Heterogeneous Dynasties and Long-Run Mobility

Jess Benhabib
NYU

Alberto Bisin
NYU

Ricardo T. Fernholz
Claremont McKenna College

May 29, 2020

Abstract

Recent empirical work has demonstrated a positive correlation between grandparent-child wealth-rank, even after controlling for parent-child wealth-rank, as well as a positive correlation between dynastic wealth-ranks across almost 600 years. We show that a simple heterogeneous agents model with idiosyncratic returns to wealth generates a realistic wealth distribution but fails to capture these long-run patterns of wealth mobility. However, an extension of the basic model which includes persistent heterogeneity in returns to wealth is able to simultaneously match the wealth distribution, short-run wealth mobility, and long-run wealth mobility.

JEL Codes: E21, E24

Keywords: Wealth Distribution, Inequality, Mobility, Incomplete Markets

1 Introduction

Recent heterogeneous agents modeling of consumption-saving decisions has successfully been able to identify the main drivers of wealth inequality (Benhabib et al., 2019; Cagetti and De Nardi, 2006; Castañeda et al., 2003; Hubmer et al., 2016; Kindermann and Krueger, 2015; Krusell and Smith, 2015; Quadrini, 2000). This literature shows that heterogeneous agent models with stochastic labor earnings and idiosyncratic returns to wealth can produce fat-tailed distributions of wealth which match the data well.[1] These models can also fit reasonably well the inter-generational social mobility of wealth, producing a realistic parent-child wealth-rank correlation (Benhabib et al., 2019).

More recently, however, empirical results suggest significant grandparent-child wealth-rank correlations (Boserup et al., 2014) even after controlling for the effects of parent wealth on child wealth. Furthermore, even long-run wealth-rank correlations appear to persist across generations (Barone and Mocetti, 2016; Clark, 2014; Clark and Cummins, 2015). With respect to this dimension of inter-generational mobility, the heterogeneous agents models in the literature are bound not to fare well, in that they can hardly generate a large enough coefficient for grandparent-child wealth-rank nor a large enough correlation for dynastic wealth-ranks over very long time periods.[2] It is not difficult to envision extensions of these models which could produce significant grandparent-child and even long-run wealth-rank correlations, e.g., postulating intergenerational autocorrelation in earnings and in the rate of return to wealth. First, however, the evidence is not favorable to the existence of independent direct causal effects across generations, beyond parent-child effects.[3] Secondly, large intergenerational autocorrelations e.g., of returns to wealth are bound to generate even fatter tails in the wealth distributions than we observe.[4] Finally, extending the model along these lines would require postulating very long intergenerational autocorrelation in earnings and in the rate of return to wealth to capture long-run persistence, which does not appear plausible and certainly not parsimonious as an explanation.

In this paper we instead study a simple heterogeneous agents model and extend it to introduce persistent heterogeneity in the rate of return to wealth across generations. In

[1]See Benhabib et al. (2017) for a discussion of the relative role of earnings and returns to wealth.

[2]We discuss the theoretical reasons why this class of models produces limited long-run rank-wealth correlations in Section 2.1. In Section 3 we confirm this by means of simulation analysis.

[3]We discuss this evidence in the next section.

[4]We confirm this in Section 3, by means of simulation analysis.

other words, we allow households in some dynasties to have their wealth grow faster on average than households in other dynasties. While we do not take a stand on the precise interpretation of this form of persistent heterogeneity, we note that it can be seen as a formalization of a latent factor representation of various dynastic characteristics suggested in the literature. We summarize the findings of this literature next.

1.1 Long-Run Rank-Wealth Correlation

In this section we briefly discuss the evidence documenting wealth-rank correlations across generations and its interpretation in the literature. First of all, a positive correlation between grandparent-child wealth rank, even after controlling for parent-child wealth rank, is documented in Boserup et al. (2014), using three generations of Danish wealth data. Since parent and grandparent wealth are correlated, and also possibly measured with error, they implement a two stage least squares procedure to identify direct grandparent effects. Though grandparent effects do not necessarily go through parents, they conclude in favor of indirect effects, which they interpret as "social status." Relatedly, Braun and Stuhler (2018) identify a possible direct causal effect of grandparent-child interactions exploiting quasi-exogenous variation in the time of grandparents' death during World War II." They find no effects of direct contacts between grandparents and grandchildren; concluding in favor of grandparent effects operating through indirect mechanisms. Finally, Warren and Hauser (1997), using data from the Wisconsin Longitudinal Study, find no evidence for an independent influence of grandparents once they condition on the status of both parents.

The evidence on long-run dynastic wealth-rank correlation is striking. Clark (2014) and Clark and Cummins (2015) find high persistence of wealth across five generations using data on rare surnames in England and Wales between 1858 and 2012, and Barone and Mocetti (2016) find significant positive wealth elasticities as well as occupational persistence for families in Florence between 1427 and 2011.[5] While these data are less amenable to statistical inference, Clark (2014) and Clark and Cummins (2015) argue that wealth, education, or occupational status are transmitted via an underlying and unobserved latent factor which is a representation of abilities, preferences, dynastic network connections, or other relevant characteristics. This interpretation is also consistent with the evidence on grandparent-child

[5]Long-run persistence is also documented by Lindahl et al. (2015), Modin et al. (2013), Long and Ferrie (2013), and Braun and Stuhler (2018) on occupational and educational attainment, and by Chan and Boliver (2013) and Hertel and Groh-Samberg (2014) on social class.

correlations, as argued in Stuhler (2012) and Braun and Stuhler (2018). Whether and how this latent factor is affected by the environment and by historical and institutional dimensions is hard to identify with the available data (Mare, 2011; Braun and Stuhler, 2018).

2 Models of Wealth Dynamics

In this section we develop the theory behind our analysis of long-run persistence in rank-wealth correlation. We study rank-based models of wealth dynamics, that is, models in which the growth rate of wealth depends on the wealth-rank rather than e.g., the wealth level. These models are convenient for our analysis as they allow for an analytic characterization of asymptotic wealth-ranks and approximate standard heterogeneous agents models well. In the following, we first introduce a standard rank-based model and relate it to heterogeneous agents models. We then introduce persistent heterogeneity in the rate of return to wealth across generations into the standard rank-based model. Finally, we derive theoretical results about long-run persistence of wealth-rank correlations.

2.1 Rank-Based Models

Consider an economy populated by N households, indexed by $i = 1, \ldots, N$. Ranking households by their wealth, let $\rho_t(i)$ denote the wealth-rank of household i at time t, so that $\rho_t(i) < \rho_t(j)$ if and only if $w_i(t) > w_j(t)$ or $w_i(t) = w_j(t)$ and $i < j$. We define the ranked wealth processes $w_{(1)} \geq \cdots \geq w_{(N)}$ by $w_{(\rho_t(i))}(t) = w_i(t)$. The aggregate wealth of the economy is then $w(t) = w_1(t) + \cdots + w_N(t)$.

For each household $i = 1, \ldots, N$, wealth dynamics are given by

$$d \log w_i(t) = \alpha_{\rho_t(i)}\, dt + \sigma_{\rho_t(i)}\, dB_i(t), \tag{2.1}$$

where B_i is a Brownian motion. We normalize, without loss of generality, the average growth rate of the economy to zero; that is, $\alpha_1 + \cdots + \alpha_N = 0$. Banner et al. (2005) show that the rank-based model (2.1) admits a stationary distribution if the parameters α_k satisfy

$$\alpha_1 + \cdots + \alpha_k < 0, \text{ for } k < N. \tag{2.2}$$

Condition (2.2) suffices to guarantee that no household in the top ranks grows faster than in the lower ranks, which would cause it to break away from the average population wealth. We will show in the next section that this condition is consistent with rate of returns to wealth which are constant or even increasing in wealth in a standard heterogeneous agent model of wealth dynamics.

