Heterogeneous Dynasties and Long-Run Mobility

Jess Benhabib
NYU

Alberto Bisin
NYU

Ricardo T. Fernholz
Claremont McKenna College

May 14, 2021

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. An auto-correlated returns specification of this model also fails to capture both short and long-run mobility. However, an extension of the heterogeneous agents model which includes permanent 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, 2021; Krusell and Smith, 2015; Quadrini, 2000). This literature shows that heterogeneous agents 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 a significant grandparent-child wealth-rank correlation even after controlling for the effects of parent wealth on child wealth (Boserup et al., 2014). 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 do not fare well, in that they cannot generate a large enough coefficient for grandparent-child wealth-rank nor a large enough correlation for dynastic wealth-ranks over very long time periods. 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.

In this paper we extend a simple heterogeneous agents model to introduce permanent heterogeneity in the rate of return to wealth across generations. In other words, we allow households in some dynasties to have their wealth grow faster on average than households in other dynasties. This can be seen as a formalization of a latent factor representation of persistent cultural and institutional factors suggested in the literature in political economy and sociology, along the lines of Bisin and Verdier (2001); Acemoglu and Robinson (2008); Bourdieu (1984, 1998). In Section 2.1 we show theoretically that such a model has the potential to generate a strong inter-generational rank-correlation of wealth, also in the long-run. In Section 3 we confirm that a calibrated permanently heterogeneous rank-based model can produce both a fat-tailed distribution of wealth which matches the data well as well as strong inter-generational correlations akin to those documented in the data.

It is not difficult to envision other extensions of simple heterogeneous agents models

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

which could produce significant grandparent-child and even long-run wealth-rank correlations, e.g., postulating inter-generational auto-correlation in earnings and in the rate of return to wealth. However, the evidence is not favorable to the existence of independent direct causal effects across generations, beyond parent-child effects. We discuss this evidence in the next section. Extending the model along these lines requires postulating very strong inter-generational auto-correlation 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. Indeed, in Section 3.2.1 we show by means of simulation analysis that the long-run auto-correlations of wealth-ranks in the data can be generated in principle by models specifying auto-correlated returns to wealth, but at the cost of excessively high parent-child and grandparent-child wealth-rank correlations with respect to the data.

Finally, we compare the implications of the permanently heterogeneous rank-based model and the auto-correlated returns to wealth model with respect to parent-child return-rank correlation. We show in Section 3.2.1 that the model with permanent heterogeneity produces in our calibration a small parent-child correlation of returns, close to the one documented by Fagereng et al. (2020). The auto-correlated returns to wealth model on the other hand also produces an excessively high parent-child return-rank correlation. We argue that this is suggestive of persistent institutional factors as mechanisms for sustaining long-run wealth persistence, rather than of direct inter-generational mechanisms like cultural transmission.

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. They find that grandparent effects do not necessarily go through parents and 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 also find no effects of direct contacts between grandparents and grandchildren and conclude 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 noteworthy. 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.[2] These data are necessarily plagued by noise, to the point of being hardly amenable to statistical inference to identify any latent factors responsible for the persistence of wealth, education, or occupational status in the long-run (Mare, 2011; Braun and Stuhler, 2018). Nonetheless this documented persistence is consistent with the evidence on grandparent-child correlations, as argued in Stuhler (2012) and Braun and Stuhler (2018). It is also consistent with recent empirical and theoretical studies identifying long-run persistence in cultural traits; see Voth (2021); Bisin and Moro (2021); Bisin and Verdier (2001) for surveys; and with the evidence of long-run persistence of the effects of institutions, especially of those institutional factors which perpetuate political and economic elites and hence wealth inequality; see Acemoglu and Robinson (2008); Bisin and Verdier (2017, 2021) and the work by Pierre Bourdieu, e.g., Bourdieu and Passeron (1970); Bourdieu (1984, 1998).

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 permanent heterogeneity in the rate of return to wealth across generations into the standard rank-based model. Finally, we derive theoretical results

[2] 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.

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 $t \in \mathbb{R}$, 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. The parameters α_k and σ_k measure the average and variance of the growth rate of wealth at each rank k. We normalize, without loss of generality, the average growth rate of the economy to zero; that is, $\alpha_1 + \cdots + \alpha_N = 0$. The parameters α_k capture then the average *relative growth rates* of wealth with respect to the growth rate of the economy. According to Proposition 2.3 of Banner et al. (2005), the rank-based model (2.1) admits a stationary distribution if

$$\alpha_1 + \cdots + \alpha_k < 0, \tag{2.2}$$

for all $k = 1, \ldots, N-1$. Condition (2.2) on the parameters α_k 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 rates of return to wealth which are constant or even increasing in wealth in a standard heterogeneous agents model of wealth dynamics.

