

The Historical Evolution of the Wealth Distribution: A Quantitative-Theoretic Investigation

Joachim Hubmer, Per Krusell, and Anthony A. Smith, Jr.*

August 9, 2017

Abstract

This paper employs the benchmark heterogeneous-agent macroeconomic model to examine drivers of the rise in wealth inequality in the U.S. over the last thirty years. By far the most important driver is the significant drop in tax progressivity starting in the late 1970s. The sharp observed increases in earnings inequality and the falling labor share over the recent decades, on the other hand, fall far short of accounting for the data. Changes in asset returns and in the inflation rate help to account for the shorter-run dynamics in wealth inequality.

1 Introduction

The distribution of wealth in most countries for which there is reliable data is strikingly uneven. There is also recent work suggesting that the wealth distribution has undergone significant movements over time, most recently with a large upward swing in dispersion in several Anglo-Saxon countries.¹ For example, according to the estimates in Saez & Zucman (2016) for the United States, the share of overall wealth held by the top 1% has increased from around 25% in 1980 to over 40% today; for the top 0.1% it has increased from less than 10% to over 20% over the same time period.

The observed developments have generated strong reactions across the political spectrum. In his 2014 book, *Capital in the Twenty-First Century*, Piketty is obviously motivated by the growing inequality in itself, but he also suggests that further increases in wealth concentration may lead to both economic and democratic instability. Conservatives in the U.S. have expressed worries as well: is the American Dream really still alive, or might it be that a large fraction of the population simply will no longer be able to productively contribute to society? Given, for example, that parental wealth and well-being are important determinants behind children's human capital accumulation, this appears to be a legitimate concern regardless of one's political views. As a result of these concerns, a number of policy changes have

*The authors' affiliations are, respectively, Yale University; Institute for International Economic Studies, NBER, and CEPR; and Yale University and NBER. For helpful comments the authors would like to thank Chris Carroll, Harald Uhlig, and seminar participants at the 2015 SED Meetings, the 2015 Hydra Workshop on Dynamic Macroeconomics, the Seventh Meeting of the Society for the Study of Economic Inequality, the 2017 NBER Summer Institute, Johns Hopkins, Indiana, M.I.T., Penn State, University of Pennsylvania, SOFI, and Yale.

¹See, e.g., Piketty (2014) and Saez & Zucman (2016).

been proposed and discussed. The primary aim of the present paper is to understand the determinants of the observed movements in wealth inequality. This aim is basic but well-motivated: to compare different policy actions, we need a framework for thinking about what causes inequality and for addressing how inequality—and other variables—are influenced by any policy proposal at hand.

In an effort to understand the movements in wealth inequality, Piketty (2014) and its online appendix suggest specific mathematical theories and as part of the present study we examine those theories.² Our aim, however, is to depart instead from a more general, and by now rather standard, quantitatively oriented theory used in the heterogeneous-agent literature within macroeconomics: the Bewley-Huggett-Aiyagari model. This is a very natural setting for the study of inequality. This model incorporates rich detail on the household level along the lines of the applied work in the consumption literature, allowing several sources of heterogeneity among consumers. It is based on incomplete markets and, hence, does not feature the “infinite elasticity of capital supply” of dynastic models with complete markets.³ This model also involves equilibrium interaction: inequality is determined not only by the individual household’s reactions to changes in the economic environment in which they operate but also by their interaction, such as in the equilibrium formation of wages and interest rates, two key prices determining the returns to labor and holding wealth, respectively. Our aim is to see to what extent a reasonably calibrated model can account for the movements in wealth inequality from the mid-1960s and on as a function of a number of drivers, the importance of each of which we then evaluate in separate counterfactuals.⁴ In this endeavor, we proceed as follows.

We build on the model studied in Aiyagari (1994), i.e., we use the core setting of the recent literature on heterogenous agents in macroeconomics.⁵ This kind of theoretical model is quantitative in nature: it is constructed as an aggregate version of the applied work on consumption. Moreover, in it, inequality plays a central role. We calibrate some key parameters of this model to match the wealth and income distributions in the United States in the mid-1960s and treat these distributions as representing a long-run steady state. In the 1960s, too, the dispersion of wealth was striking, and it is not immediate how to make the basic model match the data in this respect. Building on the formulation in Krusell & Smith (1998), we use preference heterogeneity—in particular, stochastic discount rates that vary across the population at a point in time—to generate individual behavior among the very richest characterized by propensities to save that are stochastic but (almost completely) independent of wealth. We also incorporate idiosyncratic random asset returns, for which recent work by Fagereng et al. (2015) and Bach et al. (2015) uncovers evidence in panel data from Norway and Sweden. Hence, our setting can be viewed as a microfoundation for the kind of models entertained in Piketty & Zucman (2015) (who assume linear laws of motion for wealth accumulation and either random saving propensities or random returns). These models generate a wealth distribution whose right tail is Pareto-shaped, a feature shown to characterize the data; we discuss

²The appendix is available here: <http://piketty.pse.ens.fr/files/capital21c/en/Piketty2014TechnicalAppendix.pdf>. See also Piketty (1995) and Piketty (1997) which develop theories of the dynamics of the wealth distribution.

³This elasticity refers to the long-run response of a household’s savings to a change in the interest rate: in particular, with infinitely-lived consumers and complete markets the equilibrium interest rate is pinned down by the rate of time preference.

⁴We do not specifically study Piketty’s “Second Fundamental Law”, which is not a theory about inequality per se but about the aggregate capital-output ratio and which has also been extensively examined in Krusell & Smith (2015).

⁵The first application in this literature was one to asset pricing (the risk-free rate): Huggett (1993). Aiyagari (1994) addresses the long-run level of precautionary saving, whereas Krusell & Smith (1998) look at business cycles.

this finding and the relation to a number of other papers building on the same kind of reduced form in detail in the paper.

With the resulting realistic starting wealth distribution, we then examine a number of potential drivers of wealth inequality over the subsequent period. One is tax rates: beginning around 1980 tax rates fell significantly for top incomes, so that tax progressivity in particular fell substantially. Thus, higher returns to saving in the upper brackets since that time can potentially explain increased wealth gaps between the rich and the poor. Another potential explanation for increased wealth inequality is the rather striking increases in wage/earnings inequality witnessed since the mid-1970s. Since at least Katz & Murphy (1992) it has been well-documented that the education skill premium has risen. Moreover, numerous studies have since documented that the premia associated with other measures of skill have also risen, as have measures of residual, or frictional, wage dispersion.⁶ In terms of the very highest earners, Piketty & Saez (2003) document significant movements toward thicker tails in the upper parts of the distribution. So to the extent that this increased income inequality has translated into savings and wealth inequality, it could explain some of the changes we set out to analyze. Relatedly, the share of total income paid to capital has increased recently, potentially contributing to increased wealth inequality (see, e.g., Karabarbounis & Neiman (2014b)). We consider this factor as well in this study.

Thus, the overall methodology we follow is to attempt to quantify the mechanisms just mentioned and then to examine their individual (and joint) effects on the evolution of wealth inequality from the 1960s. For the time period considered, we find that the benchmark model does account for a significant share of the increase in wealth inequality. The model is more or less successful depending on what aspect of the wealth distribution is in focus. The shares of wealth held by the top 10% or top 1% exhibit net increases that are very similar in the model and in the data, though for the top 0.1% and 0.01% the model does not deliver enough of an increase, especially for the very top group. For the bottom 50%, the model's fit is also good. Furthermore, the model delivers a time path for the ratio of capital to net output that is similar to the one in the data. As for the timing of the changes, the model delivers a rather smooth increase in inequality, whereas the data shows faster swings, first down and then up (the model generates a visible, though gentle, kink of this sort too, but in growth rates).

Turning to which specific features explain the largest fractions of the increase in wealth inequality, the marked decrease in tax progressivity is by far the most powerful force for increasing wealth inequality.⁷ First, other things equal, decreasing tax progressivity spreads out the distribution of after-tax resources available for consumption and saving. Second, decreasing tax progressivity increases the returns on savings, leading to higher wealth accumulation, especially among the rich for whom wages (earnings) is a smaller part of wealth.

Wage inequality, on the other hand, on net contributes *negatively* to wealth inequality: it increases by more in a model with changes in progressivity unaccompanied by increases in wage inequality than in a model with both types of changes. We follow Heathcote et al. (2010) in modelling increased wage inequality as an increase in the riskiness of wage realizations around a mean. In a standard additive permanent-plus-transitory model of wages, we use the estimated time series in Heathcote et al. (2010) for

⁶See, e.g., Acemoglu (2002), Hornstein et al. (2005), and Quadrini & Rios-Rull (2015).

⁷These conclusions are line with two studies of France and the U.S.: Piketty (2003) and Piketty & Saez (2003).

the variances of the permanent and transitory shocks to wages. Both of those variances have increased over time, leading to a reduction in wealth inequality for two reasons. First, increasing wage risk dampens the tendency of heterogeneity in discount rates to drive apart the distribution of wealth.⁸ In particular, as wage risk increases, poorer and less patient consumers—who are less well-insured against this risk through their own savings—engage in additional precautionary saving, compressing the distribution of wealth at the low end. Second, with more risk aggregate precautionary savings increase, reducing the equilibrium interest rate and reducing the relative wealth accumulation of the rich, for whom wage risk is also not so important. In sum, the increasing riskiness of wages compresses the wealth distribution at both ends.⁹

In addition, we follow Piketty & Saez (2003) by adding a Pareto-shaped tail to the wage distribution so as to match the concentration of earnings at the top of the earning distribution; the standard wage process (as in Heathcote et al. (2010)) does not match this extreme right tail well. Moreover, the right tail has thickened over this period, and accordingly we model this thickening as a gradually decreasing Pareto coefficient, based on the estimates in Piketty & Saez (2003). This element of increased wage inequality does generate more wealth inequality—because it occurs in a segment of the population where most workers are already rather well-insured through their own savings—but it is not so potent as to produce a net overall increase in wealth inequality from higher wage inequality. To allow for an increasing capital share over time we conduct an experiment using a CES production function with a somewhat higher than unitary elasticity between capital and labor. The resulting paths in this experiment differ only marginally from the case with unitary elasticity.

Given that the model predicts the within-period swings in the wealth shares less well than over the full period, we also begin a preliminary examination of the effects of systematic return differences on the overall portfolio between the poor and the rich. We look at both stock-market valuation effects—the idea being that the rich hold a larger fraction of stock than do the poor—and inflation effects, where we point out that progressivity jointly with a tax schedule that is not indexed to inflation reduces the returns to saving of the wealthy more than it does for the poor if inflation rises.¹⁰ These factors both turn out to have direct effects that are sizable, so this line of research seems promising. We restrict attention here, however, to hard-wired portfolio-share differences and do not allow a nominal-vs.-real asset choice, and we moreover take returns as given and thus use a partial-equilibrium setting. Hence, a deeper foray into these issues seems promising but must be left for future work.

What are the implications of our dynamic model of wealth inequality for the future? Quite strikingly, if the progressivity of taxes remains at today’s historically low level, then wealth inequality will continue to climb and reach very high levels by, say, 2100: the top 10% will have an additional 10% of all of wealth, as will (approximately) the top 1%. Thus, decreasing the progressivity of taxes is a rather powerful mechanism for wealth concentration. In this context, we also consider a possible long-run decline in the

⁸As Becker (1980) shows, if discount rates are permanently different and there is no wage risk at all, then in the long-run steady state the most patient consumer owns all of the economy’s wealth.

⁹Similar forces are at play in Krusell et al. (2009), but in the opposite direction: they find that reductions in wage risk that accompany the elimination of business cycles lead to higher wealth inequality.

¹⁰This channel is thus not the same as general bracket creep but rather appears due to the interaction between nominal taxation and progressivity: even if inflation makes no single consumer creep up a bracket, it makes the net-of-tax real return fall more for consumers in higher tax brackets.

rate of growth, g —a determinant in Piketty’s $r - g$ story behind inequality in line with a recent popular belief of “secular stagnation”—and find that, although interesting in its own right, it does not affect these conclusions appreciably.

Our paper begins in Section 2 with a brief literature review, the purpose of which is to put our modeling in a historical perspective. We discuss the data on wealth inequality and its recent trends in Section 3. We describe the basic model in Section 4 and the implied behavior of the very richest in Section 5. Section 6 discusses the calibration in detail and Section 7 the benchmark results. A number of extensions are then included in Section 8. We conclude our paper in Section 9 with a brief discussion of potential other candidate explanations behind the increased wealth inequality and, hence, of possible future avenues for research.

2 Connections to the recent macro-inequality literature

The study of inequality in wealth using structural macroeconomic modeling can be said to have started with Bewley (undated), though in Bewley’s paper the focus was not on inequality per se.¹¹ Bewley’s paper was not completed—it stops abruptly in the middle—and the first papers to provide a complete analysis of frameworks like his are Huggett (1993) and Aiyagari (1994). A defining characteristic of these models is that long-run household wealth responds smoothly to the interest rate, so long as the interest rate is not too high (higher than the discount rate in the case without growth).

In their early papers, neither Bewley nor Huggett nor Aiyagari focused on inequality per se but rather on other phenomena related to inequality (asset pricing and aggregate precautionary saving in the latter two cases, respectively). Soon after, however, the macroeconomic literature that arose from these analyses began to address inequality directly. There were several reasons for this development. One was the interest in building macroeconomic models with microeconomic foundations in which heterogeneity could influence aggregates, i.e., cases that are in some sense far from aggregation and the typical permanent-income behavior that characterize the complete-markets model.¹² Another was an interest in wealth inequality per se and the challenge it posed: the difficulty that these models have in generating significant equilibrium wealth inequality. The difficulty is apparent in Aiyagari (1994), where the wage process is calibrated to PSID data (as an AR(1) in logs): the resulting wealth distribution is slightly more skewed than the wage distribution the model uses as an input, but not by much. The Gini index for wealth, in the stationary distribution of Aiyagari’s model, is only around 0.4, whereas it is around 0.8 in the data. The purpose here is not to go over the entire literature aiming at matching the wealth distribution but several different extensions of the model have been proposed in order to match the data better. On some general level, successful paths forward involve introducing “more heterogeneity”: typically in preferences (such as discount factors, as in Krusell & Smith (1998)), in the wage/earnings process (as in Castañeda

¹¹This model is of course not the first one with theoretical implications for inequality. An early example is Stiglitz (1969) who, building on his 1966 Ph.D. dissertation, studies the dynamics of the distributions of income and wealth in a neoclassical growth model with exogenous linear savings functions. A defining characteristic of the literature in focus here is that consumers face problems much like those studied in the applied consumption literature: they are risk-averse and choose optimal saving in the presence of earnings shocks for which there is not a full set of state-contingent markets.

¹²See, e.g., Krusell & Smith (1998) and Guerrieri & Lorenzoni (2011) for this line of work.

et al. (2003)), or in occupation (as in Cagetti & De Nardi (2006) or Quadrini (2000)).

More recently, a literature evolved that focuses on explaining the observed Pareto tail at the top of the wealth distribution. Benhabib et al. (2011) show analytically that the stationary wealth distribution in an overlapping-generations (OLG) economy with idiosyncratic capital return risk has a Pareto tail. Analogously, they provide analytical results for an infinite-horizon economy (Benhabib et al., 2015b). In Benhabib et al. (2015a), they conduct a quantitative investigation of social mobility and the wealth distribution in an OLG economy with idiosyncratic returns, which are fixed over a life-time. In a stylized model, Gabaix et al. (2016) demonstrate that the random growth mechanism that can generate the Pareto tail in the wealth distribution (either through idiosyncratic capital return risk or random discount factors) implies very slow transitional dynamics. Furthermore, Nirei & Aoki (2016) consider a stationary Bewley economy with investment risk. In that setting they find that decreasing top tax rates can explain the increasing concentration of wealth at the top.

Most of the literature on Bewley models has considered only the stationary (long-run) wealth distribution. Two recent exceptions are Kaymak & Poschke (2016), who in line with our analysis here aim to quantify the contribution of changes in taxes and transfers and in the earnings distribution to changes in the U.S. wealth distribution, and Aoki & Nirei (forthcoming) who study how a one-time drop in tax rates affects transitional dynamics in a setting with investment risk.

Relative to these recent contributions, the present paper builds directly on Aiyagari (1994) and matches the wealth distribution with the aid of stochastic, heterogeneous discount rates and idiosyncratic asset returns. As we show below, the randomness in discount rates and rates of return generates capital accumulation dynamics for the very richest that are similar to those in the recent theoretical studies on Pareto tails just cited, including the very slow transitional dynamics. For earnings, we follow Aiyagari (1994) but add a transitory shock to earnings as well as an exogenous Pareto-shaped tail in earnings. Because we also consider transitional dynamics, it is important to investigate how our results might depend on the extent to which agents can foresee the changes in taxes and other exogenous factors; here we consider both perfect foresight and a “myopic” alternative. We do not incorporate assets like land, housing, or stock-market equity but focus on physical capital only. This is potentially an important omission insofar as the returns on these assets are random and have experienced a growing variance over time, as discussed in our concluding remarks in Section 9.

3 Measuring wealth inequality over time

Over the last century, the distribution of wealth in the United States has undergone drastic changes and we very briefly review data from some key studies here. Throughout the time period considered, wealth was heavily concentrated at the top. Figure 1 shows the evolution of the share of total wealth held by the top 1% and the top 0.1%, as measured using different estimation methods.¹³ Considering all three

¹³In Figure 1, the lines labelled “SCF” display findings from the Survey of Consumer Finances, as reported in Saez & Zucman (2016). The lines labelled “Capitalization” display findings from Saez & Zucman (2016), who back out the stock of wealth held by a tax unit from observed capital income tax data. Finally, the lines labelled “Estate tax multiplier” display findings from Kopczuk & Saez (2004), who use observed estate tax data to make inferences about the distribution of wealth. See Kopczuk (2015) for a detailed comparison of the different measurement methods.

