Can Taxation Predict US Top-Wealth Share Dynamics?

Gregor Boehl[a], Thomas Fischer[b]

[a] *Institute for Monetary and Financial Stability, Goethe University Frankfurt, Germany*
[b] *Department of Economics and Knut Wicksell Centre for Financial Studies, Lund University, Sweden*

Abstract

The degree of capital gains taxation can retrace the dynamics of wealth inequality in the US since the 1920s. Precisely matching up- and downturns and levels of top shares, it has high overall explanatory power. This result is drawn from an estimated, micro-founded portfolio-choice model where idiosyncratic return risk and disagreement in expectations on asset returns generate an analytically tractable fat-tailed Pareto distribution for the top-wealthy. Wealth concentration is dampened by the degree of capital gains taxation. The model generates good out-of-sample forecasts and is used to predict the future evolution of inequality for different tax regimes.

Keywords: Wealth inequality, US top-wealth shares, capital gains taxation, Fokker-Planck equation

JEL: D31, H23, G11

⋆We are grateful to Christiane Clemens, Alex Clymo, Herbert Dawid, Cees Diks, David Edgerton, Marten Hillebrand, Frederik Lundtofte, Ben Moll, Salvatore Morelli, Tahnee Ooms, Christian Stoltenberg and participants of several conferences and workshops for discussions and helpful comments on the contents of this paper.

Email addresses: `boehl@econ.uni-frankfurt.de` (Gregor Boehl), `thomas.fischer@nek.lu.se (corresponding author)` (Thomas Fischer)

1 Introduction

One of the oldest economic debates concerns the fundamental question of what drives economic inequality. While the topic lay dormant for a long time, it has recently been put back at the forefront of the academic debate by several researchers. They document that after a period of contraction, the concentration of economic resources again increased tremendously in the 1980s (Piketty, 2014; Saez and Zucman, 2016; Lundberg and Waldenström, 2017). This is not only true for individual income but especially for the stock measure of wealth, with levels of dispersion largely surpassing the levels of income inequality. Some economists, including Thomas Piketty, see an inherent higher order at play, suggesting that a trend of increasing wealth concentration is an *inbuilt* property of market economies. The answer we derive in this paper is far more mundane. While in a *laissez-faire* economy wealth inequality would indeed explode, the degree of wealth concentration in modern economies is shaped by the structure of the tax system. As such, economic policy is empowered to directly shape the distribution of wealth.

The key contribution of this paper is to show that a simple model is, when fed with the series of taxes, able to account for a long-run perspective and to reproduce 90 years of US-evidence on wealth concentration. This substantially extends the analysis as compared to other recent papers trying to match the US evidence, which start from the 1970s (or later) and thus solely focus on a sample that features monotonously increasing inequality (Hubmer et al., 2016; Aoki and Nirei, 2017; Cao and Luo, 2017). Our result has three strong implications. First, the distribution of wealth is mainly shaped by the policy maker deciding on the level of taxation with high taxes implying low inequality. Second, the data suggests that this effect is persistent despite the risk of tax evasion and avoidance. Finally, there is a structural component driving wealth concentration which displays little time variation. In fact, increases in wealth inequality can be linked to increased opinion dispersion in financial markets both in the data and in the model.

The source of inequality in our micro-founded portfolio-choice model is idiosyncratic return risk in combination with disagreement in expectations on asset returns. The exogenous variable driving the evolution of top wealth inequality dynamics – as suggested by Piketty (2014) – is the taxation of capital gains. This tax has higher explanatory power for top inequality than other taxes levied on individuals. Since we identify equity trading and returns as central features to explain the behavior of top wealth dynamics, we do not include characteristics which are of importance for the left tail of the distribution (the *poor*).[1] In particular for *the rich* the main source of income is not labor income, but capital income. It follows that the concentration of wealth shapes the concentration of top income and not vice versa.

To our best knowledge, this analysis is the first that can successfully cover such a long time horizon of 90 years. We capture the sharp increase until the Great Depression, the following drop, and the leveling of inequality after the Second World War. Finally, we capture the substantial increase in inequality after the 1980s. Our analysis simultaneously takes into account episodes of falling and rising inequality. The fact that transition speed for increasing and

[1] A non-exhaustive list features entry barriers into financial markets in favor of the wealthy, decreasing relative risk aversion, different marginal propensities to consume, and inter-generational wealth transmission.

decreasing episodes as well as levels of stagnant and transitory inequality are matched while relying on a highly parsimonious model further strengthens our argument.

Formally, we employ a model of the *random-growth* class which can exhibit Pareto tails for the stationary distribution under fairly general conditions (cf. Benhabib et al., 2011). A central feature of this type of model is its ability to cast the law of motion in closed form. This type of model usually comprises transition dynamics that are too slow to match the empirical evidence (Gabaix et al., 2016). The previous literature in this field emphasizes the effect of multiplicative idiosyncratic capital income risk as a driver of inequality, as opposed to labor income risk, which is additive. We further supplement our model with heterogeneity about future asset prospects as a novel mechanism that boosts inequality. As we argue in the following section, expectations disagreement is well documented in surveys and behavioral experiments and hence constitutes a natural candidate driving trading dynamics. This feature enables us to overcome the problem of slow convergence dynamics and enables us to fit the dynamics without having to rely on ad-hoc features such as superstar shocks.

Equipped with a closed form solution, we are able to distinguish periods of steady state stability in the distribution from transition phases, while other papers simply impose the assumption of a steady state at some arbitrary start date. As our model reveals, the wealth distribution was at a steady state only in the 1950s and the early 1960s, while at any other time (slowly) transitioning to either higher or lower levels of concentration. Additionally to the mapping from taxation to levels of concentration we analyse the effects of changes in uncertainty and fluctuations in financial variables. In the presence of risk-free assets higher idiosyncratic wealth risk – established in the literature as a mechanism creating fat tailed wealth distributions – actually lowers wealth concentration since it is internalized in agents' portfolio decisions by lowering their exposure to risky assets. In terms of dynamics, both higher taxes and a larger disagreement increase the convergence speed. From a policy perspective, this implies that the reduction of inequality after a tax hike is faster than the increase following a tax cut of the same magnitude. We further report explosive wealth inequality in the absence of redistributive taxation while higher taxation lowers the degree of steady state inequality. A further advantage of the closed-form solution is that estimation of our model is straightforward. This stands in contrast to much of the existing literature, which uses standard parameter values from the literature for their calibration and relies on eyeball goodness-of-fit tests.

In line with the novel predictions of the model, periods of substantial disagreement in the asset market – such as the Great Depression or the Dot-com boom – are also characterized by rapidly increasing wealth inequality. This also suggests that financial instability and wealth inequality do emerge as twins (Kumhof et al., 2015). Given the accuracy of the estimated model in terms of out-of-sample forecasts, we also make forecasts for alternative tax regimes. Remaining at the level of taxation initiated by the government under President Obama would in fact considerably decrease the degree of wealth concentration, whereas tax cuts back to the pre-Obama level would further increase wealth inequality in the USA. In the latter case, concentration has not yet converged to its new steady state.

The remainder of this work is structured as follows. In Section 2 we provide an overview of the empirical and theoretical literature on wealth inequality with a focus on recent papers that attempt to fit the empirical evidence. In the following section we present the micro-foundations for our formal model and discuss analytically statistical properties in Section 4. Section 5 uses the model to generate forecasts about the future evolution of wealth inequality and presents robustness checks. Finally, Section 6 concludes the paper.

2 Literature

Following the publication of the work of Piketty (2014), the empirical evidence on economic inequality has substantially improved. Cross-country evidence is assembled and made freely available on the *World Income & Wealth Database* maintained by the collaborative effort of many researchers. Despite this effort, much of the focus still lies on the income distribution as opposed to the distribution of wealth and the availability of consistent and long-run measures of wealth inequality is still limited. The database provides long-run data for the United States of America, France and the United Kingdom. The US data – important for our paper – was recently updated by Saez and Zucman (2016).[2] A recent comprehensive survey on the overall empirical evidence regarding wealth inequality is given in Roine and Waldenström (2015).

[2]The data is available at `wid.world`. For the UK and France, the latest data update was conducted by Alvaredo et al. (2018) and Garbinti et al. (2017). Evidence for Sweden is compiled by Lundberg and Waldenström (2017) and is freely available on the authors homepage.

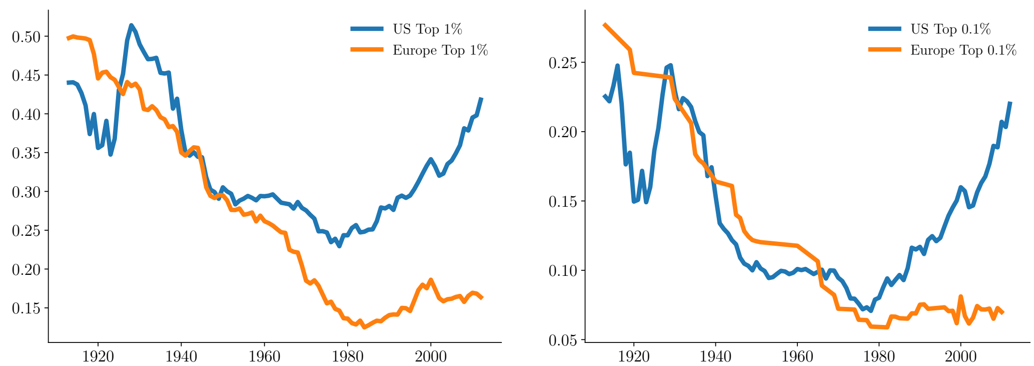

Figure 1: Top wealth shares for the USA and selected European countries

Notes: Wealth inequality decreased until the 1980s. While it leveled in Europe, it increased in the USA afterwards. Left panel includes the UK, France and Sweden, right panel only France and Sweden. Missing values are interpolated. Data source: `wid.world` and Lundberg and Waldenström (2017) for Sweden.

Figure 1 presents collected evidence on the top-shares for the USA and the European Countries – France, Sweden and the UK – in the long run. It is important to point out the deviation of the US economy from the (average) European behavior exhibiting a L-shaped pattern. The dynamics of inequality in the USA are more pronounced featuring both periods of substantial decreases in inequality as well as increases. One peak in inequality is measured at the height of the Great Depression; a period in which European countries exhibited a trend of declining inequality. The 1980s represent another turning point in the history with a substantial increase in top inequality. In comparison this increase is modest in the European countries, but highly pronounced in the USA and in particular emerges for the top wealth holders.

