

# Do the Rich Save Too Much? How to Explain the Top Tail of the Wealth Distribution\*

Michael Reiter, Universitat Pompeu Fabra, Barcelona

July 2004

PRELIMINARY DRAFT

## Abstract

The aim of this paper is to explain the wealth distribution in the US economy, in particular the high concentration of wealth at the very top end of the distribution. In the literature it was conjectured that rich households accumulate more wealth (given their income) than what can be explained by standard consumption/saving models. To address this issue, I develop a general equilibrium heterogenous agent model and confront it with data about the top income and wealth distribution. It turns out that an off-the-shelf model fails to explain some important characteristics of the data. Standard forms of altruistic bequest motives do not help either. I find two elements that help to bring the model in line with the data. First, modelling the idiosyncratic return risk that rich household face from closely held businesses. Second, modelling a capitalist spirit motive, where capital provides utility services directly, not just through consumption, as was proposed by Carroll (1998).

Address of the author:

Michael Reiter  
Universitat Pompeu Fabra  
Department of Economics and Business  
Ramon Trias Fargas, 25–27  
08005 Barcelona  
Spain  
e-mail: michael.reiter@upf.edu

---

\*I am grateful to Chris Carroll, Kai Konrad, Albert Marcet, Ramon Marimon, Jaume Ventura, Ernesto Villanueva and Joachim Voth for helpful comments and discussions. Support from the Spanish Ministry of Education under the grant PB98-1066 is gratefully acknowledged.

# 1 Introduction

This paper is about the savings behavior of rich and super-rich households in the US. Can it be explained by the standard model of consumption and saving, the life cycle model, perhaps after allowing for altruistic bequests? Based on several US micro data sets (SCF, Consumer Expenditure Survey, PSID), several recent papers have addressed this issue. Dynan, Skinner and Zeldes (2000) find that the rich save more, in the sense that the ratio of saving to permanent income increases with permanent income. In addition, they find little evidence that the old dissave. Carroll (1998), using SCF data, documents the latter point especially for old rich households. They do not seem to dissave after retirement in the way that the life cycle model predicts.<sup>1</sup> To explain the continued capital accumulation of the old rich, Carroll proposes a specification of the utility function where households derive utility not just from consumption, but also directly from wealth holdings ('capitalist spirit'). While the literature has addressed this issue using partial equilibrium models and testing them on micro data sets, the purpose of this paper is to embed these theories into a general equilibrium model, and see whether macro data can shed more light on this issue. I will exploit time series data on the income and wealth distribution, which have been made available recently.

It is well known that wealth is more unequally distributed than income, in the US and also in other developed countries. For example, Diaz-Gimenez, Quadrini and Rios-Rull (1997) report that the top 1 percent earners obtain 14.76 percent of all labor earnings. In contrast, the top 1 percent wealth owners own 30 percent of all wealth (data based on the 1992 Survey of Consumer Finances). Recent papers (Castaneda, Diaz-Gimenez and Rios-Rull, 2002, De Nardi, 2002) can explain this fact in the framework of general equilibrium incomplete markets models. Two elements have been identified as important for the explanation. First, the existence of a public pension system, which reduces the life cycle savings incentives of low and medium income families. Second, an operative bequest motive, which raises the savings incentives especially of rich households, and prevents old rich households from dissaving heavily.

What has not yet received much attention is the fact that *even within the group of rich households*, wealth is more concentrated than income. This can be seen from data have been made available recently (Piketty and Saez, 2003, Kopczuk and Saez, 2004), which are based on huge micro data bases of tax returns. They provide a detailed picture of the upper part of the income and wealth distribution, for most of the 20th century. They document drastic changes in the income and wealth distribution over the last 90 years. In this paper, I will test the life cycle model by its ability to explain some stylized facts that emerge from these and similar data. I will focus not only on the distribution at a given point in time, but also on the time series changes. To that end, I embed the standard life cycle model, with and without bequest motive, as well as the capitalist spirit model of Carroll into a general equilibrium incomplete markets models (in the tradition of Krusell and Smith (1998) and the papers cited above). I find that the life cycle model, with or without bequest motive, cannot account for the salient features of the data if we model the income risk of the rich as labor income risk. I find two specifications that do well in explaining the macro data. One is a model where households face substantial return risk on their assets, and where the richest households in particular hold a very risky portfolio. The other specification assumes more moderate

---

<sup>1</sup>Carroll discards an altruistic bequest motive as an explanation for this phenomenon, because a strong bequest motive is at odds with much of the microeconomic evidence. Researchers generally fail to find a difference in the saving behavior of households with and without children, cf. the discussion in Carroll (1998, Section 3). Altonji and Villanueva (2002) find that a 1 dollar increase in parents' life time income increases bequests and transfers by only 5 cents, much less than would the standard altruistic bequest model would predict. The latter result, however, is for a median household, and may be higher for rich households.

levels of return risk, but households have a capitalist spirit motive. The data on the wealth distribution pose strong restrictions on the size of the capitalist spirit motive or the amount of idiosyncratic return risk. However, it turns out that high return risk and a capitalist spirit motive have similar effects on capital accumulation, and are difficult to distinguish with the available macro data. In any case, the results point out that, if we want to explain the wealth distribution at the top, it is necessary to model carefully the return risk that wealth owners face.<sup>2</sup>

The plan of the paper is as follows. Section 2 establishes some stylized facts in the data that the model should explain. Section 3 presents the model. Section 4 presents the details on the calibration of the various versions of the model. Section 5 discusses the results. Section 6 concludes.

## 2 Some stylized facts

The following figures should illustrate a number of stylized facts about the distribution of income and wealth in the US. They display interesting patterns about the relative concentration of different variables at a given point in time, as well as about the evolution of these measures over the 20th century. The most important data source are the data on wealth and income that are obtained from US tax data and made available by Emmanuel Saez and coauthors (cf. Piketty and Saez, 2003, Kopczuk and Saez, 2004 and the excel files on the web page of Emmanuel Saez). Other important data sources are the Survey of Consumer Finances (henceforth SCF), which have been used in many studies, and the famous list of the wealthiest 400 Americans compiled since 1982 by the Forbes magazine.

As a measure of inequality of wealth and different types of income, Figure 1 displays the share of each variable that goes to the top 1 percent individuals. For example, the solid line shows that the one percent of the population with the highest wage income gets about 8 percent of total wage income in 1930, about 5 percent in the 1960s and 70s, and more than 10 percent in 2000. The figure also shows that total income (excluding capital gains) is more unequally distributed than wage income; income including capital gains is more unequally distributed than income excluding capital gains; and wealth is more unequally distributed than any income measure. The figure reports measures for two different data sources on wealth: the estimates by Kopczuk and Saez (2004) obtained from estate tax data, and the Scholz (2003) estimates based on the SCF, as reported in Kopczuk and Saez (2004, Table C1). As has been discussed in the literature (cf. Kopczuk and Saez, 2004, Section 5.3.2), top wealth shares from the tax data are substantially lower than the ones from the SCF data. Some partial explanations for this have been put forward; very likely, underreporting in the tax data is part of the story. For the purpose of our study, this is not a big problem, because I will not focus on the absolute numbers of the top shares, but mostly on the relative shares of the rich (top 1%) and the super-rich (top 0.01%). The ratios are probably much more reliable than the absolute value of the top shares. The latter are obtained by forming a ratio of two estimates obtained by very different methods (the wealth of the rich, obtained from tax data, compared to overall wealth obtained from the national accounts). In contrast, the ratios of the top shares compare two numbers obtained by the same method (wealth of top 1% compared to wealth of top 0.01%, both from tax data). One indication that the wealth shares obtained from the tax data are too low is that in 2000, the top 400 individuals (representing roughly 0.0002 percent of US adult individuals) own more than 3.7 percent of wealth according to the Forbes list, while the top 0.01 percent individuals own 3.82 percent according to tax data. This is clearly incompatible. The data in

---

<sup>2</sup>This aspect has been neglected in many models of life cycle saving, although there are some exceptions, for example Bertaut and Haliassos (1997).

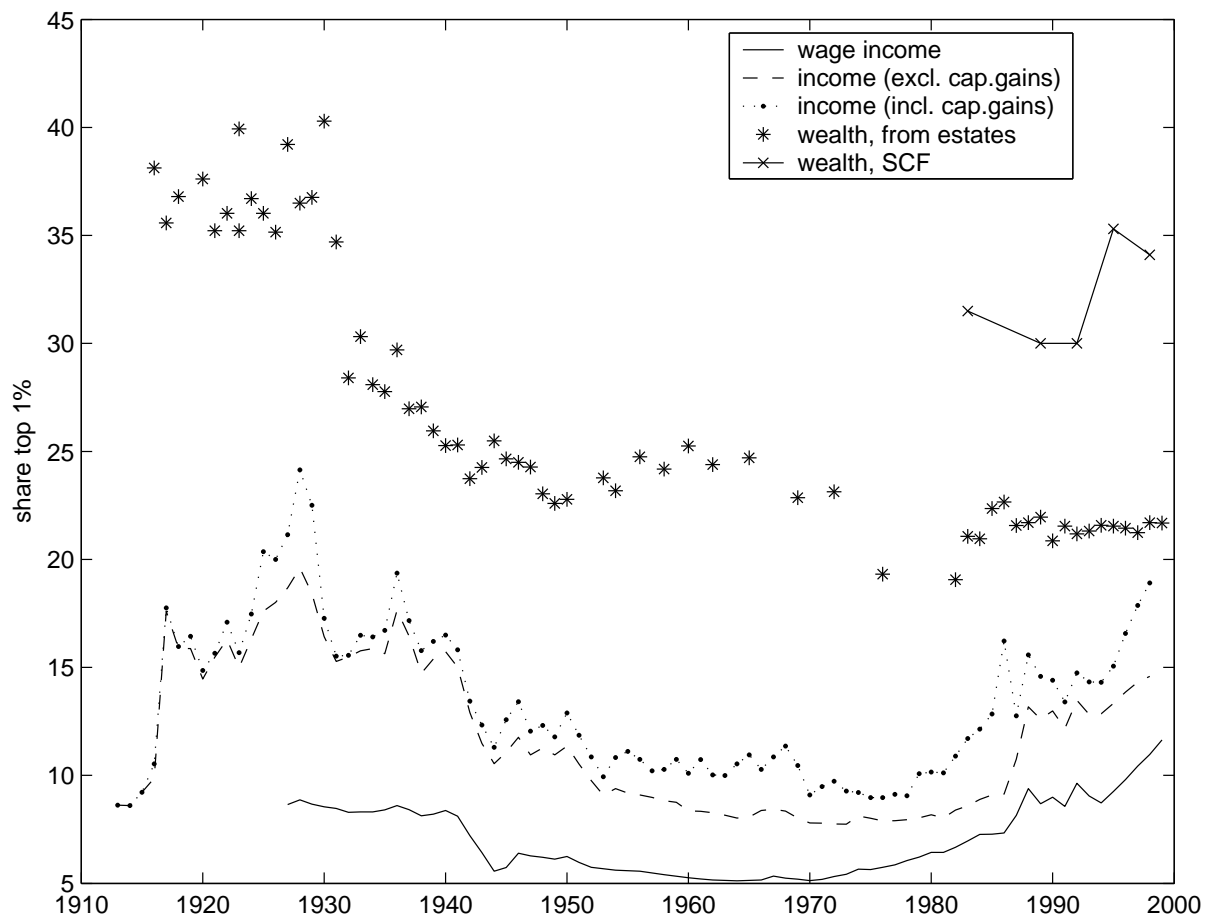


Figure 1: Top 1% shares; source: US tax data (made available by E. Saez)

the Forbes list seem to be much more compatible with the SCF data, and I will use a comparison of wealth shares based on the SCF data and of the Forbes data in Figure 2.