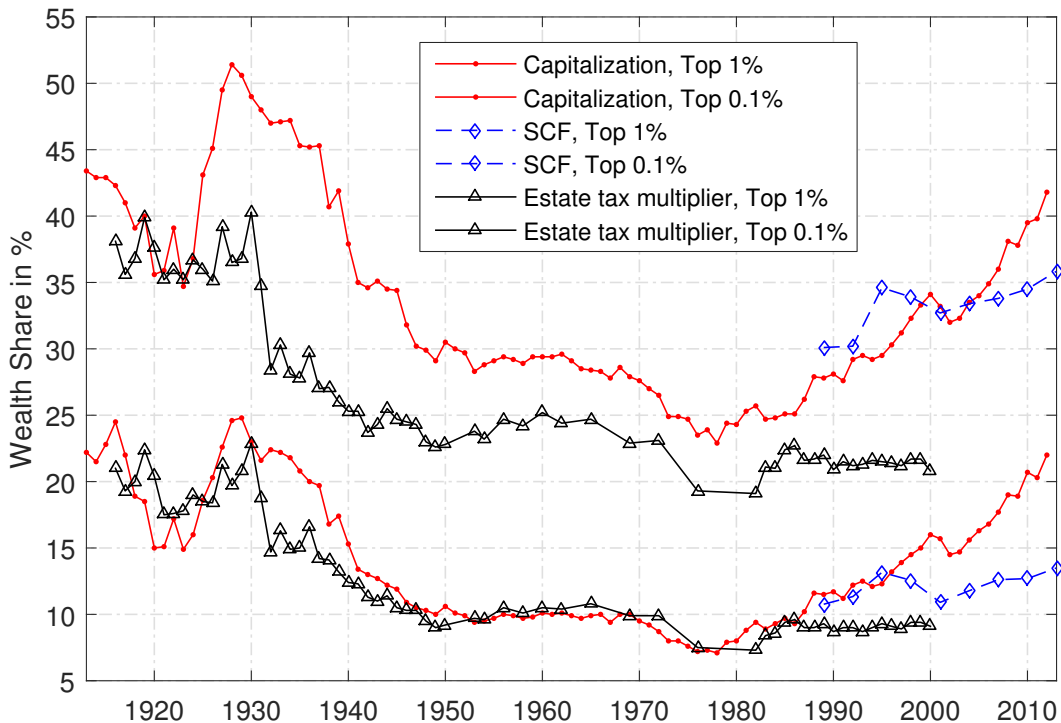


Figure 1: Top wealth share measurements over time

methods jointly, top wealth inequality exhibits a U-shaped pattern in the twentieth century. Yet, the magnitude of the increase in wealth concentration in the last thirty years differs substantially among estimation methodologies. We will calibrate the initial steady state of our model to the wealth shares estimated by Saez & Zucman (2016) and consequently compare the model transition to their estimates. Their estimates are especially useful for us as they allow for considering a group as small as the top 0.01%. Furthermore, they cover a long time period. While the capitalization method that they use to back out wealth estimates does not suffer from the shortcomings of the SCF data (such as concerns about response-rate bias and exclusion of the Forbes 400), it is an indirect way of measuring wealth and as such has other drawbacks. For example, the tax data allows only for a coarse partitioning of capital income in asset classes and within each class returns are effectively assumed to be homogeneous. Since recent evidence based on both Norwegian and Swedish data (Fagereng et al. (2015) and Bach et al. (2015), respectively) shows significantly higher returns for the high-wealth groups, the capitalization method suggests an over-prediction of wealth levels for the richest group. Therefore, we will in addition contrast our findings to estimates from the Survey of Consumer Finances.¹⁴

Another takeaway from Figure 1 is that the wealth distribution was quite stable in the 1950s and 1960s. As, in addition, some of the time series estimates we feed into our model start in 1967, we take this year as the initial steady state in our model.

¹⁴Bricker et al. (2016) make adjustments to the SCF data, including incorporating the Forbes 400. For the top 0.1% wealth shares these adjustments roughly cancel. For the top 1% shares these adjustments shift the corresponding line in Figure 1 down by approximately 2 to 3 percentage points.

4 Model framework

In this section, we describe the model economy. We depart from the framework studied by Aiyagari (1994). To generate realistic income and wealth heterogeneity, the model features stochastic discount rates and returns to capital as well as an earnings process centered around a persistent and a temporary component.

4.1 Consumers

Time is discrete and there is a continuum of infinitely lived, ex ante identical consumers (dynasties). Preferences are defined over infinite streams of consumption with von Neumann-Morgenstern utility in constant relative risk aversion (CRRA) form:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}. \quad (1)$$

In period t , a consumer discounts the future with an idiosyncratic stochastic factor β_t that is the realization of a Markov process characterized by the conditional distribution $\Gamma_\beta(\beta_{t+1}|\beta_t)$, giving rise to the following objective:

$$\max_{(c_t)_{t=0}^\infty} \left\{ u(c_0) + \mathbb{E}_0 \left[\sum_{t=1}^\infty \prod_{s=0}^{t-1} \beta_s u(c_t) \right] \right\}. \quad (2)$$

Labor supply is exogenous. Each period t , a consumer supplies a stochastic amount $l_t = l_t(p_t, \nu_t)$ of efficiency units of labor to the market that depends on a persistent component $p_t \sim \Gamma_p(p_t|p_{t-1})$ and a transitory component $\nu_t \sim \Gamma_\nu(\nu_t)$. Taking as given a competitive wage rate w_t , her earnings are $w_t l_t$.

Asset markets are incomplete: consumers cannot fully insure against idiosyncratic shocks, but instead have access only to a single asset that pays a gross return $(1 + r_t \eta_t)$, where r_t is the average market return and $\eta_t \sim \Gamma_\eta(\eta_t)$ is a transitory idiosyncratic shock.¹⁵ We briefly discuss the challenges in endogenizing portfolio behavior in general, and in obtaining differences in returns across consumers in particular, in Section 8.3 below.

The decision problem of the consumer can be stated parsimoniously in recursive form:

$$V_t(x_t, p_t, \beta_t) = \max_{a_{t+1} \geq a} \{u(x_t - a_{t+1}) + \beta_t \mathbb{E}[V_{t+1}(x_{t+1}, p_{t+1}, \beta_{t+1})|p_t, \beta_t]\} \quad (3)$$

$$\text{subject to } x_{t+1} = a_{t+1} + y_{t+1} - \tau_{t+1}(y_{t+1}) + T_{t+1} \quad (4)$$

$$y_{t+1} = r_{t+1} \eta_{t+1} a_{t+1} + w_{t+1} l_{t+1}(p_{t+1}, \nu_{t+1}) \quad (5)$$

¹⁵Fagereng et al. (2015) and Bach et al. (2015) find not only heterogeneity but persistence in idiosyncratic asset returns but a good portion of this persistence stems from richer consumers bearing more aggregate risk, which we do not model here. Furthermore, given that we allow for persistence in discount factors, we find below that we can replicate the wealth distribution in 1967, even in its remotest tails, quite accurately without persistence in idiosyncratic returns.

Given cash-on-hand x_t (all resources available in period t), the optimal savings decision and the resulting value function depend solely on the persistent component in the earnings process p_t and the current discount factor β_t . Conditional on (p_t, β_t) , the expectation is taken over (p_{t+1}, β_{t+1}) as well as the transitory shocks to earnings ν_{t+1} and the return on capital η_{t+1} . Gross income y_t is subject to an income tax $\tau_t(\cdot)$ and each consumer receives a uniform lump-sum transfer T_t .

4.2 Production, government, and equilibrium

Firms are perfectly competitive and can be described by an aggregate constant returns to scale production function $F(K_t, L)$ that yields a wage rate per efficiency unit of labor $w_t = \frac{\partial F(K_t, L)}{\partial L}$ as well as an (average) market return on capital $r_t = \frac{\partial F(K_t, L)}{\partial K} - \delta$, where $\delta \in (0, 1)$ is the depreciation rate. Aggregate labor supply L is normalized to one throughout.

The government redistributes aggregate income by means of a uniform lump-sum payment, which amounts to a constant fraction $\lambda \in [0, 1]$ of aggregate tax revenues. The remainder is spent in a way such that marginal utilities of agents are not affected.

A steady-state equilibrium of this economy is characterized by a market clearing level of capital K^* and a lump-sum transfer T^* such that:

- (i) factor prices are given by their respective marginal products $w^* = \frac{\partial F(K^*, 1)}{\partial L}$ and $r^* = \frac{\partial F(K^*, 1)}{\partial K} - \delta$;
- (ii) given r^* , w^* , and T^* , consumers solve the stationary version of their decision problem, giving rise to an invariant distribution $\Gamma(a, p, \beta, \nu, \eta)$;
- (iii) the government redistributes a fraction λ of total tax revenues, i.e.,

$$T^* = \lambda \int \tau(r^* \eta a + w^* l(p, \nu)) d\Gamma(a, p, \beta, \nu, \eta);$$

- (iv) and capital markets clear, i.e.,

$$K^* = \int a d\Gamma(a, p, \beta, \nu, \eta).$$

In the benchmark *perfect-foresight* transition experiment, we start the economy in period t_0 in some initial steady state, described by a vector θ^* that parametrizes the tax schedule and earnings process and by the equilibrium objects (K^*, T^*) . Agents are fully surprised and learn about a new exogenous environment $(\theta_t)_{t=t_0+1}^{t_1}$ that will prevail over some transition period $t = t_0 + 1, t_0 + 2, \dots, t_1$. From t_1 onwards, the exogenous environment will once again be constant and equal to θ_{t_1} . In a perfect-foresight equilibrium, agents are fully informed about future equilibrium objects $(K_t, T_t)_{t=t_0+1}^{\infty}$ too and optimize accordingly. Capital markets clear and the fraction of tax revenues λ that is redistributed is fixed.

In an alternative *myopic* transition experiment, agents are surprised about the new exogenous environment and equilibrium prices every period. That is, in period $t = t_0, t_0 + 1, \dots, t_1 - 1$, given a distribution $\Gamma_t(x_t, p_t, \beta_t)$, they choose a savings decision rule, $a_{t+1} = g_t(x_t, p_t, \beta_t)$, assuming that both θ_t and (r_t, w_t, T_t) will prevail forever. In period $t + 1$, they are accordingly surprised that: one, the exogenous environment has changed to θ_{t+1} ; and, two, that equilibrium factor returns (r_{t+1}, w_{t+1}) and transfers T_{t+1} result from

capital-market clearing and government-budget balance in period $t + 1$.¹⁶ These two informational structures are, of course, extreme. We chose them because we expect them to bracket a range of informational assumptions. Given that the results, as will be reported below, turn out to be very similar across the two structures, we are confident that our findings are robust to other variations in this dimension.

5 The right tail of the wealth distribution: approximately Pareto

In this section, we briefly explain the main mechanism that leads to a “fat” Pareto-shaped right tail in the wealth distribution. The same mechanism is at play in the much simpler stochastic- β model originally proposed in Krusell & Smith (1998).

Formally, we make use of a mathematical result on random growth by Kesten (1973): consider a stochastic process

$$a_t = s_t a_{t-1} + \epsilon_t, \tag{6}$$

where s_t and ϵ_t are (for our purposes positive) i.i.d. random variables. If there exists some $\zeta > 0$ such that $\mathbb{E}[s^\zeta] = 1$ as well as $\mathbb{E}[\epsilon^\zeta] < \infty$, then a_t converges in probability to a random variable A that satisfies $\lim_{a \rightarrow \infty} \text{Prob}(A > a) \propto a^{-\zeta}$, i.e., the right tail of the stationary distribution has a Pareto shape.¹⁷

In a setup like ours, it turns out—as we discuss in some more detail below—that s is the asymptotic marginal propensity to save out of initial-period asset holdings. Moreover, this propensity is random, whence it obtains time subscript. In a basic model with only discount-factor randomness, s varies precisely with β ; this turns out to be a property already of the model in Krusell & Smith (1998) designed to match the wealth distribution, though the β distribution there is quite stripped down. In the present somewhat augmented model, s_t also varies with the idiosyncratic return to wealth, η_t . Random earnings appear in the linear approximation through the error term ϵ_t . Crucially, in this class of models, optimal saving decisions are asymptotically, with increasing wealth, linear in economies with idiosyncratic risk and incomplete markets.¹⁸

Assuming a fixed discount rate, Carroll & Kimball (1996) prove in a finite-horizon setting that the consumption function is concave under hyperbolic absolute risk aversion, which comprises most commonly used utility functions (e.g., CRRA). Hence, the savings rule is convex. However, as household wealth increases, the convexity in the savings rule becomes weaker and weaker.¹⁹ Intuitively, as wealth grows

¹⁶That is, (r_{t+1}, w_{t+1}) are the marginal products of the net production function $F(K_{t+1}, 1) - \delta K_{t+1}$, where

$$K_{t+1} = \int g_t(x_t, p_t, \beta_t) d\Gamma_t(a_t, p_t, \beta_t, \nu_t, \eta_t),$$

and

$$T_{t+1} = \lambda \int \tau_{t+1}(r_{t+1} \eta a_{t+1} + w_{t+1} l_{t+1}(p_{t+1}, \nu_{t+1})) d\Gamma_{t+1}(a_t, p_t, \beta_t, \nu_t, \eta_t),$$

where Γ_{t+1} is the distribution in period $t + 1$ generated by the period- t distribution Γ_t and the decision rule g_t .

¹⁷The exact conditions as well as a very accessible treatment can be found in Gabaix (2009).

¹⁸In fact, the decision rules are almost linear for all but the very poorest agents, i.e., those close to the borrowing constraint. For this reason, approximate aggregation as introduced in Krusell & Smith (1998) typically works very well.

¹⁹A direct proof for a two-period problem can be found in Krusell & Smith (2006); Carroll (2012) proves the asymptotic

large consumers can smooth consumption more and more effectively. Moreover, with CRRA preferences decisions rules are exactly linear in the absence of risk (or with complete markets against such risk). The slope is then larger (smaller) than one as the discount rate is smaller (larger) than the interest rate. In the recent literature on the Pareto tail in the wealth distribution, either saving rates or returns to capital (or both, as in this paper) are assumed to vary randomly across consumers. Saving rules are then asymptotically linear with random coefficients: Benhabib et al. (2015b) show analytically that in this case the unique ergodic wealth distribution has a Pareto distribution in its right tail.

Figure 2 shows the marginal propensity to save out of capital holdings (denoted k in the figure) arising from the stochastic- β model under study in the present paper.²⁰ As discussed above, the marginal propensity to save increases in wealth, holding earnings constant, and asymptotes to a constant that depends on the consumer’s discount factor. Figure 3 displays the tail behavior of the stationary wealth distribution. In line with the theoretical results in Benhabib et al. (2015b), the logarithm of its counter-cumulative distribution function becomes linear in the logarithm of assets as assets grow large, indicating that the right tail of the distribution follows a Pareto distribution.

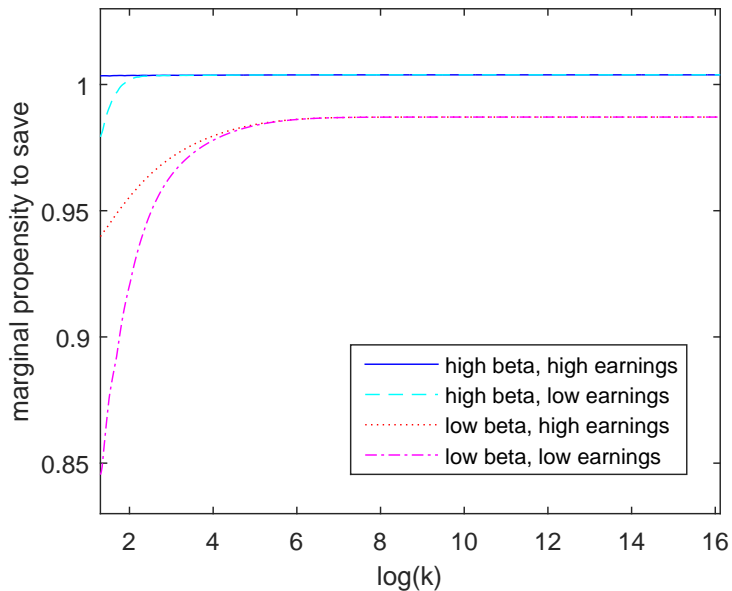


Figure 2: Asymptotic marginal propensity to save

In light of this result, it is worth noting that the model in Castañeda et al. (2003)—which generates substantial wealth inequality using an earnings process featuring a low-probability but transient very-high-earnings state—does not deliver a Pareto tail in wealth. In this model, in which consumers have a common discount rate, marginal propensities to save do not vary but instead converge to the same constant, independently of the level of earnings and as a result the steady-state distribution of wealth does not feature a Pareto tail. This model can deliver such a Pareto tail, however, if the earning process

linearity of the savings rule in a finite-horizon problem as the horizon grows large.

²⁰The graphs in this section are derived from a simplified model with a flat tax, to focus on the main mechanism.

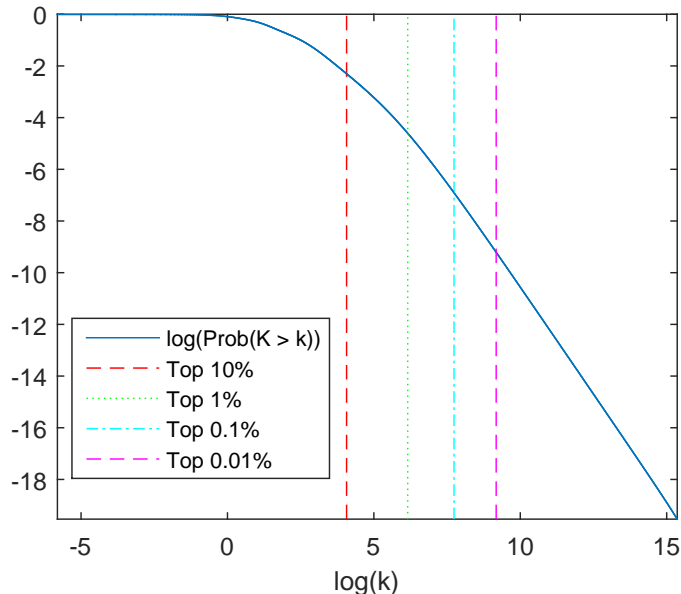


Figure 3: Pareto tail of the wealth distribution

itself has a Pareto tail. In the absence of randomness in either discount rates or returns, however, the wealth distribution inherits not only the Pareto tail of the earnings distribution but also its Pareto coefficient. Because earnings are considerably less concentrated than wealth, the resulting tail in wealth is too thin to match the data in such an alternative model.

6 Calibration

In this section, we describe how we calibrate our model economy. As indicated in Figure 1, the U.S. wealth distribution was roughly stable in the 1950s and 1960s, as was tax progressivity. This, together with the fact that some of our time series estimates start in 1967, make this year a natural initial steady state. We set the model period to a year to conform to the tax system.