Proposition 2.1. *Consider a rank-based model* (2.1) *that satisfies* (2.2) *and also*

$$\sigma_{k+1}^2 - \sigma_k^2 = \sigma_k^2 - \sigma_{k-1}^2, \tag{2.3}$$

for all $k = 2, \ldots, N-1$. The ranked wealth processes satisfy

$$\mathbb{E}\left[\log w_{(k)}(t) - \log w_{(k+1)}(t)\right] = \frac{\sigma_k^2 + \sigma_{k+1}^2}{-4(\alpha_1 + \cdots + \alpha_k)}, \tag{2.4}$$

for all $k = 1, \ldots, N-1$, where the expectation is taken with respect to the stationary distribution.

It follows that wealth inequality at the stationary distribution is increasing in the volatility parameters σ_k and decreasing in the sums of the relative growth parameters α_k.[6]

Rank-Based Model as Approximation

We introduce a simple heterogeneous agents consumption-saving model, along the lines of Benhabib et al. (2019) and Benhabib et al. (2011), and show that it can be formally mapped into an approximated rank-based model such as (2.1). In Section 3 we will then show that an appropriate calibration of this model indeed approximates the heterogeneous agents consumption-saving model well.[7]

Each household $i = 1, \ldots, N$ has CRRA preferences and a joy-of-giving bequest motive. In equilibrium, the intergenerational wealth dynamics for each household $i = 1, \ldots, N$ follow

$$w_i(t+1) = \lambda(r_{i,t})w_i(t) + \beta(r_{i,t}, y_{i,t}), \tag{2.5}$$

[6] Ichiba et al. (2011) show that the ratio of each pair of adjacent ranked wealth processes $w_{(k)}/w_{(k+1)}$ follows a Pareto distribution, with the Pareto parameter for each k depending on the parameters α_k and σ_k according to (2.4).

[7] More generally, the model (2.1) can be calibrated to approximate many different dynamic models and real-world phenomena that exhibit Pareto-like distributions (Fernholz, 2017).

where $y_{i,t}$ and $r_{i,t}$ denote, respectively, the labor income and the return on wealth for household i in generation t, and $w_i(t)$ denotes the wealth holdings of household i in generation t. The functions λ and β are the same for all households $i = 1, \ldots, N$ and represent, respectively, the average return on wealth and the average labor income, adjusted for equilibrium household behavior.[8] We refer to the model (2.5) as the Standard model.

Let the function $\pi_t(k)$ identify the index i of the k-th ranked household at time t, so that $\pi_t(k) = i$ if and only if $\rho_t(i) = k$. The *rank-based approximation* of the Standard model (2.5) is the rank-based model (2.1) where the parameters α_k and σ_k are defined by

$$\begin{aligned} \alpha_k &= \mathbb{E}\left[\log\left(w_{\pi_t(k)}(t+1)/w(t+1)\right) - \log\left(w_{\pi_t(k)}(t)/w(t)\right)\right], \\ \sigma_k^2 &= \mathrm{Var}\left[\log\left(w_{\pi_t(k)}(t+1)/w(t+1)\right) - \log\left(w_{\pi_t(k)}(t)/w(t)\right)\right], \end{aligned} \tag{2.6}$$

for each rank $k = 1, \ldots, N$.[9] The parameters α_k and σ_k measure the average and variance of the growth rate of wealth relative to the aggregate at each rank k of the distribution. The relative growth rate parameters α_k represent the main link between the rank-based model (2.1) and the Standard model (2.5): at the stationary distribution of Standard model, the rank-based relative growth rate parameters α_k satisfy, for each rank $k = 1, \ldots, N$,

$$\begin{aligned} \alpha_k &= \mathbb{E}\left[\log\left(w_{\pi_t(k)}(t+1)/w(t+1)\right) - \log\left(w_{\pi_t(k)}(t)/w(t)\right)\right] \\ &= \mathbb{E}\left[\log\left(w_{\pi_t(k)}(t+1)/w_{\pi_t(k)}(t)\right)\right] \\ &= \mathbb{E}\left[\log\left(\lambda(r_{\pi_t(k),t}) + \beta(r_{\pi_t(k),t}, y_{\pi_t(k),t})/w_{\pi_t(k)}(t)\right)\right], \end{aligned} \tag{2.7}$$

[8]The functions β and λ depend on i) the life-span T and the growth rate of labor income over time g; ii) preference parameters η, ψ, and χ, representing the time discount rate, the elasticity of substitution, and the bequest motive, respectively; iii) policy parameters b and ζ, denoting the estate tax on bequests of wealth and the capital income tax rate. More precisely, from Benhabib et al. (2011):

$$\begin{aligned} \lambda(r_{i,t}) &= (1-b)e^{r_{i,t}T}\frac{A(r_{i,t})B(b)}{e^{A(r_{i,t})T} + A(r_{i,t})B(b) - 1}, \\ \beta(r_{i,t}, y_{i,t}) &= (1-b)y_{i,t}\frac{e^{(g-r_{i,t})T} - 1}{g - r_{i,t}}e^{r_{i,t}T}\frac{A(r_{i,t})B(b)}{e^{A(r_{i,t})T} + A(r_{i,t})B(b) - 1}, \end{aligned}$$

with

$$A(r_{i,t}) = r_{i,t} - \frac{r_{i,t} - \eta}{\psi} \qquad \text{and} \qquad B(b) = \chi^{1/\psi}(1-b)^{(1-\psi)/\psi}.$$

[9]The expectations in (2.6) are calculated under the stationary distribution of model (2.5). Note that model (2.5) has a Brownian motion continuous time limit; see Saporta and Yao (2005).

since the expected value of aggregate wealth w satisfies $\mathbb{E}[\log w(t+1)] = \mathbb{E}[\log w(t)]$ by stationarity. From (2.7), we have

$$\begin{aligned}\alpha_k &= \mathbb{E}\left[\log\left(\lambda(r_{\pi_t(k),t})\left(1+\frac{\beta(r_{\pi_t(k),t}, y_{\pi_t(k),t})}{\lambda(r_{\pi_t(k),t})w_{\pi_t(k)}(t)}\right)\right)\right] \\ &= \mathbb{E}\left[\log\left(\lambda(r_{\pi_t(k),t})\right)\right] + \mathbb{E}\left[\log\left(1+\frac{\beta(r_{\pi_t(k),t}, y_{\pi_t(k),t})}{\lambda(r_{\pi_t(k),t})w_{\pi_t(k)}(t)}\right)\right], \end{aligned} \tag{2.8}$$

which yields a simple decomposition of the rank-based relative growth rates α_k into the two main parametric functions characterizing the model, β and λ. The first term measures the average log return to wealth adjusted for equilibrium behavior, λ, at rank k in the wealth distribution. The second term measures the average log ratio of labor income adjusted for equilibrium household behavior, β, to capital income λw, at rank k in the distribution.

The decomposition (2.8) allows us to map the stability condition for rank-based models in (2.2) into a condition in terms of β and λ in the Standard model (2.5). In the Standard model (2.5), wealth returns $r_{i,t}$ and labor income $y_{i,t}$ are both independent of the wealth rank of household i at time t. Consequently, the first component of the decomposition (2.8), λ, is independent of wealth rank while the second component of this decomposition, $\frac{\beta}{\lambda w}$, is decreasing in wealth rank since wealth w is increasing in rank. In the Standard model, then, $\alpha_1 < \alpha_2 < \cdots < \alpha_N$ because the ratio of labor income to capital income is lower at higher ranks in the wealth distribution. A negative relationship between wealth and returns is not required to satisfy the stability condition (2.2).[10]

Proposition 2.2. *If the Standard model* (2.5) *is stationary, then its rank-based approximation defined by* (2.1) *and* (2.6) *is also stationary.*

2.2 Persistently Heterogeneous Rank-Based Model

We introduce a form of persistent heterogeneity in the average growth rates of households in the rank-based model (2.1). For each household $i = 1, ..., N$, wealth dynamics are given by

$$d\log w_i(t) = \left(\gamma_i + \hat{\alpha}_{\rho_t(i)}\right)\, dt + \sigma_{\rho_t(i)}\, dB_i(t), \tag{2.9}$$

[10] In fact, if follows from this argument that even a positive relationship between wealth and returns in the Standard model could be consistent with condition (2.2).

for each household $i = 1, \ldots, N$, with $\gamma_i \in \{\gamma_\ell, \gamma_h\}$ and $\gamma_h > \gamma_\ell$. In other words, households are of two types: household i is a low (resp. high) type, if $\gamma_i = \gamma_\ell$ (resp. $\gamma_i = \gamma_h$) and household wealth grows more slowly (resp. quickly) on average over time. We assume that n of the households are high types with $\gamma_i = \gamma_h$, and $N - n$ of the households are low types with $\gamma_i = \gamma_\ell$. We keep normalizing the average growth rate of wealth to zero, which in this economy requires $\sum_{k=1}^{N} \hat{\alpha}_k + \sum_{i=1}^{N} \gamma_i = \sum_{k=1}^{N} \hat{\alpha}_k + (N-n)\gamma_\ell + n\gamma_h = 0$.