Different theoretical models compete in order to explain the observed degree of inequality. Usually, models in the Bewley-type tradition are considered in order to discuss inequality (Bewley, 1986; Huggett, 1993; Aiyagari, 1994). Yet, it has been formally shown by Benhabib et al. (2011) that these types of models – built around the notion of additive idiosyncratic labor income risk – will fail to generate the fat tails in the wealth distribution and thus match the shares of the top wealth holders. Benhabib et al. (2011) propose a model with multiplicative idiosyncratic capital income risk in order to replicate the current state of wealth inequality in the USA. They follow a rationale laid out as early as Wold and Whittle (1957), building on random growth (hence the term *random growth models*). A tax that addresses these multiplicative shocks on capital returns – i.e. the capital gain tax – is also crucial to understand top dynamics. In fact, the absence of taxation in a standard model with idiosyncratic investment risk implies that wealth inequality does explode (Fernholz and Fernholz, 2014).

Castaneda et al. (2003) propose extreme so-called *superstar* income states to rationalize the measured extreme top shares in a Bewley type economy. Yet, as documented in Saez and Zucman (2016), the saving rate increases for the top wealth holders. Accordingly, the prime share of their income originates from saved wealth rather than labor income (also cf. Piketty, 2014).[3] Hence, at the top tails the impact of labor income substantially reduces. In contrast the distribution of total income and the stock measure of wealth at the top are shaped by the financial decision making process. Hence, in our model we put the emphasis on the portfolio structuring and asset trading.

[3]The reader is especially referred to figures 8.4 and 8.10 in Piketty (2014) for French and US evidence.

While heterogeneous portfolios are often motivated by different degrees of risk aversion and marginal propensities to consume, this argument does not hold for the very rich since they should have a relatively similar portfolio structure. But even if the portfolio *structure* would be identical, given the large supply of similar assets within classes this would not imply all portfolios contain the same assets. However, due to heterogeneous individual expectations about future prospects, agents still hold different positions among asset classes. Greenwood and Shleifer (2014) document a considerable degree of disagreement on future returns in a survey over six data sets on investor expectations of future stock market returns. Further, Pfajfar and Santoro (2008, 2010) provide empirical evidence in support of heterogeneous expectations using survey data on inflation expectations. Additional evidence from the lab has shown that individuals generally do not perform well when forming expectations and these expectations are furthermore largely heterogeneous, which is summarized in e.g. Hommes (2013). Boehl (2018) shows that expectations will be heterogeneous even if a considerable fraction of traders is *super-rational*. Recently the heterogeneous expectations hypothesis has made its way into macroeconomics (Mankiw et al., 2004; Branch, 2004). For that reason we explicitly motivate heterogeneous portfolios by marginal disagreement on future returns.

Given the new data evidence, similar projects have been undertaken. Most prominently, Kaymak and Poschke (2016) use the evidence for the United States from 1960 to the most present date to present a calibrated model in the Bewley tradition. As these models in general fail in matching the top tails of the wealth distribution (Aiyagari, 1994), they require for a modification. Using the modification of Castaneda et al. (2003) and allowing for *superstar* shocks to reproduce high levels of income inequality, Kaymak and Poschke (2016) are able to match the data relatively well. Aiming to identify the contributing factors with a very detailed modeling of the US-tax system (including income, corporate, and estate taxes as well as the pension system), the authors argue that the structure of the taxation and transfer system is key to explaining the evolution of wealth inequality. Hubmer et al. (2016) extend an otherwise standard Bewley-type model with heterogeneous rates of time preference (Krusell and Smith, 1998), Pareto tails in the income distribution and idiosyncratic investment risk (Benhabib et al., 2011). They reproduce wealth inequality dynamics as a result of substantial tax changes, with a data scope starting from the 1970s. While concentration has started to increase in this period, this omits the relatively stable period of the 1960s and the period of decreasing wealth concentration in the 1970s.They conclude that the substantial increase in income inequality, the change in labor share, the gap between the interest rate and the growth rate $r > g$ (Piketty, 2014) all fall short of explaining these dynamics.[4]

A different approach is presented in Aoki and Nirei (2017), featuring a rich model in continuous time. In line with empirical evidence and due to idiosyncratic firm shocks the emerging stationary distribution of firms is given by Zipf's law.[5] The firm's returns translate into income for private households, implying a realistic distribution of both income and wealth for

[4]Piketty argues that wealth inequality will not converge as long as the rate of interest r is larger than the growth rate of labor income g. For a detailed formal and critical discussion see Fischer (2017).

[5]The latter is a Power-law with an exponent $\alpha = 1$.

private households. Combining this framework with a set of tax rates they are able to match both the dynamics and the state of inequality in the USA. Again, the authors focus on the data from the 1970s to the most recent years. Cao and Luo (2017) introduce idiosyncratic return risk into an otherwise standard neoclassical growth model to account for US wealth inequality and also show that the latter is accompanied by increasing capital-output-ratios and decreasing labor shares. In contrast to their work, our paper focuses on the distributional impact of taxation and does not make a statement about the macroeconomic impact of the wealth tax or even its optimal level.[6]

It is interesting to point out that the rich models discussed so far require some specific modeling assumptions in order to match top inequality such as extreme *superstar* earnings or preference heterogeneity in discount rates. These measures map back primarily to the labor earnings process or impose ex-ante heterogeneity. In contrast, our tractable model pinpoints to the role of the financial decision process rather than labor income (negligible for the top shares) in order to understand top wealth inequality. In particular, we abstain from any form of ex-ante inequality that is sometimes brought as an explanation of the witnessed inequality basically imposing it exogenously.

Moreover, our model extends the literature by enlarging the time perspective of the analysis substantially and also covering periods of decreasing inequality and the Great Depression. In the following, we present the structure of our model.

[6]Judd (1985) shows that in standard models the optimal tax on the stock value of wealth is zero. In Bewley-type models this result fails to hold and optimal taxes need to be positive in order to counteract excessive savings (Aiyagari, 1995). Also see Castaneda et al. (2003), Domeij and Heathcote (2004), and Cagetti and DeNardi (2009).

3 Model

We assume an economy with a very large number of individuals indexed by i and that time is discrete.[7] Their only income consists of investment returns and they are free to choose between a risk-free asset paying a constant gross return of R and a continuum of ex-ante identical risky assets of which each pays an idiosyncratic, stochastic dividend $d_{i,t}$ every period t. To maximize their intertemporal consumption over an infinite time horizon the agents accumulate wealth $w_{i,t}$. Hence, each agent i faces the question of which amount $c_{i,t}$ to consume and which amount $x_{i,t} = z_{i,t}w_{i,t}$ of the risky asset to purchase. In this case $z_{i,t}$ relates the demand for risky assets as a share of individual wealth $w_{i,t}$. As pointed out earlier, such a model provides a realistic representation for the behavior of wealthy agents, but – in the absence of features such as borrowing constraints and labor income – naturally falls short in the context of the lower 50% share of wealth holders. Since we aim to explain the dynamics of top-shares such simplification is justifiable because, trivially, the wealth share of the bottom $(1-x)\%$ can be seen as the residual wealth not owned by the top $x\%$.

Assuming log-preferences, the individual problem is then given by

$$\max_{c,z} E_t \sum_{t=0}^{\infty} \beta_t^t \ln c_{i,t}$$

[7] This assumption is in line with the standard literature and simplifies derivation of the model. In order to find the cross-sectional distribution (cf. Section 4) we have to transfer to a continuous time approach. Given the discrete nature of data, in our empirical application (Section 5) we afterwards return to a discrete time setting.

subject to the two constraints

$$w_{i,t} = (R + [d_{i,t} + p_t - Rp_{t-1}]z_{i,t-1})\, s_{i,t-1}(1-\tau)w_{i,t-1}, \tag{1}$$

$$c_{i,t} = (1 - s_{i,t})w_{i,t}. \tag{2}$$

Here we denote by $\beta_t = \beta + \epsilon_t^\beta$ the intertemporal discount rate with an i.i.d. zero-mean preference shock and by $s_{i,t}$ the savings rate. p_t is the price for an asset of the class of risky assets in t and $d_{i,t} = d + \epsilon_{i,t}^d$ its dividend with an idiosyncratic stochastic term $\epsilon_{i,t}^d \sim N(0, \sigma_t^d)$ where $\sigma_t^d = \sigma_d + \epsilon_t^d$ is again subject to an i.i.d. time-varying aggregate shock. Since those assets are ex-ante identical, their price is likewise the same. The value τ_t captures a tax on the stock level of wealth. For the empirical part it is crucial that taxes vary in time. However, for the sake of readability the time index is suppressed in this section. We want to assume that our taxation is a redistributive transfer towards the bottom shares of society. Given that we model the shares of the top wealthy, any positive lump-sum transfers are negligible. The above problem does not directly entail a closed form solution, but can be separated into two stages that both are relatively standard in the literature. Let us first solve the consumption problem.

Levhari and Srinivasan (1969) show that for log-utility, which is a particular case of Constant Relative Risk Aversion (CRRA) preferences, in equilibrium agents consume $1 - \beta_t$ of their wealth at the end of each period, i.e. $s_{i,t} = \beta_t \; \forall i, t$.[8] It is important to point out that this result holds despite the

[8]Levhari and Srinivasan (1969) derive a more general result for CRRA utility ($u(c) = \frac{c^{1-\gamma}}{1-\gamma}$) and i.i.d. returns. Depending on whether the income ($\gamma > 1$) or substitution effect

tax rate. Due to the exact offsetting of income and substitution effects for log-utility the savings rate is not distorted by the tax rate.[9] Note that the assumption of CRRA also explicitly avoids inequality dynamics induced by a different marginal propensity to consume. Thus, the law of motion for each individual's wealth follows

$$w_{i,t} = (1-\tau)\beta_t R^z_{i,t}(z_{i,t-1})w_{t-1}, \tag{3}$$

for which $R^z_{i,t}(z_{i,t-1})$ summarizes the individuals gross return on investment.

For the second stage, in which we solve for the optimal demand for risky asset $x_{i,t}$, let us use Equation (1) to rewrite the maximization problem as

$$\max_z E_t \sum_{k=0}^{\infty} {\beta_{t+k}}^{t+k} \ln\{(1-\beta_{t+k})w_{i,t+k}\} \quad \text{s.t. } w_{i,t} = (1-\tau)\beta_t R^z_{i,t}(z_{i,t-1})w_{i,t-1} \forall t \in \mathbb{N}$$

which is equivalent to

$$\max_z E_t \sum_{k=0}^{\infty} {\beta_{t+k}}^{t+k} \ln\left\{(1-\beta_{t+k})\beta_{t+k} w_{i,t}(1-\tau)^k \prod_{l=0}^{k} R^z_{i,t+l}(z_{i,t+l-1})\right\}.$$

Due to the logarithmic laws the term $\ln\{\prod_{l=0}^{k} R^z_{i,t+l}(z_{t+l-1})\}$ can be separated

$(0 < \gamma < 1)$ dominates, individuals adjust their savings by taking into account the risky savings technology. The special case of perfectly offsetting income and substitution effects $(\gamma = 1)$ assumed here implies that the nature of the stochastic returns has no impact on the savings decision.