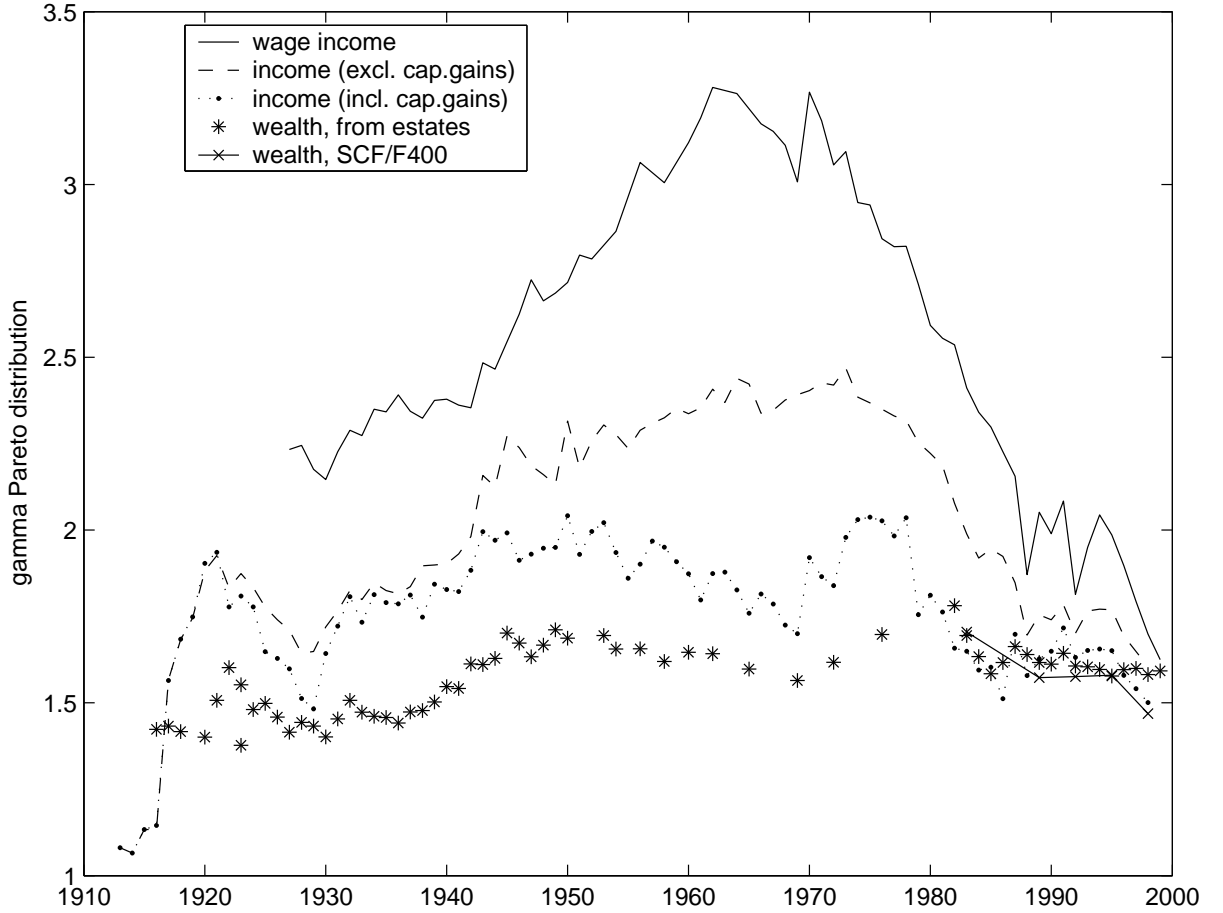


Figure 2: Inequality measures; source: same as Fig. 1

Figure 2 provides measures of inequality at the top of the income and wealth distribution. While Figure 1 compares the rich (top 1%) to the total population, Figure 2 compares the rich (top 1%) to the super-rich (top 0.01% or even top 400 families). It has been known for a long time (since the work of Pareto) that the top tails of these distributions can be well described by a Pareto distribution.<sup>3</sup> A Pareto distribution with parameters  $C$  and  $\gamma$  has the distribution function

$$F(x) = 1 - Cx^{-\gamma}, \quad C > 0, \gamma > 1 \quad (1)$$

Its density  $f(x) = \gamma Cx^{-\gamma-1}$  is monotonically decreasing, which means that it is suitable only to describe the upper part of the income or wealth distribution (the lower part is better modelled by a lognormal distribution). With the Pareto distribution, the share that goes to the upper  $\epsilon$  percentile of the distribution is given by

$$\int_{q(\epsilon)}^{\infty} x f(x) dx \sim \epsilon^{\frac{\gamma-1}{\gamma}} \quad (2)$$

where  $q(\epsilon)$  is defined such that  $F(q(\epsilon)) = 1 - \epsilon$ . From (2) it is clear that we can identify the parameter  $\gamma$  by comparing the shares of two different percentiles. Figure 2 provides

<sup>3</sup>Two recent references: Saez (2001) shows that the upper tail of the income distribution is very well described by a Pareto distribution up to very high levels of income. Levy and Solomon (1997) show that the wealth distribution in the Forbes 400 list follow very nicely a Pareto distribution.

an estimate of the parameter  $\gamma$  for various income and wealth measures, obtained by comparing the top 1% share to the top 0.01% percent share. The line that is the lowest of all lines in the graph (which appears only for the 1980s and 1990s) was obtained by comparing the top 1% wealth share estimates from the SCF to the wealth share of the Forbes 400 list.

To interpret the figure, note from (2) that with higher  $\gamma$ , the share that goes to the upper  $\epsilon$  fractile decreases more rapidly in  $\epsilon$ , which means that the distribution is less unequal. The graph shows that inequality among the rich was high at the beginning of the century, then decreased considerably from the 1940s until about 1970, and then it started to increase again. That is the same pattern that we observed in Figure 1 for the top 1% shares. Comparing the different series, we obtain the same inequality ranking as in Figure 1. What is surprising here is that there are very big changes in the distribution of wage income, and to a lesser extent in all income series, while changes in the distribution of wealth are much smaller.

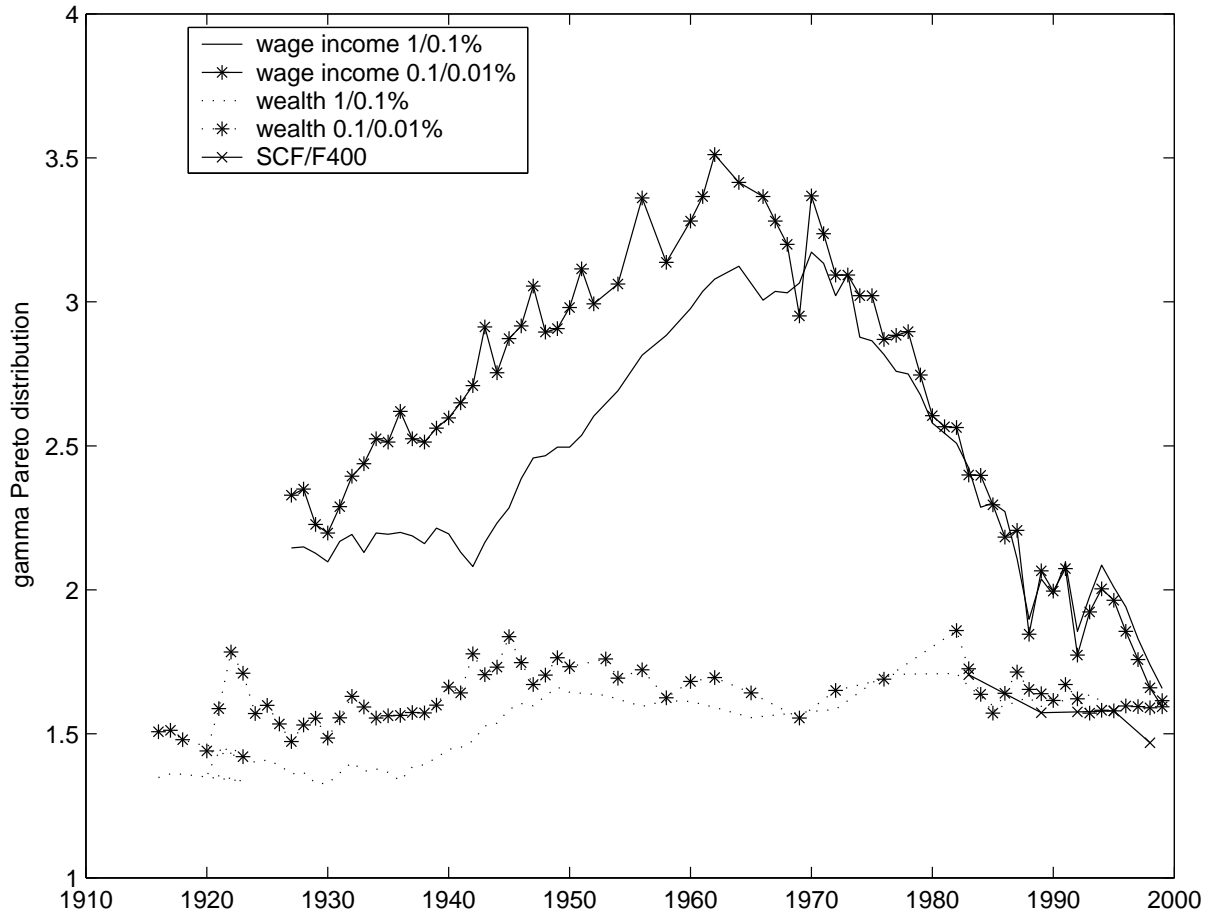


Figure 3: Inequality measures; source: same as Fig. 1

The purpose of Figure 3 is to show that the Pareto distribution is indeed a good description of the top part of the distribution of the variables under consideration. Both for wage income and for wealth, the figure compares two estimates of  $\gamma$ , one obtained by comparing the top 1% and the top 0.1% share, the other one obtained by comparing the top 0.1% to the top 0.01% share. If the distribution were exactly Paretian, the two estimates would be identical. We see that both for wage income and for wealth, the two estimates are very close since 1970. Between 1930 and 1970, the estimates for labor income look different. Then, labor income was even more compressed at the very top end (0.1%/0.01%). The estimate of  $\gamma$  obtained from comparing the top 1% in the SCF to the Forbes 400 are about 1.45, which indicates a significant deviation from the Pareto distribution, with even higher wealth inequality at the very top. This is in line with Levy and Solomon (1997), who find a value of  $\gamma = 1.36$  for wealth within the group of the Forbes 400 list of the year 1996. In fact, the 1990s witnessed an extraordinary increase in wealth concentration at the very top. Kopczuk and Saez (2004, Fig. 12) report that the wealth share of the richest 400 individuals increases from about 1.2 percent in the early 1980s to about 3.7 percent in 2000.

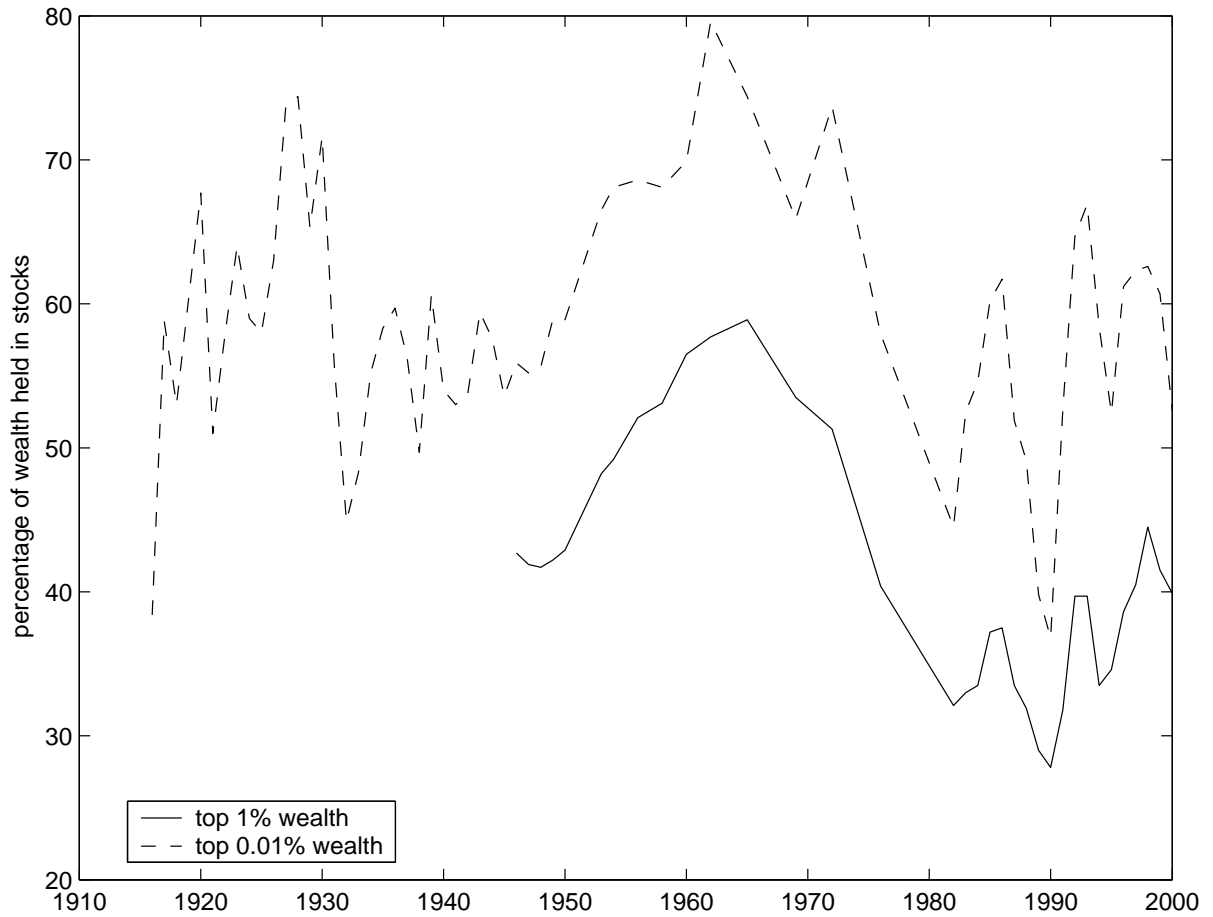


Figure 4: Wealth composition; source: same as Fig. 1

Figure 4 serves to illustrate an important fact about the *composition* of wealth across different wealth groups: the richer people are, the more of their wealth they hold in the form of stocks.<sup>4</sup> The definition of stocks underlying this figure does not only include publicly traded stock, but also the share in closely held businesses. Kopczuk and Saez (2004, footnote 13) also report that the wealthy own a disproportionate share of closely held stock, while the wider public owns more publicly traded stock. This means that many of the wealthy have a poorly diversified portfolio and are subject to a considerable extent to idiosyncratic return risk. We will later see that this is an important element in explaining the data.