To admit a stationary distribution, the persistently heterogeneous setup (2.9) must satisfy a stability condition that generalizes the condition (2.2) for the standard rank-based model (2.1) with no heterogeneity. Following Ichiba et al. (2011), this condition states that

$$\sum_{k=1}^{m} \hat{\alpha}_k + \tilde{m}\gamma_h + (m - \tilde{m})\gamma_\ell < 0, \text{ for all } m = 1, \ldots, N-1; \ \tilde{m} = \min(m, n). \tag{2.10}$$

Condition (2.10) generalizes condition (2.2) for the standard rank-based model and it ensures that no top subset of households grows faster than the aggregate. This is sufficient to guarantee that the high-type, high-growth households in the top ranks do not break away from the rest of the population. In particular, this condition ensures that the average relative growth rate of high-type households when occupying the top m ranks of the wealth distribution is negative.

2.3 Long-Run Wealth-Rank Correlations

In this section we provide a theoretical characterization of asymptotic wealth-rank for both the standard rank-based model and the model with persistent heterogeneity. We show that persistent heterogeneity is required to generate long-run wealth-rank correlations.

We start with the implications of the rank-based model (2.1) for mobility. We define occupation times $\xi_{i,k}$, for all i, k, as the fraction of time household i occupies rank k, $\xi_{i,k} = \lim_{T\to\infty} \frac{1}{T}\int_0^T 1_{\{\rho_t(i)=k\}}\, dt$. Note that, by definition, the occupation times must add up to one, so that $\sum_{i=1}^{N} \xi_{i,k} = \sum_{k=1}^{N} \xi_{i,k} = 1$. We can now show the following,

Proposition 2.3. *Occupation times $\xi_{i,k}$ in the standard rank-based model* (2.1) *satisfy*

$$\xi_{i,k} = \frac{1}{N}, \quad \text{a.s., } \textit{for all } i, k. \tag{2.11}$$

Furthermore, for each household i, the asymptotic wealth-rank satisfies

$$\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i)] = \frac{N+1}{2}. \tag{2.12}$$

This result is a consequence of the fact that all households in the model (2.1) display identical expected wealth dynamics. Therefore, i) they will spend equal time in all ranks, (2.11); and ii) they must on average approach the same rank asymptotically; hence, necessarily the median of the distribution, (2.12). In other words, (2.11)-(2.12) imply that higher-ranked households today do not occupy on average higher ranks in the future as well. As a consequence, the standard rank-based model (2.1) cannot produce long-run wealth-rank correlations.

This is not the case when persistent heterogeneity is added to the standard rank-based model, as in (2.9). We turn now to analyze the implications of this model for asymptotic wealth-rank. If household i is a low-type household with $\gamma_i = \gamma_\ell$, then, by symmetry, the fraction of time household i spends in each rank k is equal to the fraction of time any other low-type household spends in each rank k. Thus, we can define the low-type household occupation times $\xi_{\ell,k}$ such that $\xi_{\ell,k} = \xi_{i,k}$, for all ranks $k = 1,\dots,N$. Similarly, if we suppose that household j is a high-type household with $\gamma_j = \gamma_h$, then we can define the high-type household occupation times $\xi_{h,k}$ such that $\xi_{h,k} = \xi_{j,k}$, for all ranks $k = 1,\dots,N$. Because the sum of occupation times across all ranks or individual households must equal one, it follows that the low- and high-type occupation times $\xi_{\ell,k}$ and $\xi_{h,k}$ must satisfy

$$(N-n)\xi_{\ell,k} + n\xi_{h,k} = 1, \tag{2.13}$$

for all $k = 1,\dots,N$.

Proposition 2.4. *Consider a persistently heterogeneous rank-based model* (2.9) *that satisfies* (2.3) *and* (2.10)*. Then, the low- and high-type occupation times $\xi_{\ell,k}$ and $\xi_{h,k}$ satisfy*

$$0 < \xi_{\ell,1} < \xi_{\ell,2} < \cdots < \xi_{\ell,N} < \frac{1}{N-n}, \quad \text{a.s.}, \tag{2.14}$$

and

$$\frac{1}{n} > \xi_{h,1} > \xi_{h,2} > \cdots > \xi_{h,N} > 0, \quad \text{a.s.} \tag{2.15}$$

Because the occupation times for both low- and high-type households satisfy $\xi_{i,1} + \cdots + \xi_{i,N} = 1$, Proposition 2.4 implies that $\xi_{\ell,1} < \xi_{h,1}$ and $\xi_{h,N} < \xi_{\ell,N}$. This means that low-type households spend more time at the lowest ranks of the wealth distribution than high-type households. The following theorem uses this result to show that the heterogeneous rank-based model (2.9) will feature persistence in wealth-ranks over infinitely long time horizons.

Theorem 2.5. *Consider a persistently heterogeneous rank-based model* (2.9) *that satisfies* (2.3) *and* (2.10)*. Then,*

$$\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i)] < \lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(j)] \qquad \textit{if and only if} \qquad \rho_t(i) < \rho_t(j), \tag{2.16}$$

for all households $i, j = 1, \ldots, N$, where the expectations are taken with respect to the stationary distribution.

Theorem 2.5 implies that the long-run asymptotic household rank correlation will be positive in the heterogeneous rank-based model. This is because the higher-ranked households today are expected to occupy higher ranks in the future as well. It is important to emphasize that the expectations in (2.16) are taken unconditional on the types of households i and j — the values of the parameters γ_i and γ_j are unknown and not part of the information set.

The result in Theorem 2.5 is intuitive. Because all high-type households are ex-ante identical, the expected asymptotic rank of these households is the median of the top n ranks of the wealth distribution, since high-type households are expected to occupy higher ranks than low-type households. Similarly, the expected asymptotic rank of low-type households is the median of the bottom $N - n$ ranks. Without knowing its type, the expected asymptotic rank of some household i is thus a weighted average of the medians of the top n and bottom $N - n$ ranks, with the weights equal to the respective probabilities that household i is a high type and that it is a low type. Because higher-ranked households are more likely to be high-type households, it follows that the weight on the median of the top n ranks is greater for such high-ranked households and hence the expected asymptotic rank is also higher.

Because all high-type households are ex-ante identical, the expected asymptotic rank of these households is the median of the top n ranks of the wealth distribution, since high-type households are expected to occupy higher ranks than low-type households. Similarly, the expected asymptotic rank of low-type households is the median of the bottom $N - n$ ranks. Without knowing its type, the expected asymptotic rank of some household i is thus a weighted average of the medians of the top n and bottom $N - n$ ranks, with the weights equal to the respective probabilities that household i is a high type and that it is a low type.

Because higher-ranked households are more likely to be high-type households, it follows that the weight on the median of the top n ranks is greater for such high-ranked households and hence the expected asymptotic rank is also higher.

3 Simulations

In this section we present a simulation analysis of wealth dynamics. We calibrate each of the models of Section 2 and compare their simulated wealth dynamics along various relevant empirical dimensions regarding the wealth distribution and wealth-rank persistence over generations. More precisely, we consider i) the approximated rank-based model (2.1) calibrated using the Standard model (2.5); ii) the persistently heterogeneous rank-based model (2.9), the extension of the rank-based model that includes persistent heterogeneity; iii) an extension of the Standard model (2.5) in which returns to wealth are auto-correlated across generations.