6.1 Basic parameters

We parameterize the production technology and utility function using standard functional forms and parameters. The (gross) production function is given by $F(K, L) = K^\alpha L^{1-\alpha}$. The capital share is set to $\alpha = 0.36$ and depreciation to $\delta = 0.048$ annually. In an extension (see Section 8.1), we check the sensitivity of our results to using a constant-elasticity-of-substitution production function with (gross) elasticity greater than one. The coefficient of relative risk aversion, γ , is set to 1.5.

6.2 The earnings process

The earnings process is based on the traditional log-normal framework with $l_t(p_t, \nu_t) = \exp(p_t + \nu_t)$. That is, we assume that the persistent component p_t of the earnings process follows a Gaussian AR(1)

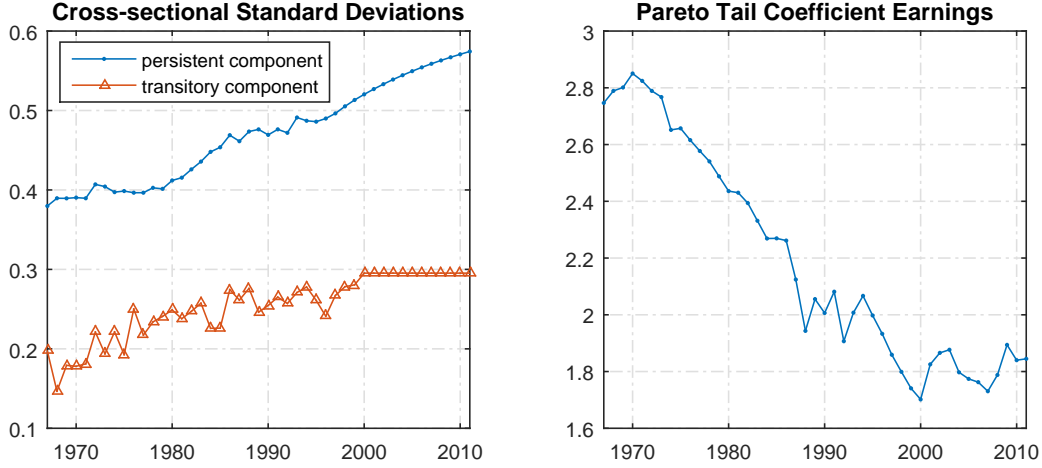


Figure 4: Earnings process ingredients

process with parameters (ρ^P, σ_t^P) . The autocorrelation coefficient, ρ^P , is fixed over time, while the innovation standard deviation varies. Likewise, the transitory component ν_t is also assumed to be normally distributed with standard deviation σ_t^T . We use estimates by Heathcote et al. (2010) that span the period 1967–2000 and assume that the time-varying variances of the innovations are constant thereafter. The left panel of Figure 4 displays the resulting cross-sectional dispersion. The estimates show a significant increase in earnings risk for both components.

As is well known, the resulting log-normal cross-sectional distribution of earnings understates the concentration of top labor income quite severely. Because the observed increase in top labor income shares is potentially an important explanation for the observed increase in wealth inequality at the top, we augment the framework for the top 10% earners in such a way that we can directly match the fraction of labor income going to the top 10%, top 1%, top 0.1% and top 0.01%. In concrete terms, we posit $l_t(p_t, \nu_t) = \psi_t(p_t) \exp(\nu_t)$, where

$$\psi_t(p_t) = \begin{cases} \exp(p_t) & \text{if } F_{p_t}(p_t) \leq 0.9, \\ F_{Pareto(\kappa_t)}^{-1} \left(\frac{F_{p_t}(p_t) - 0.9}{1 - 0.9} \right) & \text{if } F_{p_t}(p_t) > 0.9. \end{cases} \quad (7)$$

$F_{p_t}(\cdot)$ is the cdf of p_t and $F_{Pareto(\kappa_t)}^{-1}(\cdot)$ the inverse cdf for a Pareto distribution with lower bound $F_{p_t}^{-1}(0.9)$ and shape coefficient κ_t . Effectively, we thus assume that top earnings are spread out according to a (scaled) Pareto distribution, while earnings for the majority of workers are distributed according to a log-normal distribution. The Pareto tail coefficient on labor income κ_t is then one additional free parameter to calibrate in each year. We use estimates on top wage shares from an updated series by Piketty & Saez (2003) spanning 1967–2011 as calibration targets. The right panel of Figure 4 displays the calibrated Pareto tail coefficient κ_t and Figure 5 displays the resulting top labor income shares. That we can match top labor income shares very well using just a single parameter in each year (i.e., the tail coefficient) simply reflects the fact that the Pareto distribution is a very good description of the cross-sectional

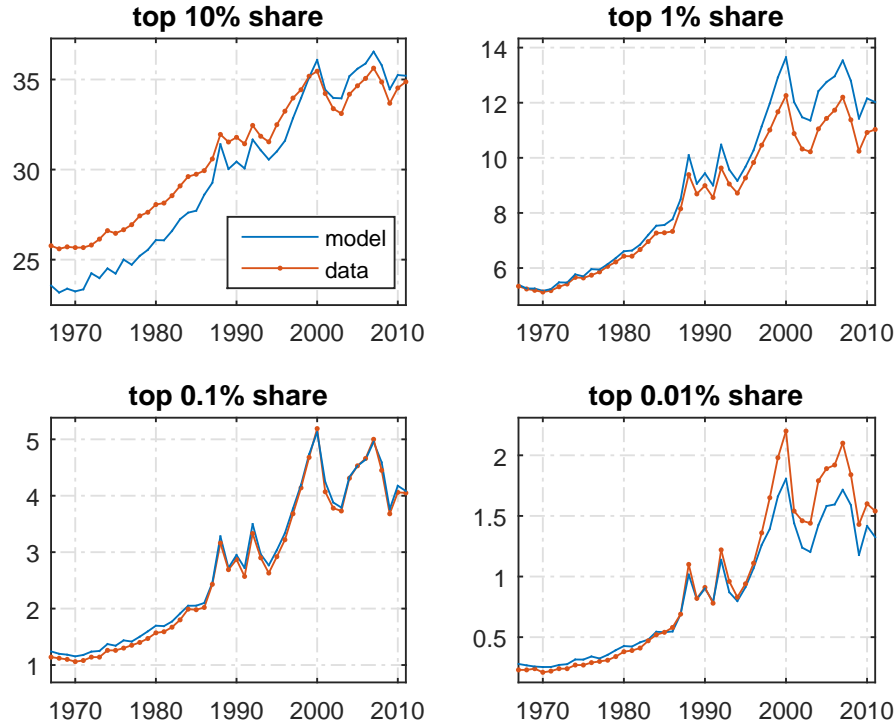


Figure 5: Top labor income shares in %

earnings distribution at the top.

We do not explicitly model unemployment, nor voluntary non-employment or retirement. We do, however, introduce a zero-earnings state, occurring with probability $\chi_0 = 0.075$ independently of (p_t, ν_t) and over time, reflecting both long-term unemployment and shocks that trigger temporary exit from the labor force. This probability is calibrated, together with a borrowing constraint amounting to roughly one yearly lump-sum transfer, so that the initial steady-state wealth distribution matches both the share of wealth held by the bottom 50% and the fraction of the population with negative net wealth.

6.3 Tax system

The progressivity of the U.S. tax system has decreased substantially over the model period. To account for these changes, we use estimates on federal effective tax rates by Piketty & Saez (2007) for the period 1967–2000, keeping them constant thereafter. These comprise the four major federal taxes: individual income, corporate income, estate and gift, and payroll taxes.²¹ Piketty & Saez (2007) calculate effective average tax rates for eleven income brackets, with a particularly detailed decomposition for top income

²¹Given that our model abstracts from the life cycle, it is appropriate to include the estate tax in the tax on total income, thus effectively smoothing out the incidence of this tax over the life cycle. Ignoring the estate tax would mean omitting a major source of decreasing tax progressivity. Piketty & Saez (2007) assume further that the corporate income tax burden falls entirely (and uniformly) on capital income. They argue that this is a middle-ground assumption (regarding the resulting tax progressivity) between assuming that the tax falls solely on shareholders at one extreme and assuming that it is effectively born by labor income at the other extreme.

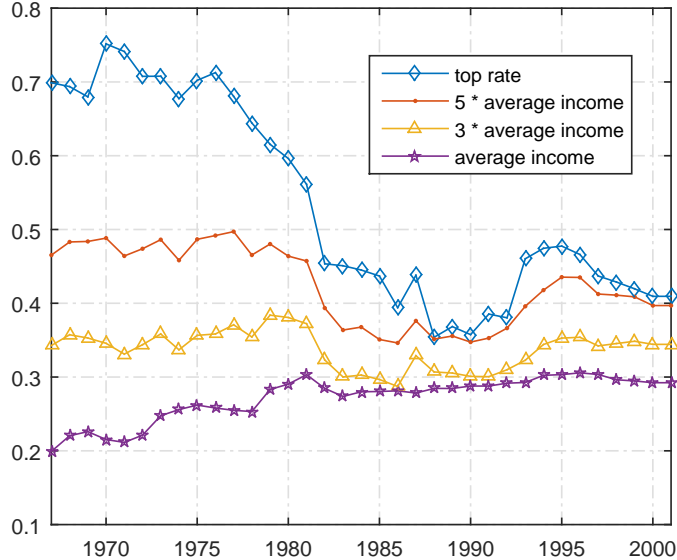


Figure 6: Imputed marginal tax rates for selected total income levels

groups (up to the top 0.01%). We translate this data to our model by means of a step-wise tax function $\tau_t(\cdot)$ with eleven steps. For each bracket, the threshold is set to match its income share in the data and the marginal tax rate such that the resulting average tax rate aligns with the data. Figure 6 shows that the U.S. tax system has indeed become much less progressive over the model period.

Note that in our model taxes $\tau_t(y_t)$ are a function of total income y_t , consistent with the measurement. A weakness of our calibration is that we do not have separate tax rates for different sources of income, but a strength is that we use effective tax rates, thereby accounting for tax avoidance and changing portfolio composition to the extent that these vary systematically with income. To account for government transfers, we introduce a social safety net in the simplest possible way by assuming that each agent receives an (untaxed) lump-sum transfer T_t every period, its size being a constant fraction $\lambda = 0.6$ of tax revenues.²²

Note that the income tax does not distort labor supply in our setting, since we assume the latter is exogenous. This simplification is obviously not a good one for understanding the welfare consequences of changes in tax rates, but because our current focus is on wealth accumulation and its distribution in the population we do not think that it is a major shortcoming.

6.4 Idiosyncratic discount rates and returns to capital

Finally, we calibrate the processes for the discount factor (β) and for the returns to capital (η) to match the right tail of the wealth distribution in the initial steady state. Intuitively, the discount-factor distribution

²²About 60% of total federal outlays are mandatory spending, the bulk of it on Social Security, Medicare, Medicaid, and income security programs (CBO, 2015). The remainder is spent on the Department of Defense and other government agencies as well as on interest payments.

affects the entire asset distribution. In terms of effects on the right tail of the distribution, both discount-factor and return heterogeneity are crucial and, as discussed in 5 above, they influence its Pareto tail coefficient. Return heterogeneity does not play a crucial role for the left tail of the wealth distribution where assets are essentially zero.

To explain how we discipline our parameter selection based on the data at hand, note first that variation in either β or η generates right-tail wealth inequality. Second, persistence in these parameters is a particularly powerful force toward dispersion. Ideally, one would estimate the entire η distribution based on individual panel data on asset returns, and one would want to also use panel data on saving rates. Since we do not have U.S. data of this sort we did not follow this route in this paper, but we are hopeful to follow this strategy in future work. We do motivate our assumptions here based on the two papers using Norwegian and Swedish data cited above (Fagereng et al. (2015) and Bach et al. (2015), respectively), which both strongly argue that there is a significant idiosyncratic returns component. These papers also argue that there is persistence in returns but their interpretations of this finding differ. The possibility that different households have different skills at return finding (an interpretation made in Fagereng et al. (2015)) is radical relative to the finance literature, and although we do not want to rule out that this hypothesis is true, we opted for the more conservative assumption that idiosyncratic return differences are iid, while allowing persistence in β s.

We use an AR(1) structure for the discount factor. Thus, from the perspective of dispersion we have three key parameters to calibrate: the variance and persistence of β and the variance of η . First, we follow in selecting the persistence of the β process based on what seems a priori reasonable given a generational structure. Second, we target two wealth-distribution statistics to obtain the remaining two variance elements (for β and η): the Pareto tail coefficient and the fraction of total wealth held by the 10% richest. This identifies our parameters. We now describe the details.

We posit that β follows a Gaussian AR(1) process:

$$\beta_t = \rho^\beta \beta_{t-1} + (1 - \rho^\beta) \mu^\beta + \sigma^\beta \epsilon_t^\beta, \quad \epsilon_t^\beta \sim N(0, 1).$$

Moreover, we assume that the idiosyncratic factor in the return to capital is normally distributed: $\eta_t \sim^{i.i.d.} N(1, \sigma^\eta)$. Importantly, all these parameters are fixed over time (by varying them freely we could of course track the evolution of the wealth distribution more or less exactly). The mean discount factor determines the equilibrium capital-output ratio and we set it to $\mu^\beta = 0.92$ to match a ratio of capital to net output of about 4 in the initial steady state. The calibrated stochastic- β parameters are $\rho^\beta = 0.992$ and $\sigma^\beta = 0.0019$, implying that the standard deviation of the cross-sectional distribution of discount factors, which does not vary over time, is 0.0148. Moreover, the choice of ρ^β implies that roughly one third of the gap between a given discount factor and the average discount factor is closed within a generation. The idiosyncratic noise in the return to capital is set to equal $\sigma^\eta = 0.725$, implying that the gross (pre-tax, net of depreciation) return on capital $(1 + r^* \eta)$ lies in the interval $[0.9874, 1.1437]$ for 90% of all agents in the initial steady state. Interestingly, although these parameters were selected based on the procedure outlined above, the implied idiosyncratic variation of returns in our calibration turns out to be close to the amount found by Fagereng et al. (2015) in Norwegian data; see, for example, Panel C of Table 1 in that paper. Bach et al.

Table 1: Matching the 1967 wealth distribution as a steady state

Parameter	ρ^β	σ^β	σ^η		\underline{a}	χ_0
Value	0.992	0.0019	0.725		-0.24	0.075
Target	Top 10% share	Top 1%	Top 0.1%	Top 0.01%	Bottom 50%	Fraction $a < 0$
Data	70.8%	27.8%	9.4%	3.1%	4.0%	8.0%
Model	70.6%	28.1%	9.5%	2.9%	3.1%	7.0%

(2015), moreover, find roughly comparable amounts of variation in Swedish data.

To summarize, Table 6.4 lists the values of the five parameters (persistence and standard deviation of the discount rates; standard deviation of return shocks; the borrowing constraint; and the probability of zero income) calibrated to match as closely as possible six features of the initial steady-state wealth distribution: the shares held by the top 10%, the top 1%, the top 0.1%, the top 0.01%, and the bottom 50% as well as the fraction of the population with negative net wealth. The fit is excellent at both ends of the distribution.²³ To the extent that the right tail of the wealth distribution has a Pareto tail, we are therefore also matching the Pareto coefficient governing its thickness, because this coefficient is pinned down by the ratio of the top 0.01% share to the top 0.1% share, or the ratio of the top 0.1% share to the top 1% share, both of which are roughly one-third, both in the model and in the data.

Two comments are in order. First, when solving the model numerically we truncate the β and η distributions to ensure that the consumer’s optimization problem is well-defined (with finite present-value utility) and that a stationary distribution of wealth emerges. Unlike in a standard Aiyagari economy without heterogeneity in preferences, in our model some agents temporarily have discount rates that are smaller than the rate of return, a necessary condition for generating a Pareto tail in the wealth distribution (see the discussion in Section 5). It follows that the support of the stationary wealth distribution is not bounded from above. In practice, we use a large enough upper bound in our numerical implementation so that the resulting truncation error is negligible.²⁴

Second, if our goal were solely to match the Pareto coefficient in the right tail of the wealth distribution, it would be excessive to calibrate as many as five parameters to match features of the wealth distribution. But the tail coefficient is not a sufficient statistic for wealth inequality unless the entire distribution is (counterfactually) Pareto-shaped: even if, say, the top 1% of the wealth distribution can be described exactly by a Pareto distribution, the tail coefficient determines only the distribution of wealth within these top 1% but not the fraction of total wealth held by the top 1%. While stochastic discount factors are the main force driving the shape of the upper tail in the initial steady-state wealth distribution, to achieve our objective of replicating the distribution of wealth on its entire domain we found that introducing in addition a reasonable amount of randomness in returns helped to improve the fit. Moreover, because ownership of primary residences and poorly diversified private equity account for a sizable fraction of net

²³The data on top wealth shares in Table 6.4 is from Saez & Zucman (2016), who use a capitalization method to calculate them. Because this method is unreliable for a breakdown of the bottom 90%, the other data moments are based on survey data (SCF and precursors); see Kennickell (2011).

²⁴Appendix A describes in detail our numerical procedure.

wealth, we view randomness in returns as a realistic feature of individual asset accumulation.²⁵

7 Results

In Section 6, we showed that our model framework, when properly calibrated, can replicate wealth heterogeneity, including the Pareto-shaped right tail, as well as other macroeconomic moments in the initial steady state. We proceed in this section to report on our main result: the evolution of the wealth distribution in the our model economy contrasted with the data. Subsequently, we employ counterfactual analysis in order to decompose those overall changes and identify the key drivers of movements in the wealth distribution.

7.1 Benchmark transition experiment

We summarize the findings from our main experiment in a set of tables: Tables 2-5. The tables differ in terms of the moments of the data we look at and the particular data set we compare the model to; the use of different tables is motivated by different coverage of the different data sets and should help readability.²⁶

Let us first look at the top part of the distribution and compare to SCF data, which is available consistently from 1989. Table 2 reports the results. As illustrated above, the SCF shows increases in the wealth shares held by the top 10%, the top 1%, and the top 0.1%.

