DISCUSSION PAPER SERIES

DP19191

WHY HAS THE NUMBER OF BILLIONAIRES INCREASED SO MUCH?

Coen Teulings and Simon Toussaint

LABOUR ECONOMICS, MACROECONOMICS AND GROWTH, POLITICAL ECONOMY AND PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

WHY HAS THE NUMBER OF BILLIONAIRES INCREASED SO MUCH?

Coen Teulings and Simon Toussaint

Discussion Paper DP19191
Published 30 June 2024
Submitted 14 June 2024

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

WHY HAS THE NUMBER OF BILLIONAIRES INCREASED SO MUCH?

Abstract

We study the fourfold increase in the number of billionaires since 2001, and its regional variation. We 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 our 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 our data. We 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. We interpret our results in the context of models where market size shapes firm (and hence wealth) concentration.

JEL Classification: D3, E2, G5

Keywords:

Coen Teulings - c.n.teulings@outlook.com
Utrecht University and CEPR

Simon Toussaint - s.j.toussaint@uu.nl
Utrecht University

Acknowledgements
We are grateful toTommaso Tulkens for research assistance.We especially thank JanuszMeylahn for helpful discussions, aswell as Bas van Bavel, Richard Blundell, Xavier Gabaix, Tomer Ifergane, Rutger-Jan Lange,Wouter Leenders, BenMoll, Thomas Piketty,Maarten de Ridder, Yasmine van der Straten, and Chen Zhou, and numerous seminar participants for helpful comments. We also thank Gustav Munch and River Chen for their coding advice.

Why Has the Number of Billionaires Increased So Much?*

Coen N. Teulings† Simon J. Toussaint‡

June 14, 2024

Abstract

We study the fourfold increase in the number of billionaires since 2001, and its regional variation. We 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 our 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 our data. We 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. We interpret our results in the context of models where market size shapes firm (and hence wealth) concentration.

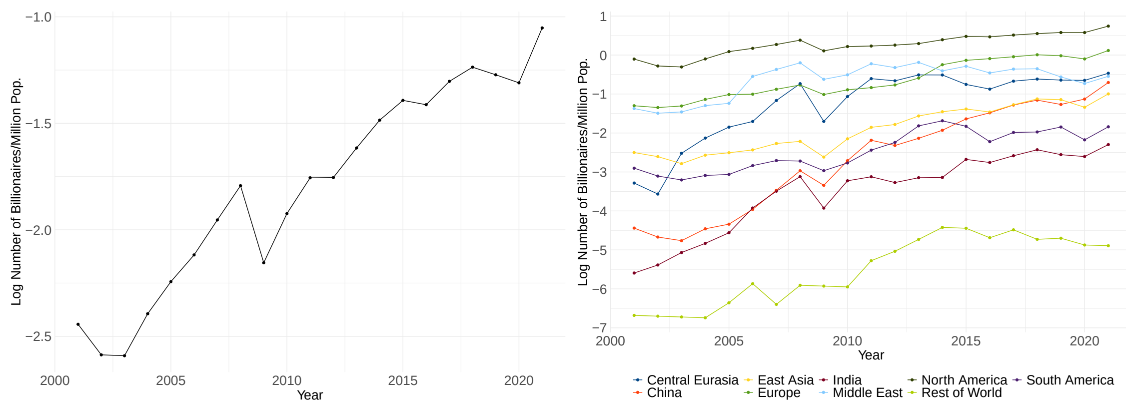
JEL Classification: D3 E2, G5
Keywords: Wealth inequality; Random networks; Global inequality.

*We are grateful to Tommaso Tulkens for research assistance. We especially thank Janusz Meylahn for helpful discussions, as well as Bas van Bavel, Richard Blundell, Xavier Gabaix, Tomer Ifergane, Rutger-Jan Lange, Wouter Leenders, Ben Moll, Thomas Piketty, Maarten de Ridder, Yasmine van der Straten, and Chen Zhou, and numerous seminar participants for helpful comments. We also thank Gustav Munch and River Chen for their coding advice.
†Utrecht School of Economics and CEPR. E-mail: c.n.teulings@outlook.com
‡Utrecht School of Economics. E-mail: s.j.toussaint@uu.nl

1 Introduction

The global ratio of billionaires to population size has more than quadrupled since 2001, see Figure 1a. The regional variation in this trend is enormous, see in Figure 1b. 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?