[9] The interested reader is also referred to Lansing (1999), who shows that the result of a zero optimal tax rate as proposed in Judd (1985) fails to hold with log-utility. Also see Straub and Werning (2014) for a more recent and general approach.

and is the only part that depends on z_t. Since we can rewrite $\prod_{l=0}^{k} R^z_{i,t+l}(z_{i,t+l-1}) = R^z_{i,t}(z_{i,t-1}) \prod_{l=1}^{k} R^z_{i,t+l}(z_{i,t+l-1})$ this portfolio problem can be well-approximated by mean-variance maximization, as laid out in Pulley (1983). The optimal demand for the risky asset $x_{i,t}$ is then, up to a second order approximation, given by

$$x_{i,t} = z_{i,t} w_{i,t} = \left(E_t[d_{t+1} + p_{t+1}] - Rp_t\right) w_{i,t} / \sigma_t^{d^2}. \tag{4}$$

Note that – identical to the optimal consumption plan – the portfolio structure is independent of the wealth tax. As presented in Stiglitz (1969) for Constant Relative Risk Aversion preferences – of which the assumed log-utility is a special case – wealth taxation does not lead to a restructuring of the portfolio.

Let us assume that return expectations are heterogeneous and each agent's expectation is a draw from the normal distribution around the rational expectation of future returns. Thus, the rational expectation operator E is replaced with a noisy individual expectation operator $\hat{E}_{i,t}$, giving

$$\hat{E}_{i,t}[d_{t+1} + p_{t+1}] = d + E_t p_{t+1} + \epsilon^E_{i,t}, \qquad \epsilon^E_{i,t} \sim N(0, \sigma^E_t),$$

where $\sigma^E_t = \sigma_E + \epsilon^\sigma_t$ as well can be subject to the i.i.d. time-varying aggregated news shock ϵ^σ_t.

Assume furthermore that no single person is rich enough or has an $\epsilon^E_{i,t}$ large enough to influence the price.[10] Market clearing requires $\sum_i x_{i,t} = X_t$,

[10]This is satisfied by the law of large numbers. This assumption enables us to provide analytic results for the law of motion of individual wealth, aggregated wealth, and prices.

with X_t being the total supply of the risky asset. Without loss of generality we can fix supply and normalize $X_t = 1$ to unity for all periods.

Keeping this in mind and aggregating over Equation (1) and (4) yields

$$p_t = R^{-1}(E_t p_{t+1} + d_{t+1} - \sigma_t^{d2} W_t^{-1}) \tag{5}$$

$$W_t = \beta(RW_{t-1} + d_t + p_t - Rp_{t-1}), \tag{6}$$

which is the law of motion for prices and aggregated wealth W_t. Note that for the stationarity of aggregate wealth, redistribution of tax proceedings is required. If this was not the case, in the long run all private wealth would be transferred to the government. If we assume that all variables are detrend, due to the law of large numbers idiosyncratic disturbances level out and aggregate wealth $W_t = W$ is constant in the absence of aggregate shocks. The values of prices and aggregated wealth thus reflect the detrend steady growth path.[11] Then, we can also normalize the price to unity without explicitly accounting for market clearing. The steady state versions of (5) and (6) are

$$\sigma_d^2/W = d + 1 - R \tag{7}$$

$$W(\beta^{-1} - R) = d + 1 - R. \tag{8}$$

[11] This assumption implies that growth in aggregate wealth can be attributed to an exogenous growth rate which does not distort distributional properties.

This implies that, given the normalization of prices,

$$W = \frac{\sigma_d}{\left(\beta^{-1} - R\right)^{0.5}} \tag{9}$$

$$d + 1 - R = \sigma_d \left(\beta^{-1} - R\right)^{0.5}. \tag{10}$$

Plugging Equation (4) into Equation (1), integrating individual forecast errors and setting prices to the steady state yields

$$w_{i,t} = \beta_t \left\{R + \left(d + 1 + \epsilon^d_{i,t} - R\right)\left(d + \epsilon^E_{i,t} + 1 - R\right)\sigma_d^{-2}\right\}(1-\tau)w_{i,t-1}. \tag{11}$$

which together with Equation (10) and some algebra can be written as the law of motion (LOM) for individual wealth

$$w_{i,t} = \beta_t \left\{\beta_t^{-1} + \left(\beta_t^{-1} - R\right)^{0.5}(\epsilon^d_{i,t} + \epsilon^E_{i,t})/\sigma^d_t + \epsilon^d_{i,t}\epsilon^E_{i,t}{\sigma^d_t}^{-2}\right\}(1-\tau)w_{i,t-1}.$$

We use annual data, so let $\beta = 0.95$. For realistic values of the mean real interest rate R we have $\left(\beta^{-1} - R\right)^{0.5} \approx 0$ to be negligibly small.[12] After defining $\gamma_t \equiv \beta_t \frac{\sigma^E_t}{\sigma^d_t}$ and $\varepsilon_{i,t} \equiv \epsilon^d_{i,t}\epsilon^E_{i,t}$ to be the product of two independent random variables that follow a standard normal distribution, the final law of

[12] A positive demand for the risky assets requires $R < 1 + d$. Stationarity of aggregate wealth further demands for $R < \beta^{-1} < 1 + d$ i.e., $\left(\beta^{-1} - R\right)^{0.5} > 0$ but small. The variance of $\epsilon^E_{i,t}$ is already of small magnitude. Rewriting $\epsilon^d_{i,t}$ in terms of a standard normal reveals that the term is negligible.

motion can be further simplified to

$$w_{i,t} = (1 + \gamma_t \varepsilon_{i,t})(1 - \tau_t) w_{i,t-1},$$

γ_t now contains the aggregate shocks ϵ_t^β, ϵ_t^d and ϵ_t^σ and the expected value γ remains as the only free parameter of our model.

4 Representation in Closed Form

This section aims to enrich our understanding of the process that generates the wealth distribution by finding a closed form solution for the stationary distribution as well as for the transition dynamics. In order to do so, we have to overcome some technical obstacles. For better readability we will omit the aggregate shocks until the end of the section as they do not have an impact on the shape of the distribution.

4.1 Cross-sectional distribution

The portfolio returns, a product of two standard normal variables, follow a so-called *product-normal distribution*. To obtain a closed form solution, we have to transfer this distribution to another distribution that can be handled analytically.

Proposition 1. *The first three moments of the product normal distribution and the Laplace distribution with shape parameter of* $\lambda = \sqrt{0.5}$ *are equal.*

Proof. See Online Appendix B. □

The Laplace distribution is very handy in our context for identifying a closed-form solution. The individual law of motion (LOM) has to be rewritten in continuous time in order to solve the Fokker-Planck equations which allows us to identify the cross-sectional distribution in terms of the free parameters γ and τ. It then reads as

$$dw_i = -\tau w_i dt + (1-\tau)\gamma w_i dNP, \tag{12}$$

for which NP is the noise following the product-normal distribution. In order to retrieve a closed-form solution we transform this to the Laplace distribution using the scaling factor λ, which we just introduced. The equation thus reads

$$dw_i = -\frac{\tau}{\lambda} w_i dt + \frac{1}{\lambda}(1-\tau)\gamma w_i dL, \tag{13}$$

for which L signifies Laplace distributed noise.

Proposition 2. *Using Itô's lemma as a second-order approximation and solving the Fokker-Planck equation, the right tails of the cross-sectional distribution (the top wealth holders) are described by a Pareto distribution with a tail parameter α,*

$$\alpha = 1 + \frac{\sqrt{2}\tau}{\gamma^2(1-\tau)^2}. \tag{14}$$

Proof. See Appendix A.

□

It is important to acknowledge the necessary conditions for this result to emerge. It requires both (i) mean reversion ($\mu > 0$) and (ii) a positive non-zero reflecting barrier ($\hat{w}_{min} > 0$). Note that we do not model the latter explicitly. Yet, one could consider that the overall proceedings of the wealth tax are redistributed to all individuals in an equivalent lump-sum manner. For a given tax rate τ and a stationary average wealth $\bar{w}$ the latter would amount to $\tau\bar{w}$. The two assumptions also have an important economic implication. Mean reversion is achieved by a positive capital tax rate that counteracts the multiplicative stochastic noise of the capital gains. The second condition also ensures that the capital tax is not a net loss for the private households. It moreover guarantees overall stationarity of private wealth.

In fact, the complete distribution is characterized by the single value α. Thus, other measures regarding inequality can be derived starting from this assumption.

Proposition 3. *The stationary ($t \to \infty$) share $s^x(\tau, \infty)$ of the top x (e.g. the top 1% implying $x = 0.01$) wealth holders is given by*

$$s^x(\tau, \infty) = x^{1-1/\alpha}, \tag{15}$$

for which α, as above, is implicitly a function of taxes τ and γ.

Proof. The result is well known and can be derived by computing the closed form value of the Lorenz curve given by $L(F) = 1 - (1-F)^{1-1/\alpha}$ and then calculating $s^x = 1 - L(1-x)$. □

The same rationale can also be used to derive a closed form expression for the Gini coefficient. In general a high tail coefficient α is accompanied by low inequality.[13] This very neat result has some strong implications for the

[13] The closed-form value for the Gini coefficient is given by $Gini(w) = \frac{1}{2\alpha-1}$ and decreasing with α for the realistic case of $\alpha > 1$. Note that then it also holds that $\frac{\partial}{\partial\alpha} s^x(\tau, \infty) < 0$.

asymptotic behavior. First of all, without taxation $\tau = 0$ the tail-coefficient is $\alpha = 1$, frequently referred to as Zipf's law. In fact, the Gini coefficient then takes the value of $Gini(w) = 1$ and $s_x(\tau) = 1$ for all $x \in (0, 1]$, implying total inequality.[14] Thus, in a laissez-faire economy without government intervention, there is no finite level of inequality. In general, inequality increases (α decreases) with γ while decreasing with taxation τ. For the extreme case of $\tau \to 1$ - which can be thought of as a completely egalitarian society - we would have $\alpha \to \infty$, and thus have a Lorenz-curve identical to the 45-degree line and thus no inequality at all.

4.2 Convergence dynamics

We can also make a statement about the convergence speed.