Table 2: Change in Top Wealth Shares (SCF)

	Data (SCF)			Model		
	Top 10%	Top 1%	Top 0.1%	Top 10%	Top 1%	Top 0.1%
1989	67.1	30.1	10.8	73.4	30.4	10.2
2013	75.3	35.8	13.5	79.2	37.5	13.5
Change	8.2	5.7	2.7	5.9	7.2	3.3
Relative Change	12.2%	19.1%	25.4%	8.0%	23.6%	32.7%
Fraction of Rel. Change Explained by Model				65.3%	123.5%	128.6%

Survey of Consumer Finance (SCF) data as reported by Saez & Zucman (2016). Wealth shares are displayed in percentage points. For example, the top 1% controlled 30.1% of all wealth in 1989. By 2013, they controlled 35.8% of all wealth, an increase of 5.7 percentage points or 19.1% in relative terms. In the model, their share increased from 30.4% to 37.5%, an increase of 7.2 percentage points or 23.6%. Thus, the model explains a fraction $\frac{23.6}{19.1} = 123.5\%$ of the cumulative increase for this group.

Here the model under-predicts the cumulated increase in inequality over the whole period for the 10% group (by about a third) but over-predicts it for the top two groups (by around 25%).

²⁵See Moskowitz & Vissing-Jørgensen (2002), who document extreme concentration of private equity (its total value being similar in magnitude to public equity): 75% of total private equity is held by households for whom it accounts for the majority of their total net worth and entrepreneurs invest on average more than 70% of their private equity holdings in a single company. See also the discussion in McKay (2013) regarding the distribution of returns on mutual funds.

²⁶We also provide time series in the Appendix; see Figure 10.

Looking at the same population percentiles and comparing the model data to the Saez & Zucman data, we can now go back and cumulate wealth increases from 1967. Table 3 gives the key numbers.

Table 3: Change in Top Wealth Shares (Saez & Zucman)

	Data (Saez & Zucman)			Model		
	Top 10%	Top 1%	Top 0.1%	Top 10%	Top 1%	Top 0.1%
1967	70.8	27.8	9.4	70.6	28.1	9.5
2012	77.2	41.8	22.0	79.0	37.3	13.4
Change	6.4	14.0	12.6	8.4	9.2	3.9
Relative Change	9.0%	50.4%	134.0%	11.9%	32.5%	41.0%
Fraction of Rel. Change Explained by Model				132.2%	64.6%	30.6%

Data based on the capitalization method estimates by Saez & Zucman (2016).

Here the model over-predicts the increase in the fraction of wealth held by the top 10% by about one third, whereas it under-predicts the increases for the two top groups, by about one and two thirds, respectively. Clearly, in terms of the model’s relative performance across groups, the two data sets give different answers, but it is comforting that the under- or over-predictions of the model are not systematic across data series. In terms of shorter-run dynamics, wealth inequality in the Saez & Zucman data set shows a U-shaped pattern within the period, with a trough in the late 1970s. The model’s dynamics deliver a U-shaped pattern in growth rates but not in levels; we will discuss the shorter-run aspects of the model-data comparison in more detail in Section 8.3 below.

The results for the bottom 50% of the population, where we do have a consistent data series from the SCF beginning in 1967, are contained in Table 4.²⁷ The bottom 50% have lost a little over two thirds; the model accounts for about two thirds of this decline in wealth.

Table 4: Change in Bottom 50% Wealth Share

	Data (SCF)	Model
1967	4.0*	3.1
2010	1.1	1.4
Change	-2.9	-1.7
Relative Change	-72.5%	-55.3%
Fraction Explained		76.2%

* Data point equals the median of estimates based on SCF precursors in the 1960s, as reported by Kennickell (2011).

Finally, looking at the very richest, and here only Saez & Zucman have data, we see in Table 5 that the model’s performance is still qualitatively correct but now the quantitative under-prediction is more sizable.

The model predicts an increase in the fraction of wealth held by the top 0.01% by about a half, whereas in the Saez & Zucman data set the increase is fivefold. Clearly, although—as suggested above—

²⁷The method of Saez & Zucman’s unfortunately does not allow for a breakdown of the bottom 90% into subgroups.

Table 5: Change in Top 0.01% Wealth Share

	Data (Saez & Zucman)	Model
1967	3.1	2.9
2012	11.2	4.2
Change	8.1	1.3
Relative Change	261.3%	45.8%
Fraction Explained		17.5%

the capitalization method underlying the data may exaggerate the increases in wealth for the richest, this discrepancy is a major one unlikely to be solely due to mismeasurement and it does not appear like the present model is well suited for capturing the bulk of how much the very, very richest have gained. In Section 8.3 we look at a key candidate model extension—the notion that the very richest have received (much) higher returns on wealth—that can help the model in the right direction for this group.

Before moving on, let us also point out that the model’s implications for aggregate wealth, though not in focus here, are broadly in line with data, thus showing a steady rise, ignoring shorter-run movements; Figure 11 in the appendix shows the time series.

7.2 Counterfactuals

Changes in three structural factors—earnings risk, top earnings inequality, and tax progressivity—drive the transitional as well as long-run dynamics in the model economy. To assess which of these is the most important quantitatively, we conducted three experiments in which only one of the three structural factors is allowed to change, the other two being held constant instead at their 1967 values. Which of these changes is the main driver of increases in wealth inequality, particularly in the upper reaches of the distribution? As we shall see, the main driver of changes in the right tail of the wealth distribution is changes in taxes. Increases in earnings risk, on the other hand, *reduce* top wealth inequality, other things equal.

Table 6 summarizes the results of the three experiments, quantifying how much each of the factors contributes to the changes in the wealth shares over the time period 1967–2012.²⁸

Table 6: Fraction of change in wealth shares explained by model: decomposition by channel

	Earnings risk	Top earnings	Taxes	Combined
Top 10%	−0.78	0.22	1.89	1.32
Top 1%	−0.19	0.05	0.82	0.65
Top 0.1%	−0.08	0.03	0.35	0.31
Top 0.01%	−0.04	0.03	0.16	0.18
Bottom 50%	−0.21	0.33	0.55	0.78

To understand the numbers in the table, focus on the share of total wealth held by the richest

²⁸The dynamics are graphed in Figure 12 in the appendix.

percentile. Saez & Zucman (2016) measure an increase in this share from 27.8% to 41.8% from 1967 to 2012. Over the same time period, allowing for changes only in earnings risk and keeping all other parameters fixed at their initial steady-state values, the model predicts a decrease from 28.0% to 25.2%. Changes in earnings risk therefore explain a fraction $\frac{25.2-28.0}{28.0}/\frac{41.8-27.8}{27.8} = -0.19$ of the actual change.²⁹ Again, the observed increases in earnings risk reduce inequality, moving it in the opposite direction from the observed changes! (Separate increases in either the persistent or transitory components of earnings risk also reduce inequality.) Instead, as can be seen for all the different distributional statistics, the main driver of the surge in wealth concentration is the changing U.S. tax system. The increase in top earnings inequality (parameterized by changes over time in the the Pareto tail coefficient κ_t on labor income) has worked in the same direction, although the effect of this channel is much smaller.

Why does an increase in earnings risk reduce wealth inequality? As noted in Section 1, persistent heterogeneity in discount rates is a powerful force driving the wealth distribution apart: with permanently different discount rates and complete markets against earnings risk, the most patient would eventually hold all the economy’s wealth.³⁰ Earnings risk, then, is a friction, or glue, that keeps the distribution from flying apart altogether as in Becker (1980)’s work cited in Section 1. This risk operates especially strongly at the low end of the wealth distribution, where poorer consumers save to move away from borrowing constraints when earnings risk is larger.

In our model higher earnings risk also generates a thinner right tail in the wealth distribution because the resulting increase in aggregate precautionary savings drives down the equilibrium interest rate. This drop in the interest rate shifts the distribution of savings propensities to the left, particularly for the well-insured wealthy consumers for whom wage risk is largely immaterial and who therefore have essentially linear decision rules. As discussed in Section 5, the Pareto tail coefficient, ζ , is defined implicitly by the equation $\mathbb{E}[s^\zeta] = 1$, where s is the (asymptotic) marginal propensity to save out of wealth. As s falls for all discount-factor types, ζ must increase to compensate, i.e., the Pareto tail becomes thinner.³¹

Why have changes in the tax system induced such large changes in wealth inequality? Note first that the average tax rate (i.e., total tax revenues as a fraction of net GDP) in our model increases from 0.23 to 0.27 over the period 1967–2012. An increase in average taxes tends to reduce effective earnings risk (because the tax is multiplicative), increasing inequality for the same reason (but in the opposite direction) that the observed increases in (pre-tax) earnings risk reduce inequality. This effect, however, is a small one unless the average tax rate changes dramatically. Much more important quantitatively is the dramatic decrease in tax progressivity, where even small changes have large effects on inequality, especially at the high end of the wealth distribution. There are both partial- and general-equilibrium effects at work here. Starting with the latter, it is well known in the context of complete-markets models without discount-factor or wage heterogeneity that progressivity in the tax rate on saving is a strong force toward long-run equality, whereas mere proportional taxes are consistent with any distribution of

²⁹Note that the fractions generally do not add up to the fraction explained when feeding in all observed changes at the same time, as in our benchmark experiment. The remainder is due to interaction effects in general equilibrium.

³⁰In our model, idiosyncratic returns are statistically independent across time, but were they persistent they would act much like persistent heterogeneity in discount rates to spread out the wealth distribution, because discount rates and returns enter similarly in consumers’ Euler equations.

³¹Nirei & Aoki (2016) observe the same effect.

wealth as a steady-state equilibrium.³² The mathematical intuition behind the force of progressivity is particularly clear in a simple case where the marginal tax rate is strictly increasing in wealth. Here, because all consumers face the same market rate under complete markets (and have the same discount rates and wage incomes), they also need to have the same net of tax return if their consumption levels are all constant (or growing at a common constant rate); hence they need to have the same wealth in the long run. This mechanism is still present in a more general model such as the present one, which has incomplete markets and differences in wages and discount rates, though with less long-run poignancy: a strictly increasing marginal tax rate is still consistent with long-run wealth inequality.

Turning to the partial equilibrium analysis, which influences the transition path too, note that the marginal saving propensity (out of initial-period assets) for a well-insured consumer with power utility is approximately $\beta(1 + r(1 - \tau'(y)))$ raised to a positive power, where β is the consumer’s current discount factor and $\tau'(y)$ is the consumer’s current marginal tax rate.³³ This tax rate varies with the consumer’s income, y , but it is persistent over time because income is persistent. Tax progressivity, therefore, generates persistent differences across consumers that act like persistent differences either in the consumers’ after-tax rates-of-return, $r(1 - \tau'(y))$, or, equivalently, in consumers’ discount factors. Consequently, *decreases* in progressivity have the same effect as increasing the spread of discount factors, a powerful force for generating differential savings behavior and as a result higher wealth inequality. In Figure 14 in the appendix we break down the effects of progressivity into a direct effect—the return differences implied by changed progressivity—and that on behavior—marginal saving propensities excluding the return effect—by showing the effects of the latter only and the effects of the former only, along with the full equilibrium response. Clearly, the former is most important for the very richest and hence for changes in top wealth inequality.

In sum, among the different drivers of wealth inequality considered in the benchmark experiment it is clear that decreasing tax progressivity is key: it spreads out the resources available to consume and invest and it increases the relative return of the rich on any given saving.

In a representative-agent model the increase in average taxes would lead in equilibrium to a decrease in the capital-to-output ratio, but it does not in our heterogeneous-agent model for three reasons. First, the (smallish) increase in average taxes does not offset the even larger increase in the riskiness of pre-tax earnings, leading to more precautionary savings in the aggregate. Second, decreasing tax progressivity increases the returns to saving, a particularly powerful force for the rich. Third, the increasingly “thick” right tail in earnings provides the rich (who tend to be those with high earnings) with additional resources for saving. These three forces combine to generate a fairly large increase in the ratio of capital to net output over the period 1967–2012.

Finally, if one looks at the wealth holdings of the bottom 50% of the population, the bulk of the decrease is again accounted for by the decrease in tax progressivity, while the movements in the aggregate capital-output ratio are mostly accounted for by the increase in earnings risk.³⁴

³²Total wealth is of course pinned down so that the return to saving equal the discount rate, abstracting from consumption growth.

³³With $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$, the power is $1/\sigma$.

³⁴Figure 13 in the appendix displays these results.

8 Extensions

We now visit a number of robustness exercises and extensions. First, we look at an aggregate production function with a non-unitary elasticity of substitution between capital and labor; our benchmark Cobb-Douglas (the unitary case) function does take a particular stand on the dynamics of the returns to capital. We find that this mechanism does not appear very promising for understanding the data at hand.

We then weaken the consumers' ability to predict changes in their environment. In particular, in our benchmark experiment we assume that consumers in 1967 could predict the future paths of the tax schedule and the degree of idiosyncratic risk, arguably a very strong assumption, so it is interesting to compare this case to one with more limited abilities to predict. Here, our finding is that a model with entirely myopic expectations (the current policy/risk environment is expected to last forever) behaves almost like our benchmark environment.

The benchmark experiment in our paper emphasized the secular changes in inequality over the sample period, i.e., the cumulative increases in the wealth fractions attributable to the richest percentiles of the population. We found that the model could account for a significant share of the observed changes (and that the discrepancies depend on which data one compares to) and that the key factor was the decrease in tax progressivity that began in the late 1970s. At the same time, as we pointed out but did not focus on, the model does not do as well in replicating some of the shorter-run movements in inequality. In particular, in the Saez & Zucman data set (see Figure 1), there is a fall in inequality from 1967 to just before 1980 and since then an increase. This U shape is not captured by our model: the model has a kink, but in the growth rate of inequality (an initial near-zero rate and a later significant positive rate). To examine further mechanisms behind a possible U shape, we consider rate of return variation driven by two factors: inflation (which has an effect via progressivity) and stock-market valuation (where richer households benefit more from market increases since they hold more risky assets). These two mechanisms are harder to embed in a choice framework, let alone one with a general equilibrium, and our extension here is therefore a first pass where we use partial-equilibrium analysis and hard-wired differences in portfolio choice between rich and poor households. We find that the valuation channel holds important promise for understanding the dynamics of inequality but that the inflation channel appears to be of second order, despite the large movements in inflation during the period.

8.1 Robustness to the elasticity of substitution in production

The stability of the fraction of income accruing to labor, for a long time a central pillar of macroeconomic models, has recently been questioned. Karabarbounis & Neiman (2014b), among others, document a visible (though not large) decline in the labor share. Using a production function with a constant elasticity of substitution (CES), they estimate an elasticity of substitution between capital and labor of 1.25. To look into the possibility of a falling labor share, we use a standard CES production function,

$$F_{CES}(K_t, L) = A_{CES} \left(\alpha_{CES} K_t^{\frac{\sigma-1}{\sigma}} + (1 - \alpha_{CES}) L^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (8)$$

where A_{CES} and α_{CES} are chosen such that the initial steady state is identical to the Cobb-Douglas benchmark. Over time, there is capital deepening, leading to a lower labor share because the elasticity of substitution is above one. We find, however, only very small differences as compared to the Cobb-Douglas benchmark (see Table 7).³⁵

Table 7: Robustness to the input substitution elasticity and to myopia

		Top 10%	Top 1%	Top 0.1%	Top 0.01%	Bottom 50%	$\frac{K}{Y}$	r
1967		70.6	28.1	9.5	2.9	3.1	3.74	6.56
	Benchmark	79.0	37.3	13.4	4.2	1.3	4.24	5.41
2012	CES	78.7	36.9	13.2	4.1	1.4	4.28	5.57
	Myopia	77.1	36.3	13.3	4.2	1.6	4.21	5.47

Wealth shares and the interest rate r are reported in %. This table compares various statistics from the benchmark model transition to alternatives. In the benchmark transition experiment, the production technology is assumed to be Cobb-Douglas and agents have perfect foresight. The row labeled 'CES' reports results from a model with CES production technology. The row labeled 'Myopia' reports results from a transition experiment in which agents are completely myopic about the future, assuming present prices, as well as the parameters of the earnings process and the tax schedule, will prevail.

Capital deepening leads to a smaller reaction of the interest rate, so the rise in the capital-output ratio is slightly larger in equilibrium and the Gini coefficient on gross income increases a small amount more (relative to the benchmark).³⁶ At the same time, we find that top wealth shares increase more slowly; unlike for the decline in tax progressivity, higher equilibrium interest rates induce more savings across the whole wealth distribution. In other words, at least over the time frame considered, the saving of the poor tends to be more elastic with respect to the interest rate than the saving of the rich. Overall, though, the message here is that the quantitative effects of considering a different elasticity of substitution are very small.

8.2 Robustness to agents' abilities to predict policy and risk

It is surely bold to assume that agents have perfect foresight on the entire path of the tax schedule, the parameters governing the earnings process, and the resulting equilibrium prices. To gauge the sensitivity of our findings to this assumption, we computed the transitional dynamics under complete myopia, i.e., a polar opposite case in terms of agents' ability to predict. That is, in every period agents believe that the current environment will prevail forever and, accordingly, they are surprised to learn about their forecasting mistake in the subsequent period.³⁷ Table 7 shows the effects of myopia in the last row.³⁸ Clearly, the differences are very small. We conclude that the perfect-foresight assumption is not critically driving the results in the benchmark experiment.

³⁵Figure 15 in the appendix shows the time series.

³⁶In addition, the gross labor share falls by about one percentage point over the period 1967–2012 in our model, though the net labor share actually rises a little. Karabarbounis & Neiman (2014a) report that since 1975 the gross labor share in the U.S. has fallen by about five percentage points and the net labor share by about two-and-a-half.

³⁷See Section 4.2 for an exact description of how this experiment is conducted.

³⁸Figure 16 in the appendix shows the time series.

8.3 Changes in returns: a preliminary look at valuation and inflation effects

In this section we look at the possibility that returns on saving differ systematically across wealth groups, as indicated in Bach et al. (2015) and Fagereng et al. (2015) as well as via the fact that the tax schedule is progressive and not inflation-indexed. As for the former effect, there seems to be some extent of “increasing returns” to wealth accumulation. The modeling in this section will be based on partial equilibrium and hard-wiring of the features observed in data, and thus the results here should be viewed as providing some directions for potential future work. We first briefly discuss why a full treatment is beyond the scope of the present paper.

First, heterogeneity in the returns to saving is incorporated in the benchmark model but in the form of hard-wired idiosyncratic differences unrelated to wealth (this feature was found to be important for matching the initial wealth distribution). It is challenging to go beyond the hard-wiring of portfolio choice. One could try to square differences in idiosyncratic returns and rational decision making (i.e., why more diversification is not chosen) by explicitly modeling asymmetric information, but to provide a quantitative version of such a theory would amount to a whole project in itself.