We can now characterize the stationary distribution of the rank-based model.

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 then that the expected value of the ratio of wealth in rank k to wealth in rank $k+1$ at the stationary distribution i) is positive for all ranks k;[3] and ii) is increasing in the volatility parameters σ_k, σ_{k+1}. As an illustration, if the relative growth parameters α_k were increasing in k and σ_k constant, $\mathbb{E}\left[\log w_{(k)}(t) - \log w_{(k+1)}(t)\right]$ would be decreasing in rank until α_k turned positive. We finally note that, by the result in Theorem 2 of Ichiba et al. (2011), $w_{(k)}/w_{(k+1)}$ follows a Pareto distribution, with the Pareto parameter for each k depending on the parameters α_k and σ_k according to (2.4).

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.[4]

Consider an economy populated by households who live for one generation, from t to $t+1$, in discrete time. Any household born at time $t \in \mathbb{N}$ has a single child entering the economy at time $t+1$, that is, at its parent's death. Generations of households are linked to form dynasties. A single generation is composed of T subperiods and each household solves a dynamic consumption-savings problem over subperiods, maximizing a present discounted CRRA utility function with a joy-of-giving bequest final term (leaving its wealth at death to its child). The household faces idiosyncratic yearly rates of return on wealth $r_{i,t}$ and yearly base earnings $y_{i,t}$ at birth; that is, $r_{i,t}$ and $y_{i,t}$ are stochastic across generations but deterministic inside each generation (more precisely the rate of return remains constant while

[3]Recall that, by the stationarity condition (2.2), the sums in the denominator of the right-hand-side of (2.4) are all negative for $k = 1, \ldots, N-1$ where the expression is defined.

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

earnings grow at a constant growth rate g). All households are ex-ante identical, except for the rate of return to wealth, earnings, and initial wealth (as bequests). In equilibrium, the intergenerational wealth dynamics for each household $i = 1, \ldots, N$ follow

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

where $y_{i,t}$ and $r_{i,t}$ denote, respectively, the yearly base labor income (growing at constant rate g in generation t) and the yearly return on wealth for household i (constant in generation t), and $w_i(t)$ denotes the wealth holdings of household i in generation t. Equation (2.5) represents wealth accumulation in reduced form, after optimal consumption has been subtracted from the right-hand side. The functions λ and β are obtained as closed-form solutions of the dynamic optimal consumption-saving problem of the household. They are the same for all households $i = 1, \ldots, N$ and represent, respectively, the inter-generational return on wealth and the present discounted value of labor income, after optimal household consumption, an affine linear function of wealth, has been netted out each period.[5] 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

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

[5]More precisely, in Benhabib et al. (2011), the functions β and λ depend on i) the generation-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. They are expressed in closed-form as:

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

with

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

for each rank $k = 1, \dots, N$.[6] At each rank k of the distribution, the parameters α_k and σ_k measure the average and variance of the growth rate of wealth relative to the aggregate for a generation in the Standard model (2.5). Because the rank-based approximation of (2.5) uses the parameters α_k and σ_k defined by (2.6), it follows that these parameters represent intergenerational relative growth rates and variances in the rank-based model. Like the Standard model, then, each new generation in the continuous-time rank-based approximation (2.1) is born at time $t \in \mathbb{N}$, and these generations are stacked.

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 the Standard model, the rank-based relative growth rate parameters α_k satisfy, for each rank $k = 1, \dots, N$,

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

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 can express the rank-based relative growth rate parameters α_k in terms of β and λ, the parametric functions characterizing the solution of the consumption-savings problem underlying the Standard model (2.5):

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

Equation (2.8) provides a simple decomposition of the rank-based relative growth rates α_k from equation (2.1) in terms of i) the inter-generational return on wealth, adjusted for equilibrium household behavior, λ, at rank k in the wealth distribution; ii) the present discounted value of labor income, adjusted for equilibrium household behavior, β, divided by a measure of generational capital income, λw, at rank k in the distribution.

The decomposition (2.8) is a fundamental interpretation tool in our analysis in that it allows us to map the stability condition for rank-based models in (2.2) into a condition in

[6] 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).

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, as e.g., in models with decreasing returns like Cagetti and De Nardi (2006), is not required to satisfy the stability condition (2.2). 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).

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.*

Benhabib et al. (2019) present a model of the form (2.5) with higher returns to wealth at higher wealth ranks and show that this model admits a stationary distribution. Therefore, Proposition 2.2 implies that the rank-based approximation of this model is also stationary and satisfies the stability condition (2.2), despite the positive relationship between wealth and returns.