3.1 Calibrations

We first discuss the details of the calibrations we adopt.

The Approximated Rank-Based model. We wish to construct the rank-based approximation of the Standard model, using (2.6) to define the rank-based parameters α_k and σ_k from (2.1). The first step is to parameterize the Standard model, which we do following Benhabib et al. (2011) and Benhabib et al. (2019). We set the household lifespan T equal to 45 years and the growth rate of labor earnings equal to 0.01. The preference parameters η, ψ, and χ are set equal to 0.04, 2, and 0.25, respectively. The estate tax and the capital income tax, b and ζ, are set equal to 0.2 and 0.15, respectively. To model lifetime labor income $y_{i,t}$, we use a six-state Markov chain calibrated to the U.S. Survey of Consumer Finances (SCF). This yields a mean of 6.4 and a standard deviation of 16.1 for $y_{i,t}$, where one unit is equal to 10,000 dollars.[11] Finally, for the idiosyncratic lifetime return on wealth, $r_{i,t}$, we use a four-state Markov chain that is calibrated so that the average and standard deviation of these returns approximately match the empirical results of Fagereng et al. (2020) for Norwegian

[11]Specifically, we have $y_{i,t} \in \{0.9, 3.6, 7.4, 18.5, 39.0, 130.0\}$, with i.i.d. transition probabilities for the six states equal to $(0.22, 0.56, 0.17, 0.02, 0.015, 0.015)$, respectively.

data.[12]

The next step is to simulate this parameterization of the Standard model, which we do for 2,000 generations with the number of households N set equal to 10,000. We then use the results of these simulations and follow the econometric procedure described by Fernholz (2017) to estimate the relative growth-rate parameters α_k, $k = 1, \ldots, N$.[13] Using our estimates of the rank-based relative growth rate parameters α_k, we can find values for rank-based variance parameters σ_k satisfying (2.3) that, according to the characterization (2.4), yield a stationary distribution for the rank-based model that best approximates the average distribution of the Standard model across the 2,000 generations.[14]

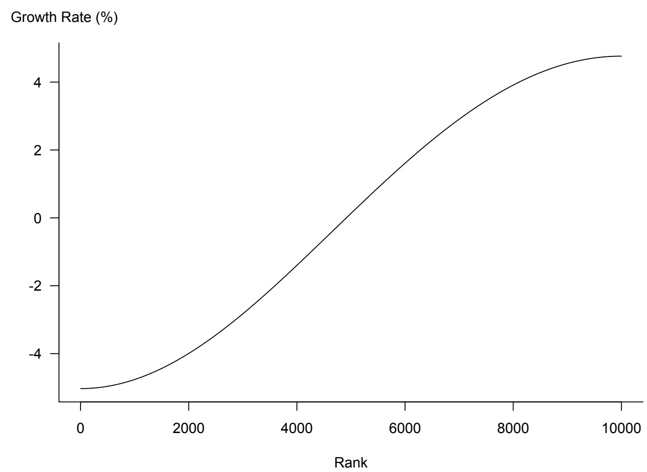

Figure 1: Annualized estimated parameters α_k for the approximated rank-based model.

Figure 1 plots the annualized estimated relative growth-rate parameters α_k for the rank-based approximation of the Standard model. The figure shows that the these parameters satisfy the stability condition (2.2), with the estimated values such that $\alpha_1 < \alpha_2 < \cdots < \alpha_N$.[15]

[12]Specifically, we have $r_{i,t} \in \{0.02, 0.05, 0.09, 0.27\}$, with i.i.d. transition probabilities for the four states equal to $(0.44, 0.45, 0.10, 0.01)$, respectively. With this parameterization, the average and standard deviation of idiosyncratic returns are 4.3% and 3.1%, respectively.

[13]Following Fernholz (2017), we apply a Gaussian kernel filter with a range of 4,000 ranks ten times to smooth the estimated parameters α_k.

[14]Specifically, we minimize the squared distance between the wealth shares reported in Table 1 for the Standard model and those predicted by (2.4) for the rank-based model.

[15]Recall from Section 2.1, however, that this does not imply that returns on wealth in the model are lower for higher-ranked, higher-wealth households.

Figure 2 plots the annualized estimated variance parameters σ_k for the rank-based approximation of the Standard model, and Figure 3 presents a log-log plot of wealth versus rank for both the Standard model and its rank-based approximation. We can see from Figure 3 that the rank-based approximation generates a smoothed version of the wealth distribution from the Standard model.

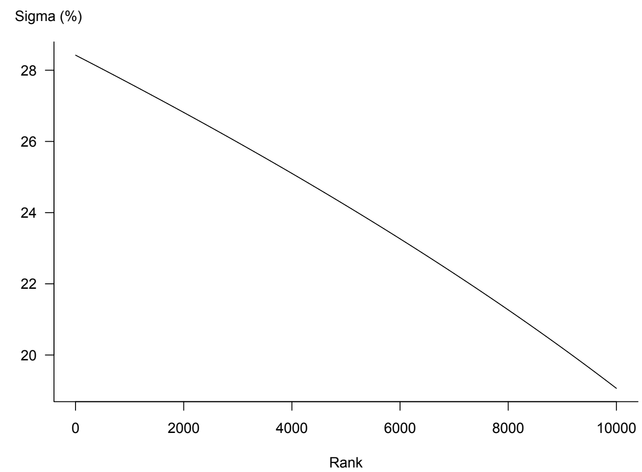

Figure 2: Annualized estimated parameters σ_k for the approximated rank-based model.

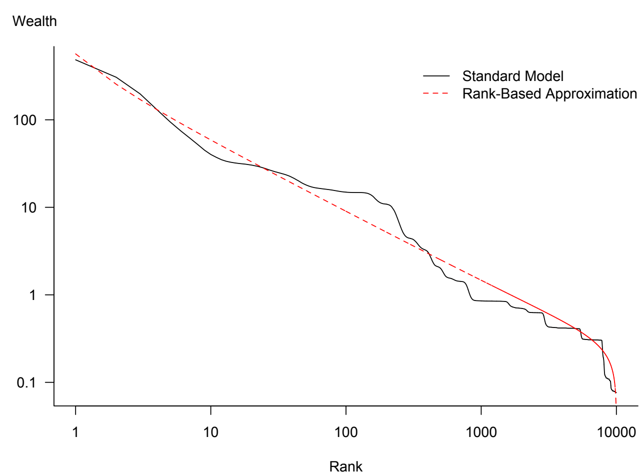

Figure 3: Log-log plot of wealth versus rank for the Standard model (average from 2,000 simulations) and its rank-based approximation.

The Persistently Heterogeneous Rank-Based model. We wish to calibrate the persistently heterogeneous rank-based model (2.9), which extends the rank-based model (2.1) to include persistent heterogeneity, so that it maintains approximately the same realistic stationary wealth distribution as the approximated rank-based model. We assume that 3,000 of the households are high-type households, with $\gamma_h = 0.02$. According to (2.10), this implies that the remaining 7,000 low-type households have $\gamma_\ell \approx -0.0086$. Furthermore, we can use the same estimated parameter values for σ_k from the approximated rank-based model (Figure 2) for the persistently heterogeneous rank-based model.

The calibration of the rank-based relative growth rates $\hat{\alpha}_k$ is more delicate, however. We cannot simply use the estimated values of α_k from the approximated rank-based model (Figure 1) for the persistently heterogeneous rank-based model since the persistently heterogeneous parameters γ_i from (2.9) lead to a more skewed stationary distribution than in the model (2.1). Thus, it is necessary to adjust the estimated rank-based relative growth rate parameters α_k in Figure 1, so that $\hat{\alpha}_k \neq \alpha_k$, to maintain a similar stationary distribution for the two rank-based models.