Second, as for allowing different portfolios in terms of aggregate returns (nominal vs. real assets, risky vs. risk-less assets), it is arguably equally challenging to provide a quantitative theory based on full portfolio choice. Under some conditions, the more well-insured are consumers to idiosyncratic shocks, the higher a portfolio share will they invest in risky assets. However, the benchmark model displays little sensitivity to risk, for reasons rooted in the causes of the equity premium puzzle. That is, consumers mostly care about returns and not about risk, so that the portfolio choices will be virtually the same across different levels of wealth.³⁹

Third, to provide a general-equilibrium theory one would need to derive return differences that are of the magnitude observed—the equity premium puzzle. There are ways to accomplish this and they rely on considering different consumer preferences (possibly in combination with different assumptions on returns, such as low-probability disaster outcomes or long-run risk components). Such preferences (with habits or non-additivity), however, have not been explored much in terms of their cross-sectional implications and thus one would need to consider a different benchmark model altogether. This is an interesting project in itself but it is also beyond the scope of the present paper. Relatedly, there are ways to computationally handle the incorporation of aggregate risk (e.g., Krusell & Smith (1998)) but to deal with aggregate risk, portfolio choice, and a nontrivial dynamic transition path is really at the frontier of our computational abilities.⁴⁰

Fourth, in reality there is also a choice between nominal and real assets and any inflation-based arguments for differences in returns would need to explain what mixes of real and nominal assets are held at different wealth levels.

For these reasons, in this section we simply assume systematic wealth-related differences in portfolio choice and we assume that the returns to risky and risk-less assets are exogenous; for the inflation

³⁹This point is made in detail, and a full portfolio-choice model is solved, in Krusell & Smith (1997).

⁴⁰Methods based on linearization with respect to aggregates, such as those in Reiter (2009) or, more recently, Ahn et al. (2017), cannot be used since they imply certainty equivalence.

experiment we simply assume that all assets are nominal.

8.3.1 Valuation effects

We thus take the dynamic path of the real interest rate as given from the benchmark transitional experiment, and we also use the equilibrium wages, transfers, and the decision rules. That is, we are given how much people save in this equilibrium at each moment in time and for each outcome of their individual states and we then simply append a return composition to this level of saving. To implement this, we use portfolio weights (for risky and risk-less assets) from Bach et al. (2015), which offers direct data on portfolios.⁴¹ These patterns are summarized in Figure 7.



Figure 7: Portfolio Weights

We then use the U.S. time series on stock market returns (using the S&P and including dividends in the returns) and fluctuations in the short-run real interest rate to generate the return outcomes. These are displayed in Figure 8.

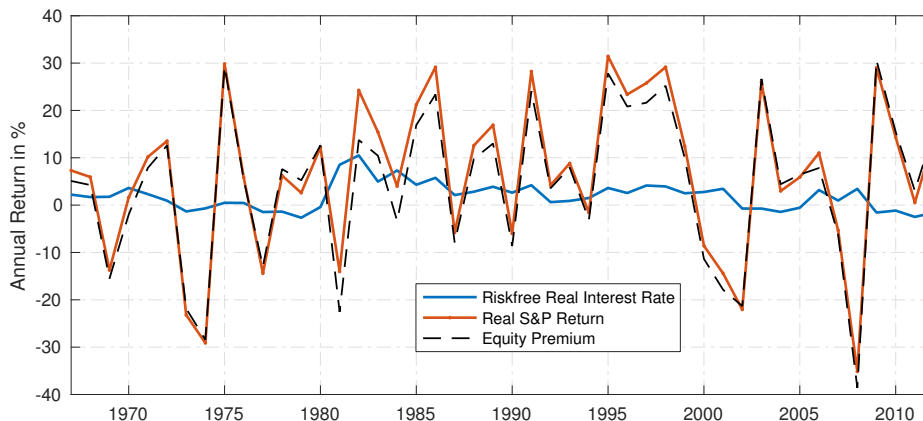


Figure 8: Annual Real Interest Rate and Real S&P Return

We now interpret the portfolios of the initial steady state as having reflected both a riskless and a risky part with the associated return differences taken as an average from the pre-1967 period and we

⁴¹Using the portfolio weights found by Fagereng et al. (2015) yields similar results.

assume a flat capital-gains tax on the risky part of the portfolio following the historical evolution of this tax; the details of the computations are contained in Appendix C.

The results on wealth shares from the valuation experiment are displayed in Table 8; Figure 17 in the appendix shows the time series. The table has two parts: on the left-hand side, we see the initial decrease in inequality (a negative entry indicates a decrease throughout the table) and on the right-hand side the cumulative over the whole period. The valuation effects are tabulated in the third row.

Table 8: Valuation and Inflation Effects

	Initial-to-Trough			Cumulative		
	Top 10%	Top 1%	Top 0.1%	Top 10%	Top 1%	Top 0.1%
Data (Saez & Zucman)	-7.2	-4.9	-2.3	6.4	14.0	12.6
Benchmark Model (BM)	0.0	0.0	-0.1	8.4	9.2	3.9
BM + Valuation Effects	-6.9	-5.4	-2.5	3.8	4.0	1.7
BM + Inflation	-2.5	-3.2	-1.6	5.1	5.5	2.6

This table reports changes in wealth shares in percentage points. The columns titled ‘Initial-to-Trough’ refer to the change in the share of wealth held by a particular group from the start of the time series in 1967 to its trough (which is around 1980). The columns titled ‘Cumulative’ refer to the cumulative change from 1967 to 2012. The data is from Saez & Zucman (2016). The second line reports wealth shares from the benchmark model transition. The third line reports shares from an out-of-equilibrium experiment that adds valuation effects to the benchmark model. The shares in the last row refer to a similar experiment that incorporates the interaction between inflation and the tax code in the benchmark model.

Looking at the cumulative effects first, there are visible effects of portfolio differences across wealth groups on inequality over the period: because the equity premium over the period has been somewhat lower than prior to 1967, inequality goes up a few percentage points less than in the benchmark model. More interestingly, however, the dynamics within the period are strongly accentuated due to valuation movements. From 1967 to 1980, when the trough is reached, the inclusion of valuation effects induces changes in the top shares that are very similar to those in the data for top 10%, top 1%, and top 0.1%. For the extreme right tail (0.01%) of the wealth distribution, the valuation effects do not suffice in generating the large recent run-up in wealth, but it is highly likely that valuation effects can play a role here too, though through various forms of private equity not captured by a broad stock index. We conclude that in order to understand shorter-run movements in the data, valuation effects seem quite promising. Given their importance in accounting for relative wealth shares, moreover, it becomes all the more urgent to explore different theories for their behavioral underpinnings in future work.

8.3.2 Inflation

As background for the table, the derivations in the appendix, Section D, show how progressivity interacts with inflation to lower the relative returns on saving as a function of wealth. The data in Figure 9 shows the rather dramatic movements in inflation over the period.

Turning to the results of these effects over the given period, Table 8 shows the results for inflation in the last row; Figure 18 in the appendix shows the time series. The effects of inflation are similar to those arising from valuation changes, though weaker. There are some cumulative effects, also lowering wealth

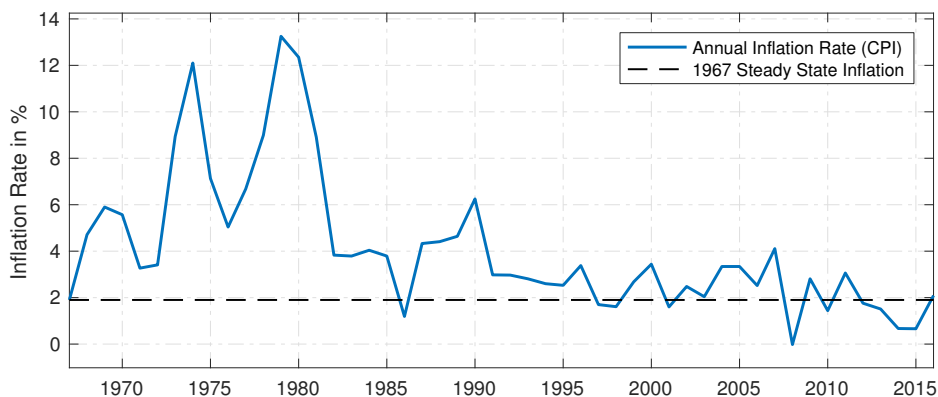


Figure 9: Annual Inflation Rate (CPI)

inequality somewhat relative to the benchmark. Second, the movements in inflation also help generate a U-shaped pattern in the wealth shares of the richest. To understand these findings, note that over the period up to around 1980, there was an upward drift in inflation and progressivity remained at a high level, thus accounting for a downward movement in the wealth shares of the richest. The Volcker disinflation era that followed then amounted to a positive effect on the shares of the richest, though these effects were limited by the fact that progressivity also fell—progressivity is a necessary component of inflation effects.⁴² Going forward, unless both inflation and progressivity rises, inflation will not likely be a major contributor to changes in the wealth shares across the population.

8.4 The long run

We have focused so far on the transitional dynamics of the wealth distribution over the period 1967–2012, but what are the longer-run implications of the changes in earnings risk and, especially, tax progressivity that have occurred over this time period? Table 9 shows the model’s prediction for the evolution of top wealth inequality in the 21st century. In the calculations underlying these results, we have assumed no further changes in either earnings risk or taxes after 2012.⁴³

Table 9: Model Predictions for the 21st Century

	Top 10%	Top 1%	Top 0.1%	Top 0.01%	Bottom 50%
1967	70.6	28.1	9.5	2.9	3.1
2017	79.9	38.5	14.0	4.4	1.2
2100	86.2	50.7	20.9	7.0	0.7
2100: secular stagnation	87.4	51.6	21.0	6.9	0.7

The predictions from the third row of the table are striking: the model suggests that the adjustment to the new fundamentals is far from completion and that wealth inequality is likely to rise even more. As pointed out before, the wealth distribution is a slow-moving object, especially in a setting with random

⁴²Notice that the effects are not continuous as the tax brackets are discrete.

⁴³Figure 19 in the appendix shows the time series.

growth in which the right tail of the wealth distribution is Pareto-shaped. Changes in fundamentals (such as the structure of taxes) that influence the consumption-savings decision differently for consumers with different wealth levels are bound, then, to have long-lasting effects. The contrast between the behavior of the wealth distribution over the transitional period and the eventual long-run steady-state wealth distribution (assuming an unchanged environment going forward) underscores the hazards of looking solely at steady states when attempting to quantify how fundamentals affect wealth inequality.

Table 9 also contains a row labeled “secular stagnation”. The results in this row are computed based on assuming a lower growth rate in output from 2017 and on. This exercise appears relevant for two reasons: (i) secular stagnation is increasingly considered a benchmark scenario for macroeconomic analysts (though we take no stand on it here) and (ii) it has an interesting connection to Piketty’s “ $r - g$ ” theory elaborated on in Piketty & Zucman (2015). The idea of the latter is that the real interest rate net of the growth rate of real GDP would increase wealth inequality because capital owners become richer the higher is the return on their savings, r , whereas workers’ economic strength increases with wage (or GDP) growth.

Specifically, we consider a permanent 0.5% decrease in TFP growth and we work out the general-equilibrium implications of the secular-stagnation scenario. A change in the growth rate of TFP has nontrivial implications for inequality in the present model. Let us first consider the long-run effects. Lower TFP growth will lower the gap between the long-run interest rate and the growth rate through optimal saving if the utility function has more than logarithmic curvature (as we assume: $\sigma = 1.5$). However, the effect is not just given by the approximation $r - g \approx 1/\beta - 1 + (\sigma - 1)g$ —a linear approximation to the long-run Euler-equation relation—because of the presence of taxes (so r in this equation should be net of the tax). It turns out, in particular, since wealth is so concentrated and taxes are progressive, that the relevant Euler equation is that of the consumers in the highest tax bracket. This means that the change in r will be larger, the higher is the marginal tax rate of the richest, and it also means that the net return to the poorest therefore changes by more than for the richest, as their marginal tax rate is lower. Hence, the fall in g implies that the net return of the poor falls by more than for the rich in the long run, hence increasing wealth inequality. It also means, for our parameter values, that $r - g$ actually *falls* somewhat in the long run. Given the rather modest 2017 marginal tax rates for the richest, this effect is not huge.⁴⁴ Second, given that the effect on the interest rate takes time, the transitional dynamics following a fall in g feature a small and direct effect in the opposite direction ($r - g$ actually rises in the short run).

The results in the table indicate a modest effect of our assumed secular stagnation on inequality: the effect on the very richest is close to nil and the effect on the 10% and 1% richest is to increase their shares by around 1 percentage point each.⁴⁵ As pointed out, in the short run the effects are smaller and, in fact, have the opposite sign for a few decades.⁴⁶

The experiments with a lower g , which deliver a lower long-run $r - g$ (given $\sigma > 1$), indicate a large

⁴⁴We also experimented with a fall in g beginning in 1967, when tax progressivity was much stronger. In that case, the effect on long-run wealth inequality was dramatic and the fall in $r - g$ sizable.

⁴⁵We have, of course, computed the entire transition path, i.e., beyond 2100, and the effects in the very long run are somewhat larger: the top shares all increase moderately (e.g., the share of wealth held by the top 10% goes up by around 5 percentage points).

⁴⁶The time series for the top shares can be found in Figure 20 in the Appendix.

difference between a general-equilibrium and a partial-equilibrium treatment. The partial-equilibrium effect of secular stagnation, i.e., abstracting from indirect effects on r , are expected also be small. We did consider a partial-equilibrium experiment with a temporary (10-year) rise in $r - g$; even here, the finding was surprising in that wealth inequality went down. The chief reason for this is that the higher return on wealth has a more potent effect on the less wealthy, as the net-of-tax return increase is larger for them.⁴⁷

Of course, we urge caution in interpreting Table 9 as a plain prediction for the future, because no doubt the economic environment will not remain unchanged going forward. Various exogenous impulses are possible (e.g., external forces affecting the U.S. interest rate, changes in demographics, and further change in earnings inequality). In addition, the model abstracts from plausible feedback mechanisms. For example, changes in wealth inequality could themselves, via the political process, lead to changes in the structure of taxes. Notwithstanding these points, the long-run analysis contained here does emphasize how powerfully tax progressivity can shape the wealth distribution, particularly in its right tail.

9 Concluding remarks

The determinants of wealth inequality, in particular its developments over the last half a century, are much-discussed recently and a number of new hypotheses have been put forth. This paper takes a “first-thing-first” perspective and asks what established quantitative theory predicts based on the behavior of a number of plausible, and observable, factors over the same period. We thus use a macroeconomic general-equilibrium model with heterogeneous agents—the Bewley-Huggett-Aiyagari setting—to more closely examine a set of candidate explanations for the increase in U.S. wealth inequality over the last 30 or so years. The method we follow is thus to (i) independently measure changes in the environment, such as in the tax code and the earnings processes facing individuals; (ii) feed these into the model assuming that the economy is in a steady state in 1967; (iii) examine the resulting wealth distribution path; and (iv) conduct counterfactuals. We find that the model generates a path for inequality that is reasonably close to that observed, the main exception being that the rise in inequality at the very top of the distribution is under-predicted. The satisfactory performance of the model in predicting the overall path for wealth inequality notwithstanding, the main contribution is the conclusion that the most important factor—by far—behind the developments is the significant decline in tax progressivity that began in the late 1970s.

Declining tax progressivity, together with increasing earnings risk and higher earnings inequality amongst top earnings, can also account for the rise in the capital-to-net-output ratio and at least some of the decline in the (gross) labor share when the elasticity of substitution between capital and labor is larger than one as in Karabarbounis & Neiman (2014b). Our model thus provides an alternative to the central mechanism—declining growth rates—to which Piketty (2014) draws attention in attempting to connect these macroeconomic trends to rising inequality.

Our findings merit several remarks. Although we find that tax progressivity has played a central role in increasing inequality, our model is designed primarily as a positive rather than a normative tool. To evaluate the pros and cons of, say, reversing the changes in tax progressivity, it is important to account

⁴⁷We discuss this experiment briefly in Section E in the appendix.

for the distortions created by labor taxation; in the present setting, labor earnings are exogenous and taxation is levied jointly on all incomes. We do not think that the introduction of distortionary labor taxation would change the model’s predictions for wealth inequality measurably, but it would be central for understanding the welfare consequences of tax changes. Further research contrasting the larger distortions of increased tax progressivity with the accompanying reductions in inequality seems very promising.

To improve the model’s ability to account for the data, we note from the work in Section 8.3 that introducing a richer asset framework could play an important role. We do not include endogenous portfolio choice nor endogenous aggregate asset-price movements in our present setting and we do not consider the choice between real and nominal assets. Asset choice is likely an important aspect for the very top of the distribution, where in addition labor and asset incomes may be inextricably linked in practice. In particular, many modern entrepreneurs hold a large part of their private assets in their own companies, giving them an unbalanced portfolio that pays off greatly in successful states. With better availability of tax records and register data more generally, there are hopes to measure how the idiosyncratic returns to assets have evolved over time in the U.S. and we suspect that the variance here has risen considerably, especially at the top. This is another great avenue for future research.⁴⁸ Thus, although this is in broad agreement with Piketty’s stylized $r - g$ theory emphasizing the rate of return on assets, r , as an important determinant of the relative growth rates of wealth (including human wealth which grows at rate g) of the rich and the poor, we would rather want to emphasize differences in asset returns across different people, net of tax rates (where progressivity is key), in our future work.

Regardless of one’s normative views on wealth inequality, there are many reasons to care about its future course, as there are now many research contributions suggesting that the macroeconomy works quite differently when there is significant heterogeneity among consumers. This goes for fiscal as well as monetary policy; for examples, see Heathcote (2005), McKay & Reis (2016), and Brinca et al. (2016) for fiscal policy and Auclert (2015), McKay et al. (2016), and Kaplan et al. (2016) for monetary policy. The prediction from the present paper is that, barring reverses in the tax code, wealth inequality will go up even further, thus potentially strengthening the case for further research on the heterogeneous-agent approach to macroeconomics.

⁴⁸As we emphasized in the paper, Fagereng et al. (2015) and Bach et al. (2015) have already embarked down this avenue using data from Sweden and Norway.

A Computational strategy

A.1 Dynamic programming problem

The consumers' dynamic programming problem is solved by value-function iteration using Carroll (2006)'s endogenous grid-point method (EGM) on a grid for cash-on-hand and the persistent idiosyncratic shocks (β, p) .