<sup>4</sup>The same point is made in Carroll (2002, Table 3), comparing the top 1% to the general population.



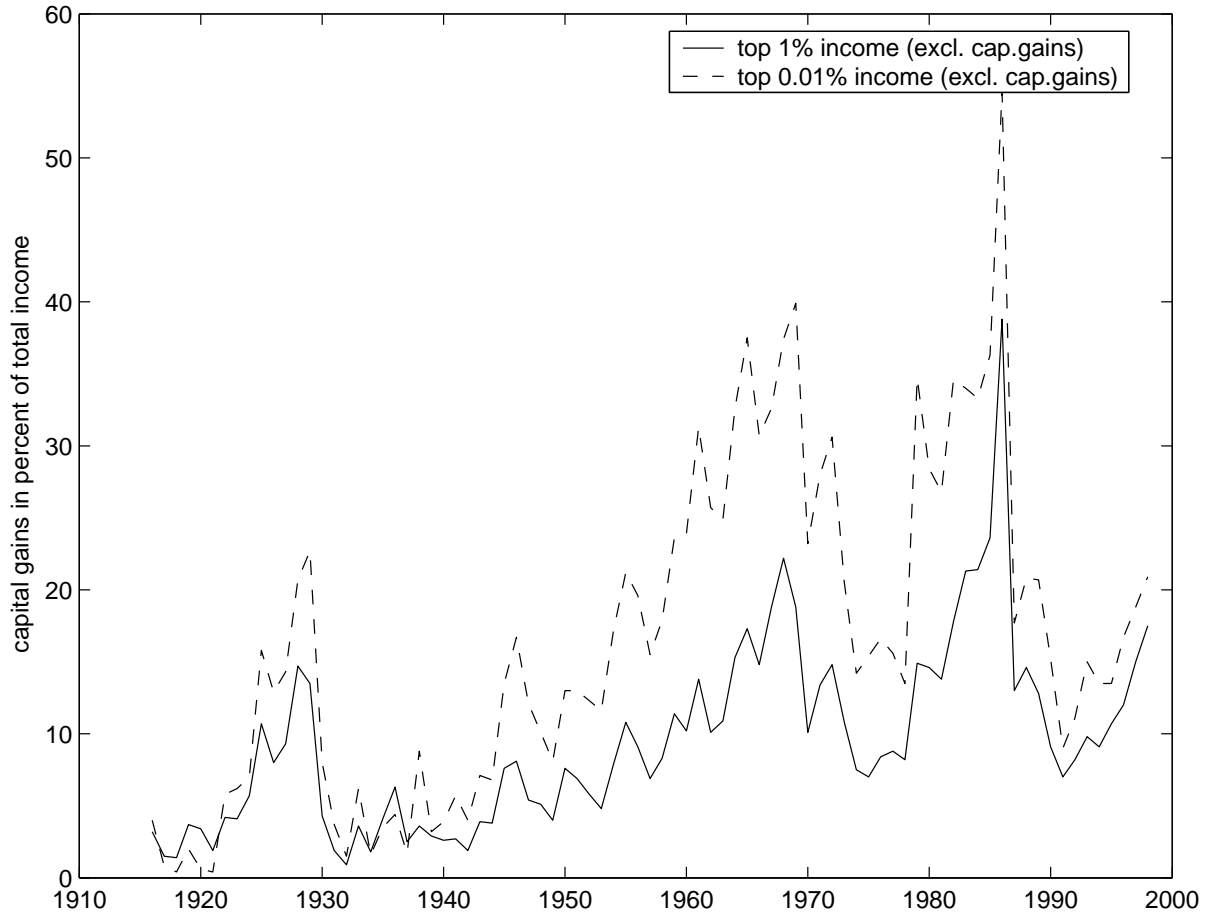


Figure 5: Capital gains; source: same as Fig. 1

Further evidence in this direction is given in Figure 5, showing that the super-rich obtain a higher share of their income in form of capital gains. Since capital gains fluctuate a lot, this means that the super-rich are more affected by return risk than the ordinary rich (the figure reflects aggregate risk, not idiosyncratic return risk). Note that the top income groups in this graphs are defined by their income *excluding* capital gains, so that the result is not an artefact of how the groups were selected.

## 2.1 Wealth accumulation and the Pareto distribution

The fact that the wealth distribution can be described over a wide range by a Pareto distribution allows us to draw some important inferences on the wealth accumulation process of households. This was shown by Levy (2003), who considers the following dynamic equation for wealth  $A_{i,t}$  of household  $i$

$$A_{i,t+1} = \lambda_{i,t} A_{i,t} \quad (3)$$

where  $\lambda_{i,t}$  is a random return on wealth that is independent over time but may differ in its distribution over households  $i$ . The interpretation of (3) is that it applies to very rich households, for which both non-capital income and consumption play a negligible role, so that wealth accumulation is solely driven by random returns to capital. Levy shows that if the distribution of  $\lambda_i$  is identical across households, then the cross sectional distribution of  $A$  converges to a Pareto distribution, with parameter  $\gamma$  implicitly given by

$$E \lambda^\gamma = 1 \quad (4)$$

If, on the contrary, different households have systematically different expected returns (because the rich are “smarter” in their investment decisions), this leads to significant deviations from the Pareto distribution.

While the model (3) is mechanical, it holds important lessons for the present exercise. If, for example, the super rich get the same expected return as other households, but consume a lower fraction of their resources than ordinary rich households, this is equivalent to (3) with a higher  $\lambda$  for the super rich, and would lead to deviations from the Pareto distribution. It turns out that this mechanism imposes tight limits on the strength of the capitalist spirit motive.

### 3 The model

We now develop a model that should explain the more important aspects of the data that we have seen in the last section. The model is a dynamic general equilibrium model with heterogenous agents in the tradition of Krusell and Smith (1998). It inherits the simple life cycle structure in Castaneda et al. (2002) (compared to the more elaborate structure in De Nardi, 2002). What distinguishes the present paper from earlier contributions is the elaborate modelling of the income process of the rich.

#### 3.1 Production

There is a continuum of firms that work in perfect capital and labor markets. Denote by  $K_t$  the aggregate quantity of capital and by  $L_t$  the aggregate amount of labor employed in these firms (see Section 3.7 on how the aggregates are obtained from individual assets and labor). All firms produce with the Cobb-Douglas net production function

$$Y_t = f(K_t, L_t, t) = Z_t K_t^\alpha L_t^{1-\alpha} - \delta K_t \quad (5)$$

$\delta$  is a constant depreciation rate of capital. Aggregate productivity  $Z(t)$  satisfies

$$Z_{t+1} = (1 + g)Z_t \quad (6)$$

where  $g$  is a constant growth rate of productivity. The interest rate and wage rate are then given by

$$r_t = f_k(K_t, L_t, t) \quad (7a)$$

$$W_t = f_l(K_t, L_t, t) \quad (7b)$$

#### 3.2 Demographics

I take population as constant. In reality, the population in the US is growing at a positive rate, but this is due to immigration, not to fertility. Since in this model it is important how wealth is transferred from parents to children, it might be misleading to assume that families are growing in size. So it seems a good compromise to keep the population constant.

In each point in time there is a continuum of households. Households can be either young or old. While young, they face a constant probability  $p^{retire}$  of becoming old. When old, they face a constant probability  $p^{death}$  of dying (this follows Castaneda et al., 2002). When a household dies at the end of period  $t$ , it is replaced in  $t + 1$  by a young household. (We will also investigate a somewhat different setup, taking into account that households have several children, who then match with other children to form new households. This might be an important mechanism of wealth dispersion.)

The numbers of young and old households at time  $t$ , which are denoted respectively by  $N_{y,t}$  and  $N_{o,t}$ , satisfy

$$N_{y,t} = N_{y,t-1}(1 - p^{retire}) + N_{o,t-1}p^{death} \quad (8a)$$

$$N_{o,t} = N_{y,t-1}p^{retire} + N_{o,t-1}(1 - p^{death}) \quad (8b)$$

The system (8) has a steady state with constant population ( $N_{y,t} = N_{y,t-1}$  and  $N_{o,t} = N_{o,t-1}$ ). Then the shares of young and old people, denoted by  $\pi^{young}$  and  $\pi^{old}$ , respectively, are given by

$$\pi^{young} = \frac{p^{death}}{p^{retire} + p^{death}} \quad (9)$$

$$\pi^{old} = \frac{p^{retire}}{p^{retire} + p^{death}} \quad (10)$$

We will always assume that the economy starts in the demographic steady state.

### 3.3 Household income

Old households are retired. They receive capital income and a government pension. Young households receive income from labor as well as capital. Among the young, there are “worker households” and “entrepreneur households”. The fraction of entrepreneur households is constant over time, and I will set this fraction to 5 percent. The term “entrepreneur” should not be taken too literally. The entrepreneurs in this model represent the top income group in the economy, which in reality also includes executives, sports stars etc.

Entrepreneurship is important to explain income and wealth inequality (Quadrini 1999), and several papers (starting with Quadrini (2000)) have developed models of entrepreneurship. Here I will model the “entrepreneurs” in a much simpler way, which seems to be sufficient for the present purpose. Entrepreneurs differ from workers in two ways: first, they have a much higher labor productivity, and second, the return on their capital is very volatile. The latter point should capture the fact that entrepreneurs hold a large part of their wealth in their own firm; they have a badly diversified portfolio and face a big idiosyncratic return risk (Moskowitz and Vissing-Jrgensen 2002). I will refer to the stochastic part of capital income as “capital gains”. A more detailed description of how they are modelled will be given in Section 4.2.

The transition probabilities between types (young worker households, young entrepreneur households of different productivity, old households) are exogenous and constant over time. We will only look at steady states where the fractions of different types in the economy is constant over time.

#### Idiosyncratic productivity

The productivity of a household at time  $t$  is influenced by two factors: its productivity state  $\xi_t^p$  (the superscript “p” stands for “permanent”), which follows a first order Markov process, and an i.i.d. shock. The lowest possible realization of  $\xi_t^p$  is normalized to one, which is called the “worker state”. In each period, a constant fraction  $\phi_w$  of young households are workers. The fraction  $\phi_e = 1 - \phi_w$  of young households are in the “entrepreneurial state” with  $\xi_t^p > 1$ . Conditional on  $\xi_t^p > 1$ , we assume that the distribution of  $\xi_t^p$  follows in each  $t$  a Pareto distribution  $P(\gamma_p, C_p)$  (recall the description of the Pareto distribution in Section 2).

The actual realization  $\xi_t$  of productivity in  $t$  is related to the household’s productivity state  $\xi_t^p$  such that

$$E[\xi_t | \xi_t^p = x] = x \quad (11)$$

The exact distribution of  $\xi_t$  depends on whether the household is in the worker state ( $\xi_t^p = 1$ ) or in one of the entrepreneurial states ( $\xi_t^p > 1$ ). Conditional on being a worker,  $\xi_t$  follows a lognormal distribution  $\log \xi_t \sim N(-\sigma^2/2, \sigma^2)$ , which is consistent with (11). Conditional on being an entrepreneur, I assume that  $\xi_t$  follows a Paretian distribution  $P(\gamma, C)$ . The parameters  $(C, \gamma)$  of the distribution of  $\xi_t$  will in general differ from the parameters  $(C_p, \gamma_p)$  of the distribution of  $\xi_t^p$ . To satisfy the constraint (11), we must require that  $\gamma_p \geq \gamma$ , which can be seen from the following

**Lemma 1.** *If  $x \sim P(\gamma_x, c_x)$  and  $y \sim P(\gamma_y, c_y)$  and  $E[y|x] = x$ , then  $\gamma_y \leq \gamma_x$ .*

*Proof.* Cf. Appendix B. □

The time series properties of  $\xi_t^p$  and  $\xi_t$  will be described in Section 4.1. For the computation, we will approximate  $\xi_t^p$  and  $\xi_t$  by finite state processes which approximate the above distributions, cf. Appendix A.