Figure 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 1.

Given the often heated debate surrounding billionaire wealth (e.g., Saez and Zucman 2019), it is surprising that this ques–tion has not been extensively studied. The purpose of this paper 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 our companion paper (Teulings and Toussaint 2023), 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 paper 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 $p \rightarrow 0$, 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 market size, this means 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 t 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.

Our highly stylized model yields strong and testable predictions. 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 popula–tion size (Jones 2022). 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.

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

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 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 2019; Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi 2012; Carvalho and Gabaix 2013; Liu and Tsyvinski 2024). While the use of networks in this paper 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.

Paper Outline: Section 2 recapitulates the main arguments in Teulings and Toussaint (2023). Section 3 presents our model, while Section 4 analyzes its identification and estimation. In Section 5, we discuss our data. Section 6 presents the results. We use our model for counterfactual exercises in Section 7. We discuss and interpret our results in Section 8.

2 Stylized Facts

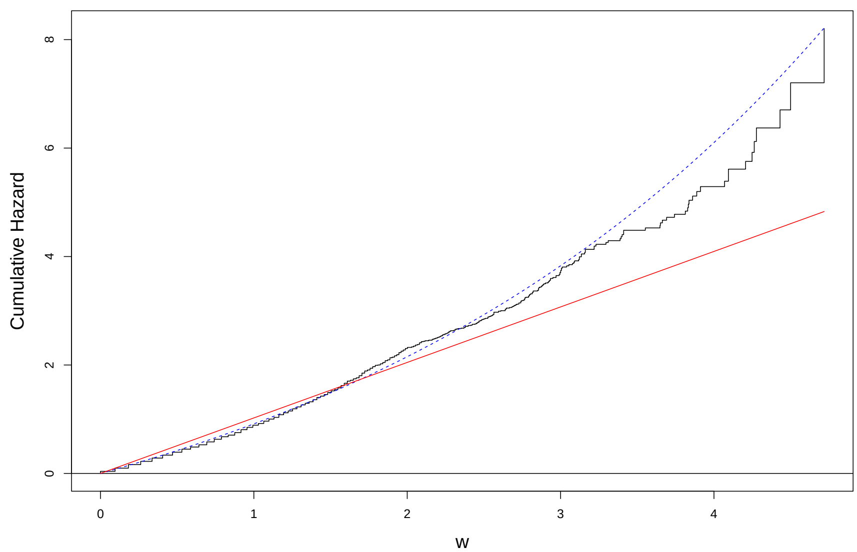
We need a model to match the following facts:

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 Teulings and Toussaint (2023) 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 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.

Figure 2: Empirical and Parametric Hazard Rates, Billionaire Log Wealth Distribution

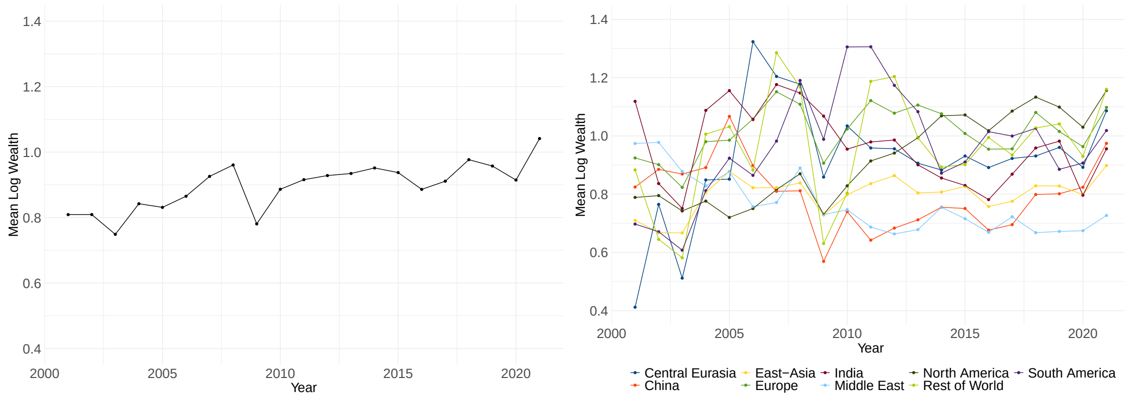


Notes: Figure plots the Kaplan-Meier estimator of the empirical cumulative hazard of the billionaire log wealth distribution, using the 2018 *Forbes List of Billionaires* and pooling all observations. The blue dotted line is a fitted Gompertz hazard, and solid red is an Exponential hazard.