Proposition 4. *The convergence of the Laplace-transformed pdf ($\mathcal{L}f(\hat{w}, t) = F(s, t)$) is given by*

$$F(s, t) - F(s, \infty) \sim \exp(-\phi t), \tag{16}$$

with an average convergence speed of

$$\phi = (0.5\gamma(1 - \tau)\alpha)^2 . \tag{17}$$

For a one period gap it is given by

$$F(s, t) = F(s, t - 1) \exp(-\phi) + F(s, \infty)(1 - \exp(-\phi)), \tag{18}$$

[14] Or put differently, the overall process is a unit-root process which not only lacks a finite variance, but is also characterized by an exploding mean. Thus, in the long run the Lorenz curve is a flat line at zero implying maximum inequality.

Proof. See Online Appendix C. □

This implies a half-life of $t_{0.5} = \frac{\ln(2)}{\phi}$. In fact the effective taxation τ not only decreases steady-state inequality, but also increases the speed of convergence to the latter. This also means that there is an asymmetry in the convergence. The increase of inequality for low taxes is slower than the decrease after high tax rates. Thus, the implication for the policy maker is that it is faster to come down to lower inequality rather than to increase the level of inequality. Finally, we use this general result in order to make an approximate statement about the evolution of top-shares which are the focus of the recent empirical literature and thus also take in a central position in this paper.

4.3 Estimation model

Note that so far we ignored the time dimension. Let us assume that the value of γ is constant in time. Yet, due to policy changes the tax rate τ_t is varying in time. Thus, not only the stationary level of inequality as modeled by the Pareto-tail α_t varies in time t, but so does the convergence speed ϕ_t.

Proposition 5. *Ignoring aggregated shocks, the top-shares approximately evolve according to an autoregressive process of first-order with*

$$s_t^x = \rho_t s_{t-1}^x + (1-\rho_t) s^x(\tau_t, \infty), \tag{19}$$

and $\rho_t = \exp(-\phi_t)$ *for the average convergence speed* $\phi_t = \phi(\tau_t)$ *as defined in Equation* (17)*.*

Hence, the share owned by the fraction x of the population is a linear combination of the last period's share and the share of the top x of the stationary distribution given the tax rate τ_t at each time t. Let us now reintroduce the aggregated shocks, which are important for the estimation. The cross-sectional shocks σ_i^E and σ_i^d – indicated by the subscript i – drive the cross-sectional distribution in the first place and are represented through the transformations above. Aggregated shocks affect *the behavior* of all agents equally and hence can introduce aggregate temporary fluctuations to our law of motion for top shares. We summarize these in a composite shock term ε_t^s.

Proposition 6. *As a first-order approximation around the stochastic steady state and using the central limit theorem, the law of motion including aggregate time-varying shocks can be summarized by*

$$s_t^x = \rho_t s_{t-1}^x + (1-\rho_t) s^x(\tau_t, \infty) + \varepsilon_t^s \tag{20}$$

$$\varepsilon_t^s \sim \mathcal{N}(0, \sigma_s). \tag{21}$$

Proof. Let us write out that $s_t^x(\gamma(\epsilon_t^\beta, \epsilon_t^d, \epsilon_t^\sigma), \tau_t))$ is a function of the three aggregate unobserved i.i.d. shocks. These shocks operate on the idiosyncratic risk ϵ^d, the standard deviation of disagreement ϵ^σ, and the rate of time preference β. Using the multivariate Taylor approximation around the expected value where all shocks are zero yields

$$s_t^x(\gamma(\epsilon_t^\beta, \epsilon_t^d, \epsilon_t^E), \tau_t) \approx s_t^x(\gamma, \tau_t) + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^\beta}\epsilon_t^\beta + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^d}\epsilon_t^d + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^E}\epsilon_t^E$$

for which

$$\varepsilon_t^s := \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^\beta}\epsilon_t^\beta + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^d}\epsilon_t^d + \frac{\partial s_t^x}{\partial \gamma}\frac{\partial \gamma}{\partial \epsilon_t^E}\epsilon_t^E$$

is the sum of zero-mean i.i.d. random variables each multiplied by a constant. Applying the central limit theorem this is approximately normally distributed and the result in the proposition follows.

□

4.4 Comparative statics

So far it was assumed that there was a pure substance tax on the stock level of wealth, which is not in place in the US. Yet, the stock level of wealth is subject to other more subtle forms of taxation. In particular, for the case of the USA – our empirical application in the next section – net capital gains are taxed. Of course taxes are only levied on positive measures – i.e. capital gains – and not losses. We thus have to translate between the measures.

Proposition 7. *The gross-wealth tax τ given a capital gains tax θ_r can be approximated by finding a τ such that the expected value of after tax returns from a capital gains tax and after tax returns of a gross-wealth tax are equal. Gross-wealth taxes are then given by*

$$\tau = \frac{1}{2}\gamma\lambda\theta_r. \tag{22}$$

Proof. Dropping time subscripts for taxes, the after-tax returns given the capital income tax θ_r are

$$\bar{R}_{\theta_r} = 1 + \begin{cases} (1-\theta_r)\gamma\varepsilon_{i,t} & \text{if } \varepsilon_{i,t} > 0 \\ \gamma\varepsilon_{i,t} & \text{if } \varepsilon_{i,t} \leq 0. \end{cases}$$

To use the LOM in Equation (19) we approximate τ given θ_r by finding a τ such that the expected value of $\bar{R}_\tau$ equals the expected value of $\bar{R}_{\theta_r}$. Then, given that $\varepsilon_{i,t}$ approximately follows a Laplace distribution with scale $\lambda = \sqrt{0.5}$, the expected value $E[\varepsilon_{i,t}|\varepsilon_{i,t} \leq 0]$ is the mean of an exponential distribution with inverse scale λ, which is again λ. Then

$$
\begin{aligned}
E\bar{R}_\tau &= E\bar{R}_{\theta_r} \\
E\left\{(1-\tau)\left(1+\gamma\varepsilon_{i,t}\right)\right\} &= 1 + \gamma P(\varepsilon_{i,t} \leq 0)E[\varepsilon_{i,t}|\varepsilon_{i,t} \leq 0] + (1-\theta_r)\gamma P(\varepsilon_{i,t} > 0)E[\varepsilon_{i,t}|\varepsilon_{i,t} > 0] \\
1-\tau &= 1 - 0.5\gamma\lambda + 0.5(1-\theta_r)\gamma\lambda \\
\tau &= \frac{1}{2}\theta_r\gamma\lambda,
\end{aligned}
$$

where $P(\varepsilon_{i,t} > 0)$ denotes the probability that $\varepsilon_{i,t}$ is positive. □

Finally, our model is fully specified, allowing us to conduct some comparative statics. To obtain some intuition, let us plug Equation (22) into the closed-form solution from Equation (19), and for simplicity take $(1-\tau) \approx 1$. Then

$$
s_t^x = \exp(-\phi_t)s_{t-1}^x + (1-\exp(-\phi_t))x^{1-1/\alpha_t},
$$

with

$$
\begin{aligned}
\alpha_t(\theta_{r,t}) \approx& 1 + \frac{0.5\theta_{r,t}}{\gamma}, \\
\phi_t(\theta_{r,t}) \approx& \frac{1}{4}\gamma^2 + \frac{1}{4}\theta_{r,t}\gamma + \frac{1}{16}\theta_{r,t}^2,
\end{aligned}
$$

for which follows that *if the system is at the steady state*

$$\frac{\partial s_t^x}{\partial \gamma} > 0 > \frac{\partial s_t^x}{\partial \theta_{r,t}} \qquad \text{and} \qquad \frac{\partial |\Delta s_t^x|}{\partial \theta_{r,t}}, \frac{\partial |\Delta s_t^x|}{\partial \gamma} > 0.$$

The weight on the most recent value s_{t-1}^x decreases in the transition speed, which depends positively on taxes $\theta_{r,t}$ and dispersion γ. This means that in terms of inequality dynamics, an increase in dispersion γ is a complement to an increase in taxes and will speed up dynamics. However, in terms of inequality levels these two have opposing effects: an increase in taxes $\theta_{r,t}$ decreases the stationary level of inequality, while a higher value of γ will increase it.

It is also insightful to keep in mind the definition of $\gamma = \beta \frac{\sigma^E}{\sigma^d}$ to decompose the effects. We have a higher degree of wealth inequality (as measured by top shares) for high disagreement σ^E. Meanwhile – and somewhat surprisingly – wealth concentration decreases for high idiosyncratic risk σ^d. The latter is due to the fact that individuals incorporate risk into their portfolio decision by increasing the share of risk-free assets. Finally, the inequality increases with the discount factor β which for the assumed case of log-utility is equal to the savings rate. Thus, high savings are accompanied by higher degrees of wealth inequality. In terms of dynamics, both the higher savings rates and higher expectation disagreement increases the dynamics, whereas higher idiosyncratic risk slows down the dynamics.

5 A Quantitative Exploration for the USA

In this section we make use of Maximum Likelihood Estimation (MLE) to fit the model specified in Equation (20) to the empirical top-wealth shares of the US economy, while feeding-in the series of taxes as the sole input. We perform out-of-sample forecasts to test the model against the data. In the last step we forecast the concentration of wealth given different scenarios of taxation.

For the top-wealth data we rely on the recent study of Saez and Zucman (2016) which is made available at `wid.world` employing income tax data and using a capitalization technique to translate this to the stock measure of wealth. For the data on capital gains top-tax rates we use the estimates conducted in Sialm (2009).[15] The First World War and its (financial) aftermath was accompanied by a brief but very substantial raise from an extremely low top tax rates of 7% up to values as high as 73%. Since the extreme circumstances surrounding 1921 complicate the analysis, we start our investigation after World War I from 1922 in which tax rates reduced to a stable level of 12.5%. The last observation on the top 1% is obtainable for 2012. The tax series is transformed as suggested in Proposition 7.

In the United States, individuals generally pay income tax on the net of their capital gains. There are a considerable number of exemptions, depending on investment duration, net-worth, and general status. The series in use here represents the maximum tax rate on returns with positive net capital gains which prevail for the top wealthiest individuals. This has the great advantage that it is a tax explicitly and only on capital gains, which are the focus of our model. Yet, we note that reducing a complicated system of progressive personal taxation to just one number bears the risk of misalign-

[15]The original paper provides estimates until 2006. We are very grateful to Salvatore Morelli for providing us with his updated estimates for the more recent years.

ment. It is further worth noticing that the specific tax rates (for the US case especially the income and dividend taxes) are highly positively correlated. From the given set we consider the capital gain tax to be the one with the highest explanatory power. In contrast to a bequest tax only collected at an individual's death[16], it is regularly imposed. As opposed to the tax on (labor) income, which is only constitutes a minor share overall income for top wealthy individuals, the capital gain tax is directly related to the stock measure of wealth. For the US case taxes on dividends are higher or equal than capital gains taxes (Sialm, 2009), but these higher taxes can be avoided by paying out business income associated with stocks in new shares rather than in dividends. Thus, general capital gains taxes are expected to have the highest explanatory power as they form a lower bound.