2.2 Permanently Heterogeneous Rank-Based Model

We introduce a form of permanent 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

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

for each household $i = 1, \ldots, N$, with $\gamma_i \in \{\gamma_\ell, \gamma_h\}$ and $\gamma_h > \gamma_\ell$. The parameter γ_i acts as a permanent additive factor to the mean of the growth rate of wealth: $\gamma_i = \gamma_\ell$ (resp. $\gamma_i = \gamma_h$), household wealth grows more slowly (resp. quickly) on average over time. We assume that n of the households are characterized by $\gamma_i = \gamma_h$, and $N - n$ of the households by $\gamma_i = \gamma_\ell$. We keep normalizing the average growth rate of wealth to zero, which in this economy requires $\sum_{k=1}^{N} \hat{\alpha}_k + \sum_{i=1}^{N} \gamma_i = \sum_{k=1}^{N} \hat{\alpha}_k + (N - n)\gamma_\ell + n\gamma_h = 0$. The growth rate of wealth of each household i is nonetheless stochastic, due to the Brownian motion term $\sigma_{\rho_t(i)} \, dB_i(t)$, whose volatility depends on the wealth-rank of the household $k = \rho_t(i)$.

To admit a stationary distribution, the permanently heterogeneous model (2.9) must satisfy a condition that generalizes condition (2.2) for the 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) ensures that, accounting for the permanent heterogeneity in the average growth rates of households, no top subset of households grows faster than the aggregate. This is sufficient to guarantee that the high-growth households (with $\gamma_i = \gamma_h$) in the top ranks do not break away from the rest of the population.

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 permanent heterogeneity. We show that permanent 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 permanent 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 has $\gamma_i = \gamma_\ell$, then, by symmetry, the fraction of time household i spends in each rank k is equal to the fraction of time any other low-growth household spends in each rank k. Thus, we can define the low-growth household occupation times $\xi_{\ell,k}$ such that $\xi_{\ell,k} = \xi_{i,k}$, for all ranks $k = 1, \ldots, N$. Similarly, if we suppose that household j is a high-growth 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 $k = 1, \ldots, N$. Because the sum of occupation times across all ranks or individual households must equal one, it follows that the low- and high-growth 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 $k = 1, \ldots, N$.

Proposition 2.4. *Consider a permanently heterogeneous rank-based model* (2.9) *that satisfies* (2.3) *and* (2.10). *Then, the low- and high-growth 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-growth 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-growth households spend more time at the lowest ranks of the wealth distribution across generations than high-growth 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 permanently 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 wealth-rank correlation will be positive in the heterogeneous rank-based model. This is because higher-ranked households occupy higher ranks in expectation, due to the underlying persistence heterogeneity (the expectations in (2.16) are unconditional with respect to whether households i and j are high- or low-growth).

The intuition for the result in Theorem 2.5 is worth presenting in some detail as it underlies some of the simulation results in the next section. Because all high-growth households are ex-ante identical, the expected asymptotic rank of these households is the median of the top n ranks of the wealth distribution; that is, high-growth households occupy higher ranks in expectation across generations than low-growth households. Similarly, the expected asymptotic rank of low-growth households is the median of the bottom $N - n$ ranks. Without knowing whether a household i is high- or low-growth, its expected asymptotic rank 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 high-growth and that it is low-growth. Because higher-ranked households are more likely to be high-growth 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.

To better understand the simulation results in the next section it is important to emphasize, however, that while the long-run asymptotic household wealth-rank correlation depends on the permanent heterogeneity, the parent-child correlation of the growth rate of wealth is affected negatively in a crucial manner by the volatility of the Brownian motion term in the wealth dynamics equation (2.9).

3 Simulations

In this section we present a simulation analysis of inter-generational 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); and ii) the permanently heterogeneous rank-based model (2.9), the extension of the rank-based model that includes permanent heterogeneity.

3.1 Calibration

In this section we discuss the details of the calibrations we adopt. We start from a parameterization of the Standard model mostly based on the one in Benhabib et al. (2019). This parametrization is the outcome of a Simulated Method of Moments estimation procedure to match the wealth distribution and intergenerational social mobility data for the U.S.[7] According to the estimates, 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 yearly base labor income $y_{i,t}$, we use a six-state Markov chain calibrated using inter-generational persistence in labor income data from Chetty et al. (2014) together with the U.S. Survey of Consumer Finances (SCF).[8] Finally, for the idiosyncratic yearly 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 data.[9]

The Approximated Rank-Based model. We then construct the rank-based approximation of the parametrization of the Standard model we just described. This approximation is obtained using (2.6) to define the rank-based parameters α_k and σ_k from (2.1). We first simulate the parameterization of the Standard model, which we do for 2,000 generations with the number of households N set equal to 10,000. Importantly, this parametrization induces by construction a stationary wealth distribution which matches the data for the U.S. well. We then use the results of these simulations and follow the econometric procedure described

[7]Wealth shares data is from the Survey of Consumer Finances (SCF), while intergenerational mobility data is from Charles and Hurst (2003).