We can summarize the above assumptions by writing productivity  $\xi_t$  as a function of current productivity state  $s_t$  and an i.i.d. shock  $z_t$ ,  $\xi_t = \xi(s_t, z_t)$ . We denote by  $\Phi_t(s, A)$  the joint cross-sectional distribution of  $s$  and assets  $A$ , and by  $\tilde{\Phi}_t(s, z, A)$  the joint cross-sectional distribution of  $s$ ,  $z$  and  $A$ .

### Taxes and the household budget constraint

Household  $i$  receives the before tax interest rate  $r_t$  on its assets  $A_{i,t}$ . Young households also receive labor income, which is the product of the wage  $w_t$  and their effective labor supply  $\xi_{i,t}$ .

The government raises taxes on labor income and on capital income. Labor is taxed at the flat rate  $\tau_{l,t}$ , which varies over time. Capital income above an exemption level of  $r_t A_t^*$  is taxed at the flat rate  $\tau_a$ , which is constant over time. The exemption level will be chosen to be rather high and should reflect the fact that a large part of wealth is held in the form of owner-occupied housing which gets a favourable tax treatment.

Pension transfers to the old  $T_t$  are lump sum; they depend on average wages in the economy. In addition to regular capital income  $r_t A_{i,t}$ , entrepreneurs may also have capital gains  $\tilde{r}_{i,t} A_{i,t}$ . The rate of gains  $\tilde{r}_{i,t}$  is specific to the household, and is intended to measure the effect of imperfectly diversified portfolios, in particular private businesses and closely held stock. I assume that capital gains are not taxed. This is certainly an extreme assumption; in reality, the tax on capital gains varies a lot over time, but the fact that they have to be paid on realization, not on accrual, reduces the effective tax burden, and it allows many households to accumulate capital over a long time without paying taxes. Note that taxable business income is modelled here as labor income; entrepreneurs have a very high labor productivity.

With these assumptions, the budget constraint of the household is

$$A_{i,t+1} = (1 + r_t + \tilde{r}_{i,t})A_{i,t} - \tau_a r_t \max\{A_{i,t} - A_t^*, 0\} + \xi_{i,t} W_t (1 - \tau_{l,t}) + T_t - C_{i,t} \quad (12a)$$

We also assume that households cannot borrow:

$$A_{i,t} \geq 0 \quad (12b)$$

### 3.4 Household utility

Instantaneous utility of the household is described by the function

$$U(C, A) = \frac{C^{1-\eta}}{1-\eta} + b_{A,t} \frac{(A + m_{A,t})^{1-\eta_k}}{1-\eta_k} \quad b_{A,t} > 0, \quad m_{A,t} > 0, \quad \eta_k < \eta \quad (13)$$

The first term is the usual CRRA utility term in consumption  $C$ . The second term has been proposed by Carroll (1998), who calls it the “capitalist spirit” hypothesis. It says that households derive utility not just from consumption, but also from the ownership of capital directly (this may be because ownership of capital increases social reputation, because it gives people more political power, because it measures work success etc.). Marginal utility is decreasing both in consumption and in asset holdings, but the restriction that  $\eta_k < \eta$  means that it diminishes less rapidly in asset holdings. Choosing the right values for  $b_{A,t} > 0$  and  $m_{A,t} > 0$  (see Section 4.2 on the calibration), we get that people with normal income and wealth seek income almost exclusively for consumption purposes, while for super-rich people the capital accumulation motive becomes dominant.

Since the utility function (13) is not homothetic in  $C$  and  $A$ , the model can only have a steady state if the parameters  $b_{A,t}$  and  $m_{A,t}$  vary systematically over time. Concretely, we have to assume that they depend on aggregate productivity in the following way:

$$m_{A,t} = Z_t m_A \quad (14a)$$

$$b_{A,t} = Z_t^{\eta_k - \eta} b_A \quad (14b)$$

where  $b_A$  and  $m_A$  are constants.

In each  $t_0$ , household  $i$  then maximizes

$$\mathbb{E} \sum_{t=t_0}^{\infty} B(i, t_0, t) U(C_{i,t}, A_{i,t}) \quad (15)$$

where the stochastic discount factor is given by

$$B(i, t_0, t_1) = \prod_{t=t_0}^{t_1-1} \hat{\beta}(s_{i,t}) \quad (16)$$

$$\hat{\beta}(s) = \begin{cases} \beta & \text{if } s \text{ represents young} \\ \beta(1 - p^{death}(1 - \beta^{altr})) & \text{if } s \text{ represents old} \end{cases} \quad (17)$$

The discount factor  $\hat{\beta}(s)$  reflects three elements: absolute time preference  $\beta$ , death probability  $p^{death}$ , and the preference towards the utility of offspring. The case  $\beta^{altr} = 1$  gives the standard model of altruistic (dynastic) bequests. The case  $\beta^{altr} = 0$ , which is our benchmark specification, is the pure life-cycle model without a bequest motive, where all bequests are accidental (length of life is stochastic, and there are no annuity markets).

Summarizing the foregoing sections, we can now formulate the household problem. A household is characterized by its demographic and employment state (which is a discrete, exogenous variable) and by its wealth, which is continuous and endogenous. The household maximizes (15) subject to (12), taking as given the stochastic process of prices in general equilibrium. It is straightforward to show that the household problem is convex.

### 3.5 The government

The government spends resources on (useless) government consumption  $G_t$ , which we model as proportional to GDP, and on lump sum pension transfers to the old, which are a constant fraction of before tax average wage earnings of workers (not including the entrepreneurs). Since the average labor productivity of workers was normalized to 1, transfers are given by

$$T_t = \bar{t} W_t \quad (18)$$

The government gets revenues from taxes on labor (including entrepreneurial labor), capital (excluding capital gains), and estates. There is an exemption level for capital and estate taxes. Budget balance then implies that

$$G_t + \bar{t}\pi^{old}W_t = \int \left[ \tau_l w_t \xi(s, z) + \tau_a r_t \max\{A - A_t^*, 0\} + \tau_e \mathcal{I}(s \in \text{old}) p^{death} \max\{B(s, z, A) - B_t^*, 0\} \right] d\tilde{\Phi}_t(s, z, A) \quad (19)$$

Here,  $\mathcal{I}(s \in \text{old})$  is an indicator function that takes the value 1 iff the state  $s$  refers to retired households, 0 otherwise.  $\xi(s, z) = 0$  if  $s$  refers to retired households.  $B(s, z, A)$  is the estate that an old person leaves (in case of dying), which is given by formula (12a), since estates are equal to next period's assets. We assume that the tax rate on labor income adjusts endogenously, such that the government budget is balanced in each period.

### 3.6 Shocks to the income distribution

In this paper I will present many results referring to the steady state of the described model. However, the data presented in Section 2 make clear that we cannot assume to be in a steady state, as far as the top income distribution is concerned. This applies both to the share of income that goes to the top 1% income group, as well as the distribution of income within this group (comparing the top 0.01% to top 1%). In order to model the secular shifts in the earnings distribution that we observe in the data, I assume that the parameter  $\gamma_{\xi,t}$  (of the Pareto distribution of individual productivity) follows a two-state Markov process with very high persistence. This is described in more detail in Section 5.2.

The fluctuations in earnings inequality are the only source of aggregate uncertainty in the model.

### 3.7 General equilibrium

Let us define the vector of aggregate state variables  $S_t \equiv \Phi_t(s, A), \gamma_{\xi,t}$ . It consists of the cross-sectional asset distribution  $\Phi(s, A)$  and the parameter  $\gamma_{\xi,t}$  of the productivity distribution. As is common in the literature, I assume there is a stationary Markov equilibrium in which all aggregate variables at time  $t$  are a function of  $S_t$ .

An equilibrium is defined as a tuple consisting of

- pricing functions  $r(S)$  and  $w(S)$
- an aggregate transition law  $S_{t+1} = \mathcal{H}(S_t, \gamma_{\xi,t+1})$
- an individual savings function  $C(a, s, z; S)$  where  $a$  is individual assets,  $s$  is the individual productivity state, and  $z$  is the current productivity shock

which satisfies the restrictions

- Eqs. (7) and

$$K_t = \int A d\Phi_t(s, A) \quad (20)$$

$$L_t = \int \xi(s, z) d\tilde{\Phi}_t(s, z, A) \quad (21)$$

(Note that aggregate labor supply  $L_t$  is measured in units of individual productivity, not including aggregate productivity  $Z_t$ . It changes over time due to exogenous changes in the productivity distribution.)

- the transition law  $S_{t+1} = \mathcal{H}(S_t, \gamma_{\xi, t+1})$  is compatible with the individual consumption function  $C(a, s, z)$  (that means,  $\mathcal{H}(\cdot)$  results from integrating over individual savings functions).
- Given  $\mathcal{H}(\cdot)$ ,  $r(S)$  and  $w(S)$ , the consumption function  $C(a, s, z)$  solves the household optimization problem.

The numerical approximation to the equilibrium will use the assumption that individuals base their current decisions not on the whole distribution (which is an infinite-dimensional object) but on a finite (in fact very small) number of statistics of these distributions, for example moments (this is common procedure since Krusell and Smith, 1998 and den Haan, 1997; the method used here is described in Reiter, 2001).

### Steady state

The only aggregate shock in this model is on the productivity distribution. Keeping this distribution constant, we can solve for a deterministic steady state. In each steady state experiment, the discount factor  $\beta$  and the labor tax rate  $\tau_l$  are adjusted such that the capital output ratio is equal to 3.13 and the government budget is balanced.

## 4 Calibration

The model is calibrated to US data. I first describe the calibration for the benchmark model, which is a standard life cycle model without bequest motive, capitalist spirit, or idiosyncratic return risk (we think of households as having a perfectly diversified portfolio). Afterwards I describe the variations on this benchmark economy that I will investigate numerically.

### 4.1 Benchmark model

#### Production function:

I set annual productivity growth  $g$  to 1.586 percent, which is the average growth rate of real GDP per total number of civilian employees in the period 1960-2000. I take from Castaneda et al. (2002) the parameters of the production function,  $\alpha = 0.376$  and  $\delta = 0.059$ , as well as the capital output ratio of 3.13. In all the experiments, I adjust the discount factor  $\beta$  such that we obtain this ratio in steady state. This leads to a real interest rate of 6.39 percent before taxes, which is very high, given that this return is not risky. We maintain this parameterization for comparability with the literature. In any case, changes in the average interest rate should not matter too much if compensated by changes in the discount factor.

#### Fiscal policy:

Government expenditures (not including pensions) are set to 20 percent of production, which approximately equals the average ratio of government consumption expenditures and investment to GDP over the period 1960-2000. I assume that the pension transfer is 20 percent of before tax wages; this is less than what we observe in the US (about 40 percent of after tax wages, equivalent to roughly 30 percent of before tax wages). I choose a low rate for this parameter, because otherwise the worker households save even less, and I get a too high wealth concentration. The discrepancy of the model with the data in this respect is probably due to the fact that the savings incentives to low- and medium income people are not correctly modelled here (return to owner

occupied housing, rates of return to “low-income entrepreneurs” such as small shop keepers, precautionary savings motives due for example to health risks).

Following again Castaneda et al. (2002), I choose a tax rate on capital income of 30 percent. The level of assets on which the returns are tax exempt is taken to be half of average capital per capita, that means 1.56 years of average production, which should conform roughly to the average value of an owner occupied home. Estates are taxed at a flat rate of 16 percent, with an exemption level equal to 10 years of per capita production. The tax on wage income adjusts endogenously to guarantee a balanced budget in each period; in all the experiments, this rate is approximately 30 percent.

### Utility:

I set the risk aversion parameter for consumption to  $\eta = 2$ , which seems to be in the middle of recent estimates.

### Demographics

For the demographics, I follow Castaneda et al. (2002) and set the probability of switching from young to old as  $p^{retire} = 0.022$ , and the probability of dying for the old as  $p^{death} = 0.066$ . At each point in time, 5% of the young population are entrepreneurs, which means that they form part of the Pareto distribution. This is broadly consistent with the evidence reported in Saez (2001, Fig. 2), which indicates that the Pareto part of the wage distribution (in 1992) sets in at an annual wage income of about 100,000 dollars (1992\$). Since we focus on the people in the top 1% income group, it means that the focus group is a subgroup of the “entrepreneurs”.