The discussion of the causality between wealth concentration and taxation naturally involves the question of tax evasion. An increase in taxation could in particular motivate wealthy households to move their wealth offshore, leading to a lower aggregated stock of capital but also to lower concentration of wealth. In our dataset an estimate of offshore wealth is already incorporated in order to capture tax evasion.[17] Note that if tax evasion would increase with the level of taxation, this would imply a decrease in the wealth-income ratio when taxes are high, which cannot be confirmed by the data from Piketty and Zucman (2014).

Let us first estimate the two model parameters for the complete sample

[16]Inheritance taxes are important for intergenerational mobility. A discussion on inheritance taxes can be found e.g. in Benhabib et al. (2014).

[17]The topic is also treated in detail in Alstadsaeter et al. (2017).

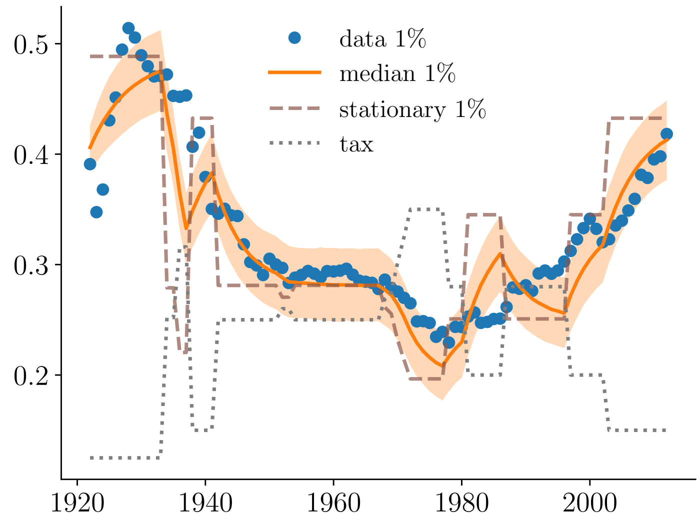

Figure 2: The model-implied mean of the top-1% together with the empirical data (dots) and the tax series.
Notes: Input and output data from year 1922 – 2012. The shaded area present 95% confidence bands. Estimated values: $\hat{\gamma} = 0.346$; Std. dev. of shocks $\hat{\sigma}_s = 1.08\%$.

length and focus on the share of the wealthiest 1% of the USA. In Figure 2, we show the performance of the estimated model by initializing a set of 10,000 batch runs with the first observation of the data and only feeding in the tax series. The dashed line stands for the analytic result of the stationary cross-sectional distribution implied by the tax rate $\theta_{r,t}$ at time t, i.e. the share to which the distribution will converge if time goes to infinity. For this reason the shares of the stationary distribution *jump* with each change in the tax rate. The median of the simulation follows more slowly and, in line with the data, slowly converges towards the stationary value.

The corresponding estimate of $\gamma = \beta \frac{\sigma^E}{\sigma^D}$ is $\hat{\gamma} = 0.346$. The finance literature usually uses a value of σ^d in the range from 0.08 to 0.3 (Campbell and Viceira, 2002) on a quarterly basis. Greenwood and Shleifer (2014) estimate a quarterly standard deviation of return disagreement ranging between 1% and 4%. The annual discount rate is assumed to be 0.95. Combining and annualizing these values suggests that our estimate for γ lies in the reasonable range from about 0.02 to 0.45 and implies furthermore that the standard deviation of agents' forecasting errors is yet quite small compared to the standard deviations of returns. Thus, the estimated value is in line with relatively low disagreement variance as compared to return variance. We also report 95% confidence bands.

Our long-run time series captures very different episodes in the US history. Despite its very parsimonious nature the model matches both the level and dynamics of wealth inequality for period of 90 years very well. Starting after the First World War the inequality substantially increased to peak shortly before the outburst of the Great Depression. Eventually, our model slightly understates the peak level of inequality by assuming a constant value of γ. The speculative boom leading up to the Great Depression suggests a (temporary) increase in disagreement σ^E implying a higher value of γ for that period, going along with higher inequality. After the Great Depression top tax rates markedly increased to a new level after the end of the Second World War which is accompanied by a decline in wealth concentration, as predicted by the model. Following an increase of capital gains taxes starting in the late 1960s, inequality decreases until the Reagan period, in which taxes return to the previous level.

In the period of deregulation in the late 1980s taxes return to their relatively low level from the 1950s, making inequality catch up to its postwar-level. Finally, the substantial tax decreases in the late 1990s and the early 2000s initiate a sizable increase in wealth inequality. Similar to the Great Depression the model underestimates the level of inequality prevailing in the Dot-com asset boom (early 2000s) by (implicitly) assuming a constant value for γ. Like the Great Depression the Dot-com asset boom was accompanied by a major dispersion in investor opinion σ^E implying a higher (temporary) level of inequality.

The presence of a closed-form solution also allows to disentangle the stationary distribution (as displayed by the dashed line in Figure 2) and the slow time convergence. As shown it was only in the years of both stable taxation and inequality from the 1950s to the early 1960s that the stationary distribution coincides with the actual level of wealth concentration. This is of notable importance since these periods are omitted in the work of others discussed earlier (Hubmer et al., 2016; Aoki and Nirei, 2017; Cao and Luo, 2017). Thus, a stationary wealth distribution is imposed in the starting year which is a highly problematic assumption for the years ranging between 1970 and 1990. This then might also lead to misleading predictions about the future evolution of wealth inequality.

As stated in Proposition 6 the model requires that the error terms of the residuals exhibit the i.i.d. property. Figure 3 plots the residuals at each point of time. With a formal Ljung-Box test the hypothesis that the autocorrelation parameter of the residuals is zero, cannot be rejected. Eyeballing however suggests that residuals might exhibit heteroscedasticity with a substantially larger variance in the period until the end of the Second World

War as compared to the following periods as observable in Figure 3. Thus, a formal Anderson-Darling test for normality is rejected for the entire sample at a significance level of 1%. Meanwhile, the same test for the subperiods until 1945 respectively afterwards cannot reject the normality of the subsamples. The higher variance of model residuals in the early periods can be due to several reasons. First, the quality of very old data is generally inferior. Thus, the measure of wealth inequality in the early periods is subject to higher measurement error. Secondly, the change in variance might be due to structural changes exogenous to our model, given that these time periods are characterized by major (temporary) disruptions such as the Great Depression and the Second World War (also posing a measuring challenge). Lastly, it may be that in the early years of our sample different structural reasons had impact on the dynamics of wealth concentration which over time lost relevance. Yet, overall the estimated standard deviation of the error term, $\hat{\sigma}_s = 1.08\%$, is relatively small given the parsimonious nature of our model.

So far the model was estimated and evaluated for the complete sample period. To further assess the robustness of the model we conduct out-of-sample forecasts for the top 1% share that are shown in Figures 4, 5 and 6. We use the data until a given end point as the estimation period to estimate γ and σ_s and then run a batch of 10,000 simulations starting from this end date, while again feeding in the respective time series of taxes. Thus, in Figure 4 the model uses 18 observations (until 1940) in order to predict the remaining 72 observations (i.e. 20% of the overall sample). In line with the higher variance of residuals in the earlier years is the increase in the estimated standard error of shocks of $\hat{\sigma}_s = 2.67\%$. Again the model captures well both

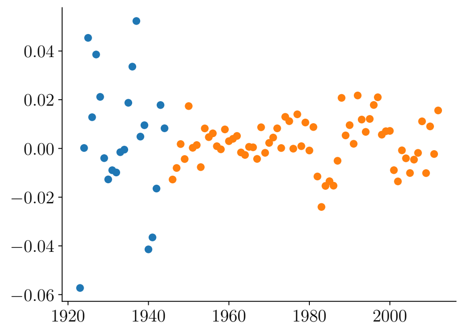

Figure 3: Model residuals

Notes: Formal tests reject the hypothesis of autocorrelation. There is substantial heteroscedasticity in the error terms. Taken separately, normality holds for both the pre and post World War II sub-samples.

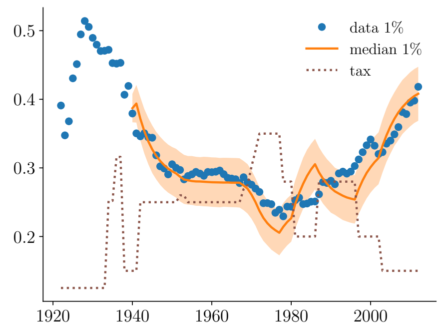

Figure 4: Out-of-sample testing for the time series from years 1940 – 2012.

Notes: The model estimation uses the periods 1922 – 1940 as sample, where the remaining periods are out-of-sample forecasts. Estimation value: $\hat{\gamma} = 0.358$. Std. dev. of shocks $\hat{\sigma}_s = 2.67\%$.

the level and the dynamics of wealth inequality.

As a further robustness check we change the composition of in- and out-of-sample data and estimate the parameters using the empirical evidence until 1975 (cf. Figure 5) respectively until 1990 (cf. Figure 6). As expected, averaging over a sample period that includes the later years reduces the estimated standard error of shocks $\hat{\sigma}_s$.

Interestingly, starting the out-of-sample period in 1975 (cf. Figure 5) gives a lower value of $\hat{\gamma}$ as compared to the one with the full sample period (cf. Figure 2) and especially the one estimated with the period until 1940 (cf. Figure 4). By excluding the increase in the recent years (as compared to the full sample, Figure 2) and also diminishing the impact of the volatile evolution in the earlier periods (captured in Figure 4), the evolution of inequality is rather tranquil. This is captured by a low value of $\hat{\gamma}$ also implying smoother dynamics. This is also important because – if anything – the model slightly overstates the dynamics. The latter fact is actually encouraging since – as shown in Gabaix et al. (2016) – models trying to capturing the dynamics of inequality (especially for income) usually massively understate its dynamics. In the logic of our model higher values for γ leading to faster dynamics can be rationalized by higher general disagreement σ^E which – suggested by the model was especially prevailing in the turbulent times until the end of the Second World War and after the 2000s with the Dot-com boom and the Great Financial Crisis.