[8]Earnings persistence by itself, without stochastic returns across generations, cannot induce the wealth inequality observed in the data. This is because the distribution of earnings has a much thinner right tail than the distribution of wealth; see Benhabib et al. (2017); Benhabib and Bisin (2018).

[9]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. The estimates of the process for $r_{i,t}$ in Benhabib et al. (2019) are very close to these.

by Fernholz (2017) to estimate the relative growth-rate parameters α_k, $k = 1, \ldots, N$.[10] 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.[11]

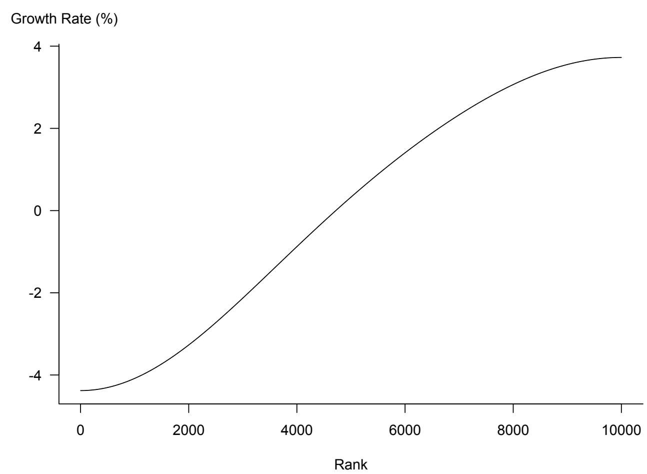

Figure 1: Annualized estimated parameters α_k for the approximated rank-based model.

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 $\alpha_1 < \alpha_2 < \cdots < \alpha_N$.[12] Table 1 reports the wealth distribution shares and wealth-rank correlations in the data and those implied by the two models at their stationary distribution. The rank-based model closely approximates the Standard model and both fit the data relatively well.

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

[10] Following Fernholz (2017), we apply a Gaussian kernel filter with a range of 3,000 ranks ten times to smooth the estimated parameters α_k.

[11] Specifically, we minimize the squared distance between the wealth shares reported in Table 2 for the Standard model and those predicted by (2.4) for the rank-based model.

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

from the Standard model.

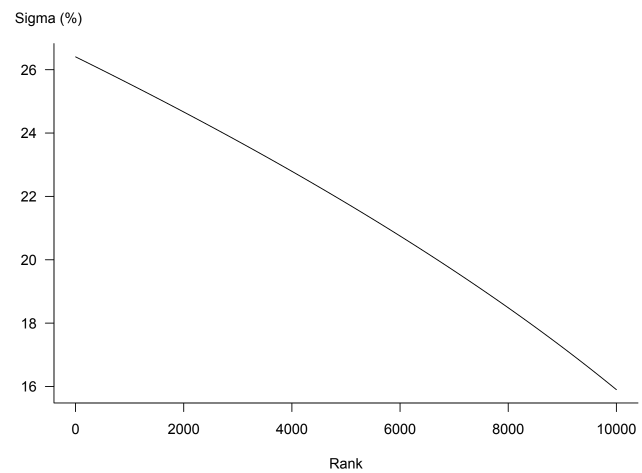

Figure 2: Annualized estimated parameters σ_k for the approximated rank-based model.

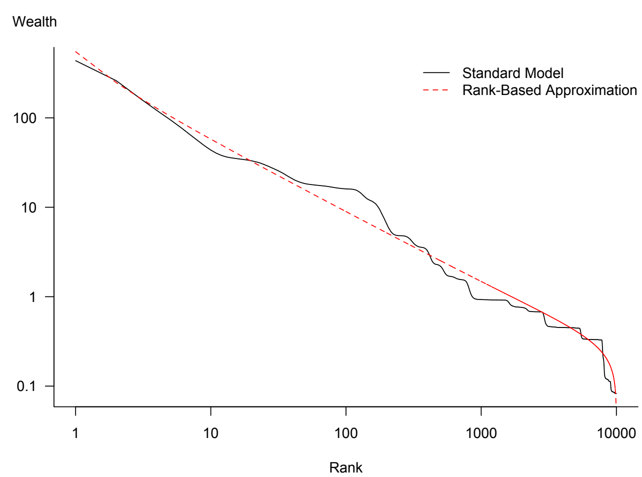

Figure 3: Log-log plot of wealth versus rank for the Standard model (average from 2,000 simulations) and its rank-based approximation.