### Individual productivity

Since my main interest is the top of the income and wealth distribution, I choose a rather simple description of the income process for low- and medium income households. The income process of the young worker households is modelled as lognormal, with a standard deviation of 0.4. The high standard deviation of 0.4 should somewhat compensate for the fact that the persistency of the shocks is not modelled, as in Carroll (2001). Experiments not reported here show that changing the standard deviation has a negligible effect on the results that we focus on.

The persistence of the earnings process for entrepreneurs is captured by

$$E [\log(\xi_{t+1}^p) | \log(\xi_t^p)] = \rho_0 \log(\xi_t^p) - \rho_1 \quad (22)$$

In the benchmark version of this model, I set  $\rho_0 = 1$  and  $\rho_1 = 0.05$ . That means that earnings are very persistent, the log of productivity decreases in expectation by 0.05. Several parameter variations are tried to check robustness.

The productivity state of young households is approximated by an 8-state Markov chain,  $\xi_{i,t}^p \in \{\Xi_0^p, \dots, \Xi_7^p\}$ . The state  $\Xi_0^p = 1$  represents a worker family, states  $\Xi_1^p, \dots, \Xi_7^p$  represent entrepreneur families of different productivity. I assume a Pareto parameter  $\gamma_p = 2$  for this part, slightly above (cf. Lemma 1) the value  $\gamma_\xi = 1.8$  for current productivity realizations, cf. below. Table 4.1 shows the results of the approximation. It reports the level of productivity for workers and entrepreneurs. The 7 entrepreneurial states are characterized by the the lower limit of their quantiles; for example, in the column with quantile number 0.99 we have the results for the quantile from 0.99 to 0.9997 (the next quantile number). Results are reported both in model units and what this means in year 2000 US dollars (for details on how model units are translated in dollars, cf. below).



Workers		Entrepreneurs						
Quantile		0	0.3	0.7	0.9	0.99	0.9997	$1-10^{-5}$
Units	1.00	4.47	5.93	9.51	19.73	70.01	400.99	2597.30
1000\$	51	229	304	487	1011	3586	20542	133053

Table 1: Distribution of productivity state

The probability state  $\xi_t^p$  represents the *expected value* of a family’s productivity at  $t$ . Conditional on being a worker, the current realization of the productivity is represented by a ten point distribution that approximates the lognormal distribution. Conditional on being in one of the entrepreneurial states, the productivity distribution is modelled as a 15-point distribution that approximates a Paretian distribution. The parameter  $\gamma_\xi$  of this distribution was set to 1.8 in the benchmark case. This is a little bit lower (more unequal distribution) than what would be suggested by the wage income distribution in the 1990s (cf. Figure 2), which appears adequate since it includes the labor income of entrepreneurs. The parameter  $C$  was chosen such that the top 1% earnings group obtains a share of 14.76 percent of total earnings, which is the estimate obtained by Diaz-Gimenez et al. (1997) from 1992 SCF data, and was also used in Castaneda et al. (2002).

Quantile	0	0.2	0.4	0.6	0.7
Units	3.88	4.46	5.40	6.56	7.93
1000\$	198.813	228.532	276.698	336.137	406.336
Quantile	0.8	0.9	0.953	0.978	0.99
Units	10.65	15.92	24.40	37.38	62.14
1000\$	545.71	815.747	1250.26	1915.36	3184.08
Quantile	0.9968	0.999	$1-10^{-4}$	$1-10^{-5}$	$1-10^{-6}$
Units	117.81	271.33	975.10	3504.32	17693.05
1000\$	6036.63	13903	49964.5	179563	906598

Table 2: Distribution of entrepreneurs’ productivity realizations

### Interpreting the model results in dollar values:

To allow a more intuitive interpretation, I report some of the model results, such as fractile thresholds, not just in model units but also in absolute dollar values. These should be interpreted as year 2000 dollars. In 2000, population in the US was 282.4 million people. The average family size in the SCF (which is the kind of family we are modelling here) has 2.41 members, so we assume that the US consists of roughly 110 million “SCF equivalent families” (average family size in the total population is less than that of SCF families). The Forbes 400 therefore represent roughly 0.00036 percent of US families.

In 2000, US GDP was \$ 9817 bn. With 282.4/2.41 million families, of the average size in the SCF sample, this gives a family income of \$83,778. Average production in the model (excluding capital gains) is 1.635 (both the wage rate and the average productivity of a worker are normalized to unity), so all income and wealth measures are translated into dollars by applying a factor of 83,778/1.635.

## 4.2 Variations of the model

### Capital gains (CG)

In the benchmark model there are no capital gains  $\tilde{r}_{i,t} = 0$ . They are now introduced to account for the fact that many rich households have invested a large part (or all) of their money in one business.

It is clear that many rich individuals have a large fraction of their wealth invested in one company. Moskowitz and Vissing-Jrgensen (2002, Table 2) report that households who own private equity have on average around 40 percent of their net wealth invested in private equity. They also report that household who have own-company publicly traded stocks have on average only about 10 percent of their net worth invested in their own stock. However, at the top end of the wealth distribution there seem to be many people who have a large proportion of their net wealth invested in one publicly traded stock. The best known example is Bill Gates, who has now about 60 percent of his net worth invested in Microsoft (other examples are the owners of Wal-Mart; many more examples can be found).

For whatever reason, these individuals have a badly diversified portfolio, which probably carries a much higher risk than that of the national stock market. The exact size of this risk is hard to quantify, and obviously it differs very much between households, but we can get some indications from the following numbers. Note first that the standard deviation of log annual changes in a broad stock market index (Standard&Poor 500 index in the period 1950-2003) is 0.148, about a 15% annual change. This is a broadly based index; it is clear that the return of an individual company fluctuates much more. The return to an individual, publicly traded stock has an annualized standard deviation of about 50 percent according to Campbell, Lettau, Malkiel and Xu (2001) (this number has varied over time; see for example their Figure 6B). Moskowitz and Vissing-Jrgensen (2002, Section 4) establish that the idiosyncratic risk of private equity is also very high, although precise numbers are not available, and Bitler, Moskowitz and Vissing-Jrgensen (2004, Appendix A) calibrate their model such that the return on private equity is a bit lower than that for public equity. Since the available data do not allow me to calibrate the model with any precision, I will go the opposite way and ask which level of risk (if any) allows me to explain the observed wealth inequality.

Explaining why rich households have a non diversified portfolio is beyond the scope of this paper. I will take a shortcut and assume that different types of households (workers, entrepreneurs of different productivity) face different levels of return risk. For simplicity, I assume in the following that the capital gains shock has a two-point distribution. With probability 0.5, the household loses a certain percentage of its capital, and with probability 0.5, it gains another percentage. I assume the households do not get a risk premium for their riskier portfolio, so the capital gains integrate to zero. A household that faces return risk with standard deviation  $\sigma$  gets

$$\tilde{r}_{i,t} = \begin{cases} -\sigma & \text{with probability 0.5} \\ +\sigma & \text{with probability 0.5} \end{cases} \quad (23)$$

The two-point distribution obviously underpredicts the probabilities of catastrophic failures or fantastic successes; we will probably have to compensate for that by an increased standard deviation of capital gains.

In a first specification (called 'E' in the following tables), all entrepreneurs face the same proportional return risk  $\sigma$ , worker households have a lower return risk  $\sigma^W$ . More precisely, a household that becomes an entrepreneur at any point in time will face the same capital gains risk now and in the future, even if it returns to a worker state or retires. The child household, however, will start again as a worker household with no capital gains.

In the second specification (called 'T'), the two top groups of entrepreneurs (representing only 0.3% of all entrepreneurs) face the higher return risk  $\sigma$ , while all other young households have the return risk  $\sigma^W$ . This specification was tailored so as to explain the increase in wealth inequality at the very top of the distribution. Again, top entrepreneurs who retire will be subject to the high capital gains for the rest of their lifetime.

I will try different values of  $\sigma$  and  $\sigma^W$  to find out which level is necessary to explain the data.

### Intergenerational earnings persistence (IGP)

The empirical literature (for example, Zimmerman, 1992) has found considerable persistence of earnings across generations. The dependence of children's earnings on parents' earnings is typically modelled as

$$\log(Y^P(\text{child})) = \rho_{IG} \log(Y^P(\text{parent})) \quad (24)$$

and the estimate of the parameter  $\rho_{IG}$  is usually around 0.4. One has to say, however, that these estimates refer to the population as a whole; the usual data set has few rich people. We therefore do not have any specific evidence of the earnings persistence at the top end. However, specification (24), which is linear in logs, seems rather plausible for high income parents. I have written (24) without a constant; think about the median income as normalized to 1, so its log is zero. Then a household with five times the median income can expect its children to have income  $\exp(0.4 \log(5)) \approx 1.9$ . So its children have an income substantially higher than normal, and still comparable to the parent's income; this reduces the parents' bequest motive. A super-rich household with, say, 1000 times the median income, expects its children to have income  $\exp(0.4 \log(1000)) \approx 16$ , which is negligible compared to its own income. Super-rich households therefore have a stronger bequest motive than rich household, which might help to explain the observed wealth concentration.

I try parameter values of  $\rho_{IG} = 0.4$  and  $\rho_{IG} = 0.7$ .

### Capitalist spirit (CS)

The capitalist spirit part of the utility function has three free parameters, which are difficult to calibrate. As a consequence, I have solved the model for many parameter constellations, to find out which are suitable to explain the model. The basic idea is that the marginal utility of capital decreases more slowly than that of consumption, which means  $\eta_k < \eta$ . Since we use  $\eta = 2$ , I tried values  $\eta_k = 1$  and  $\eta_k = 1.8$ . It turned out that  $m_A = 100000$  is a good value to make sure that the capitalist spirit motive is significant only for the very top households, which is in line with the fact that the wealth distribution is Paretian for a wide range of wealth levels, except the very top. The parameter  $b_A$  was then adapted in each case so as to give a capitalist spirit motive of the strength necessary to explain the data.

### Splitting of estates (SPL)

The dynastic model, where at the time of death the total wealth of an agent is passed on to one child household, is routinely used in economics, but very implausible. In reality, wealth is usually split between different heirs, who are linked to spouses who themselves might come with an inheritance. This may be an important mechanism of wealth dispersion.

Since our model has non-overlapping generations, generations succeed each other too infrequently (every 60 years rather than every 30 years, as in reality). I account for this

difference by splitting the estates not between 2 heirs, which is average in reality, but between 4 heirs. The wealth splitting is only applied in the stochastic simulations of the model. The steady state results refer to the dynastic model.

## 5 Results

I first present results for the deterministic steady state of several variants of the model. Then I report results for the model with an aggregate shock to the earnings distribution.

### 5.1 Steady state results

		top shares		$\gamma$ from ratios of top shares		
		1%	top400	1/0.1	0.1/0.01	top400/top100
Wages	model	14.76	0.40	1.81	1.85	1.89
	US 1992	14.76	-	1.86	1.77	-
Wealth	model	25.54	0.60	1.90	1.98	2.06
	US 1992	30.00	1.66	1.59	1.62	1.36

Total estates, percent of GDP: 1.82 (US: 2)

Percentage decedents filing: 5.06 (US: 2.88)

Table 3: Results benchmark model

Table 3 displays some results for the steady state version of the benchmark model. The parameter values for this version are described in Section 4.1. We see that this calibration of the model slightly underpredicts the top 1% wealth share; this could be easily fixed by a small change of parameters.

The key qualitative difference is that in the model, wealth is more equally distributed than wage income (higher  $\gamma$ ), while in the data, the opposite is true. The result in the model is easy to understand: since households face income fluctuations, accumulated wealth reflects an *average* over past earnings, and as such should be more equally distributed than earnings themselves. Is the discrepancy between model and data big enough to worry about? The ratio of the top 0.01% wealth share to the top 1% wealth share is 0.114 in the model (not reported in the table) and 0.176 in the data. It is more than 50% higher in the data than in the model, which is a significant difference, given the precision of the data set.