If log wealth is Gompertz, wealth in levels is (truncated-)Weibull. In Teulings and Toussaint (2023), we provide further evidence based on cross-equation restrictions that Weibull provides a better fit to the data than Pareto. We also find this re–jection of Pareto and the better fit of Weibull for city size and, importantly for this paper, firm size. The model in the present paper 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 1. The number of billionaires has quadrupled, with large variation between broad 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 also investigate whether billionaires' wealth increased. Figure 3 shows global and regional trends in mean log wealth, analogous to Figure 1.

Figure 3: Global and Regional Trends in Billionaire Mean Log Wealth



(a) Global (b) Regional Variation

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).

3 The Model

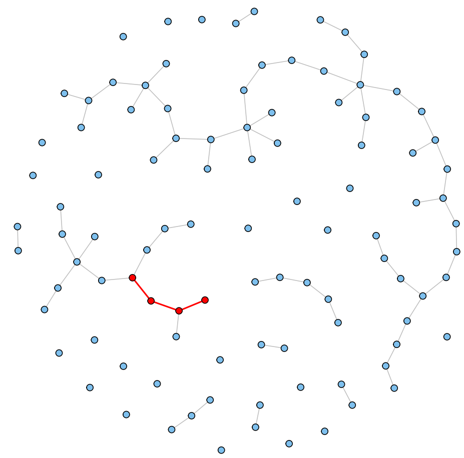
Consider an economy in region j in year t with population $L_{jt} > 0$. This economy consists of a number of Erdős-Rényi networks $\text{ER}[L_{jt}, p_{jt}]$ where L_{jt} is the number of nodes in this network and where $p_{jt} \in (0, 1)$ is probability that there exists a link between two nodes. The expected number of links of a node to other nodes c_{jt} satisfies:

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

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. Our favourite interpretation of these networks is that each network represents the customers of a particular industry. A link represent for example either a shared technology of two nodes which require 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 $\text{ER}[L_{jt}, p_{jt}]$ 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 paper 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 4 for an illustration.[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 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.

If p_{jt} were independent of log population size $\ell_{jt} := \ln L_{jt}$, equation (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 proportional 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}:

$$\ln p_{jt} = \ln p\left(\ell_{jt}\right) = \gamma_0 - \gamma \ell_{jt}, \tag{2}$$

where $\gamma_0 \in \mathbb{R}$ and $\gamma \in (0,1)$ are parameters (throughout the paper, 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:

$$\Pr\left[\underline{z} \geq z\right] = \exp\left(-\frac{e^{p(\ell_{jt})z} - e^{p(\ell_{jt})d(\ell_{jt})}}{p\left(\ell_{jt}\right)} e^{-p(\ell_{jt})h(\ell_{jt})}\right). \tag{4}$$

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

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

$$\Pr\left[\underline{c} = c\right] = \frac{e^{-\mu}\mu^c}{c!} \tag{3}$$

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.

of $\underline{z}$ is drawn from this right tail is:

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

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 realisation of $\underline{z} < d^*\left(\ell_{jt}\right)$, for whom 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-to-income ratio K_t. Individual i's wealth W_{ijt} is therefore equal to:

$$W_{ijt} = e^{z_{ijt}} Y_{jt} K_t \Rightarrow z_{ijt} = w_{ijt} - y_{jt} - k_t. \tag{5}$$

Substitution of equation (5) in equation (4) yields:

$$\Pr\left[\underline{w} \geq w_{ijt}\right] = \exp\left(-\frac{e^{p_{jt}w_{ijt}} - e^{p_{jt}d_{jt}}}{p_{jt}} e^{-p_{jt}h_{jt}}\right) \tag{6}$$

where

$$h_{jt} := h\left(\ell_{jt}\right) + y_{jt} + k_t = \eta_0 + y_{jt} + k_t + \eta\ell_{jt}, \tag{7}$$

$$d_{jt} := d\left(\ell_{jt}\right) + y_{jt} + k_t = \delta_0 + y_{jt} + k_t + \delta\ell_{jt}. \tag{8}$$