	Data	Standard Model	Approximated Rank-Based Model
Wealth Distribution			
Top 1%	33.6%	33.0%	31.9%
Top 1-5%	26.7%	23.0%	17.1%
Top 5-10%	11.1%	6.9%	9.5%
Top 10-20%	12.0%	8.6%	11.2%
Top 20-40%	11.2%	11.4%	13.1%
Top 40-60%	4.5%	8.4%	8.3%
Bottom 40%	-0.1%	8.6%	8.9%

Table 1: Average wealth shares from 1,000 simulations of the different models - data from the Survey of Consumer Finances.

The Permanently Heterogeneous Rank-Based model. We calibrate the permanently heterogeneous rank-based model (2.9), which extends the rank-based model (2.1) to include permanent heterogeneity, so as to maintain approximately the same realistic stationary wealth distribution as the approximated rank-based model. We assume that 3,000 of the households are high-growth households, with $\gamma_h = 0.015$. According to (2.10), this implies that the remaining 7,000 low-growth households have $\gamma_\ell \approx -0.0064$. Furthermore, we can use the same estimated parameter values for σ_k from the approximated rank-based model (Figure 2) for the permanently heterogeneous rank-based model.

Given the postulated distribution of γ_i and σ_k, the calibration of the rank-based relative growth rates $\hat{\alpha}_k$ is chosen to produce a stationary distribution similar to the one produced by the Standard and the Approximated rank-based models (which, in turn, match well the distribution in the data). Indeed, we cannot simply use the estimated values of α_k from the approximated rank-based model (Figure 1) for the permanently heterogeneous rank-based model since the permanently heterogeneous parameters γ_i from (2.9) lead to a more skewed stationary distribution than in the model (2.1).

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 $k = 1, \ldots, N$. 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 permanently 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 $k = 1, \ldots, N$. Next, we simulate the permanently 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 sum of squared errors 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 errors of $\alpha'_k - \alpha_k$. 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 errors of $\alpha'_k - \alpha_k$ 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.

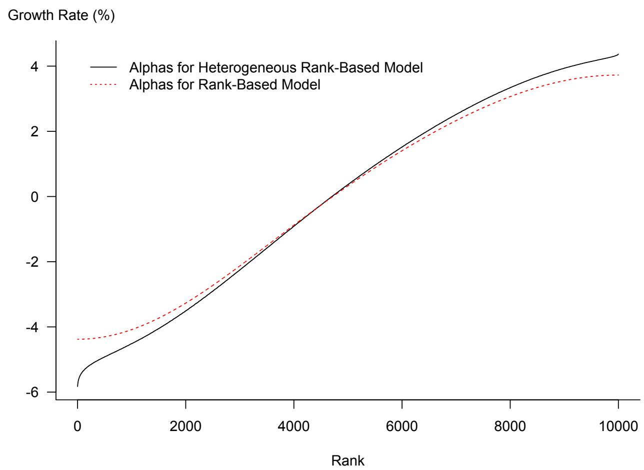

Figure 4: Annualized estimated parameters $\hat{\alpha}_k$ for the permanently heterogeneous rank-based model, and annualized estimated parameters α_k for the approximated rank-based model.

3.2 Results

All the models we calibrate are stationary and hence can be simulated to generate stationary distributions of wealth which can be compared with the SCF data on wealth shares by percentile. The results of these simulations are reported in the upper part of Table 2. Since these models are calibrated from a parameterization of the Standard model constructed to match these wealth shares, they all do relatively well at this, especially for the top 1% wealth share.

	Data	Approximated Rank-Based Model	Perman. Heterog. Rank-Based Model
Wealth Distribution			
Top 1%	33.6%	31.9%	34.0%
Top 1-5%	26.7%	17.1%	16.6%
Top 5-10%	11.1%	9.5%	9.2%
Top 10-20%	12.0%	11.2%	10.8%
Top 20-40%	11.2%	13.1%	12.7 %
Top 40-60%	4.5%	8.3%	8.1%
Bottom 40%	-0.1%	8.9%	8.5%
Wealth-Rank Correlations			
Parent-Child Rank Coeff.	0.191	0.229	0.255
Grandparent-Child Rank Coeff.	0.116	0.018	0.077
Long-Run Persistence Coeff.	0.105	0.000	0.100

Table 2: 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 models - 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 i) parent-child and grandparent-child wealth-rank correlations (average coefficients from regressions of child rank on parent rank and grandparent rank — see the note to Table 1) which we compare to those of Boserup et al. (2014); and ii) the long-run link of dynastic wealth-ranks which we compare to those of Barone and Mocetti (2016). The results of these comparisons are reported in the lower part of Table 2.[13] Both calibrated models tend to generate parent-