Table 4 performs a number of robustness checks. The variations of the persistency of income shocks, staying in the benchmark (BM) version of the model, does not bring significant changes. Interestingly, introducing altruistic bequests (AB), both without and with (IGP) intergenerational earnings persistence, bring only a very small increase in wealth concentration, far less than what is needed to reconcile model and data. Things change once we introduce capital gains, in a way that is very nonlinear in the standard deviation  $\sigma$  of the return risk (The letters 'E' and 'T' refer to the two specifications of return risk, cf. Section 4.2). With specification 'E' and  $\sigma = 0.1$ , no big change is visible, with  $\sigma = 0.2$ , we come much closer to the data, with  $\sigma = 0.22$ , we explain  $\gamma W$  well, with  $\sigma = 0.25$ , we exaggerate already. Note that this specification leads to a substantial increase in the top 1% wealth share, more than what we observe in the data, but that should not bother us too much, since it appears that the decision problem of ordinary households is not very well modelled, so that the savings incentives of normal households may be severely underestimated. Specification 'E' explains well the wealth distribution in the range of the top 1% to the top 0.01%, but fails at the very top. Let's now turn to specification 'T', which was tailored to account for that group. We see that with a parameter combination of  $\sigma = 0.35$  and  $\sigma^W = 0.15$ , the model does a good

Model	$\rho_0$	$\rho_1$	$\beta^{altr}$	$\rho_{IG}$	$\sigma$	$\sigma^W$	W1%	$\gamma W$	$\gamma W_{400}$	$\gamma Y$	$\gamma Y_{CG}$
US92							30.0	1.60	1.36	1.70	1.63
BM	1.00	0.05	0	0.0	0.00	0.00	25.5	1.94	2.06	1.88	1.88
		0.10					24.2	1.92	2.09	1.87	1.87
		0.02					25.9	1.95	2.05	1.89	1.89
	0.98	0.00					26.1	1.93	2.19	1.89	1.89
AB	1.00	0.05	1				29.0	1.91	2.12	1.87	1.88
IG			1	0.4			28.8	1.89	2.13	1.88	1.88
			1	0.7			28.6	1.90	2.14	1.88	1.88
CG			0	0.0	E0.10		28.4	1.90	2.06	1.88	1.88
					E0.20		37.5	1.65	1.83	1.80	1.72
					E0.22		40.2	1.57	1.69	1.75	1.64
					E0.25		44.6	1.46	1.52	1.65	1.53
					E0.30		52.6	1.31	1.37	1.49	1.37
					T0.20	0.15	34.1	1.76	1.83	1.84	1.79
					T0.25	0.15	34.3	1.73	1.64	1.82	1.76
					T0.30	0.15	34.5	1.68	1.48	1.80	1.71
				T0.35	0.15	34.9	1.63	1.37	1.77	1.66	
				T0.40	0.15	35.2	1.59	1.30	1.74	1.61	

Notes:

The row US92 refers to US data of 1992

Models: BM: benchmark; AB: altruistic bequests; IGP: altruistic bequ. with intergenerational earnings persistence; CG: model with capital gains

Parameters:  $\rho_0, \rho_1$ : Equ. (22);  $\beta^{altr}$ : Equ. (17);  $\rho_{IG}$ : Equ. (24);  $\sigma$  and  $\sigma^W$ : Equ. (23)

W1%: top 1 percent wealth share;  $\gamma W, \gamma Y, \gamma Y_{CG}$ : Parameter  $\gamma$  of Pareto distribution for wealth, income (excl. capital gains) and income incl. capital gains; obtained from ratio top 1% to top 0.01% share

$\gamma W_{400}$ , Parameter  $\gamma$  for wealth, obtained from ratio top 400 to top 100 wealth owners

Table 4: Steady state results

job at explaining all the steady state characteristics that we consider. Table 5 provides some more details for this parameter combination. A standard deviation of 35 percent annually for the very top group may appear high, but is not totally implausible if we consider that the standard deviation of a single stock is about 50 percent annually, and many people at the very top have invested heavily in their own company. However, at the moment I don't know any systematic microeconomic evidence to pin down this parameter more precisely.

		top shares		$\gamma$ from ratios of top shares		
		1%	top400	1/0.1	0.1/0.01	top400/top100
Wages	model	14.76	0.40	1.81	1.85	1.89
	US 1992	14.76	-	1.86	1.77	-
Wealth	model	34.86	2.32	1.66	1.61	1.37
	US 1992	30.00	1.66	1.59	1.62	1.36

Total estates, percent of GDP: 2.10 (US: 2)  
Percentage decedents filing: 4.34 (US: 2.88)

Table 5: Results benchmark model

As I have argued above, this kind of steady state exercise is problematic, given that the we observe large changes in the income distribution over the last decades. The next section looks at stochastic simulations.

## 5.2 The evolution of the distribution over time

In this section I show some graphs that present simulation results for the full version of the model with aggregate distribution shocks. To solve the model, I have assume that there are two aggregate states with different income inequality. In the high inequality state, we have  $\gamma_\xi = 1.8$ , as described in Section 4.1. In the low inequality state, we have  $\gamma_\xi = 2.7$ . The transition probabilities are symmetric, with a probability of staying in the same state of 0.98.

In the experiments that are reported in each of the following graphs, the aggregate shock is not simulated randomly, but was chosen deliberately to mimic the historical development of the 20th century. We start early in history (call the starting year 1700) with a wealth distribution that is relatively equal at the top (the steady state distribution of the benchmark model). Then we give the model 220 years (until 1920) of high earnings inequality, to see how long the model takes to build up the high wealth inequality that we observe at the beginning of the 20th century. Then we switch for 50 years (1920-1970) to the state with low earnings earnings inequality, and then we switch back to high in equality, and follow the system for another 130 years.

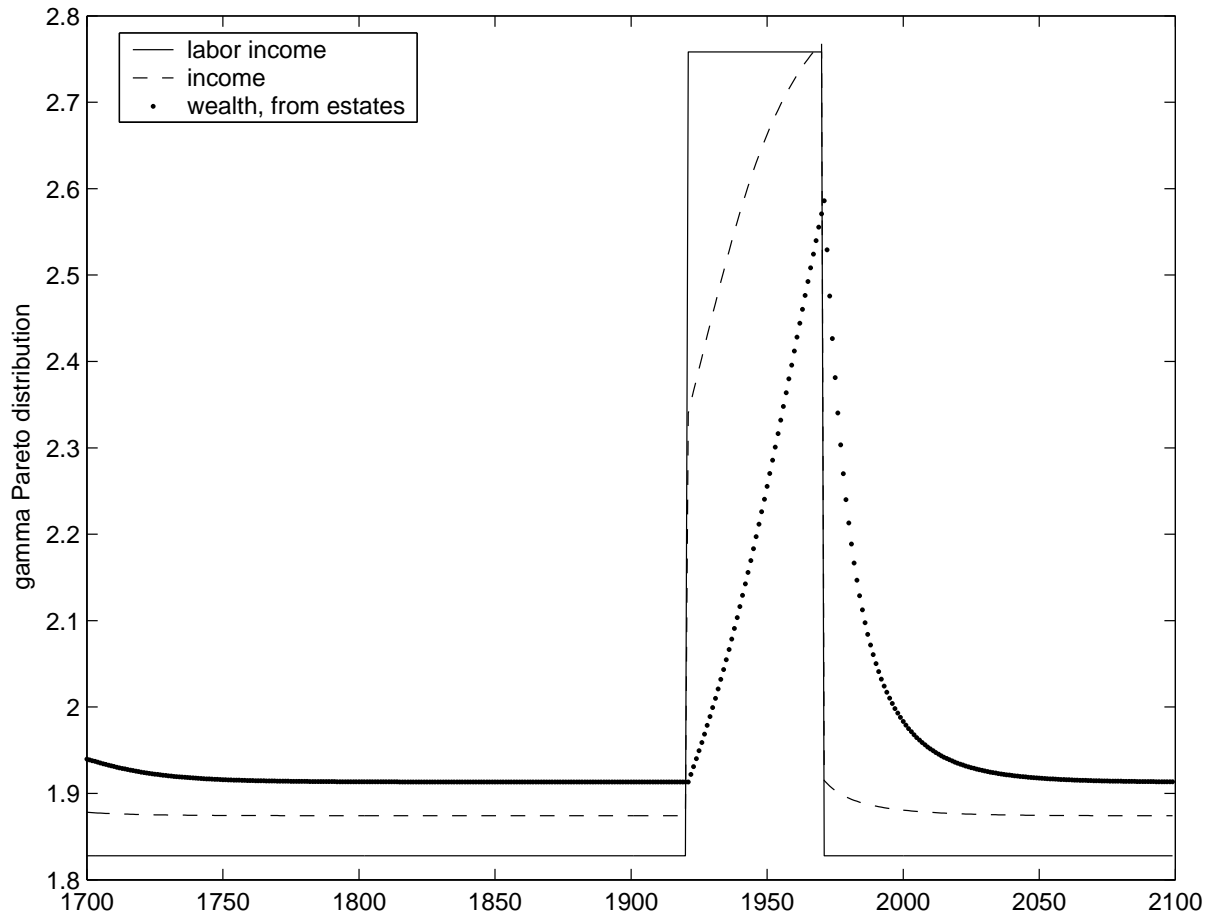


Figure 6: Simulation benchmark model

Figure 6 refers to the benchmark calibration of the model. It displays the parameter  $\gamma$  of the Pareto distribution of labor earnings, total income and wealth. (For earnings, this parameter is exogenous and displayed for reference). In each case, the  $\gamma$  was computed by comparing the share of the top 1% to the top 0.01% of the respective distribution.

In addition to the fact that the model predicts too little wealth concentration (which we already knew), we see another shortcoming of this model: after a change in the earnings distribution, wealth concentration starts to decline immediately, and strongly. This is not in line with what we see in Figure 2, where the decline in the  $\gamma$  of wage income leads to a very small change in the  $\gamma$  of wealth.

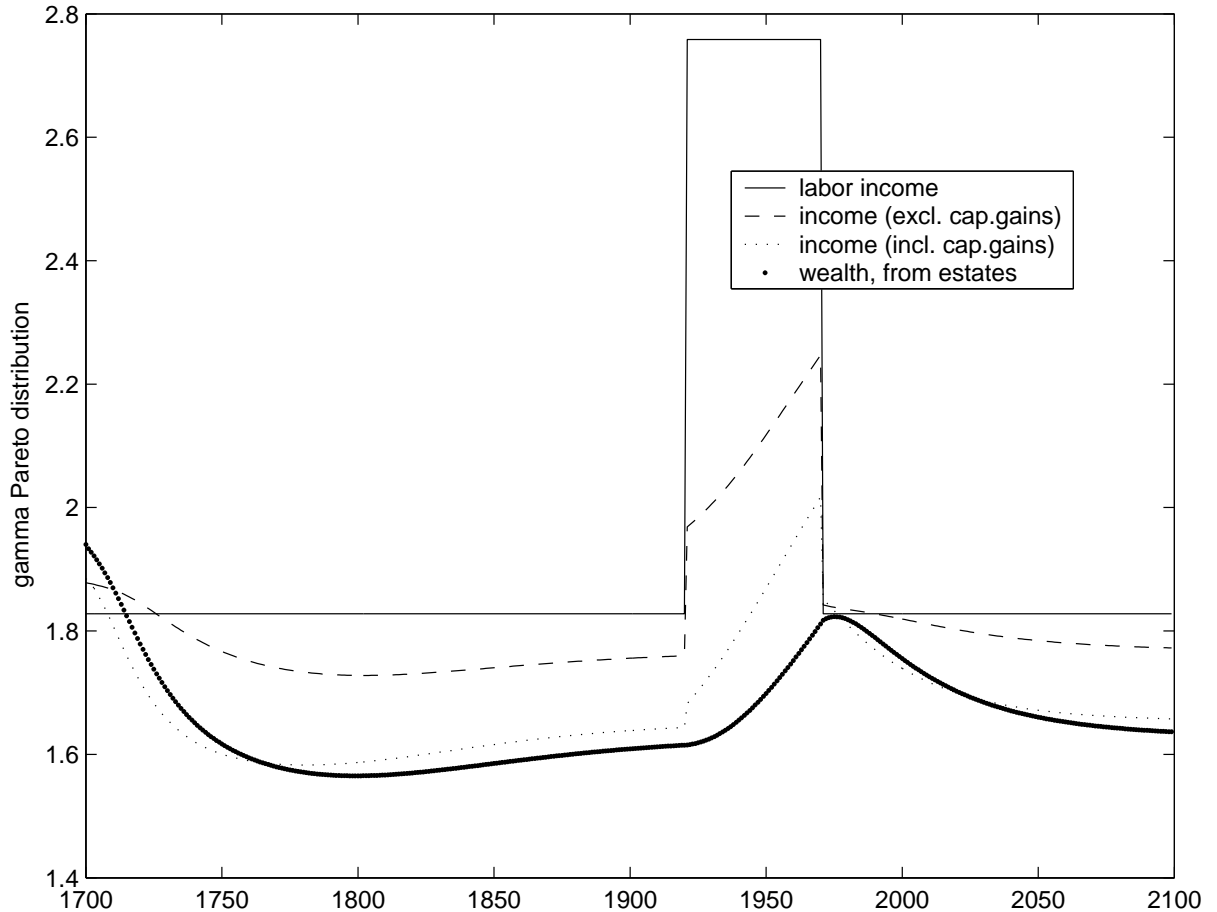


Figure 7: Simulation model with capital gains, wealth distribution

Figure 7 displays the same series for the model with capital gains; I chose the specification with  $\sigma = 0.15$  for everybody and  $\sigma = 0.35$  for top entrepreneurs (before last line in Table 4 which performed best in the steady state experiment). We see that the model does well not just in explaining the steady state distribution, but also the time series evidence. After the change in earnings distribution, the wealth distribution changes only very sluggishly. Once a high level of wealth inequality is reached, as was the case in the beginning of the 20th century, the return risk keeps it unequal for a long time, even if the earnings inequality decreases a lot. If earnings inequality then reverts back to its earlier level, it brings only a small change in wealth inequality, which is precisely what we have seen in the last 20 years, and which has puzzled a number of researchers (Kopczuk and Saez, 2004; Scholz, 2003).