Unlike in the plain Aiyagari (1994) model, the support of the ergodic wealth distribution is unbounded in this framework. We use a log-spaced grid with 100 points for cash-on-hand $(x_i)_{i=1}^{100}$ with a very large upper bound (one million times average wealth) to minimize the truncation error.⁴⁹ Cubic splines are used to interpolate the value function along the wealth dimension.

The grid for the persistent component of individual productivity $(p_j)_{j=1}^{17}$ is chosen to account for the long right tail in earnings. First, we chose the grid points as the 0.0001, 0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 0.925, 0.95, 0.975, 0.99, 0.999, ..., and 0.99999999 quantiles of the unconditional (i.e., cross-sectional) p -distribution (which is a normal). Second, we compute the corresponding grid in actual efficiency units of labor $(\psi(p_1), \dots, \psi(p_{17}))$. Third, given that in the current period $p = p_j$ for $j = 1, \dots, 17$, we use Gauss-Hermite quadrature to integrate over $p'|p$, the value of idiosyncratic productivity in the next period, when updating the value function. In doing so, we use linear interpolation in $\psi(p)$ -space to evaluate the value function off the grid (the value function is much more non-linear in p -space than in $\psi(p)$ -space).⁵⁰

Regarding the discount factor, we choose the grid points $(\beta_m)_{m=1}^{15}$ as the Gauss-Hermite quadrature points of the unconditional (i.e., cross-sectional) β -distribution (this will turn out to be useful when integrating over the joint distribution to compute aggregate wealth). Again, when updating the value function, we integrate over $\beta'|\beta$ using Gauss-Hermite quadrature and linear interpolation in β -space.

In addition to the these three state variables, the setup requires numerical integration over the two idiosyncratic i.i.d. shocks to earnings ν' and capital returns η' (as they affect next period's cash-on-hand x'). As both shocks are normally distributed, we use Gauss-Hermite quadrature once again.

A.2 Computing the ergodic distribution

The focus on tiny population groups such as the top 0.01% of the wealth distribution implies that solving for the ergodic distribution directly is more efficient than simulating a large number of agents and applying the ergodic theorem. In doing so, simulation error is eliminated; instead one can directly control the numerical error by updating the distribution until convergence is reached.

Specifically, note that the EGM entails using a grid for assets $(a_i)_{i=1}^{100}$. Given p_j and β_m , saving a_i is optimal with cash-on-hand $x(a_i; p_j, \beta_m)$ that solves

$$u'(x(a_i; p_j, \beta_m) - a_i) = \beta_m \mathbb{E} \left[(1 + r\eta' (1 - \tau'(y'))) V_1(a_i + y' - \tau(y') + T, p', \beta') | p_j, \beta_m \right].$$

⁴⁹Alternatively, given that the Pareto tail has stabilized at some \bar{x} , one could in principle also impute the distribution for $x > \bar{x}$. However, this did not turn out to be necessary as the log-spaced grid—which works well as the curvature of the value function is high only close to the borrowing constraint—allows for selecting a very large upper bound while keeping the number of grid points computationally feasible.

⁵⁰Note that these linear interpolation coefficients can be pre-computed, resulting in a 17×17 - matrix w^p , where w_j^p are the integration weights for evaluating next period's value function on (p_1, \dots, p_{17}) given that in the current period $p = p_j$.

While the main advantage of the EGM is efficiency ($x(a_i; p_j, \beta_m)$ can be found without maximizing the right-hand side of the Bellman equation), it is also convenient that the savings function is already inverted. First, for all $p_j, \beta_m, \nu_q, \eta_h$ and for all $a_i, i = 1, \dots, 100$, there exists a unique level of asset holdings $a = s^{-1}(a_i; p_j, \beta_m, \nu_q, \eta_h)$ such that saving a_i is optimal.⁵¹ Second, we define a finer grid for asset holdings $(k_i)_{i=1}^{1000}$ and interpolate (using a cubic spline) to find the inverse savings function $s^{-1}(k_i; p_j, \beta_m, \nu_q, \eta_h)$. Note that the borrowing constraint is binding for all $k \leq s^{-1}(k_1; p_j, \beta_m, \nu_q, \eta_h)$. Finally, we can solve for the ergodic distribution $G(k_i; p_j, \beta_m) \equiv \text{Prob}(k \leq k_i | p = p_j, \beta = \beta_m)$ at the grid points $(k_i)_{i=1}^{1000}, (p_j)_{j=1}^{17}$ and $(\beta_m)_{m=1}^{15}$. To simplify notation, we will denote by $G_{j,m}(k_i)$ this conditional cdf evaluated at grid points (p_j, β_m) . This distribution has to satisfy

$$G_{j,m}(k_i) = \int_p \int_\beta \int_\nu \int_\eta G(s^{-1}(k_i; p, \beta, \nu, \eta); p, \beta) d\Gamma_\eta(\eta) d\Gamma_\nu(\nu) d\Gamma_\beta(\beta | \beta_m) d\Gamma_p(p | p_j). \quad (9)$$

Note that p_j and β_m are the realizations of the shock in period $t + 1$ and the integration is over the shock values in period t . Nevertheless, e.g., $\Gamma_\beta(\beta | \beta_m)$ is the correct distribution as for any stationary Gaussian AR(1) process z_t the conditional random variables $z_t | z_{t+1}$ and $z_{t+1} | z_t$ have the same distribution.⁵² Starting from some initial distribution $G_{j,m}^0(k_i)$ and using the short-hand notation $s_{j,m,q,h}^{-1}(k_i) = s^{-1}(k_i; p_j, \beta_m, \nu_q, \eta_h)$, we update until convergence according to

$$G_{j',m'}^1(k_i) = \sum_j w_{j',j}^p \sum_m w_{m',m}^\beta \sum_q w_q^\nu \sum_h w_h^\eta \hat{G}_{j,m}^0(s_{j,m,q,h}^{-1}(k_i)). \quad (10)$$

In (10), w_q^ν and w_h^η are the Gauss-Hermite quadrature weights for the transitory shocks ν and η (normalized to sum to one). The construction of the integration weights for the persistent shocks p and β is based on linear interpolation in $\psi(p)$ - and β -space, respectively (see details below). $\hat{G}_{j,m}^0(\cdot)$ linearly interpolates $G_{j,m}^0(k_i)$ off the grid in the k -dimension.

Integration weights $w_{j',j}^p$ and $w_{m',m}^\beta$. Consider the persistent earnings shock p . Conditional on its value in the next period being $p' = p_j$ for some fixed $j' \in \{1, \dots, 17\}$, the integration over the current period value p is with respect to the distribution of p , conditional on p' , where $p | p' \sim N(\rho^P p' + (1 - \rho^P)\mu^P, \sigma^P)$. Gauss-Hermite quadrature, here with ten sample points, entails evaluating the function of interest $G(s^{-1}(k_i; p, \beta, \nu, \eta); p, \beta)$ at $(\tilde{p}_n)_{n=1}^{10}$, where $\tilde{p}_n = \rho^P p' + (1 - \rho^P)\mu^P + \sqrt{2}\sigma^P \tilde{x}_n$ and $(\tilde{x}_n)_{n=1}^{10}$ are the roots of the Hermite polynomial, and approximating the integral using the associated weights $(\tilde{w}_n)_{n=1}^{10}$ as

$$\approx \frac{1}{\sqrt{\pi}} \sum_{n=1}^{10} \tilde{w}_n G(s^{-1}(k_i; \tilde{p}_n, \beta, \nu, \eta); \tilde{p}_n, \beta).$$

Of course, \tilde{p}_n will in general not lie on the p_j -grid, where the function value is known. Hence, we have to interpolate. Using linear interpolation, we can pre-compute the integration weights $(w_{j',j}^p)_{j=1}^{17}$ we

⁵¹ $s^{-1}(a_i; p_j, \beta_m, \nu_q, \eta_h)$ is defined as the unique a that solves

$$x(a_i; p_j, \beta_m) = a + y - \tau(y) + T,$$

where $y = r\eta_h a + wl(p_j, \nu_q)$.

⁵² That is, the densities satisfy $f_{z_t | z_{t+1}}(x | y) = f_{z_{t+1} | z_t}(x | y)$.

put on evaluating the function of interest at $(G(s^{-1}(k_i; p_j, \beta, \nu, \eta); p_j, \beta))_{j=1}^{17}$ in an efficient manner: for $n = 1, \dots, 10$, locate $j(n)$ such that $p_{j(n)} \leq \tilde{p}_n \leq p_{j(n)+1}$ and compute the linear interpolation coefficient in $\psi(p)$ -space λ_n as

$$\lambda_n = \frac{\psi(\tilde{p}_n) - \psi(p_{j(n)})}{\psi(p_{j(n)+1}) - \psi(p_{j(n)})}.$$

Then, looping over $n = 1, \dots, 10$, add $(1 - \lambda_n) \frac{1}{\sqrt{\pi}} \tilde{w}_n$ to $w_{j', j(n)}^p$ and $\lambda_n \frac{1}{\sqrt{\pi}} \tilde{w}_n$ to $w_{j', j(n)+1}^p$. The construction of the integration weights for β is analogous, except that linear interpolation can be performed directly in β -space.

Computing moments of the distribution. For example, aggregate wealth is given by

$$K = \int_p \int_\beta \left(\int_k k dG(k|p, \beta) \right) f_p(p) f_\beta(\beta) dp d\beta,$$

where $f_p(\cdot)$ and $f_\beta(\cdot)$ are the unconditional (i.e., cross-sectional) normal densities of the persistent shocks p and β . We integrate numerically according to

$$\hat{K} = \sum_{j=1}^{17} \bar{w}_j^p \sum_{m=1}^{15} \bar{w}_m^\beta \left(k_1 G_{j,m}(k_1) + \sum_{i=2}^{1000} \frac{k_{i-1} + k_i}{2} (G_{j,m}(k_i) - G_{j,m}(k_{i-1})) \right). \quad (11)$$

As the discount factor grid $(\beta_m)_{m=1}^{15}$ was chosen as the Gauss-Hermite sample points, we set $(\bar{w}_m^\beta)_{m=1}^{15}$ to be the associated Gauss-Hermite quadrature weights. Recall that the Pareto tail transformation of the persistent earnings component p prompted us to define a grid $(p_j)_{j=1}^{17}$ with a particular emphasis on the right tail. Hence, we (pre-)compute the integration weights $(\bar{w}_j^p)_{j=1}^{17}$ manually: (i) define a very fine equally spaced grid $(\hat{p}_n)_{n=1}^N$ (if, say, $N = 100,000$, this has to be carried out only once) that covers the coarser grid $(p_j)_{j=1}^{17}$; (ii) for all $n = 1, \dots, N$, locate $j(n)$ and compute λ_n as above; (iii) looping over $n = 1, \dots, N$, add $(1 - \lambda_n) f_p(\hat{p}_n)$ to $\bar{w}_{j(n)}^p$ and $\lambda_n f_p(\hat{p}_n)$ to $\bar{w}_{j(n)+1}^p$ ($f_p(\cdot)$ is the pdf of $p \sim N(\mu^P, \frac{\sigma^P}{1-\rho^P})$); and (iv) finally, normalize such that $\sum_{j=1}^{17} \bar{w}_j^p = 1$.⁵³

A.3 Transition experiments

The perfect-foresight transition experiment is computationally straightforward. Given the calibrated initial steady state (K^*, T^*) , the new steady state (K^{**}, T^{**}) is computed under the new exogenous environment. We then search for a fixed point in $(K_t, T_t)_{t=t_0+1}^{t_1}$ -space where $t_1 - t_0$ is chosen to be large enough that $(K_{t_1}, T_{t_1}) \approx (K^{**}, T^{**})$. For each iteration, we first solve for the value functions and corresponding (inverse) savings decisions backwards and subsequently roll the distribution forward, as described in the previous sections for the steady state. Note that now the grids and integration weights for the earnings process components are time-varying.⁵⁴

⁵³Of course one could also use Gauss-Hermite quadrature here, as the corresponding weights and results coincide for all practical purposes.

⁵⁴In particular, as the variance of the innovation term of the persistent earnings component σ_t^P is time-varying, $p_{t|t+1}$ is no longer equal to $p_{t+1|t}$ in distribution (but still normal); hence the integration weights for the decision problem (forward-looking) and the cross-sectional distribution (backward-looking) differ.

The myopic transition experiment is conceptually very different. Given a period t distribution $G_{j,m}^t(k_i)$ and savings decisions $s_{j,m,q,h}^t(k)$ (reflecting factor prices r_t , w_t , transfers T_t and exogenous environment θ_t , all naively assumed to persist forever), $G_{j,m}^{t+1}(k_i)$ is obtained as in (10).⁵⁵ In turn, $G_{j,m}^{t+1}(k_i)$ and θ_{t+1} determine K_{t+1} (thus r_{t+1} , w_{t+1}) and T_{t+1} . The surprised agents expect this new endogenous and exogenous environment to prevail forever and hence we solve the dynamic programming problem given this environment and accordingly obtain $s_{j,m,q,h}^{t+1}(k)$. Note that no fixed point problem has to be solved and the capital stock converges to the same new steady state as under perfect foresight. Theoretically, this strategy could give rise to oscillatory paths of capital. However, this turns out not to be the case in our application.

⁵⁵Again, the grids and integration weights for the earnings process components are time-varying.

B Additional figures

This section contains additional figures and results referred to in the main text.

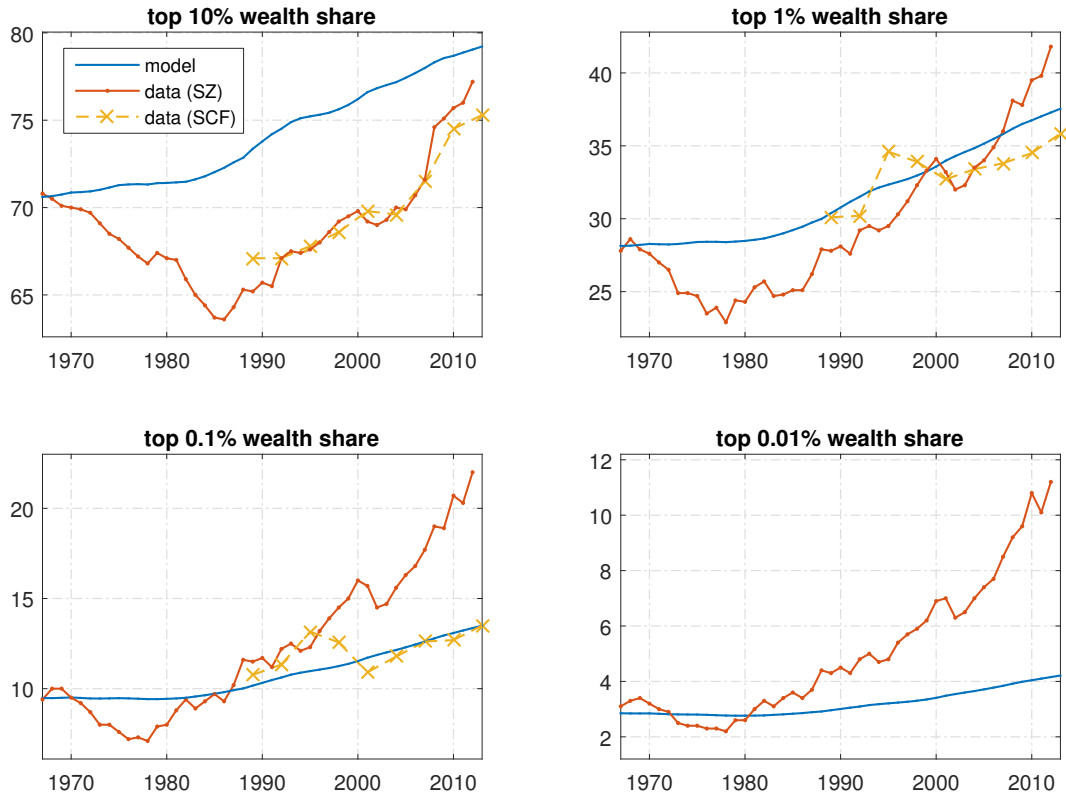


Figure 10: Top wealth shares in %, 1967–2012

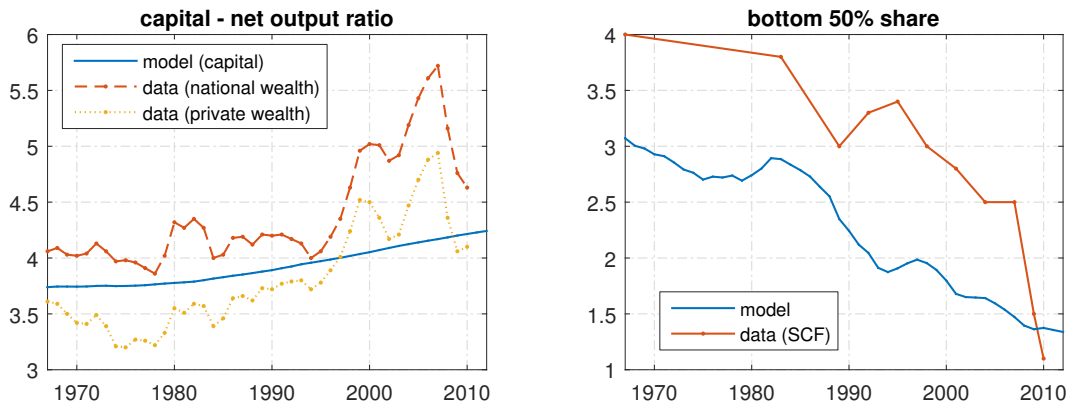


Figure 11: Capital-output ratio and bottom 50% share (in %), 1967–2012



Figure 12: Counterfactual top wealth shares in %, 1967–2012

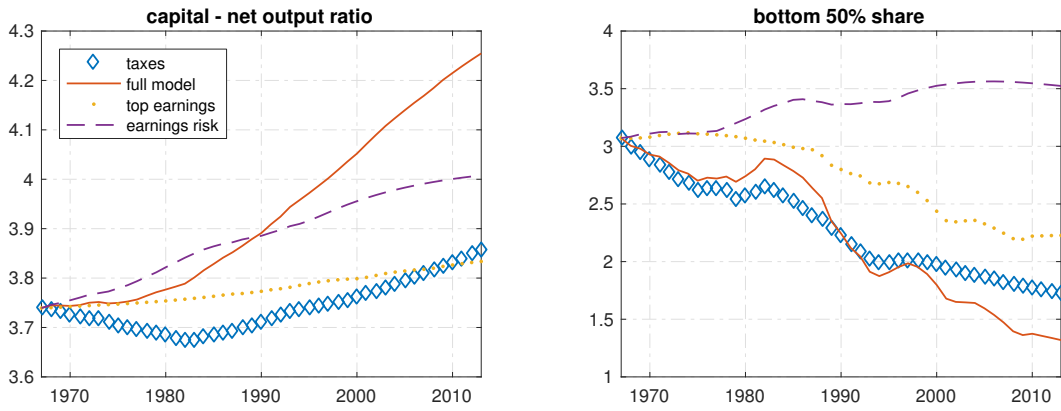


Figure 13: Counterfactual capital-output ratio and bottom 50% share (in %), 1967–2012