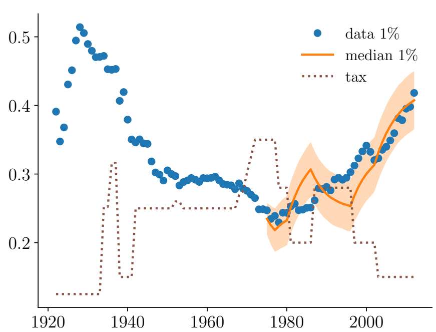

Figure 5: Out-of-sample testing for the time series from years 1975 – 2012.
Notes: The model estimation uses the periods 1922 – 1975 as sample, where the remaining periods are out-of-sample forecasts. Estimation value: $\hat{\gamma} = 0.344$. Std. dev. of shocks $\hat{\sigma}_s = 1.12\%$.

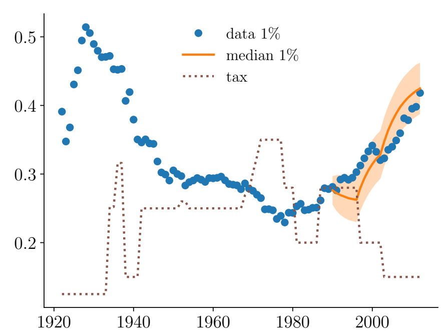

Figure 6: Out-of-sample testing for the time series from years 1990 – 2012.
Notes: The model estimation uses the periods 1922 – 1990 as sample, where the remaining periods are out-of-sample forecasts. Estimation value: $\hat{\gamma} = 0.362$. Std. dev. of shocks $\hat{\sigma}_s = 1.1\%$.

So far we considered – following the protests of the so-called 99% movement – the almost proverbial top 1%. Saez and Zucman (2016) also provide more narrow share of the top 0.1% and 0.01%. Although the upper tail of wealth is often approximated to follow a Pareto distribution, as discussed in Blanchet et al. (2017), this does not match the actual distribution precisely. As presented in Figure 8 the local Pareto coefficient is larger the more we go into the tails of the distribution i.e. for lower values of x.[18]

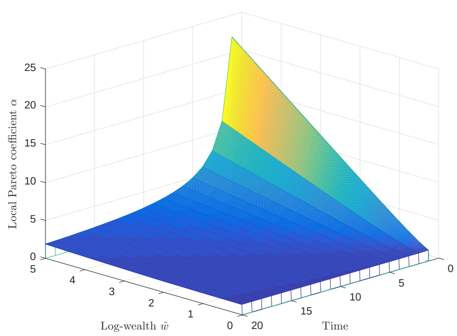

Figure 7: Local Pareto coefficient for log-wealth $\hat{w}$ in time for the model.
Notes: During transition the coefficient increases for the top shares, after convergence the coefficient is independent of the level of wealth.

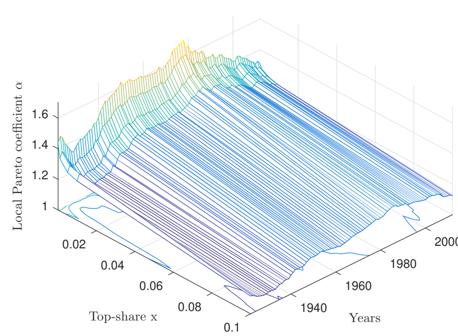

Figure 8: Local Pareto coefficient for top-shares x in time for the US evidence.
Notes: The coefficient does not only vary across time, but also persistently increases towards the tail.

A true Pareto distribution, however, is scale-free and thus exhibits the identical Pareto coefficient regardless of the level of x. Thus, when estimating the model for a more narrow share of top inequality we would get a lower value of γ implying a larger value of the (local) Pareto coefficient $\alpha(x)$ (cf. Table

[18]We compute the local Pareto coefficient using the equation $s^x = x^{1-1/\alpha} \leftrightarrow \alpha = 1/(1-\ln(s^x - x))$. Similar evidence, is reported in Saez and Stantcheva (2018) who show that the local Pareto coefficient of capital income increases in the tails. There is still disagreement on whether wealth follows a true Pareto distribution or not (Clauset et al., 2009; Chan et al., 2017; Vermeulen, 2016). To summarize the debate briefly, the Pareto distribution is a good approximation which has some pleasant analytic properties, yet fails to match the data precisely with the limitations pointed out above.

1). Formally, we could capture this by assuming ex-ante heterogeneity of individuals from different shares of the wealth quantiles having different $\gamma(x)$. An interpretation of this would be that the top rich have riskier portfolios (i.e. higher σ^D)[19] or lower expectation disagreement (low σ^E) due to better financial knowledge or advisory.[20] Gabaix et al. (2016) propose this *type dependency* or *superstar effects* as a solution to the puzzle of failing to match the dynamics of wealth inequality. In this case, we could maintain our simple regression equation and also fit more narrow shares. However, this is not in line with the assumption of ex-ante identical agents (cf. Section 3).

Table 1: Parameter estimates for the filter using different series of top-shares from 1945 as input.

Top-shares	$\hat{\gamma}$	$\hat{\sigma}_s$
1	0.358	2.67%
0.1	0.28	0.6%
0.01	0.253	0.4%

Notes: The estimated standard errors are also a measure of the goodness of fit. In line with Figure 8 the value of γ decreases for higher quantiles.

In fact, we do not have to make an assumption of ex-ante heterogeneity as the behavior of increasing Pareto coefficients in the tail is coherent with the overall model. In Online Appendix D we discuss the (non-stationary) closed-form solution of the Fokker-Planck equation. It turns out that in the short run the distribution resembles a log-normal distribution with increasing

[19]Note this mechanism is also at play in Kasa and Lei (2018). Combined with higher savings rates of the rich, the growth rate of wealth increases with the level of wealth. This *scale dependency* (as opposed to ex-ante type dependency) is micro-founded by the presence of Knightian uncertainty.

[20]A third option would be the assumption of heterogeneous thriftiness (time preference β) as e.g. entertained in Krusell and Smith (1998) in order to match top wealth inequality.

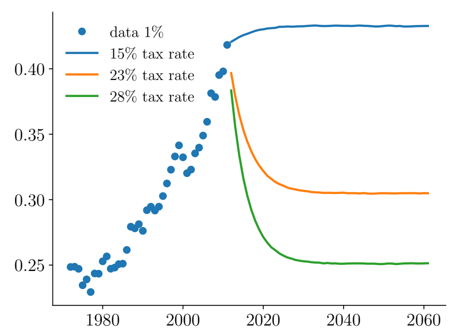

Figure 9: Projections using tax rates of 15%, 23% and 28%, estimation based on the post-1945 sample

Notes: For the prevailing tax rate (23%) we expect a decrease to a level comparable to the early 2000s. If the tax rate was increased to the level of the end of our sample in 2012 (15%), wealth inequality is expected to increase and to level off close to 45%. A substantial reduction to the level of 1980, on the other hand, will decrease inequality to the level of the late 1990s.

Pareto coefficients in the tails.[21] The increase in inequality is slowly transmitted to the fat ends of the tails. In line with Gabaix et al. (2016), in this type of model the convergence is slower in the tails. In fact, in the short run for the non-stationary distribution the local Pareto coefficient increases in the tail for high values of log-wealth $\hat{w}$ (cf. Figure 7).[22] This is in line with the empirical results as presented in Figure 8. Unfortunately, the overall evolution of the model – in particular the local convergence speed – is highly non-linear making it unfeasible to estimate the complete dynamics jointly.

Finally, we employ the model for some policy analysis. In Figure 9 we

[21] For a true log-normal distribution the estimated Pareto tail in the tails would diverge, i.e. $\lim_{\hat{w}\to\infty} \hat{\alpha}(\hat{w}) = \infty$.

[22] The figure is generated using the time-varying solution of the Fokker-Planck equation, as presented in Appendix D.

show forecasts using the estimated parameters and different tax regimes. As of 2017 the actual tax rate calculated as above is at 23.8%. This episode is not included in our series since Saez and Zucman (2016) only provide data for top inequality until 2012. An unchanged tax regime would be sufficient to reverse the trend and bring inequality back to the level of the early 2000s. A further increase to 28%, which is the level from 1980, would lever the concentration back to its value from the 1990s. In this case the share of the top 1% would almost fall down to the level currently held by the top 0.1%, implying a considerable level of redistribution. On the other hand, a decrease of taxation back to the level of 2003-2012 (15%) would result in a further increase of inequality as seen before the Second World War. The simulations also confirm our analytic result that the decrease in inequality after tax increases is faster than the ascent of inequality following tax reductions.

6 Conclusion

The main purpose of this work is to develop a simple, yet microfounded portfolio selection model that allows us to study the relationship between the dynamics of wealth inequality and empirical tax series. Although a quite straightforward approach, this stands in contrast to the majority of the theoretical literature on wealth inequality which takes income inequality as a starting point. We emphasize the role of equity trading and marginal heterogeneous expectations about asset prospects as a mechanism explaining top wealth inequality.

We apply this model to explain long-run dynamics of wealth inequality in the USA, which experienced periods of both substantial decrease and more recently a rapid increase of inequality. Due to the parsimonious nature of our model the degree of freedom to fit the empirical evidence is very limited. Nevertheless, our model matches the data well, both in levels and also in transition speed. In particular, and in contrast to much of the rest of the literature, we are able to match both, up and downturns in the concentration of wealth. Our analytic results emphasize that the level and the transition speed of wealth inequality depend crucially on the degree of capital gains taxation, which is quantitatively and qualitatively in line with of the estimated model. We conclude that the given tax series have a very high explanatory power regarding the dynamics of US wealth distribution over a very long time period of 90 years.

This also implies that one answer on the policy question on how to influence the distribution of wealth – and potentially reverse the recent increase in wealth inequality observed in developed economies – can be given by looking at the tax system. An increase in capital gains taxes, or alternatively a gross tax on wealth as suggested in Piketty (2014), will very likely reduce wealth concentration and has the potential to upturn the observed trends. Our projections predict that, for the USA – continuing on the present path of capital taxation – the gap between rich and poor is expected to shrink whereas substantial tax cuts will further increase the degree of wealth concentration. Our analyses further suggests that effects of tax avoidance or evasion are, from the macro perspective, of second order.

There are two implications for future research. Although our model fits the data quite well, there are periods where it falls short of accounting for the data. First, we consider it important to identify whether the those shortcomings are due to measurement errors, or to reasons that are exogenous to our model. Second, if these reasons are exogenous it is crucial to investigate them further.

Further, the quality of the model's result severely hinges on the quality of the data. A better availability of data on wealth dispersion at higher frequencies and for different countries would give better means to improve our model and enhance the understanding of the issue of wealth inequality in the 21st century.

References

Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics 109*(3), 659–84.