[13]The data we use to evaluate the models refers to different countries, though all are developed market economies. This is due to a lack of comparable evidence for the U.S. It is arguably not problematic in that we simply aim at a general theoretical and empirical understanding of the fundamental elements of a model of wealth dynamics rather than at a formal estimation procedure. Interestingly, estimates of returns to wealth in the U.S. (Benhabib et al., 2019) and in Norway (Fagereng et al., 2020) are very close; and ii) several of these developed market economies in the West tend to share comparable wealth distributions, at least in terms of their inequality (measured by the Gini coefficient); see Benhabib et al. (2017).

child wealth-rank correlations slightly higher than in the data. They also tend to generate grandparent-child wealth-rank correlations lower than in the data — though the permanently heterogeneous rank-based model fares much better in this respect. With regards to long-run correlations, the results we obtain are consistent with the theoretical results of Section 2. As implied by Proposition 2.3, household wealth-ranks are uncorrelated over very long time periods in the rank-based approximation of the Standard model. Household wealth-ranks are instead positively correlated over arbitrarily long time periods in a rank-based model that features permanent heterogeneity, as allowed by Theorem 2.5.[14] In conclusion, the permanent heterogeneity in the rank-based model helps to match rather well 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 3 shows the composition of the top 1% and top 5% wealth-ranked households in terms of low- and high-growth households. This table also shows the composition of the bottom 50% and bottom 25% ranked households. According to the table, high-growth households make up the great majority of the top 1% and 5%, but there is still a non-negligible minority of low-growth households in these top subsets. The results in the table also suggest that low-growth households are more common in top subsets of the wealth distribution than high-growth households are in bottom subsets of the wealth distribution. Indeed, the fraction of low-growth households in the bottom 25% approximately matches the fraction of high-growth households in the top 1%, even though the latter is a much smaller and more exclusive subset of the wealth distribution.

	Top 1%	Top 5%	Bottom 50%	Bottom 25%
High-Growth Households	82.6%	68.8%	18.3%	12.8%
Low-Growth Households	17.4%	31.2%	81.7%	87.2%

Table 3: 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

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

distribution for, respectively, high- and low-growth 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-growth households and 7,000 low-growth households, the maximum average occupation time for a high-growth household in any percentile of the wealth distribution is $1/3000 \approx 0.033\%$, while the maximum occupation time for a low-growth household in any percentile is $1/7000 \approx 0.014\%$.[15]

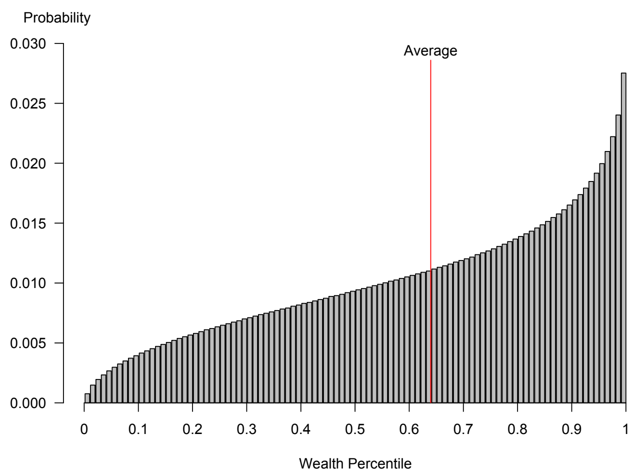

Figure 5: Average high-growth household occupation times for different percentiles of the wealth distribution from 1,000 simulations of the heterogeneous rank-based model.

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

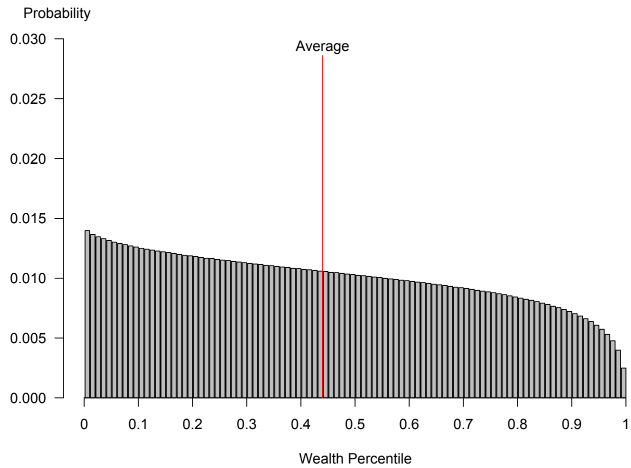

Figure 6: Average low-growth household occupation times for different percentiles of the wealth distribution from 1,000 simulations of the heterogeneous rank-based model.