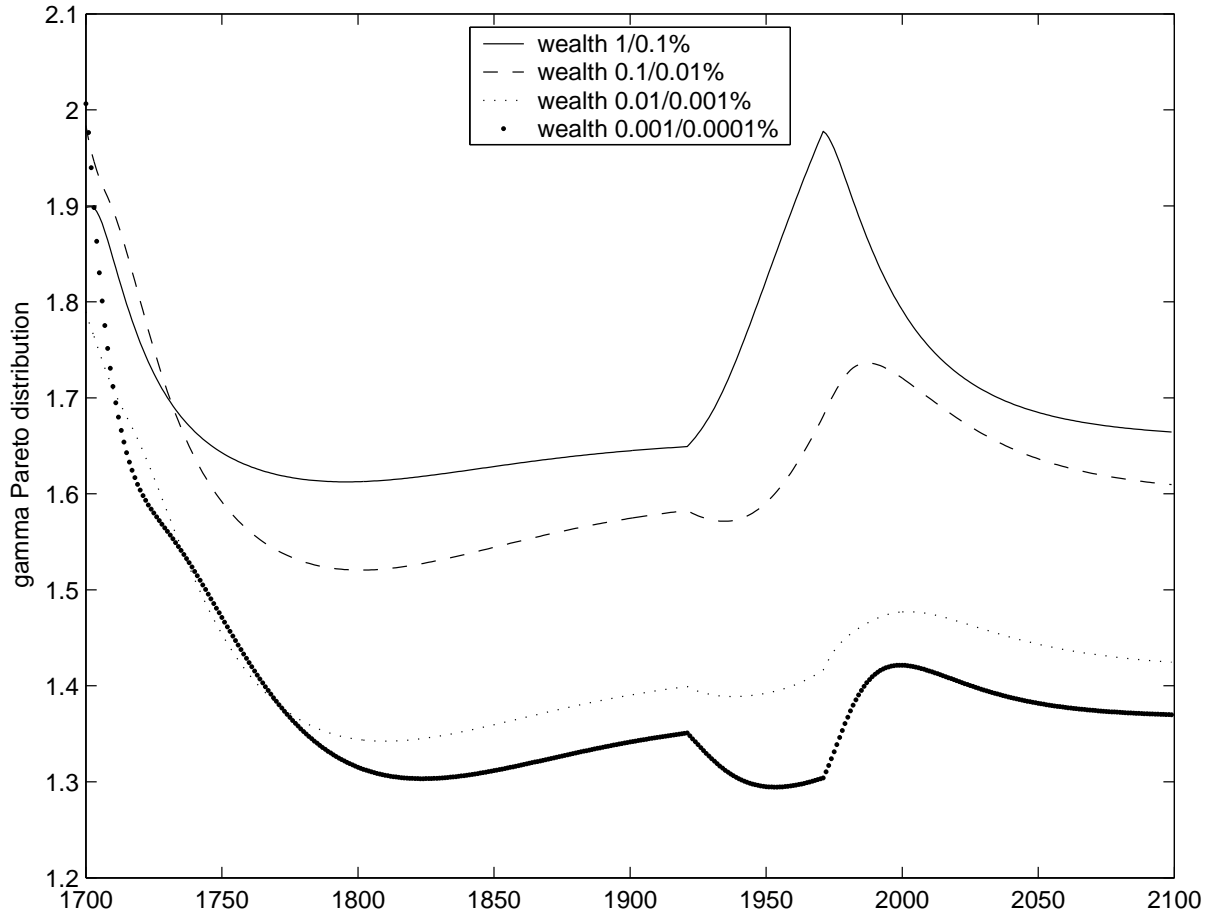


Figure 8: Simulation model with capital gains, wealth distribution

Figure 8 provides more detailed information about the wealth distribution, showing the Parameter  $\gamma$  relating to different quantiles. We see that there are systematic deviations from the Pareto distribution. That the Parameter  $\gamma$  is lower for the very top groups is in line with the data. That the  $\gamma(1\%/0.1\%)$  is lower than the  $\gamma(0.1\%/0.01\%)$  is not observed in the data (cf. Figure 3), but this effect is rather small (a difference of about 0.05). I do not know at the moment whether the fit can be improved by fine-tuning the parameters.

Obviously, this model does much better in explaining the macro facts than the benchmark distribution. How about the micro evidence? What does this model predict about the savings behavior of the old and the rich? Figures 9 and 10 displays consumption as a function of cash on hand for retired households. We compare the retired households in the benchmark calibration to the model with capital gains, in which we have to distinguish the old with high return risk (because they were top entrepreneurs) from other retired households.

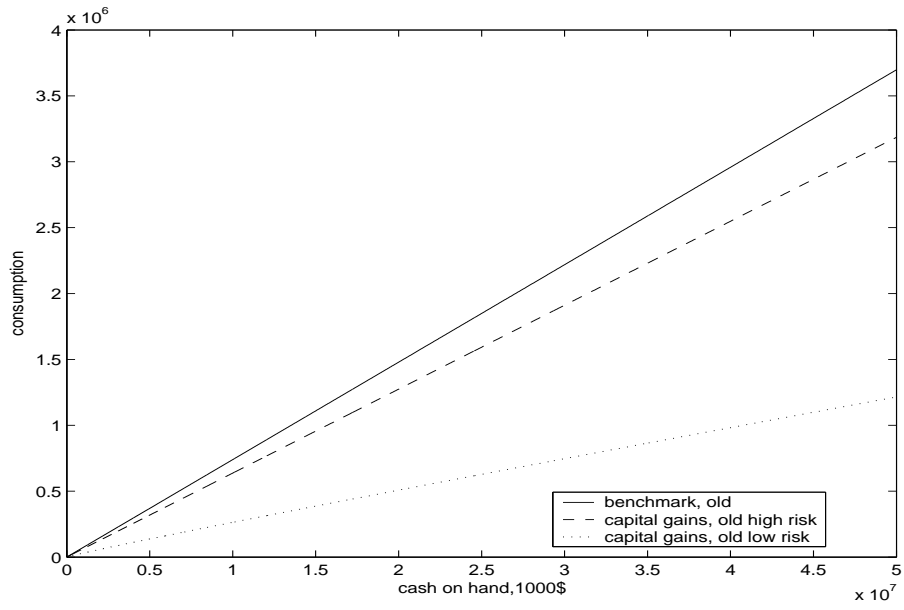


Figure 9: Consumption function, low assets

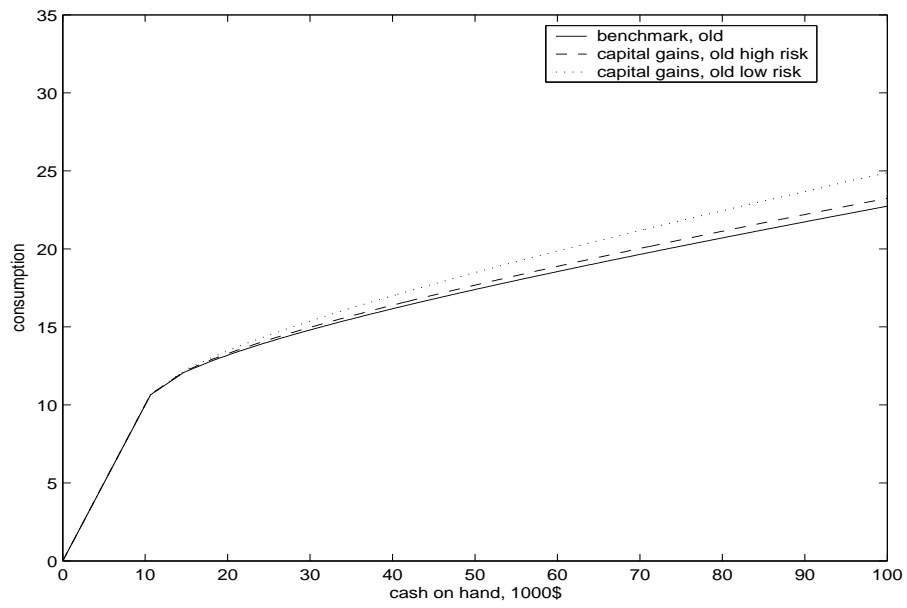


Figure 10: Consumption function, high assets

Figure 10 shows that the consumption function is similar for low wealth levels, but Figure 9 shows that the high wealth risk reduces the consumption of the old substantially, due to a kind of precautionary savings motive. In the high-risk specification, an old individual with the wealth of Bill Gates (net worth of 50 billion dollars, upper end of the x-axis) would consume a bit more than 1 billion dollars annually. Given that the interest rate in this economy is more than 6 percent, these households would on average still accumulate assets. Consumption of more than 1 billion dollars annually may still appear unrealistically high, but one should see that for the purpose of wealth accumulation it does not matter whether households really consume this amount, or rather give it away for charitable purposes.

Figures 11 and 12 present the same graph for simulations of the capitalist spirit model, for the cases  $\eta_k = 1$  and  $\eta_k = 1.8$ . In both cases, the parameter  $m_A$  was set to 100000, so that the capitalist spirit motive is relevant only for very rich households, and the parameter  $b_A$  was adjusted such that we reach the right level of inequality in 1920. We see that both versions of the model do a good job in explaining the available data. In the range of the top 1% to top 0.01% we are close to a Pareto distribution, while at the very top,  $\gamma$  is even lower, in line with the data. For the time period under consideration, it is difficult to distinguish this model from the model with high capital gains. In the *very* long run, several centuries later, the capitalist spirit model would have caused even higher inequality. The consumption level of the very rich old households here is comparable to the level of the high-risk old households of the earlier calibration. The advantage is that now we need not rely on high, perhaps exaggerated, levels of return risk.

## 6 Conclusions

To explain the wealth distribution at the top, we must take care to model correctly the return risk faced by many wealth owners. Substantial wealth risk leads to a big reduction in consumption, and helps explaining the wealth accumulation of the richest households. However, from macro data it is not easy to distinguish such a specification from a model with a capitalist spirit motive. It may be necessary to go back to micro data to estimate the right degree of return risk.

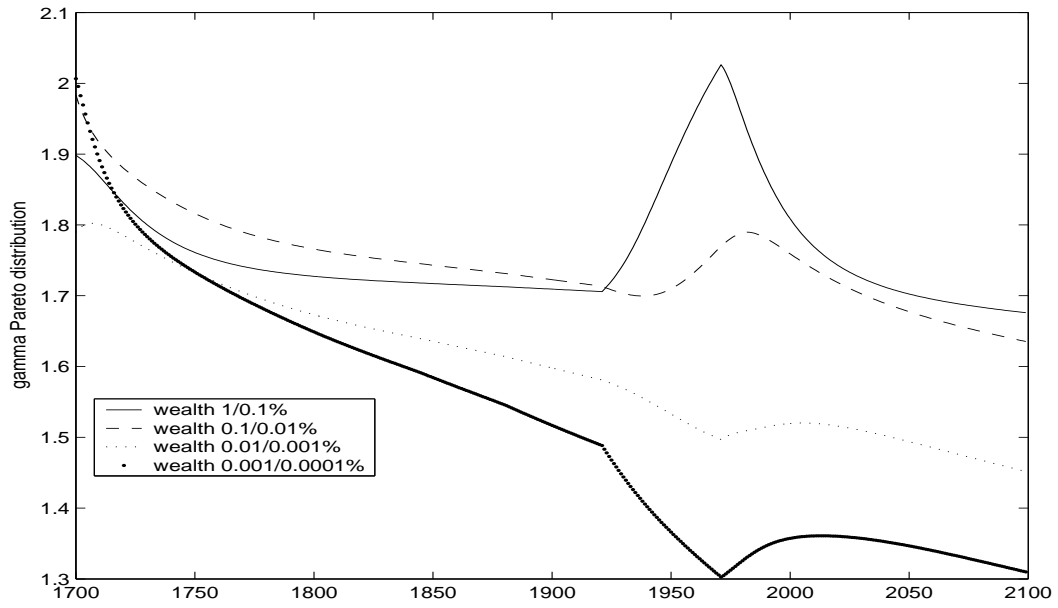


Figure 11: Simulation model with capitalist spirit,  $\eta_k = 1$ , wealth distribution