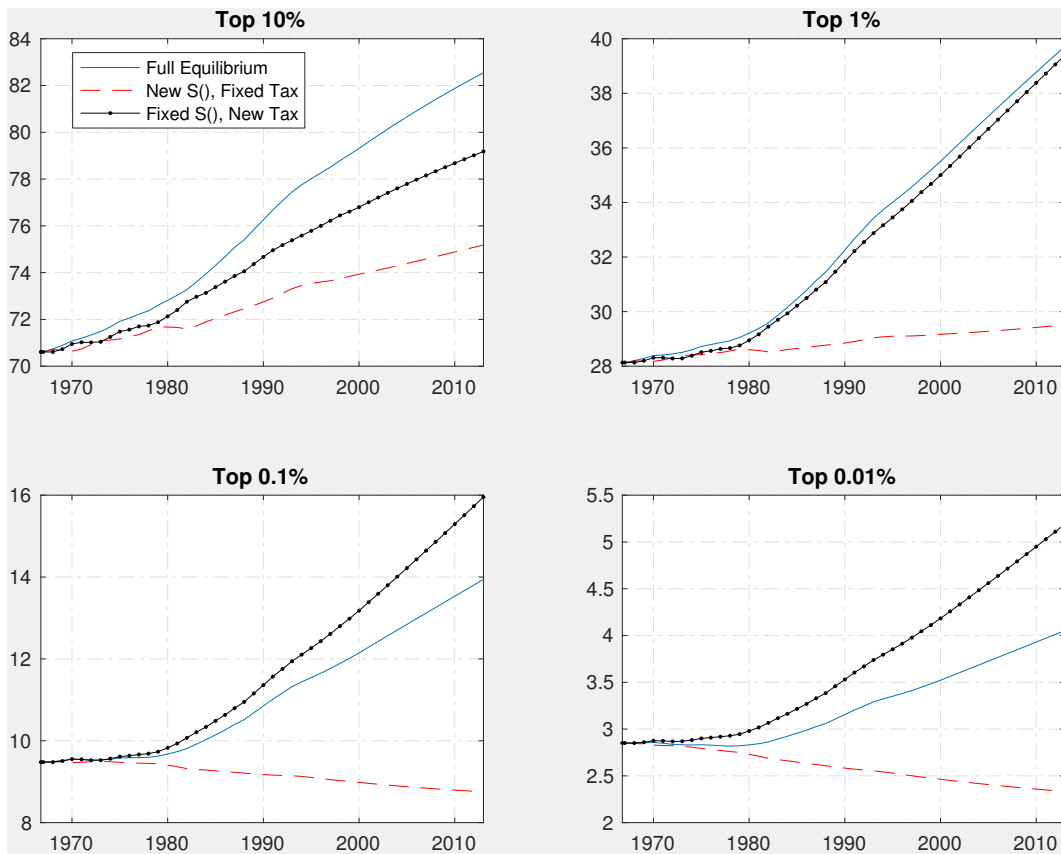


Figure 14: Tax-change decomposition: top wealth shares

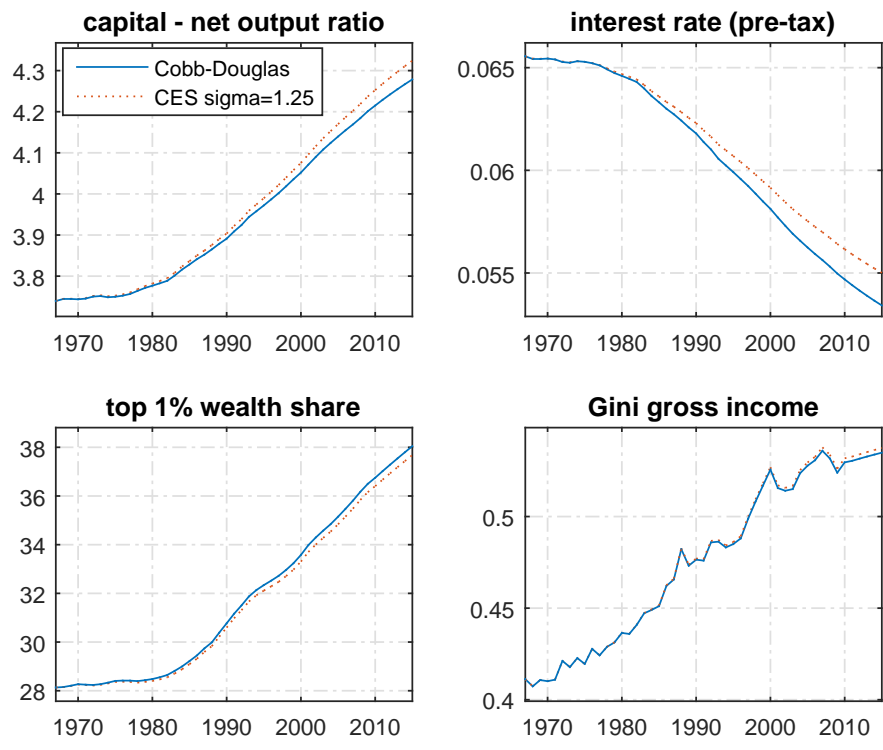


Figure 15: Cobb-Douglas vs. CES production function with elasticity $\sigma = 1.25$

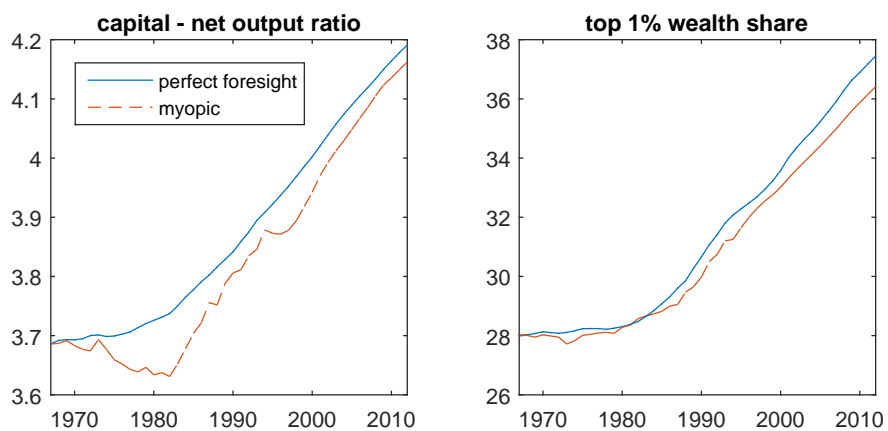


Figure 16: Contrasting prediction ability assumptions

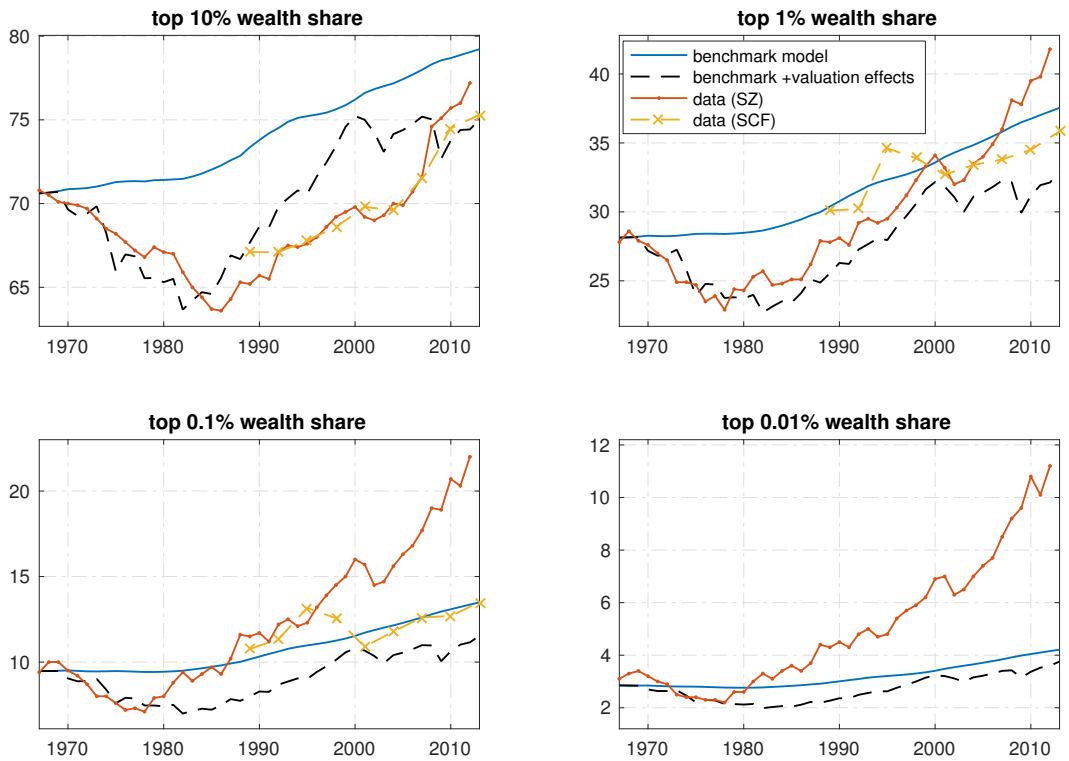


Figure 17: Valuation effects experiment

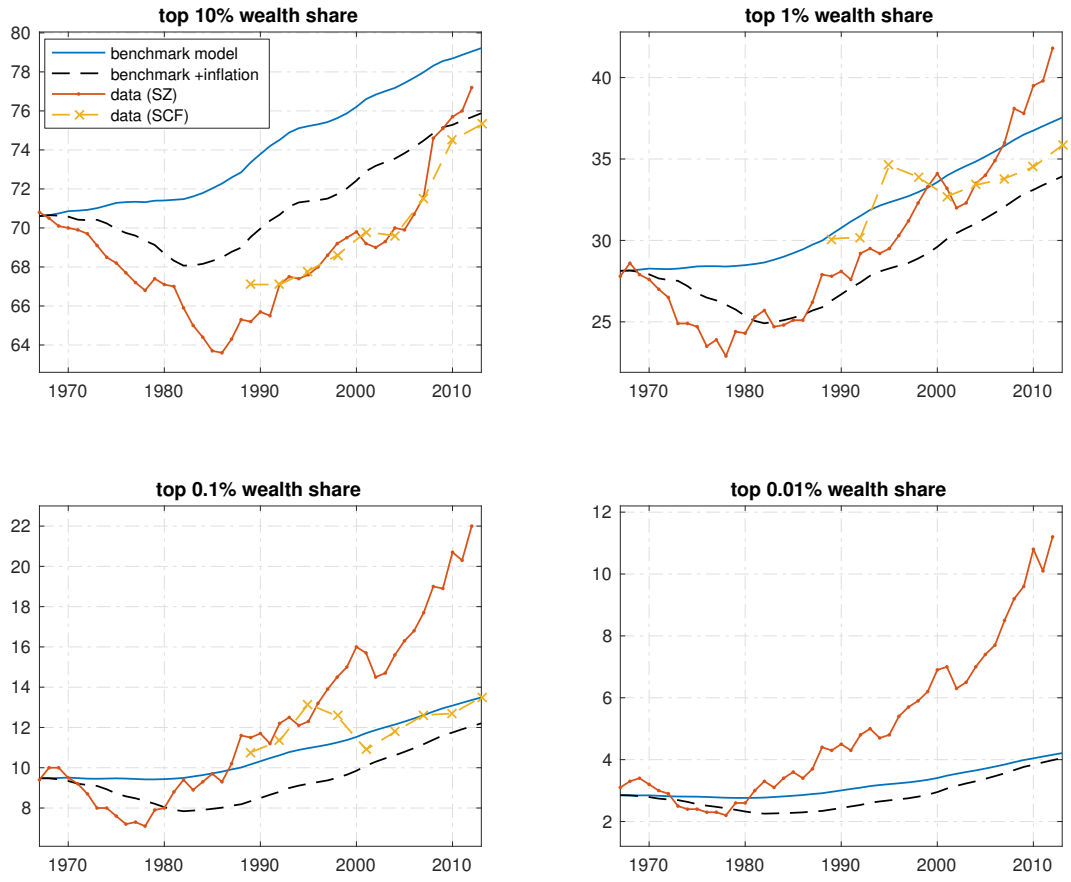


Figure 18: Inflation experiment

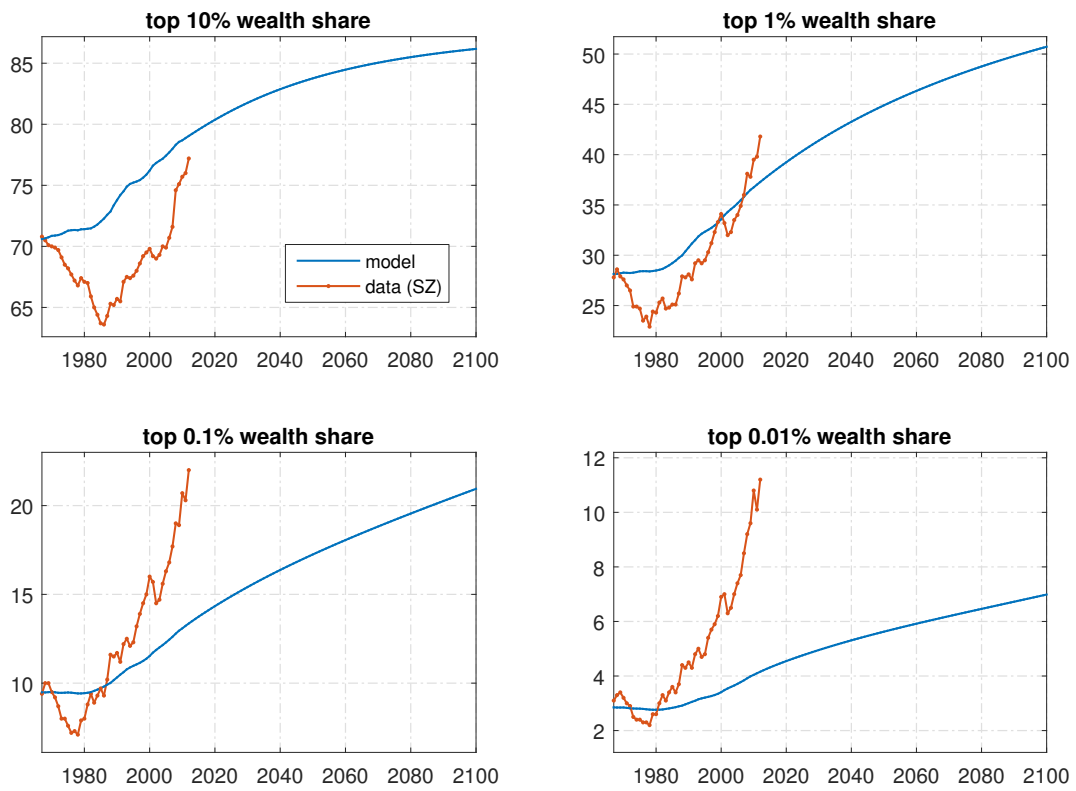


Figure 19: Top wealth shares in %, long run

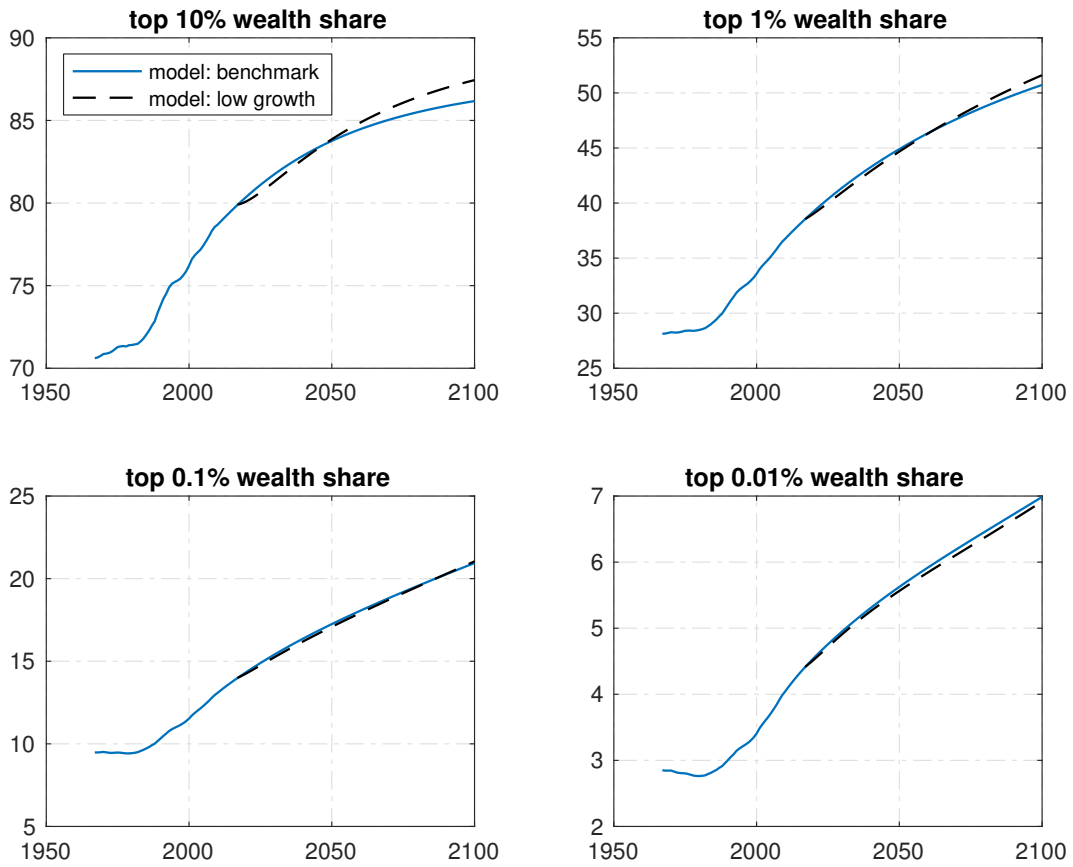


Figure 20: Top wealth shares in %, long run with secular stagnation

C Valuation

In the valuation effects scenario, pre-tax income y_t is given by

$$y_t = a_t(\eta_t r_t + \tilde{r}_t(a_t)) + w_t l_t(p_t, \nu_t),$$

where a_t is the asset level at the beginning of period, η_t the idiosyncratic return shock, and r_t the economy-wide interest rate (which is the equilibrium interest rate if $\tilde{r}_t(\cdot) = 0$ as in the benchmark transition). $\tilde{r}_t(a_t)$ is the differential return term unique to this out-of-equilibrium experiment, defined as

$$\tilde{r}_t(a_t) = \chi(F_t^{-1}(a_t))(r_t^{\text{S\&P}} - r_{1960-67}^{\text{S\&P}}) + (1 - \chi(F_t^{-1}(a_t)))(r_t^{\text{riskfree}} - r_{1960-67}^{\text{riskfree}}),$$

where $\chi(\cdot)$ is the risky portfolio weight, a function of the wealth distribution quantile. For taxation purposes, we split income into ordinary income y_t^{ord} and capital gains y_t^{cg} :

$$\begin{aligned} y_t^{\text{ord}} &= a_t \eta_t r_t + w_t l_t(p_t, \nu_t) \\ y_t^{\text{cg}} &= a_t \tilde{r}_t(a_t) \end{aligned}$$

Ordinary income is subject to the progressive tax function $\tau_t(\cdot)$ based on Piketty & Saez (2007), as in the benchmark transition. Capital gains are subject to a time-varying flat tax τ_t^{cg} . In particular, we use the average effective tax rate on capital gains, for which an annual time series is available from U.S. Department of the Treasury (2016).⁵⁶ Summarizing, cash-on-hand x_t is given by

$$x_t = a_t + y_t^{\text{ord}} - \tau_t(y_t^{\text{ord}}) + T_t + (1 - \tau_t^{\text{cg}})y_t^{\text{cg}}.$$

For robustness, we perform a check using the maximum capital gains tax rate, yielding very similar results.