3.2.1 Sensitivity analysis: Auto-correlated returns

To better identify the role of permanent heterogeneity in capturing both the wealth distribution and wealth-rank persistence over generations, in this section we compare it with a different form of imperfect social mobility, inter-generationally auto-correlated returns to wealth.[16] More specifically we report on the simulations of an extension of the Standard model (2.5) in which there is no permanent component to the growth rate of wealth but returns to wealth are highly auto-correlated across generations.

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 0.041 and the persistence parameter θ is equal to 0.95.[17] The standard deviation of $\epsilon_{i,t}$ is chosen to match the U.S. wealth distribution

[16]The persistent heterogeneity in household saving behavior is the mechanism exploited by Degan and Thibault (2016) to induce long correlations across generations. Such a mechanism is also at work in both the permanently heterogeneous and the auto-correlated returns models, as dynasties with higher returns endogenously display a higher savings rate.

[17]Results are very similar for $\theta = 0.90$.

data according to the SCF. For symmetry, we assume that labor earnings $\log y_{i,t}$ also follow an AR-1 process with persistence equal to 0.3, which matches the intergenerational persistence of income in the U.S. according to Chetty et al. (2014), and with mean and standard deviation chosen to match the calibration of income in the Standard model.[18] In all other respects, the auto-correlated returns model is identical to the Standard model which was used to calibrate the rank-based model.

	Data	Auto-Correlated Returns Model ($\theta = 0.95$)	Perman. Heterog. Rank-Based Model
Wealth Distribution			
Top 1%	33.6%	31.5%	34.0%
Top 1-5%	26.7%	20.6%	16.6%
Top 5-10%	11.1%	12.3%	9.2%
Top 10-20%	12.0%	13.5%	10.8%
Top 20-40%	11.2%	12.8%	12.7%
Top 40-60%	4.5%	5.8%	8.1%
Bottom 40%	-0.1%	3.5%	8.5%
Wealth-Rank Correlations			
Parent-Child Rank Coeff.	0.191	0.407	0.255
Grandparent-Child Rank Coeff.	0.116	0.044	0.077
Long-Run Persistence Coeff.	0.105	0.041	0.100

Table 4: See the notes to Table 2.

We report the results of these simulations in Table 4. These results show that the version of the Standard model with highly auto-correlated returns is able to match rather well the wealth shares in the data. Interestingly, it is able to generate a significant grandparent-child wealth-rank correlation, but to do so it requires a much too strong correlation between the parent and the child wealth-ranks. Fundamentally, however, the auto-correlated returns model fails to match the long-run persistence coefficient in wealth-ranks which instead is quite precisely captured by the permanently heterogeneous rank-based model.

To better understand these results, the long-run persistence of wealth-ranks implied by

[18]Note that the Standard model is calibrated so that income matches the U.S. data according to the SCF.

the permanently heterogeneous rank-based model and the auto-correlated returns model are compared 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 permanently 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.

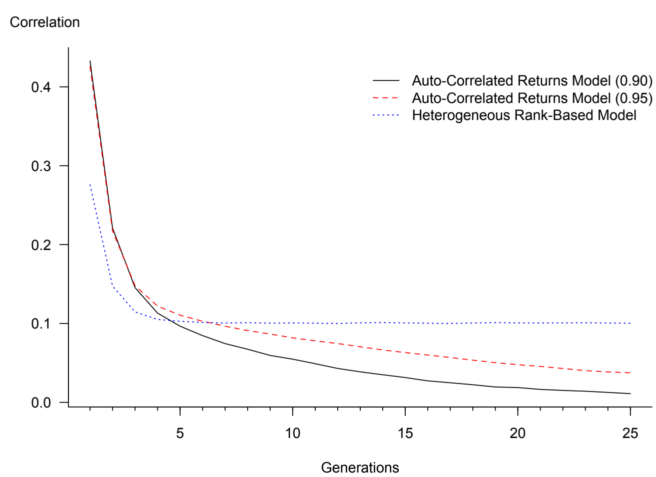

Figure 7: Rank correlations across multiple generations from 1,000 simulations of the permanently heterogeneous rank-based and auto-correlated returns models.

Another important dimension along which it is useful to compare the predictions of the permanently heterogeneous rank-based model and the auto-correlated returns models is parent-child return-rank coefficients. These correlation coefficients are interesting also as possible indicators of the mechanisms behind long-run wealth persistence. A relatively high parent-child return-rank coefficient suggests direct inter-generational mechanisms, like cultural transmission. A low coefficient suggests, on the contrary, institutional factors, like the perpetuation of the political and economic elites, which are extremely persistent but do not run directly from parent to child (Bourdieu, 1984, 1998; Acemoglu and Robinson, 2008; Bisin and Verdier, 2017). Using Norwegian data, Fagereng et al. (2020) report a small coefficient, 0.16, from a regression of child return rank on parent return rank.