Consider the rank-based approximation (2.1) of the heterogeneous rank-based model (2.9), where the parameters α_k are defined as in (2.6). In this case, Fernholz et al. (2013) show that the relative growth rate parameters α'_k for the rank-based approximation are given by

$$\alpha'_k = \hat{\alpha}_k + (N-n)\xi_{\ell,k}\gamma_\ell + n\xi_{h,k}\gamma_h, \tag{3.1}$$

for all $k = 1, \dots, N$. According to Proposition 2.1, the stationary distributions of the rank-based approximation of the model (2.9) and the rank-based model (2.1) will be the same if we choose $\hat{\alpha}_k$ such that $\alpha'_k = \alpha_k$, for each rank k. However, solving for the parameters $\hat{\alpha}_k$ that achieve this equality is complicated by the fact that we cannot directly solve for the occupation times $\xi_{\ell,k}$ and $\xi_{h,k}$ in (3.1), but instead must rely on simulations of the persistently heterogeneous rank-based model to generate estimates of these parameters.

We use a simple procedure to generate estimates of the parameters $\hat{\alpha}_k$ from the model (2.9) such that α'_k is approximately equal α_k, for each rank k. First, we use (3.1) to guess values of the parameters $\hat{\alpha}_k$ such that $\alpha'_k - \alpha_k \approx 0$, for all $k = 1, \dots, N$. Next, we simulate the persistently heterogeneous rank-based model with these parameters $\hat{\alpha}_k$ to generate estimates of the rank-based approximation parameters α'_k, and then calculate the standard deviation of $\alpha'_k - \alpha_k$. Once this error term is calculated, we incrementally alter the values of $\hat{\alpha}_k$ by setting each equal to $x\hat{\alpha}_k$, where x is slightly less than or slightly greater than one. We then re-estimate the parameters α'_k and again calculate the sum of squared values $\alpha'_k - \alpha_k$. If the squared error with the parameter values $x\hat{\alpha}_k$ is smaller, then we keep the new parameter values and repeat the procedure by altering the new parameter values in the same way. If not, then we consider a different value of x and repeat the procedure. This procedure is repeated until the sum of squared values $\alpha_k' - \alpha_k$ is larger for the parameter values $x\hat{\alpha}_k$, for both $x = 1.001$ and $x = 0.999$. The annualized estimated parameters $\hat{\alpha}_k$ found using this procedure are shown in Figure 4.

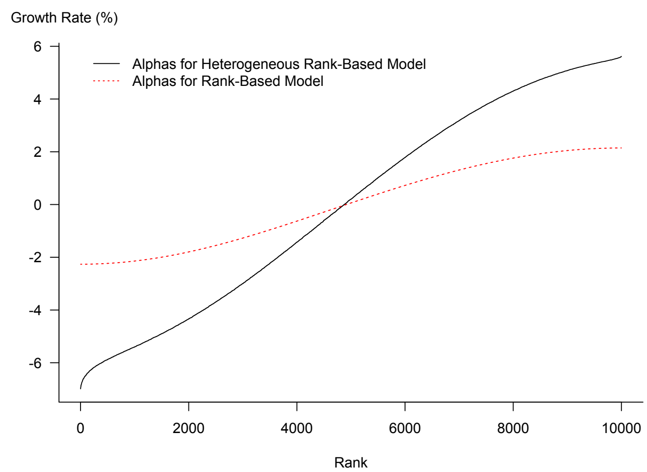

Figure 4: Annualized estimated parameters $\hat{\alpha}_k$ for the persistently heterogeneous rank-based model, and annualized estimated parameters α_k for the approximated rank-based model.

The Auto-Correlated Returns model. We introduce autocorrelated returns into the Standard model (2.5) to capture imperfect social mobility as in Benhabib et al. (2011) and Benhabib et al. (2019) and to investigate how such returns impact long-run mobility. In this version of the model, we assume that wealth returns follow a highly persistent AR-1 process, with

$$\log(1 + r_{i,t+1}) = \theta \log(1 + r_{i,t}) + \epsilon_{i,t}, \tag{3.2}$$

where $\epsilon_{i,t}$ is normally distributed with mean equal to 0.0375 and standard deviation equal to 0.025 and the persistence parameter θ is equal to 0.9. For symmetry, we also assume that labor earnings $\log y_{i,t}$ are i.i.d. and drawn from a normal distribution with mean equal to 0.85 and standard deviation equal to 1.416.[16] In all other respects, the auto-correlated

[16]These values imply that the mean of $y_{i,t}$ is 6.4 and the standard deviation is 16.1, exactly matching the calibration of the Standard model which is based on the SCF.

returns model is identical to the Standard model which was used to calibrate the rank-based model.

3.2 Results

The wealth shares of different subsets of households in the SCF data can be compared with those generated by the different calibrated models in the upper part of Table 1. All models other than the auto-correlated returns model are calibrated to match these wealth shares and do relatively well at this, especially for the top 1% wealth share.

	Data	Standard Model	Approximated Rank-Based Model	Persist. Heter. Rank-Based Model	Auto-Correlated Returns Model
Wealth Distribution					
Top 1%	33.6%	33.1%	32.6%	35.1%	98.5%
Top 1-5%	26.7%	26.4%	17.1%	16.6%	0.6%
Top 5-10%	11.1%	6.3%	9.4%	9.1%	0.3%
Top 10-20%	12.0%	8.0%	11.0%	10.6%	0.3%
Top 20-40%	11.2%	10.5%	12.9%	12.4%	0.2%
Top 40-60%	4.5%	7.7%	8.2%	7.9%	0.1%
Bottom 60-100%	-0.1%	8.0%	8.7%	8.3%	0.1%
Wealth-Rank Correlations					
Parent-Child Rank Coeff.	0.191	0.177	0.187	0.218	0.252
Grandparent-Child Rank Coeff.	0.116	0.015	-0.004	0.089	0.147
Long-Run Persistence Coeff.	0.105	0.001	0.000	0.116	0.022

Table 1: Upper part: Average wealth shares from 1,000 simulations of the different models - data from the Survey of Consumer Finances. Lower part: Average coefficients from regressions of child rank on parent rank and grandparent rank from 1,000 simulations of the different model - data from Danish wealth holdings for three generations in Boserup et al. (2014). Average coefficient from regressions of household rank in generation t on household rank in generation $t - 23$ (585 years) from 1,000 simulations of the different models - data from estimates of very long-run (585 years) dynastic wealth holdings in Florence, Italy, in Barone and Mocetti (2016).

In addition to realistic wealth distributions, these models generate parent-child wealth-rank correlations comparable to those of Boserup et al. (2014), as reported in the lower part of Table 1.

The numerical results on the rank coefficient in the lower part of Table 1 are consistent with the theoretical results of Section 2. Proposition 2.3 implies that household wealth ranks will be uncorrelated over very long time periods in the rank-based approximation of the Standard model, while Theorem 2.5, in contrast, implies that household wealth ranks will be positively correlated over arbitrarily long time periods in a rank-based model that features persistent heterogeneity. The simulation results confirm these results.[17] Furthermore, these results show that the version of the Standard model with highly auto-correlated returns is able to approximately match the empirical results of Boserup et al. (2014) and generate a significant link between child and grandparent wealth ranks, after controlling for parent wealth rank. This auto-correlated returns model, however, fails to match the long-run link between dynastic wealth ranks reported by Barone and Mocetti (2016), and also generates an implausibly skewed wealth distribution as shown in the last column of Table 1. In conclusion, it is only the heterogeneous rank-based model that can match all aspects of the data simultaneously — the wealth distribution, the link between child, parent, and grandparent wealth ranks, and the positive correlation of dynastic wealth ranks over very long time periods.

It is useful, then, to study the properties of the heterogeneous rank-based model more closely. Table 2 shows the composition of the top 1% and top 5% wealth-ranked households in terms of low- and high-type households. This table also shows the composition of the bottom 50% and bottom 25% ranked households. According to the table, high-type households make up the great majority of the top 1% and 5%, but there is still a non-negligible minority of low-type households in these top subsets. The results in the table also suggest that low-type households are more common in top subsets of the wealth distribution than high-type households are in bottom subsets of the wealth distribution. Indeed, the fraction of low-type households in the bottom 25% approximately matches the fraction of high-type households in the top 1%, even though the latter is a much smaller and more exclusive subset of the wealth distribution.