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 (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 (6) implies:

$$\frac{\Pr\left[\underline{w} = w_{ijt}\right]}{\Pr\left[\underline{w} \geq w_{ijt}\right]} = e^{p_{jt}\left(w_{ijt} - h_{jt}\right)}, \tag{9}$$

$$\Pr\left[\underline{w} \geq 0\right] = \exp\left(\frac{e^{p_{jt}d_{jt}} - 1}{p_{jt}} e^{-p_{jt}h_{jt}}\right). \tag{10}$$

Equation (9) is the hazard rate of the distribution of top wealth, equation (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 semi-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 $a_{jt} := e^{-p_{jt}h_{jt}}$. The distribution function for the exponential distribution is obtained by taking the limit $p_{jt} \to 0$ while keeping $a_{jt} = e^{-p_{jt}h_{jt}}$ constant:

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

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

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

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 — comes therefore as a surprise, since the common wisdom holds that top wealth is fat tailed, while the empirical evidence in Teulings and Toussaint (2023) 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:

$$\mathrm{E}\left[\underline{W}^{k}\right]=G_{jt}^{k/p_{jt}}\mathrm{e}^{G_{jt}^{-1}}\Gamma\left(1+k/p_{jt},G_{jt}^{-1}\right), \tag{11}$$

$$\mathrm{E}\left[\underline{w}^{k}\right]=p_{jt}^{-k}h\left(G_{jt},k\right), \tag{12}$$

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

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

$$h\left(G,k\right):=k\mathrm{e}^{G^{-1}}\int_{0}^{\infty}q^{k-1}\exp\left(-G^{-1}\mathrm{e}^{-q}\right)\mathrm{d}q,$$

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

$\mathrm{Ei}\left(\cdot\right)$ is the exponential integral, see Teulings and Toussaint (2023).

4 Estimation

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 Teulings and Toussaint (2023) for a more detailed discussion. The density function of the conditional Gompertz distribution for billionaires reads:

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

Hence, the log likelihood reads:

$$N_{jt}^{-1}\ln\mathcal{L}\left(a_{jt}, p_{jt}\right) = \ln a_{jt} + p_{jt}\overline{w_{jt}} - a_{jt}\frac{\overline{e^{p_{jt}w_{jt}}} - 1}{p_{jt}}, \tag{13}$$

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

$$N_{jt}^{-1}\frac{\partial\ln\mathcal{L}\left(a_{jt}, p_{jt}\right)}{\partial a_{jt}} = a_{jt}^{-1} - p_{jt}^{-1}\left(\overline{e^{p_{jt}w_{jt}}} - 1\right) = 0 \Rightarrow \widehat{a}_{jt} = \widehat{p}_{jt}\left(\overline{e^{\widehat{p}_{jt}w_{jt}}} - 1\right)^{-1}, \tag{14}$$

where a hat on top of a parameter denotes its maximum likelihood estimator. Substitution of equation (14) in equation (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

$$N_{jt}^{-1}\frac{\mathrm{d}\ln\mathcal{L}\left(p_{jt}\right)}{\mathrm{d}p_{jt}} = p_{jt}^{-1} - \frac{\overline{w_{jt}}}{\overline{e^{p_{jt}w_{jt}}} - 1} = 0 \Rightarrow \widehat{p}_{jt} = \frac{\overline{e^{\widehat{p}_{jt}w_{jt}}} - 1}{\overline{w_{jt}}}. \tag{15}$$

We first calculate $\widehat{p}_{jt}$ as the solution to equation (15), then use this solution to calculate $\widehat{a}_{jt}$ as the solution to equation (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 (10)

$$\widehat{d}_{jt} = \widehat{p}_{jt}^{-1}\ln\left(1 + \frac{\widehat{p}_{jt}}{\widehat{a}_{jt}}\ln\left(\frac{N_{jt}}{L_{jt}}\right)\right), \tag{16}$$

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 $\gamma_0, \gamma, \eta_0, \eta, \delta_0$, and δ in equations (2), (7) and (8) by means of standard linear techniques like OLS or WLS.