Aiyagari, S. R. (1995). Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting. *Journal of Political Economy 103*(6), 1158–75.

Alstadsaeter, A., N. Johannesen, and G. Zucman (2017). *Tax Evasion and Inequality.* Working Paper.

Alvaredo, F., A. B. Atkinson, and S. Morelli (2018). Top wealth shares in the UK over more than a century. *Journal of Public Economics 162*(C), 26–47.

Aoki, S. and M. Nirei (2017, July). Zipf's Law, Pareto's Law, and the Evolution of Top Incomes in the United States. *American Economic Journal: Macroeconomics 9*(3), 36–71.

Benhabib, J., A. Bisin, and S. Zhu (2011). The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents. *Econometrica 79*(1), 123–157.

Benhabib, J., A. Bisin, and S. Zhu (2014). The distribution of wealth in the blanchard–yaari model. *Macroeconomic Dynamics FirstView*, 1–16.

Bewley, T. (1986). Stationary Monetary Equilibrium with a Continuum of Independently Fluctuating Consumers. In W. Hildenbrand and A. Mas-Collel (Eds.), *Contributions to Mathematical Economics in Honor of Gerard Debreu*. Amsterdam: North-Holland.

Blanchet, T., J. Fournier, and T. Piketty (2017). Generalized Pareto Curves : Theory and Applications. Working Papers 201703, World Inequality Lab.

Boehl, G. (2018). *On the Evolutionary Fitness of Rational Expectations. In preparation.*

Branch, W. A. (2004). The theory of rationally heterogeneous expectations: Evidence from survey data on inflation expectations*. *The Economic Journal 114*(497), 592–621.

Cagetti, M. and M. DeNardi (2009). Estate Taxation, Entrepreneurship, and Wealth. *American Economic Review 99*(1), 85–111.

Campbell, J. Y. and L. M. Viceira (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Number 9780198296942 in OUP Catalogue. Oxford University Press.

Cao, D. and W. Luo (2017). Persistent Heterogeneous Returns and Top End Wealth Inequality. *Review of Economic Dynamics 26*, 301–326.

Castaneda, A., J. Diaz-Gimenez, and J.-V. Rios-Rull (2003). Accounting for the U.S. Earnings and Wealth Inequality. *Journal of Political Economy 111*(4), 818–857.

Chan, S., J. Chu, and S. Nadarajah (2017, 3). Is the wealth of the forbes 400 lists really pareto distributed? *Economics Letters 152*, 9–14.

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-law distributions in empirical data. *SIAM Review 51*(4), 661–703.

Craig, C. C. (1936, 03). On the frequency function of xy. *Ann. Math. Statist. 7*(1), 1–15.

Domeij, D. and J. Heathcote (2004). On The Distributional Effects Of Reducing Capital Taxes. *International Economic Review 45*(2), 523–554.

Fernholz, R. and R. Fernholz (2014). Instability and concentration in the distribution of wealth. *Journal of Economic Dynamics and Control 44*(C), 251–269.

Fischer, T. (2017). Thomas Piketty and the rate of time preference. *Journal of Economic Dynamics and Control 77*(C), 111–133.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The dynamics of inequality. *Econometrica 84*(6), 2071–2111.

Garbinti, B., J. Goupille-Lebret, and T. Piketty (2017). Accounting for wealth inequality dynamics: Methods, estimates and simulations for france (1800-2014). CEPR Discussion Papers 11848, C.E.P.R. Discussion Papers.

Greenwood, R. and A. Shleifer (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies 27*(3), 714–746.

Hommes, C. (2013). *Behavioral rationality and heterogeneous expectations in complex economic systems*. Cambridge University Press.

Hubmer, J., P. Krusell, and J. Anthony A. Smith (2016). The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation. NBER Working Papers 23011, National Bureau of Economic Research, Inc.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control 17*(5-6), 953–969.

Judd, K. L. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics 28*(1), 59–83.

Karlin, S. and H. Taylor (1981). *A Second Course in Stochastic Processes*. Number 2. Academic Press.

Kasa, K. and X. Lei (2018). Risk, uncertainty, and the dynamics of inequality. *Journal of Monetary Economics 94*(C), 60–78.

Kaymak, B. and M. Poschke (2016). The evolution of wealth inequality over half a century: The role of taxes, transfers and technology. *Journal of Monetary Economics 77*(C), 1–25.

Krusell, P. and T. Smith (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Kumhof, M., R. Rancière, and P. Winant (2015). Inequality, Leverage, and Crises. *American Economic Review 105*(3), 1217–1245.

Lansing, K. J. (1999). Optimal redistributive capital taxation in a neoclassical growth model. *Journal of Public Economics 73*(3), 423–453.

Levhari, D. and T. N. Srinivasan (1969). Optimal savings under uncertainty. *The Review of Economic Studies 36*(2), 153–163.

Lundberg, J. and D. Waldenström (2017). Wealth inequality in sweden: What can we learn from capitalized income tax data? *Review of Income and Wealth* (77(1)), 1–25.

Mankiw, N. G., R. Reis, and J. Wolfers (2004). Disagreement about Inflation Expectations. In *NBER Macroeconomics Annual 2003, Volume 18*, NBER Chapters, pp. 209–270. National Bureau of Economic Research, Inc.

Pfajfar, D. and E. Santoro (2008). *Asymmetries in inflation expectation formation across demographic groups.* University of Cambridge, Faculty of Economics.

Pfajfar, D. and E. Santoro (2010). Heterogeneity, learning and information stickiness in inflation expectations. *Journal of Economic Behavior & Organization 75*(3), 426–444.

Piketty, T. (2014). *Capital in the twenty-first century.* Boston: Harvard University Press.

Piketty, T. and G. Zucman (2014). Capital is back: Wealth-income ratios in rich countries 1700–2010. *The Quarterly Journal of Economics 129*(3), 1255–1310.

Pulley, L. B. (1983). Mean-variance approximations to expected logarithmic utility. *Operations Research 31*(4), 685–696.

Roine, J. and D. Waldenström (2015). Chapter 7 - long-run trends in the distribution of income and wealth. In A. B. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 2 of *Handbook of Income Distribution*, pp. 469 – 592. Elsevier.

Saez, E. and S. Stantcheva (2018). A Simpler Theory of Optimal Capital Taxation . *Journal of Public Economics 162*, 120–142.

Saez, E. and G. Zucman (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *The Quarterly Journal of Economics 131*(2), 519–578.

Sialm, C. (2009). Tax Changes and Asset Pricing. *American Economic Review 99*(4), 1356–1383.

Singer, A., Z. Schuss, A. Osipov, and D. Holcman (2008). Partially reflected diffusion. *SIAM Journal on Applied Mathematics 68*(3), 844–868.

Stiglitz, J. E. (1969). The Effects of Income, Wealth, and Capital Gains Taxation on Risk-Taking. *The Quarterly Journal of Economics 83*(2), 263–283.

Straub, L. and I. Werning (2014). Positive Long Run Capital Taxation: Chamley-Judd Revisited. NBER Working Papers 20441, National Bureau of Economic Research, Inc.

Toda, A. A. (2012). The double power law in income distribution: Explanations and evidence. *Journal of Economic Behavior & Organization 84*(1), 364 – 381.

Vermeulen, P. (2016). Estimating the Top Tail of the Wealth Distribution. *American Economic Review 106*(5), 646–650.

Wold, H. O. A. and P. Whittle (1957). A model explaining the pareto distribution of wealth. *Econometrica 25*(4), 591–595.

Appendix A Proof of Proposition 2

Let us define the log of wealth $\hat{w}_{i,t} = \log(w_{i,t})$ and apply Itô's lemma as a second-order approximation. Thus the equation reads

$$d\hat{w}_i \approx \left(-\frac{\tau}{\lambda} - 0.5\frac{\gamma^2}{\lambda^2}(1-\tau)^2\right) dt + \frac{1}{\lambda}(1-\tau)\gamma dL = -\mu dt + \delta dL,$$

with a diffusion term $\delta \equiv \frac{1}{\lambda}(1-\tau)\gamma$ and a drift $\mu \equiv \frac{\tau}{\lambda} + 0.5\frac{\gamma^2}{\lambda^2}(1-\tau)^2 = \frac{\tau}{\lambda} + 0.5\delta^2$. As shown in Toda (2012), the Laplace distribution with unit standard deviation can be modeled by

$$dL = -\lambda \operatorname{sign}(L)dt + \ dB$$

with B being the standard Brownian motion and $\operatorname{sign}(x) = \frac{x}{|x|}$ representing the sign function. In essence, this is a Brownian motion which reverts to its zero mean both in the positive and the negative domain. Thus, the noise in the returns ε before taxes (approximately) follows a Laplace distribution with a zero mean

$$f(\varepsilon) = \frac{0.5}{\gamma\lambda} \exp\left(-\frac{|\varepsilon|}{\gamma\lambda}\right),$$

which can be understood as a symmetric double exponential distribution. This result will also be of use in Proposition 7. If we ignore the fat-tail properties in the returns – induced by the mean reversion – we can model the wealth evolution of the wealthiest individuals by

$$d\hat{w}_i = -\mu dt + \delta dB, \qquad \hat{w}_{i,t} >> 0. \tag{A.1}$$

The cross-sectional distribution can be found by solving the so-called Fokker-Planck equation[23]

$$\frac{\partial f(\hat{w},t)}{\partial t} = -\frac{\partial}{\partial \hat{w}} \left(\mu f(\hat{w},t)\right) + 0.5 \frac{\partial^2}{\partial \hat{w}^2} \left(\delta^2 f(\hat{w},t)\right).$$

We first consider the stationary distribution ($\frac{\partial f(\hat{w},t)}{\partial t} \overset{!}{=} 0$). The solution is well-known (Karlin and Taylor, 1981, p. 221) and given by[24]

$$f(\hat{w}) = C \exp(-\alpha \hat{w}), \tag{A.2}$$

for $\hat{w} > \hat{w}_{min} = \ln(w_{min})$ with an integration constant of $C = w_{min}^{\alpha}\alpha$ to ensure $\int_{\hat{w}_{min}}^{\infty} f(\hat{w})d\hat{w} = 1$. For our case we have

$$\alpha = \frac{2\mu}{\delta^2} = 1 + \frac{\sqrt{2}\tau}{\gamma^2(1-\tau)^2}. \tag{A.3}$$

It is easy to transfer the exponential distribution to a Pareto distribution. In fact, if $\hat{w}$ follows the described exponential distribution, wealth

[23]The latter is frequently also referred to as Kolmogorov forward equation. The terms can be used interchangeably. Note that our proof heavily relies on second-order approximations. This is not problematic for realistic values of $\alpha < 2$, for which only the first two moments exists. The Fokker-Planck equation is a second-order approximation to the more general Master equation while the use of Itô's lemma is also a second-order approximation. For noise generated by a Brownian motion (rather than the product-normal distribution) it would hold exactly.