[17] The Standard model is, like its rank-based approximation, unable to generate long-run wealth rank correlations as well.

	Top 1%	Top 5%	Bottom 50%	Bottom 25%
High-Type Households	86.6%	72.7%	17.5%	12.0%
Low-Type Households	13.4%	27.3%	82.5%	88.0%

Table 2: Average composition of the top 1%, top 5%, bottom 50%, and bottom 25% of households from 1,000 simulations of the heterogeneous rank-based model.

Figures 5 and 6 plot the estimated occupation times of different percentiles of the wealth distribution for, respectively, high-type and low-type households. The estimated occupation times presented in the figures are clearly consistent with the result in Proposition 2.4. Because there are 3,000 high-type households and 7,000 low-type households, the maximum average occupation time for a high-type household in any percentile of the wealth distribution is $1/3000 \approx 0.033\%$, while the maximum occupation time for a low-type household in any percentile is $1/7000 \approx 0.014\%$.[18]

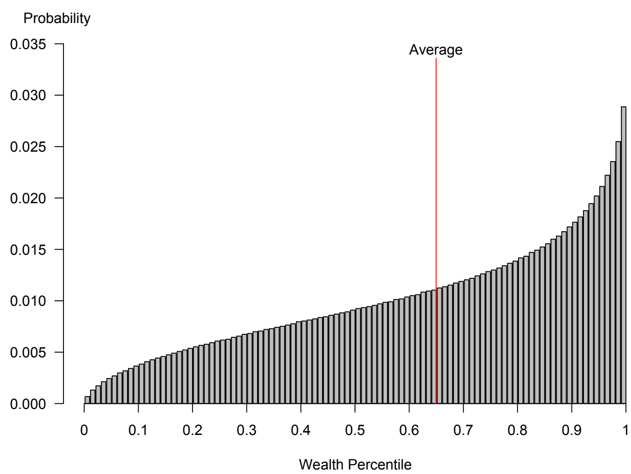

Figure 5: Average high-type household occupation times for different percentiles of the wealth distribution from 1,000 simulations of the heterogeneous rank-based model.

[18]These upper bounds for low- and high-type household occupation times also appear in Proposition 2.4, since the number of high-type households in this simulation n is equal to 3,000.

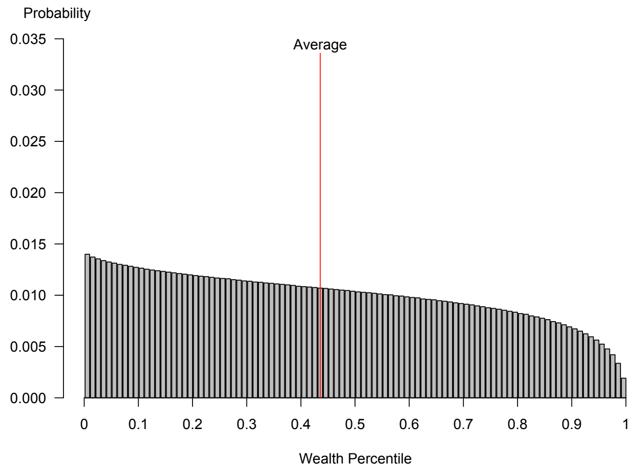

Figure 6: Average low-type household occupation times for different percentiles of the wealth distribution from 1,000 simulations of the heterogeneous rank-based model.

The very long-run persistence of wealth rank that exists in the heterogeneous rank-based model can be seen most clearly in Figure 7. In this figure, we plot the correlation between the wealth ranks of households in generation t and generation $t+x$, with values of x ranging from 1 to 25, for both the heterogeneous rank-based and auto-correlated returns models. Although the auto-correlated returns model is able to generate substantial persistence in rank across one or two generations, the rank correlation in this model quickly declines towards zero as the generational gap between households increases. In contrast, the heterogeneous rank-based model generates a more realistic but smaller persistence in wealth rank across one or two generations, and this persistence never falls below 0.1 even as the generational gap grows large. Of course, this very long-run persistence in wealth rank is exactly what is predicted by Theorem 2.5.

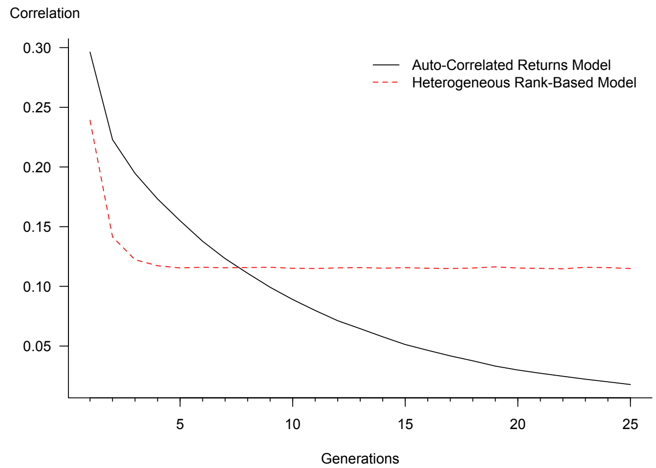

Figure 7: Rank correlations across multiple generations from 1,000 simulations of the heterogeneous rank-based and auto-correlated returns models.

4 Conclusion

We consider a simple heterogeneous agents model based on Benhabib et al. (2019) and show that such standard models fail to match recent empirical results regarding long-run wealth mobility. In particular, this type of model does not generate a positive correlation between grandparent-child wealth rank, after controlling for parent-child wealth rank, and does not generate a positive correlation between dynastic wealth ranks across very long time periods. We extend the standard model to include persistent heterogeneity in returns to wealth, and show that such an extended model is able to simultaneously match the wealth distribution, short-run wealth mobility, and long-run wealth mobility. While we do not take a stand on the precise interpretation of this form of persistent heterogeneity, we note that it can be seen as a formalization of the latent factor representation of abilities, preferences, dynastic network connections, occupational persistence or other relevant characteristics.

References

Banner, A., R. Fernholz, and I. Karatzas (2005). Atlas models of equity markets. *Annals of Applied Probability 15*(4), 2296–2330.

Barone, G. and S. Mocetti (2016, April). Intergeneraitonal mobility in the very long run: Florence 1427-2011. *Bank of Italy Working Paper No. 1060.*

Benhabib, J., A. Bisin, and M. Luo (2017, May). Eearnings inequality and other determinants of wealth inequality. *American Economic Review: Papers and Proceedings 107*(5), 593–597.

Benhabib, J., A. Bisin, and M. Luo (2019, May). Wealth distribution and social mobility in the us: A quantitative approach. *American Economic Review 109*(5), 1623–1647.

Benhabib, J., A. Bisin, and S. Zhu (2011, January). The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica 79*(1), 123–157.

Boserup, S., W. Kopczuk, and C. Kreiner (2014, October). Stability and persistence of intergenerational wealth formation: Evidence from danish wealth records of three generations. *mimeo, University of Copenhagen.*

Braun, S. T. and J. Stuhler (2018, March). The transmission of inequality across multiple generations: Testing recent theories with evidence from germany. *Economic Journal 128*(609), 576–611.

Cagetti, M. and M. De Nardi (2006, October). Entrepreneurship, frictions, and wealth. *Journal of Political Economy 114*(5), 835–870.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003, August). Accounting for U.S. earnings and wealth inequality. *Journal of Political Economy 111*(4), 818–857.

Chan, T. W. and V. Boliver (2013, August). The grandparents effect in social mobility: Evidence from british birth cohort studies. *American Sociological Review 78*(4), 662–678.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton, NJ: Princeton University Press.

Clark, G. and N. Cummins (2015, February). Intergenerational wealth mobility in england, 1858-2012: Surnames and social mobility. *Economic Journal 125*(582), 61–85.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020, January). Heterogeneity and persistence in returns to wealth. *Econometrica 88*(1), 115–170.

Fernholz, R., T. Ichiba, and I. Karatzas (2013, August). A second-order stock market model. *Annals of Finance 9*(3), 439–454.