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 $\boldsymbol{p}$, $\boldsymbol{h}$, and $\boldsymbol{d}$. By equation (2), $\ln p_{jt}$ varies negatively with log population size, ℓ_{jt}. By equations (7) and (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 $\boldsymbol{X} := \begin{bmatrix}\boldsymbol{\ell} & \boldsymbol{y} & \boldsymbol{k}\end{bmatrix}$. We might need (time or region) fixed effects as well. Compactly, our empirical model can be written as

$$\boldsymbol{\theta} = \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{\zeta} + \boldsymbol{\epsilon}, \tag{17}$$

where $\boldsymbol{\theta} := (\ln \boldsymbol{p}, \boldsymbol{h}, \boldsymbol{d})$. We potentially allow the set of covariates $\boldsymbol{X}$ to vary with the parameter. $\boldsymbol{\zeta}$ 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). $\boldsymbol{\epsilon}$ is an error term.

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

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

where $\gamma < 0$. The motivation for assuming $\gamma < 0$ comes from equation (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.

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. For most of our exercises, a classification according to origin would make little difference, except for the country-specific predictions at the very end of the paper, in Section 7. There, we discuss the impact of this data choice in more detail.

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 1bn$ 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 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

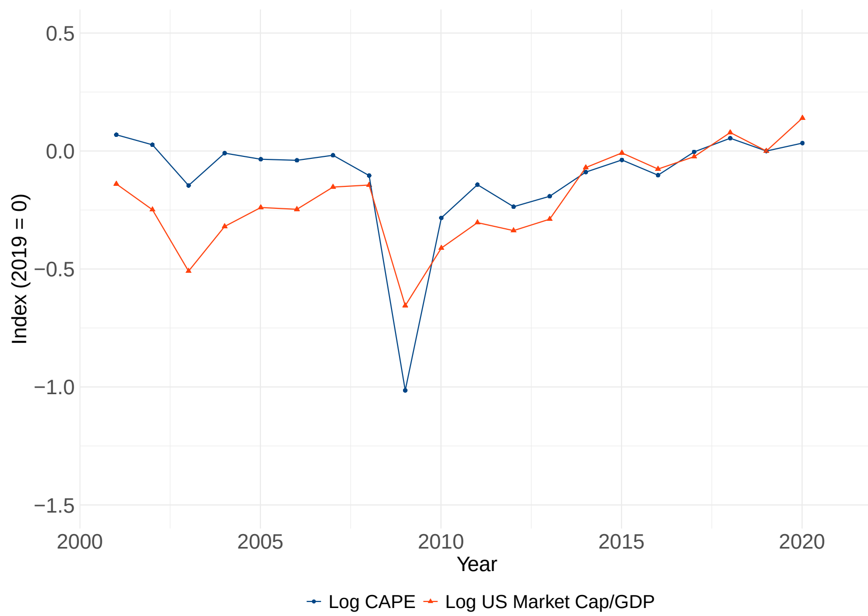
Table 1: Region Classification & Descriptive Statistics

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.

wealth-income ratio is based on the US because of this country has the largest and most developed capital market and it there–fore 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 distri–bution. 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-divi–dend 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 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 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.

Figure 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.

Table 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.

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 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 2. Column (1) begins by testing all variables in $\boldsymbol{x}$. 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 4, we use the fitted values of $\widehat{p}_{jt}$ from Table 2, column (9).

Table 3 reports the results of regressing our fitted $\widehat{h}_{jt}$ on $\boldsymbol{x}$. 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 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

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.

Table 2: Regressions on $\ln \widehat{p}_{jt}$

Dependent Variable:	$\ln \widehat{p}_{jt}$						
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables							
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: Regressions on $\widehat{h}_{jt}$

Dependent Variable:	$\widehat{h}_{jt}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
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.

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.

Table 4: Regressions on $\widehat{d}_{jt}$

Dependent Variable:	$\widehat{d}_{jt}$					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
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.

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, column (9) to calculate $\widehat{d}_{jt}$. The results are reported in Table 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, it is interesting to 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 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.

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 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 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 (10), (11) and (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 (15). However, mean wealth in levels is untargeted, so our model need not necessarily fit this moment well.