[24]A more formal derivation using Laplace-transforms is presented in Appendix C, also determining the average convergence speed.

$w = \exp(\hat{w})$ is given by the probability density function

$$\lim_{w \to \infty} f(w) \sim w^{-\alpha-1}.$$

ONLINE APPENDIX NOT INTENDED FOR PUBLICATION

Appendix B Proof of Proposition 1

The product-normal distribution is treated extensively in Craig (1936). The probability distribution function is given by

$$f(z_{PN}) = \frac{1}{\pi} K_0 \left(|z_{PN}| \right), \tag{B.1}$$

with $z_{PN} \equiv \epsilon^1 \epsilon^2$ with $\epsilon^i \sim N(0, 1)$ and K_0 being the modified Bessel-function of the second kind. The function is symmetric around the mean of zero and exhibits leptokurtic behavior. It is more appealing to write this using the Moment-Generating Function (MGF), which in this case is given by

$$M_{Z_{PM}}(t) = \frac{1}{\sqrt{1 - t^2}}. \tag{B.2}$$

Using this it is easy to show that the mean and skewness are zero, while the standard deviation is given by

$$SD(z_{PN}) = 1. \tag{B.3}$$

This distribution is highly comparable to the Laplace distribution. For a zero-mean the probability density function of the latter is given by

$$f(z_L) = \frac{1}{2\lambda} \exp \left(-\frac{|z_L|}{\lambda} \right) \tag{B.4}$$

for shape parameter $\lambda > 0$, having both a mean and a skewness of zero. The standard deviation of Laplace is

$$SD(z_L) = \sqrt{2}\lambda. \tag{B.5}$$

The Laplace distribution is also very appealing as each half takes the form of an exponential function. The moment generating function of the Laplace distribution is

$$M_{Z_L}(t) = \frac{1}{1 - \lambda^2 t^2}. \tag{B.6}$$

Comparing this with the MGF of the product-normal distribution it becomes obvious that the two are not identical. In fact, the sum of two product-normal variables follows a Laplace distribution.[25]

As a reasonable approximation we replace the product-normal with the Laplace distribution. To obtain the shape parameter λ that best approximates the standard normal product distribution we equalize the second order Taylor expansions of both MGFs around $t = 0$, which in fact is equivalent to choosing λ to match the first two moments of the function. This yields

$$\sum_{n=0}^{2} \frac{\partial^n M_{Z_{PM}}(0)}{n! \partial t^n}(t-0)^n = \sum_{n=0}^{2} \frac{\partial^n M_{Z_L}(0)}{n! \partial t^n}(t-0)^n$$
$$1 + \frac{t^2}{2} = 1 + t^2 \lambda^2$$
$$\lambda = \sqrt{\frac{1}{2}} \approx 0.707. \quad \square$$

[25] Using the MGF it is easy to show that if there are four independently distributed normal shocks with zero mean $X_i \sim N(0, \sigma_i)$ and we have $\sigma_1 \sigma_2 = \sigma_3 \sigma_4$ then $X_1 X_2 + X_3 X_4$ follows a Laplace distribution with zero mean and $\lambda = 1$.

Appendix C Proof of Proposition 4

The solution to the Fokker-Planck equation can be easily determined using the Laplace transform into the frequency domain[26] given by

$$\mathcal{L}\left\{f(\hat{w},t)\right\} \equiv F(s,t) \equiv \int_0^\infty f(\hat{w},t)\exp(-s\hat{w})d\hat{w}. \tag{C.1}$$

The latter is of particular help for solving linear differential equations as the n-th derivative is given by $\mathcal{L}\left\{f^n(\hat{w})\right\} = s^n F(s,t)$. For the right tail (index r) the characteristic equation is given by

$$\frac{\partial F(s,t)}{\partial t} = \mu s F(s,t) + 0.5\delta^2 s^2 F(s,t) = \Lambda_r(s)F(s,t) \tag{C.2}$$

with $\Lambda_r(s) = \mu s + 0.5\delta^2 s^2$. The stationary solution is found by setting $\frac{\partial F(s,t)}{\partial t} \overset{!}{=} 0$, leading to

$$\Lambda_r(s) = 0 \rightarrow s_r = -\frac{2\mu}{\delta^2} \equiv -\alpha. \tag{C.3}$$

In this case, the cross-sectional distribution of log wealth $\hat{w} \equiv \ln(w)$ is given by an exponential distribution, while wealth follows a Pareto distribution. The value α is the rate parameter of the exponential distribution respectively the Pareto coefficient.

This approach can also be employed to make a statement about the convergence rate. As our paper only considers the top shares we focus on the right tail of the distribution, as described by $\Lambda_r(s)$. In fact the convergence

[26] This procedure is also employed in Gabaix et al. (2016) and Kasa and Lei (2018) to solve similar problems.

rate of the n-th moment $E(\hat{w}^n)$ is given by $\Lambda_r(-n)$. For the example of the mean it would be

$$\Lambda_r(-1) = -\mu + 0.5\delta^2 = -\frac{\tau}{\lambda}. \tag{C.4}$$

It is well known that for the Pareto distribution only moments with $0 < n < \alpha$ exist. For the parametrization to fit the US wealth distribution we always have $\alpha < 2$. The average convergence time - as defined in Gabaix et al. (2016) - emerges for $\bar{n} = 0.5\alpha = \frac{\mu}{\delta^2}$. It is given by

$$\Lambda_r\left(s = -\frac{\mu}{\delta^2}\right) = -0.5\frac{\mu^2}{\delta^2} < 0 \tag{C.5}$$

Assume that the distribution starts at a stationary distribution $F(s, 0)$. After a shock in parameters the new stationary distribution is $F(s, \infty)$. Solving the differential equation C.2, we find the convergence in the frequency domain for some s is given by

$$F(s,t) = F(s,\infty) + [F(s,0) - F(s,\infty)]\exp(\Lambda_r(s)t). \tag{C.6}$$

In this case, we have $\Lambda_r(s) = -\phi = -\frac{\mu^2}{2\delta^2}$ as the average convergence rate. More generally we can write it as

$$F(s,t+\tau) = F(s,\infty) + [F(s,t) - F(s,\infty)]\exp(\Lambda_r(s)\tau), \tag{C.7}$$

which for our special case of $\tau = 1$ implies

$$F(s,t+1) = F(s,\infty) + [F(s,t) - F(s,\infty)]\exp(\Lambda_r(s)). \tag{C.8}$$

Appendix D Details on the time-dependent cross-sectional distribution

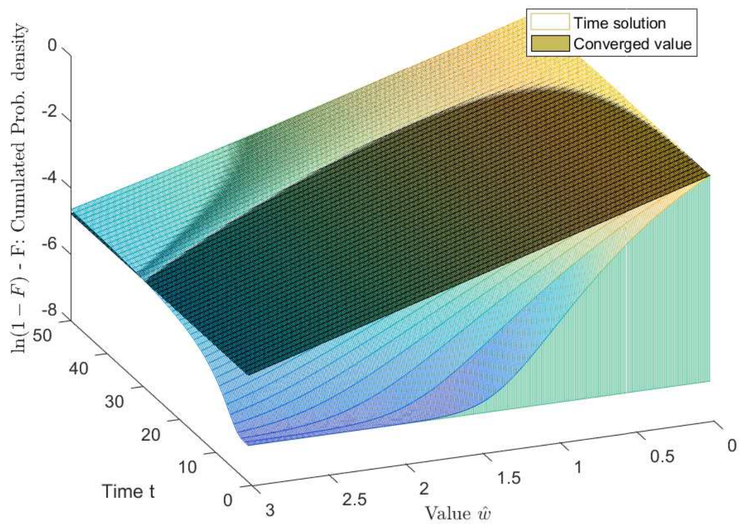

Figure D.10: Log-Cumulative probability density function: Stationary solution and time-dependent solution

(Singer et al., 2008, p. 853) provide a full solution of the underlying Fokker-Planck equation in the time domain for some given initial value $\hat{w}_0 = 0$. We also want to assume that the reflecting boundary is $\hat{w}_{min} = 0$ (i.e. $w_{min} = 1$). The solution is given by

$$f(\hat{w},t) = \frac{1}{\sqrt{\pi t 2\delta^2}} \exp\left(-\frac{\hat{w}^2}{2\delta^2 t}\right) \exp\left(-0.5\alpha\hat{w} - \frac{1}{8}\alpha^2\delta^2 t\right) + f(\hat{w},\infty)\Phi\left(-\frac{\hat{w}}{\delta\sqrt{t}} + 0.5\alpha\delta\sqrt{t}\right), \tag{D.1}$$

for which $f(\hat{w},\infty) = C\exp(-\alpha\hat{w})$ describes the long-run stationary solution and Φ is the cumulative probability density function of the normal distribution. We have $C = \alpha w_{min}^{\alpha} = \alpha$. For small time values t it is Gaussian, finally converging to an exponential distribution. In terms of transformed values $w = \exp(\hat{w})$ this implies a transformation from log-normal to Pareto.

It is evident that the solution is both a function of time t and the value of $\hat{w}$. Essentially, the function slowly *fattens out* to the tails (cf. figure D.10). Thus, the measured Pareto tail $\hat{\alpha}$ decrease in time, but increases with the value of $\hat{w}$. Technically, it never converges in the fattest tails ($\lim_{\hat{w}\to\infty} f(\hat{w}, t \to \infty) \neq f(\hat{w},\infty)$).

Acknowledging that the first part is a normal distribution with zero mean and variance $\delta^2 t$ (exploding in time) and abbreviating this with $f_0(\hat{w},t)$ as well as using the definition of the average convergence speed $\phi = \frac{\mu^2}{2\delta^2} = \frac{1}{8}\alpha^2\delta^2$, we can write:

$$f(\hat{w},t) = f_0(\hat{w},t)\exp(-0.5\alpha\hat{w})\exp(-\phi t) + f(\hat{w},\infty)\Phi\left(-\frac{\hat{w}}{\delta\sqrt{t}} + 0.5\alpha\delta\sqrt{t}\right). \tag{D.2}$$

The very last term in the equation related to the normal CPDF captures both the convergence speed $(1-\exp(-\phi t))$ and the non-linearity adjustment for $f(\hat{w},\infty)$. It is obvious that this is incorporated in a non-trivial manner. In the empirical application we choose a simplified non-linearity adjustment not least to keep the estimation feasible.