Fernholz, R. T. (2017). Nonparametric methods and local-time-based estimation for dynamic power law distributions. *Journal of Applied Econometrics 32*(7), 1244–1260.

Hertel, F. R. and O. Groh-Samberg (2014, March). Class mobility across three generations in the u.s. and germany. *Reserach in Social Stratification and Mobility 35*, 35–52.

Hubmer, J., P. Krusell, and A. A. Smith Jr. (2016, December). The historical evolution of the wealth distribution: A quantitative-theoretic investigation. *NBER Working Paper 23011*.

Ichiba, T., V. Papathanakos, A. Banner, I. Karatzas, and R. Fernholz (2011). Hybrid atlas models. *Annals of Applied Probability 21*(2), 609–644.

Kindermann, F. and D. Krueger (2015). Sigh marginal tax rates on the top 1%? lessons from a life cycle model with idiosyncratic income risk. *mimeo, University of Pennsylvania.*

Krusell, P. and A. A. Smith (2015, August). Is piketty's "second law of capitalism" fundamental? *Journal of Political Economy 123*(4), 725–748.

Lindahl, M., M. Palme, S. S. Massih, and A. Sjögren (2015, Winter). Long-term intergenerational persistence of human capital: An empirical analysis of four generations. *Journal of Human Resources 50*(1), 1–33.

Long, J. and J. Ferrie (2013, June). Intergenerational occupational mobility in great britain and the united states since 1850. *American Economic Review 103*(4), 1109–1137.

Mare, R. (2011, February). A multigenerational view of inequality. *Demography 48*, 1–23.

Modin, B., R. Erikson, and D. Vagero (2013, August). Intergenerational continuity in school performance: Do grandparents matter? *European Sociological Review 29*(4), 858–870.

Quadrini, V. (2000, January). Entrepreneurship, saving, and social mobility. *Review of Economic Dynamics 3*(1), 1–40.

Saporta, B. and J.-F. Yao (2005). Tail of a linear diffusion with markov switching. *Annals of Applied Probability 15*(1B), 992–1018.

Stuhler, J. (2012, December). Mobility across multiple generations: The iterated regression fallacy. *IZA Discussion Paper No. 7072.*

Warren, J. R. and R. M. Hauser (1997, August). Social stratification across three generations: New evidence from the wisconsin longitudinal study. *American Sociological Review 62*(4), 561–572.

Proofs

This section presents the proofs of Propositions 2.3, 2.2, and 2.4 and Theorem 2.5.

Proof of Proposition 2.3. The first part of the proposition, (2.11), follows directly from Proposition 2.3 of Banner et al. (2005). For the second part, we have, for any household $i = 1, \dots, N$,

$$\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i)] = \lim_{\tau\to\infty} \sum_{k=1}^{N} kP\left(\rho_{t+\tau}(i) = k\right) = \lim_{\tau\to\infty} \sum_{k=1}^{N} \frac{k}{N} = \frac{N(N+1)}{2N} = \frac{N+1}{2},$$

where the second equality follows from (2.11). □

Proof of Proposition 2.2. First, note that $\mathbb{E}[\log \lambda(r_{\pi_t(k),t})] = \mathbb{E}[\log(\lambda(r_{\pi_t(\ell),t}))]$ and also $\mathbb{E}[\beta(r_{\pi_t(k),t}, y_{\pi_t(k),t})] = \mathbb{E}[\beta(r_{\pi_t(\ell),t}, y_{\pi_t(\ell),t})]$ for all wealth ranks k, ℓ, because the expected value of both returns to wealth and labor income do not vary across different ranks. Then, because the ranked wealth processes satisfy $w_{(1)} \geq \cdots \geq w_{(N)}$ by definition, it follows from (2.8) that $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N$ and hence (2.2) is satisfied and the rank-based model is stationary. □

Proof of Proposition 2.4. In order to prove the proposition, it is necessary to characterize the occupation times for low- and high-type households in a heterogeneous rank-based model (2.9) that satisfies (2.3) and (2.10). We denote the symmetric group of permutations of $\{1, \dots, N\}$ by Ψ_N, where $p(k) \in \{1, \dots, N\}$ denotes the k-th element of the permutation $p \in \Psi_N$. Each permutation p describes a potential wealth ranking of the households $i = 1, \dots, N$, with $p(k)$ denoting the index of the k-th ranked household. According to Corollary 4 from Ichiba et al. (2011), for all $i, k = 1, \dots N$, the occupation time $\xi_{i,k}$ is given by

$$\xi_{i,k} = \sum_{\{p\in\Psi_N \,|\, p(k)=i\}} \prod_{j=1}^{N-1} \phi_{p,j}\, \Omega, \quad \text{a.s.,} \tag{A.1}$$

where

$$\phi_{p,j} = \frac{\sigma_j^2 + \sigma_{j+1}^2}{-4(\sum_{m=1}^{j} \hat{\alpha}_m + \gamma_{p(m)})}, \tag{A.2}$$

for any permutation $p \in \Psi_N$ and all $j = 1, \ldots, N-1$, and $\Omega = \left(\sum_{q \in \Psi_N} \prod_{j=1}^{N-1} \phi_{q,j} \right)^{-1}$.

According to (A.1), for any $k = 1, \ldots, N-1$,

$$\xi_{\ell,k} = \sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=\ell\}} \prod_{j=1}^{N-1} \phi_{p,j}\, \Omega + \sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=h\}} \prod_{j=1}^{N-1} \phi_{p,j}\, \Omega, \tag{A.3}$$

and also

$$\xi_{\ell,k+1} = \sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=\ell\}} \prod_{j=1}^{N-1} \phi_{p,j}\, \Omega + \sum_{\{p \in \Psi_N \,|\, p(k)=h,\, p(k+1)=\ell\}} \prod_{j=1}^{N-1} \phi_{p,j}\, \Omega. \tag{A.4}$$

If $k < N-1$ as well, then

$$\sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=h\}} \prod_{j=1}^{N-1} \phi_{p,j} = \sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=h\}} \phi_{p,k}\phi_{p,k+1} \prod_{j=1}^{k-1} \phi_{p,j} \prod_{j=k+1}^{N-1} \phi_{p,j}, \tag{A.5}$$

and

$$\sum_{\{p \in \Psi_N \,|\, p(k)=h,\, p(k+1)=\ell\}} \prod_{j=1}^{N-1} \phi_{p,j} = \sum_{\{p \in \Psi_N \,|\, p(k)=h,\, p(k+1)=\ell\}} \phi_{p,k}\phi_{p,k+1} \prod_{j=1}^{k-1} \phi_{p,j} \prod_{j=k+1}^{N-1} \phi_{p,j}. \tag{A.6}$$

For every $p \in \Psi_N$ with $p(k) = \ell$ and $p(k+1) = h$, there exists a $p' \in \Psi_N$ with $p'(k) = h$, $p'(k+1) = \ell$, and $p'(j) = p(j)$ for all $j \neq k, k+1$, so it follows that

$$\sum_{\{p \in \Psi_N \,|\, p(k)=h,\, p(k+1)=\ell\}} \prod_{j=1}^{k-1} \phi_{p,j} \prod_{j=k+1}^{N-1} \phi_{p,j} = \sum_{\{p \in \Psi_N \,|\, p(k)=\ell,\, p(k+1)=h\}} \prod_{j=1}^{k-1} \phi_{p,j} \prod_{j=k+1}^{N-1} \phi_{p,j}. \tag{A.7}$$

Let $p \in \Psi_N$ with $p(k) = \ell$ and $p(k+1) = h$, and $p' \in \Psi_N$ with $p'(k) = h$, $p'(k+1) = \ell$, and $p'(j) = p(j)$ for all $j \neq k, k+1$. According to (A.2), we have

$$\phi_{p,k} = \frac{\sigma_k^2 + \sigma_{k+1}^2}{-4(\sum_{m=1}^{k} \hat{\alpha}_m + \gamma_{p(m)})} < \frac{\sigma_k^2 + \sigma_{k+1}^2}{-4(\sum_{m=1}^{k} \hat{\alpha}_m + \gamma_{p'(m)})} = \phi_{p',k}, \tag{A.8}$$

since $\gamma_h > \gamma_\ell$. Together with the fact that $\phi_{p,k+1} = \phi_{p',k+1}$, (A.3), (A.4), (A.5), (A.6), (A.7), and (A.8) thus imply that $\xi_{\ell,k} < \xi_{\ell,k+1}$.