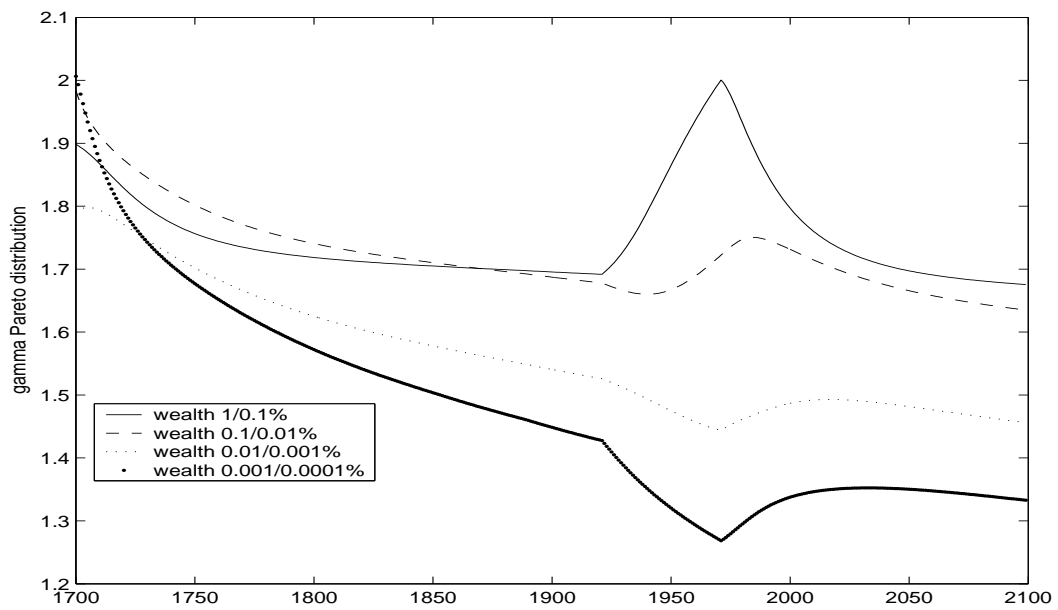


Figure 12: Simulation model with capitalist spirit,  $\eta_k = 1.8$ , wealth distribution

## A Finite state approximation to idiosyncratic productivity

If a household is young at time  $t$ , it can be in  $n + 1$  different productivity states,  $\xi_t^p \in \{\Xi_0^p, \dots, \Xi_n^p\}$ .  $\Xi_0^p$  denotes the “worker state”, States  $\Xi_i^p, i = 1, \dots, n$  are “entrepreneurial states”. They represent the upper tail of the income distribution which can be described well by a Pareto distribution.

If  $\xi_t^p > 1$ , then productivity can take on one of  $m$  different values  $\xi_t \in \{\Xi_1, \dots, \Xi_m\}$ , which are chosen to approximate a Pareto distribution.

Some notation:

- $\pi_i^y$  is the fraction of people that are young and in productivity state  $i, i = 0, \dots, n$ . We have  $\sum_{i=0}^n \pi_i^y = \pi^{young}$ .
- $\pi_i^o$  is the fraction of people that are old and whose children will start in productivity state  $i, i = 0, \dots, n$ . We have  $\sum_{i=0}^n \pi_i^o = 1 - \pi^{young}$ .
- $\pi_{ij}$  is the probability that  $\xi_t = \Xi_j$  conditional on  $\xi_t^p = \Xi_i^p$ .
- $T_{ij}^{yy}$  is the probability that a young agent switches to type  $\xi_{t+1}^p = \Xi_j^p$ , conditional on starting from  $\xi_t^p = \Xi_i^p$ .
- $\pi_j^s$  is the fraction of young people (not conditional on being entrepreneur) with current income  $\Xi_j, j = 1, \dots, m$ .

The parameter of the idiosyncratic income process are now determined by the following four steps:

1. Fix the  $\pi_i^y$  exogenously.  $\pi_0^y$  is set to  $0.95\pi^{young}$ , since 5 percent of the young are entrepreneurs.

The  $\pi_i^y, i = 1, \dots, n$  are the probabilities of the different types in the discrete approximation of the Pareto distribution (in Table 4.1, these would be 0.3, 0.4, 0.2, 0.09, 0.009, 0.00097, 0.00003, each multiplied by 0.05). Given Pareto parameters  $\gamma$  and  $C$ , we then compute  $\Xi_i^p, i = 1, \dots, n$  as the conditional expectation of the fractiles that these types represent (in this example, 0-0.3, 0.3-0.7, 0.7-0.9 etc.).

2. Find the  $\pi_i^o$ . In the benchmark case without intergenerational earnings persistence, there is only one type of retired household. It’s frequency  $\pi_0^o$  is given by  $\pi_0^o = \pi^{old} = 1 - \pi^{young}$ .

In the benchmark case *with* intergenerational earnings persistence, there are as many retired household types as there are young household types (although some of them may have a zero frequency). The type of the retired household determines the type of the young household by which it is replaced on death. This type is determined at the time of retirement; it reflects the productivity state of the household immediately before retirement, plus a random shock. (At the moment when a household retires, the uncertainty about the starting type of its offspring is resolved; in later periods, children may of course change their type.) We assume that i) the children of workers start as workers; ii) for entrepreneurs, the transition probabilities are chosen so as to minimize the variance of  $\log(\xi_{t+1}^p(\text{child in first period}))$ , conditional on  $\xi_t^p(\text{parent at retirement})$ , under the constraint

$$E \log(\xi_{t+1}^p(\text{child in first period})) = \rho_{IG} \log(\xi_t^p(\text{parent at retirement})) \quad (25)$$

The expectation in ((25)) is conditional on information at the time of retirement. This implies a regression towards the mean of productivity.

This rule gives us a probability distribution over retirement states, and thereby a probability distribution over entry states (the productivity state of newly born households).

3. We determine the  $T_{ij}^{yy}$  by solving

$$\min_{T_{ij}^{yy}} \sum_{i,j} T_{ij}^{yy} \left( \log(\Xi_j^p) \right)^2 \quad (26)$$

subject to

$$\sum_{j=0}^n T_{ij}^{yy} = 1 - p^{retire}, \quad i = 0, \dots, n \quad (27a)$$

$$\sum_{j=0}^n \pi_j^y T_{ji}^{yy} + \pi_i^o p^{death} = \pi_i^y, \quad i = 0, \dots, n \quad (27b)$$

$$\frac{\sum_{j=0}^n T_{ij}^{yy} \log(\Xi_j^p)}{1 - p^{retire}} = \rho_0 \log(\Xi_i^p) - \rho_1, \quad i = 1, \dots, n \quad (27c)$$

(27a) and (27b) guarantee that the  $T_{ij}^{yy}$  are consistent with the frequencies  $\pi_j^y$  etc. (27c) is the discrete analogue to (22). The objective(26) amounts to minimizing the sum of the conditional variances of the  $\log(\Xi_j^p)$ . I chose this criterion since the transition variances are still rather high, even when minimized.

4. We find the  $\pi_{ij}$  by minimizing

$$\min_{\pi_{ij}} \sum_{i,j} (\pi_{ij})^2 (\log(\Xi_i^p) - \log(\Xi_j^p))^2 \quad (28)$$

subject to

$$\sum_{j=1}^m \pi_{ij} = 1, \quad i = 1, \dots, n \quad (29a)$$

$$\sum_{i=1}^n \pi_i^y \pi_{ij} = \pi_j^s, \quad j = 1, \dots, m \quad (29b)$$

$$\sum_{j=1}^m \pi_{ij} \Xi_j = \Xi_i^p, \quad i = 1, \dots, n \quad (29c)$$

Again, (29a) and (29b) guarantee that the  $\pi_{ij}$  are consistent with the frequencies  $\pi_i^y$  and  $\pi_j^s$ . (29c) is the discrete analogue to (22). The idea of the objective(28) is to strongly punish deviations  $\log(\xi_t^p) - \log(\xi_t)$ .

## B Proofs

*Proof of Lemma 1.* Define  $u \equiv y - x$ , then  $E[u|x] = 0$  and we get from Jensen's inequality that  $E[(x+u)^{1+k}|x] \geq x^{1+k}$  for any  $k > 0$ . Therefore

$$E y^{1+k} = E \left[ (x+u)^{1+k} \right] = E \left[ E \left[ (x+u)^{1+k} | x \right] \right] \geq E x^{1+k} \quad (30)$$

For a variable  $z \sim P(C, \gamma)$ , a simple calculation shows that  $E z^{1+k} = \frac{\gamma}{\gamma-1-k} C^{(1+k)/\gamma}$  for all positive  $k < \gamma - 1$  (and analogously for  $y$ ). This implies that  $E [z^{1+k}] =$

$E[z]^{1+k} \frac{\gamma-1}{\gamma-1-k} \left(\frac{\gamma-1}{\gamma}\right)^k$  A straightforward calculation shows that

$$\frac{d}{d\gamma} \left[ \frac{\gamma-1}{\gamma-1-k} \left(\frac{\gamma-1}{\gamma}\right)^k \right] = -\frac{(\gamma-1)^k}{\gamma^{-k}(\gamma-1-k)^2} \frac{1+k}{\gamma} k < 0 \quad (31)$$

Since  $E(x) = E(y)$ , we obtain from the inequality (30) that  $\gamma_y \leq \gamma_x$ .  $\square$

## C Computational method

[TO BE FILLED IN.]

## References

- Altonji, J. and Villanueva, E.: 2002, The effect of parental income on wealth and bequests, mimeo, Northwestern University.
- Bertaut, C. C. and Haliassos, M.: 1997, Precautionary portfolio behavior from a life-cycle perspective, *Journal of Economic Dynamics and Control* **21**(8-9), 1511–1542.
- Bitler, M., Moskowitz, T. and Vissing-Jrgensen, A.: 2004, Testing agency theory with entrepreneur effort and wealth, *Journal of Finance*, forthcoming.
- Campbell, J. Y., Lettau, M., Malkiel, B. G. and Xu, Y.: 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *Journal of Finance* **56**(1), 1–43.
- Carroll, C. D.: 1998, Why do the rich save so much?, NBER Working Paper 6549.
- Carroll, C. D.: 2001, A theory of the consumption function, with and without liquidity constraints, *Journal of Economic Perspectives* **15**(3), 23–45.
- Carroll, C. D.: 2002, Portfolios of the rich, in L. Guiso, M. Haliassos and T. Jappelli (eds), *Household portfolios*, MIT Press, Cambridge and London.
- Castaneda, A., Diaz-Gimenez, J. and Rios-Rull, J.-V.: 2002, Accounting for the U.S. earnings and wealth inequality, forthcoming in *JPE*.
- De Nardi, M.: 2002, Wealth inequality and intergenerational links, FRB Minneapolis, Staff Report 314.
- den Haan, W. J.: 1997, Solving dynamic models with aggregate shocks and heterogenous agents, *Macroeconomic Dynamics* **1**, 355–86.
- Diaz-Gimenez, J., Quadrini, V. and Rios-Rull, J.-V.: 1997, Dimensions of inequality: Facts on the u.s. distributions of earnings, income, and wealth, *Federal Reserve Bank of Minneapolis Quarterly Review* **21**(2), 3–21.
- Dynan, K. E., Skinner, J. and Zeldes, S. P.: 2000, Do the rich save more?, NBER Working Paper 7906.
- Kopczuk, W. and Saez, E.: 2004, Top wealth shares in the united states, 1916-2000: Evidence from estate tax returns, University of Berkeley.
- Krusell, P. and Smith, Anthony A, J.: 1998, Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* **106**(5), 867–96.

- Levy, M.: 2003, Are rich people smarter?, *Journal of Economic Theory* **110**(1), 42–64.
- Levy, M. and Solomon, S.: 1997, New evidence for the power-law distribution of wealth, *Physica A* **242**, 90–94.
- Moskowitz, T. J. and Vissing-Jrgensen, A.: 2002, The returns to entrepreneurial investment: A private equity premium puzzle?, *American Economic Review* **92**(4), 745–78.
- Piketty, T. and Saez, E.: 2003, Income inequality in the united states, 1913-1998, *Quarterly Journal of Economics*.
- Quadrini, V.: 1999, The importance of entrepreneurship for wealth concentration and mobility, *Review of Income and Wealth* **45**(1), 1–19.
- Quadrini, V.: 2000, Entrepreneurship, saving, and social mobility, *Review of Economic Dynamics* **3**(1), 1–40.
- Reiter, M.: 2001, Recursive solution of heterogenous agent models, mimeo, Universitat Pompeu Fabra.
- Saez, E.: 2001, Using elasticities to derive optimal income tax rates, *Review of Economic Studies*.
- Scholz, J. K.: 2003, Wealth inequality and the wealth of cohorts, mimeo, University of Wisconsin-Maddison.
- Zimmerman, D. J.: 1992, Regression toward mediocrity in economic stature, *American Economic Review* **82**(3), 409–29.