⁵⁶This is a slight approximation to the actual, historical, U.S. tax schedule for capital gains, which features rates that vary across asset categories, amount of time the asset was held, and also overall income. The capital gains tax schedule has been slightly progressive as well, though much less so than the one on ordinary income. Note that we apply the flat tax in this scenario only on the differential return, relative to the initial steady state.

D Inflation and progressivity

Let r be the constant real interest rate, π the inflation rate and i the nominal interest rate, i.e. $(1+i) = (1+r)(1+\pi)$. Let τ be the tax rate an agent faces. Intuitively, as the tax is applied to nominal capital income, the real after-tax return on capital is given by

$$\tilde{r} \approx i(1-\tau) - \pi \approx (r+\pi)(1-\tau) - \pi = r(1-\tau) - \tau\pi. \quad (12)$$

Formally, we can derive equation (12) as follows: Consider an agent with earnings e growing at the rate of inflation. Let s be savings in period t . Cash-on-hand in period $t+1$, in real terms, after-tax, amounts to

$$\begin{aligned} \frac{s + (1-\tau)(e(1+\pi) + si)}{1+\pi} &= (1-\tau)e + \frac{s + (1-\tau)s(r+\pi+r\pi)}{1+\pi} \\ &= (1-\tau)e + s \underbrace{\left(\frac{1 + (1-\tau)(r+\pi+r\pi)}{1+\pi} \right)}_{\equiv 1+\tilde{r}} \end{aligned}$$

The effective real after-tax return \tilde{r} is thus given by

$$\begin{aligned} \tilde{r} &= \left(\frac{1 + (1-\tau)(r+\pi+r\pi)}{1+\pi} \right) - 1 \\ &= \frac{1 + (1-\tau)(r+\pi+r\pi) - (1+\pi)}{1+\pi} \\ &= \frac{1 + (1-\tau)r(1+\pi) + (1-\tau)\pi - (1+\pi)}{1+\pi} \\ &= (1-\tau)r + \frac{(1-\tau)\pi - \pi}{1+\pi} \\ &= (1-\tau)r - \tau \frac{\pi}{1+\pi} \approx r(1-\tau) - \tau\pi. \end{aligned}$$

Hence, the after-tax real return is affected by inflation. In particular, the higher the tax rate, the worse inflation is. Formally, holding real pre-tax returns r and the tax rate τ constant, for small inflation rates we have that

$$\left. \frac{\partial \tilde{r}}{\partial \pi} \right|_{\pi=0} = -\tau$$

Inflation is bad for agents in high tax brackets. It acts like a progressive increase in capital tax rates (given that the tax system is progressive to start with, which it is).

E $r - g$?

The intuitive idea underlying this hypothesis is twofold. First, the movements in the wealth position of the rich are influenced heavily by the return on wealth, r , if the rich hold most of the wealth and save most of it. Second, in contrast, the wealth of the poor is mostly made up of wage income if the poor do not save appreciably; moreover, wage growth tends to equal that of output, g . Piketty & Zucman (2015) explicitly examine the link between $r - g$ and wealth inequality: in a model with linear, random savings rules they show that a permanent increase in $r - g$ decreases the Pareto coefficient characterizing the right tail of the wealth distribution. In our model, too, an increase in $r - g$ would increase savings rates for all consumers and thus decrease the Pareto coefficient in the right tail in steady state, as shown in the analysis in Section 5. But this result leaves open both the question of how an increase $r - g$ affects overall wealth inequality and the question of how wealth inequality evolves during the transition to a steady state after an increase in $r - g$.

To investigate these issues we conducted a partial-equilibrium experiment, starting from the distribution of wealth in 1967, in which the pre-tax real interest rate increases by 50 percent for a ten-year period and then returns to its original level.⁵⁷ The resulting impulse responses are displayed in Figure 21. Surprisingly, the wealth share of the top 1% as well as the Gini coefficient for wealth initially decrease sharply.⁵⁸ Wealth inequality then rebounds and overshoots slightly before settling back to the steady-state level. As in the experiment in Section 8.1, the model predicts a short-run elasticity of savings with respect to changes in r that is larger for poorer agents. In the very long run, wealth inequality indeed increases in response to a rise in $r - g$, but as pointed out by Gabaix et al. (2016) the dynamics implied by the random growth mechanism work very slowly: it takes a long time to fill a long Pareto tail. In our framework, then, movements in $r - g$ certainly cannot explain the observed sharp changes in wealth inequality over the period 1967–2012 upon which we focus. The partial equilibrium assumption is not crucial for this finding. Finally, although wealth inequality falls on impact, income inequality increases in this experiment because capital income constitutes a large fraction of total income for high-income agents and the interest rate rises. The Gini coefficient for total income is plotted in the lower right panel of Figure 21, showing a sharp increase as the interest-rate shock hits the economy.

⁵⁷There is no growth, i.e., we set $g = 0$.

⁵⁸The dynamics are very similar for other groups such as the top 10% or the top 0.1%; conversely, for groups such as the bottom 50% wealth shares increase before returning to their original levels.

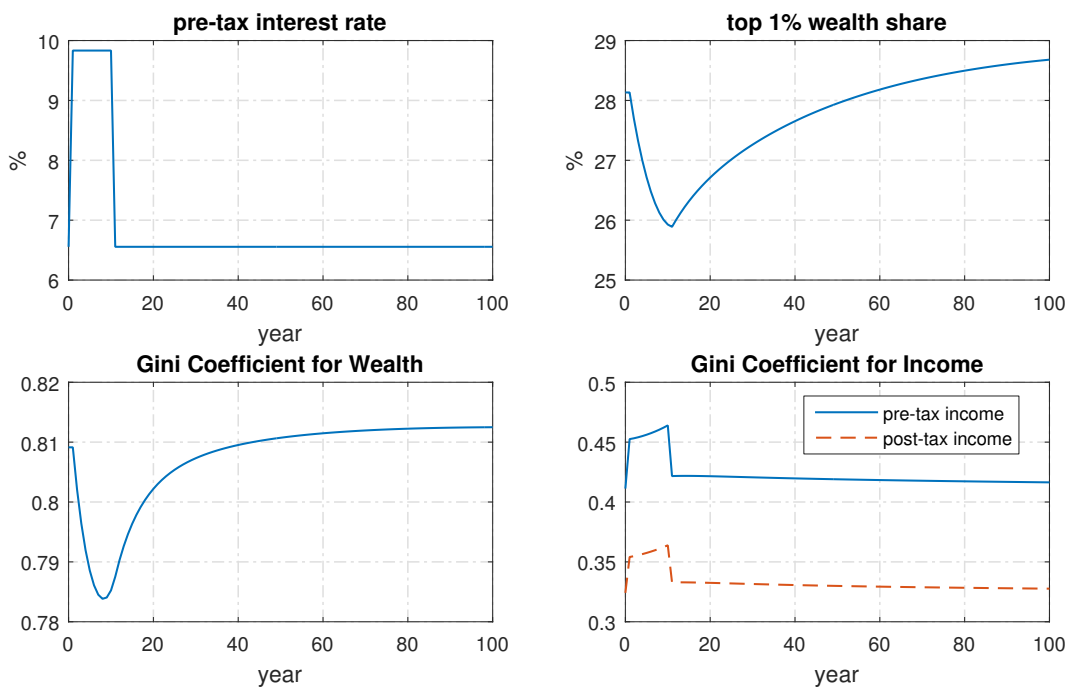


Figure 21: Impulse responses to interest-rate shock (partial equilibrium)

References

- Acemoglu, D. (2002). Technical Change, Inequality, and the Labor Market. *Journal of Economic Literature*, 40(1), 7–72.
- Ahn, S., Kaplan, G., Moll, B., Winberry, T., & Wolf, C. (2017). When inequality matters for macro and macro matters for inequality. In *NBER Macroeconomics Annual 2017, volume 32*. University of Chicago Press.
- Aiyagari, S. R. (1994). Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, 109(3), pp. 659–684.
- Aoki, S., & Nirei, M. (forthcoming). Zipf’s Law, Pareto’s Law, and the Evolution of Top Incomes in the United States. *American Economic Journal: Macroeconomics*.
- Auclert, A. (2015). Monetary Policy and the Redistribution Channel. Working paper.
- Bach, L., Calvet, L. E., & Sodini, P. (2015). Rich Pickings? Risk, Return, and Skill in the Portfolios of the Wealthy. Working paper.
- Becker, R. A. (1980). On the Long-Run Steady State in a Simple Dynamic Model of Equilibrium with Heterogeneous Households. *The Quarterly Journal of Economics*, 95(2), 375–382.
- Benhabib, J., Bisin, A., & Luo, M. (2015a). Wealth Distribution and Social Mobility in the US: A Quantitative Approach. NBER Working Paper 21721, National Bureau of Economic Research, Inc.
- Benhabib, J., Bisin, A., & Zhu, S. (2011). The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents. *Econometrica*, 79(1), 123–157.
- Benhabib, J., Bisin, A., & Zhu, S. (2015b). The Wealth Distribution in Bewley Economies with Capital Income Risk. *Journal of Economic Theory*, 159, Part A, 489 – 515.
URL <http://www.sciencedirect.com/science/article/pii/S0022053115001362>
- Bewley, T. (undated). Interest Bearing Money and the Equilibrium Stock of Capital. Manuscript.
- Bricker, J., Henriques, A., Krimmel, J., & Sabelhaus, J. (2016). Measuring Income and Wealth at the Top Using Administrative and Survey Data. *Brookings Papers on Economic Activity, Spring 2016*, 261–331.
- Brinca, P., Holter, H., Krusell, P., & Malafry, L. (2016). Fiscal Multipliers in the 21st Century. *Journal of Monetary Economics*, 77, 53–69.
- Cagetti, M., & De Nardi, M. (2006). Entrepreneurship, Frictions, and Wealth. *Journal of Political Economy*, 114(5), 835–870.
- Carroll, C. D. (2006). The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems. *Economics Letters*, 91(3), 312–320.

- Carroll, C. D. (2012). Theoretical Foundations of Buffer Stock Saving. Working paper.
- Carroll, C. D., & Kimball, M. S. (1996). On the Concavity of the Consumption Function. *Econometrica*, *64*(4), 981–92.
- Castañeda, A., Díaz-Giménez, J., & Ríos-Rull, J.-V. (2003). Accounting for the U.S. Earnings and Wealth Inequality. *Journal of Political Economy*, *111*(4), 818–857.
- CBO (2015). The Budget and Economic Outlook: 2015 to 2025. Tech. rep., Congressional Budget Office.
- Fagereng, A., Guiso, L., Malacrino, D., & Pistaferri, L. (2015). Heterogeneity and Persistence in Returns to Wealth. Working paper.
- Gabaix, X. (2009). Power Laws in Economics and Finance. *Annual Review of Economics*, *1*(1), 255–294.
- Gabaix, X., Lasry, J.-M., Lions, P.-L., & Moll, B. (2016). The Dynamics of Inequality. *Econometrica*, *84*(6), 2071–2111.
 URL <http://dx.doi.org/10.3982/ECTA13569>
- Guerrieri, V., & Lorenzoni, G. (2011). Credit Crises, Precautionary Savings, and the Liquidity Trap. NBER Working Papers 17583, National Bureau of Economic Research, Inc.
- Heathcote, J. (2005). Fiscal policy with heterogeneous agents and incomplete markets. *Review of Economic Studies*, *72*(1), 161–188.
- Heathcote, J., Storesletten, K., & Violante, G. L. (2010). The Macroeconomic Implications of Rising Wage Inequality in the United States. *Journal of Political Economy*, *118*(4), 681–722.
- Hornstein, A., Krusell, P., & Violante, G. (2005). *The Effects of Technical Change on Labor Market Inequalities*, vol. 1 of *Handbook of Economic Growth*. Elsevier.
- Huggett, M. (1993). The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies. *Journal of Economic Dynamics and Control*, *17*(5-6), 953–969.
- Kaplan, G., Moll, B., & Violante, G. L. (2016). Monetary Policy According to HANK. Working paper.
- Karabarbounis, L., & Neiman, B. (2014a). Capital Depreciation and Labor Shares Around the World: Measurement and Implications. Working paper.
- Karabarbounis, L., & Neiman, B. (2014b). The Global Decline of the Labor Share. *The Quarterly Journal of Economics*, *129*(1), 61–103.
- Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, *107*(1), 35–78.
- Kaymak, B., & Poschke, M. (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.

- Kennickell, A. B. (2011). Tossed and Turned: Wealth Dynamics of U.S. Households 2007-2009. Finance and Economics Discussion Series 2011-51, Board of Governors of the Federal Reserve System.
- Kesten, H. (1973). Random Difference Equations and Renewal Theory for Products of Random Matrices. *Acta Mathematica*, 131(1), 207–248.
- Kopczuk, W. (2015). What Do We Know about the Evolution of Top Wealth Shares in the United States? *Journal of Economic Perspectives*, 29(1), 47–66.
- Kopczuk, W., & Saez, E. (2004). Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns. *National Tax Journal*, 2, Part 2, 445–487.
- Krusell, P., Mukoyama, T., Şahin, A., & Smith, Jr., A. (2009). Revisiting the Welfare Effects of Eliminating Business Cycles. *Review of Economic Dynamics*, 12, 393–404.
- Krusell, P., & Smith, Jr., A. (1998). Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy*, 106(5), 867–896.
- Krusell, P., & Smith, Jr., A. (2006). Quantitative Macroeconomic Models with Heterogeneous Agents. In R. Blundell, W. Newey, & T. Persson (Eds.) *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Econometric Society Monographs*, 41, (pp. 298–340). Cambridge University Press.
- Krusell, P., & Smith, Jr., A. (2015). Is Piketty’s “Second Law of Capitalism” Fundamental? *Journal of Political Economy*, 123(4), 725–748.
- McKay, A. (2013). Search for Financial Returns and Social Security Privatization. *Review of Economic Dynamics*, 16(2), 253–270.
- McKay, A., Nakamura, E., & Steinsson, J. (2016). The Power of Forward Guidance Revisited. *American Economic Review*, 106(10), 3133–3158.
- McKay, A., & Reis, R. (2016). The Role of Automatic Stabilizers in the U.S. Business Cycle. *Econometrica*, 84(1), 141–194.
- Moskowitz, T. J., & Vissing-Jørgensen, A. (2002). The Returns to Entrepreneurial Investment: A Private Equity Premium Puzzle? *American Economic Review*, 92(4), 745–778.
- Nirei, M., & Aoki, S. (2016). Pareto Distribution of Income in Neoclassical Growth Models. *Review of Economic Dynamics*, 20(1), 25–42.
- Piketty, T. (1995). Social Mobility and Redistributive Politics. *The Quarterly Journal of Economics*, 110(3), 551–84.
- Piketty, T. (1997). The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing. *Review of Economic Studies*, 64, 173–189.

- Piketty, T. (2003). Income inequality in France, 1901–1998. *Journal of political economy*, 111(5), 1004–1042.
- Piketty, T. (2014). *Capital in the Twenty-First Century*. Translated by Arthur Goldhammer. Cambridge, MA: Belknap.
- Piketty, T., & Saez, E. (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics*, 118(1), 1–41.
- Piketty, T., & Saez, E. (2007). How Progressive is the U.S. Federal Tax System? A Historical and International Perspective. *Journal of Economic Perspectives*, 21(1), 3–24.
- Piketty, T., & Zucman, G. (2015). *Wealth and Inheritance in the Long Run (Chapter 15)*, vol. 2 of *Handbook of Income Distribution*. Elsevier.
- Quadrini, V. (2000). Entrepreneurship, Saving, and Social Mobility. *Review of Economic Dynamics*, 3(1), 1–40.
- Quadrini, V., & Rios-Rull, J.-V. (2015). *Inequality in Macroeconomics (Chapter 14)*, vol. 2 of *Handbook of Income Distribution*. Elsevier.
- Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3), 649–665.
- Saez, E., & Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. *Quarterly Journal of Economics*, 2, 519–578.
- Stiglitz, J. E. (1969). Distribution of Income and Wealth Among Individuals. *Econometrica*, 37(3), 382–397.
- U.S. Department of the Treasury (2016). Office of Tax Analysis: Taxes Paid on Capital Gains for Returns with Positive Net Capital Gains, 1954–2014. Tech. rep.