Table 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 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 overpre–dicts 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

Table 5: Predictions versus Realizations, 2001–2021 Average

	$\Pr\left[\underline{w} \geq 0\right]$		$E\left[\underline{w}\right]$		$E\left[\underline{W}\right]$		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	0.0516	0.0956	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	0.0335	0.0401	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 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.

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 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 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 $\natural$ 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.

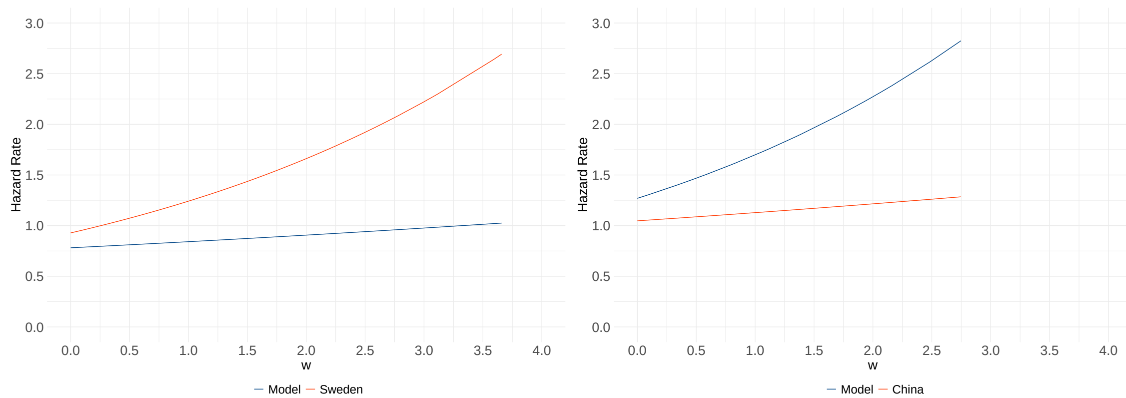
Table 6 reports the results. The change in predicted moments is almost entirely driven by the increase in global GDP per capita, explaining $0.121/0.181 = 66.8\%$ of the increase in the billionaire fraction, 67.2% of the increase in mean log wealth and 58% of the increase in mean wealth. The log market 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 6. Here, we do not re-estimate the equations and fixed effects, instead taking the values from Table 5. Nevertheless, columns (1)–(4) reveal that the predicted values remain very close to the data. The billionaire 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 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.

Figure 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

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 6, suggesting that the denominator effect is second-order.

panel 6a and slightly above 1 for counterfactual Sweden), 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.

8 Interpretation

Our most robust finding is that the semi-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 (2022), 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).

References

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.

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan. 2020. "Thickness and Information in Dynamic Matching Markets." *Journal of Political Economy* 128 (3): 783–815.

Baqaee, David Rezza, and Emmanuel Farhi. 2019. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica* 87 (4): 1155–1203.

Bauluz, Luis, Thomas Blanchet, Clara Martínez-Toledano, and Alice Sodano. 2022. "Global Wealth Dynamics: Understanding the Determinants." *Conference Draft.*

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.

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy* 114 (5): 835–870.

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.

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.

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.

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.

Gomez, Matthieu. 2023. "Decomposing the Growth of Top Wealth Shares." *Econometrica* 91 (3): 979–1024.

Goyal, Sanjeev. 2023. *Networks: An Economics Approach.* MIT Press.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–784.

———. 2022. "The End of Economic Growth? Unintended Consequences of a Declining Population." *American Economic Review* 112 (11): 3489–3527.

Jones, Charles I., and Jihee Kim. 2018. "A Schumpeterian Model of Top Income Inequality." *Journal of Political Economy* 126 (5): 1785–1826.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–1419.

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.

Liu, Ernest, and Aleh Tsyvinski. 2024. "A Dynamic Model of Input-Output Networks." *Review of Economic Studies,* rdae012.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

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.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century.* Harvard University Press.

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.

Quadrini, Vincenzo. 1999. "The Importance of Entrepreneurship for Wealth Concentration and Mobility." *Review of Income and Wealth* 45 (1): 1–19.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

Saez, Emmanuel, and Gabriel Zucman. 2019. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay.* WW Norton & Company.

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.

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.

Vermeulen, Philip. 2016. “Estimating the Top Tail of the Wealth Distribution.” *American Economic Review* 106 (5): 646–50.