If $k = N-1$, then we have

$$\sum_{\{p\in\Psi_N \,|\, p(N-1)=\ell,\, p(N)=h\}} \prod_{j=1}^{N-1} \phi_{p,j} = \sum_{\{p\in\Psi_N \,|\, p(N-1)=\ell,\, p(N)=h\}} \phi_{p,N-1}\phi_{p,N} \prod_{j=1}^{N-2} \phi_{p,j}, \tag{A.9}$$

and

$$\sum_{\{p\in\Psi_N \,|\, p(N-1)=h,\, p(N)=\ell\}} \prod_{j=1}^{N-1} \phi_{p,j} = \sum_{\{p\in\Psi_N \,|\, p(N-1)=h,\, p(N)=\ell\}} \phi_{p,N-1}\phi_{p,N} \prod_{j=1}^{N-2} \phi_{p,j}. \tag{A.10}$$

As with the previous case, for every $p \in \Psi_N$ with $p(N-1) = \ell$ and $p(N) = h$, there exists a $p' \in \Psi_N$ with $p'(N-1) = h$, $p'(N) = \ell$, and $p'(j) = p(j)$ for all $j < N-1$, so it follows that

$$\sum_{\{p\in\Psi_N \,|\, p(N-1)=\ell,\, p(N)=h\}} \prod_{j=1}^{N-2} \phi_{p,j} = \sum_{\{p\in\Psi_N \,|\, p(N-1)=h,\, p(N)=\ell\}} \prod_{j=1}^{N-2} \phi_{p,j}, \tag{A.11}$$

Similarly, we if we let $p \in \Psi_N$ with $p(N-1) = \ell$ and $p(N) = h$, and $p' \in \Psi_N$ with $p'(N-1) = h$, $p'(N) = \ell$, and $p'(j) = p(j)$ for all $j < N-1$, then (A.2) implies that

$$\phi_{p,N-1} = \frac{\sigma_{N-1}^2 + \sigma_N^2}{-4(\sum_{m=1}^{N-1} \hat{\alpha}_m + \gamma_{p(m)})} < \frac{\sigma_{N-1}^2 + \sigma_N^2}{-4(\sum_{m=1}^{N-1} \hat{\alpha}_m + \gamma_{p'(m)})} = \phi_{p',N-1}. \tag{A.12}$$

Thus, combining (A.3), (A.4), (A.9), (A.10), (A.11), (A.12) implies that $\xi_{\ell,N-1} < \xi_{\ell,N}$. All low-type households have equal occupation times at each rank, so that $\xi_{i,k} = \xi_{j,k} = \xi_{\ell,k}$ for all households i, j that are low types with $\gamma_i = \gamma_j = \gamma_\ell$, and thus because there are $N-n$ total low-type households, it must be that $\xi_{\ell,k} < 1/(N-n)$ for all $k = 1, \dots, N$. It follows, then, that $0 < \xi_{\ell,1} < \xi_{\ell,2} < \cdots < \xi_{\ell,N} < 1/(N-n)$, a.s. A similar argument establishes that $1/n > \xi_{h,1} > \xi_{h,2} > \cdots > \xi_{h,N} > 0$, a.s. □

Proof of Theorem 2.5. Suppose that household i is a low-type household with $\gamma_i = \gamma_\ell$, and household j is a high-type household with $\gamma_j = \gamma_h$. According to Proposition 2.4, we have $\xi_{\ell,1} < \xi_{\ell,2} < \cdots < \xi_{\ell,N}$, a.s. and $\xi_{h,1} > \xi_{h,2} > \cdots > \xi_{h,N}$, a.s. It follows, then, that

$$\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i) \mid \gamma_i = \gamma_\ell] = \xi_{\ell,1} + 2\xi_{\ell,2} + \cdots + N\xi_{\ell,N} > \frac{N+1}{2}, \tag{A.13}$$

where the last inequality follows because $\xi_{\ell,N} > 1/N$ and the expected value in (A.13) is increasing in the value of $\xi_{\ell,N}$ despite the constraint that $\xi_{\ell,1} + \cdots + \xi_{\ell,N} = 1$. From (A.13), we have

$$\xi_{\ell,1} + 2\xi_{\ell,2} + \cdots + N\xi_{\ell,N} > \frac{N+1}{2},$$
$$N(\xi_{\ell,1} + 2\xi_{\ell,2} + \cdots + N\xi_{\ell,N}) > 1 + 2 + \cdots + N,$$

which implies that

$$1 - N\xi_{\ell,1} + 2 - 2N\xi_{\ell,2} + \cdots + N - N^2\xi_{\ell,N}) < 0,$$
$$\frac{1}{n} - \frac{N}{n}\xi_{\ell,1} + \frac{2}{n} - \frac{2N}{n}\xi_{\ell,2} + \cdots + \frac{N}{n} - \frac{N^2}{n}\xi_{\ell,N} < 0. \tag{A.14}$$

If we write (2.13) as

$$\xi_{h,k} = \frac{1}{n}(1 - (N-n)\xi_{\ell,k}) = \frac{1}{n} - \frac{N}{n}\xi_{\ell,k} + \xi_{\ell,k},$$

for all $k = 1, \ldots, N$, then (A.14) implies that

$$\xi_{h,1} - \xi_{\ell,1} + 2(\xi_{h,2} - \xi_{\ell,2}) + \cdots + N(\xi_{h,N} - \xi_{\ell,N}) < 0,$$

and hence also

$$\xi_{h,1} + 2\xi_{h,2} + \cdots + N\xi_{h,N} < \xi_{\ell,1} + 2\xi_{\ell,2} + \cdots + N\xi_{\ell,N}. \tag{A.15}$$

Because household j is a high type, $\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(j) \mid \gamma_j = \gamma_h] = \xi_{h,1} + 2\xi_{h,2} + \cdots + N\xi_{h,N}$, and hence it follows from (A.15) that

$$\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i) \mid \gamma_i = \gamma_\ell] > \lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(j) \mid \gamma_j = \gamma_h]. \tag{A.16}$$

Having shown in (A.16) that low-type households will on average in the long run occupy lower ranks than high-type households, the last step is to show that the higher the rank of a household at time t, the more likely that household is to be a high type. If some household i occupies rank k at time t, so $\rho_t(i) = k$, for any rank $k = 1, \ldots, N$, then the probability that household i is a high type is

$$P(\gamma_i = \gamma_h \mid \rho_t(i) = k) = \frac{n\xi_{h,k}}{(N-n)\xi_{\ell,k} + n\xi_{h,k}} = n\xi_{h,k},$$

where the second equality follows from (2.13). By Proposition 2.4, it follows that

$$1 > n\xi_{h,1} > n\xi_{h,2} > \cdots > \xi_{h,N} > 0,$$
$$1 > P(\gamma_i = \gamma_h \mid \rho_t(i) = 1) > \cdots > P(\gamma_i = \gamma_h \mid \rho_t(i) = N) > 0, \tag{A.17}$$

and also that

$$0 < P(\gamma_i = \gamma_\ell \mid \rho_t(i) = 1) < \cdots < P(\gamma_i = \gamma_\ell \mid \rho_t(i) = N) < 1. \tag{A.18}$$

Taken together, (A.16), (A.17), (A.18) imply that

$$\begin{aligned}\lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i)] = \lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i) \mid \gamma_i = \gamma_h] P(\gamma_i = \gamma_h \mid \rho_t(i) = k) \\ + \lim_{\tau\to\infty} \mathbb{E}[\rho_{t+\tau}(i) \mid \gamma_i = \gamma_\ell] P(\gamma_i = \gamma_\ell \mid \rho_t(i) = k),\end{aligned}$$

is increasing in k. $\square$