinants and consequences of entrepreneurial employee activity.
USE 045	**Ian Koetsier**	(2018)	Government debt: The economic consequences of natural disasters and pension funds' herding.
USE 046	**Jordy Meekes**	(2019)	Local Labour Markets, Job Displacement And Agglomeration Economies.
USE 047	**Timur Pasch**	(2019)	Essays On The Design Of The Management Accounting System: Determinants, Components And Effects.

USE 048	**Jeroen Content**	(2019)	The role of relatedness and entrepreneurship in regional economic development
USE 049	**Franziska Heinicke**	(2019)	Essays on self-image and preferences for honesty
USE 050	**Rebean Al-silefanee**	(2019)	Entrepreneurship and Private Sector Development: The Case of Kurdistan Region of Iraq
USE 051	**Markus Meinzer**	(2019)	Countering cross-border tax evasion and avoidance: An assessment of OECD policy design from 2008 to 2018.
USE 052	**Zornitza Kambourova**	(2019)	Women's Adverse Health Events and Labor Market Participation
USE 053	**Tim van der Valk**	(2019)	Household finance in France and the Netherlands 1960-2000: An evolutionary approach
USE 054	**Milena Dinkova**	(2019)	Brace yourselves, Pension is coming: Consumption, financial literacy and tailored pension communication
USE 055	**Lisa Dumhs**	(2019)	Finding the right job: School-to-work transitions of vocational students in the Netherlands
USE 056	**Dea Tusha**	(2020)	FDI spillovers in developing countries: channels, conditions, challenges
USE 057	**Jingyang Liu**	(2020)	Money and credit dynamics in the euro area
USE 058	**An Duong**	(2020)	Financial integration, trade, and productivity
USE 059	**Katharina Weddige-Haaf**	(2021)	Real and Financial Asymmetries in the Euro Area
USE 060	**Peter Gerbrands**	(2021)	Tax Dynamics and Money Laundering. Simulating Policy Reforms in a Complex System
USE 061	**Timo Verlaat**	(2022)	Carrot and Stick: Experiments With Social Welfare Policies
USE 062	**Lucia Rossel Flores**	(2022)	A multidisciplinary analysis of tax reform: from politics to human behavior
USE 063	**Wanxiang Cai**	(2022)	Social capital and crowdfunding. A multilevel perspective

USE 064	**Vincent Schippers**	(2022)	The local economic impacts of natural disasters. A view from outer space
USE 065	**Peter D. van der Meer**	(2022)	Job insecurity and mental health. Essays on the effect of job insecurity on mental health and the moderating effect of religiousness and psychological factors
USE 066	**Thomas Gomez**	(2022)	The roles of uncertainty and beliefs in the economy.
USE 067	**Bora Lancee**	(2023)	The role of attention and information for behavioral change
USE 068	**Merve Burnazoglu**	(2023)	Inequalities Beyond the Average Man. The Political Economy of Identity-Based Stratification Mechanisms in Markets and Policy.
USE 069	**Ronja Röttger**	(2023)	Worker adjustment in the digital age.
USE 070	**Jesse Groenewegen**	(2023)	Management Practices and Firm Adaptiveness during an External Shock. The Case of Covid-19.
USE 071	**Max Mulhuijzen**	(2023)	The Unexpected Sources of Innovation.
USE 072	**Sebastiaan Tieleman**	(2023)	A Window into Economic Practice. A Study into the Practice of Macroeconomic Modeling.
USE 073	**Emiel van Bezooijen**	(2024)	Essays on Income Inequality. The Role of Institutions, Firms and Neighborhoods.
USE 074	**Mirella Schrijvers**	(2024)	Entrepreneurial Ecosystems: Measurement, mechanisms and outcomes.
USE 075	**Cäcilia Maria Lipowski**	(2024)	Technological Change: Origins and Impacts in the Labor Market
USE 076	**Xing Li**	(2024)	Institutions and Social Entrepreneurship: A Multilevel Perspective
USE 077	**Margot Leger**	(2024)	Stimuleren van Ondernemerschap: Ondersteuning van intermediairs en sociale ondernemingen in Sub-Sahara Afrika
USE 078	**Milande Busquet**	(2025)	Social Entrepreneurship in the cocoa industry in West Africa: a human-centered perspective on labor issues