Both the permanently heterogeneous and the auto-correlated returns models induce by construction some parent-child correlations in the returns to wealth.[19] In both models however this correlation is reduced by the postulated volatility of the growth rate, which is captured by the term σ_k in (2.9) for the permanently heterogeneous model and by the variance of $\epsilon_{i,t}$ in the auto-correlated model. In both models, these variances are chosen to match the distribution of wealth and wealth-rank correlations as reported in Table 4. Interestingly, we find that the permanently heterogeneous rank-based model, in our calibration, produces quite a small parent-child correlation of returns, even smaller than in the Norwegian data, while this correlation is much higher for the auto-correlated returns model. Specifically, the coefficient from a regression of child return rank on parent return rank, averaged across 1,000 simulations, is equal to 0.08 for the permanently heterogeneous model[20] and is equal to 0.95 for the auto-correlated returns model. Effectively, in our calibration, the volatility of the growth rate in the permanently heterogeneous model introduces enough churning in parent-child returns to lower their rank correlation substantially. This is not the case for the auto-correlated model, which requires a small volatility of the growth rate to match the distribution of wealth and wealth-rank correlations and as a consequence dramatically misses the low return-rank correlation documented by Fagereng et al. (2020).

[19] In order to calculate the parent-child return rank coefficient for the permanently heterogeneous rank-based model, it is necessary to decompose the parameters $\hat{\alpha}_k$ from (2.9) into log return and log labor income to capital income components, as in (2.8). Specifically, we can write (2.9) as

$$d \log w_i(t) = \left(\gamma_i + \kappa_{\rho_t(i)} + \omega_{\rho_t(i)}\right) \, dt + \sigma_{\rho_t(i)} \, dB_i(t), \tag{3.3}$$

where the parameters κ_k and ω_k measure, respectively, the log return to wealth and the log ratio of labor income to capital income for the k-th ranked household and satisfy $\hat{\alpha}_k = \kappa_k + \omega_k$, for each rank $k = 1, \ldots, N$. We assume that returns to wealth are constant across wealth ranks as in the Standard model. As a consequence, $\kappa_k = 0$ for all $k = 1, \ldots, N$ (following the normalization of the $\hat{\alpha}_k$ parameters and without loss of generality, we normalize the κ_k parameters to sum to zero).

[20] We conjecture that using a model that incorporates higher returns at higher wealth ranks as in Benhabib et al. (2019) to estimate a different permanently heterogeneous rank-based model of the form (2.9) could generate an even closer match to the 0.16 parent-child return rank coefficient in Fagereng et al. (2020).

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 permanent 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. Finally, we find that the model with permanent heterogeneity produces in our calibration a small parent-child correlation of returns, close to the one documented by Fagereng et al. (2020) for Norway. This suggests persistent institutional factors as mechanisms for sustaining long-run wealth persistence, rather than direct inter-generational mechanisms like cultural transmission. While we do not have enough structure and data to identify particular institutional channels responsible for the long-run persistence in wealth correlations, future work along these lines is needed to further this literature.

References

Acemoglu, D. and J. A. Robinson (2008). Persistence of power, elites, and institutions. *American Economic Review 98*(1), 267–93.

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). Intergenerational mobility in the very long run: Florence 1427-2011. *Bank of Italy Working Paper No. 1060.*

Benhabib, J. and A. Bisin (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature 56*(4), 1261–91.

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.

Bisin, A. and A. Moro (2021). Late for history. In A. Bisin and G. Federico (Eds.), *Handbook of Historical Economics.* Amsterdam: Elsevier North Holland.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory 97*(2), 298–319.

Bisin, A. and T. Verdier (2017). On the joint evolution of culture and institutions. Technical report, National Bureau of Economic Research.

Bisin, A. and T. Verdier (2021). Phase diagrams in historical economics: Culture and institutions. In A. Bisin and G. Federico (Eds.), *Handbook of Historical Economics.* Amsterdam: Elsevier North Holland.

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.*

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste.* Harvard university press.

Bourdieu, P. (1998). *The state nobility: Elite schools in the field of power*. Stanford University Press.

Bourdieu, P. and J. Passeron (1970). La reproduction: éléments pour une théorie du système d?enseignement (paris, ed. de minuit).

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.

Charles, K. K. and E. Hurst (2003). The correlation of wealth across generations. *Journal of political Economy 111*(6), 1155–1182.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014, November). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics 129*(4), 1553–1623.

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.

Degan, A. and E. Thibault (2016). Dynastic accumulation of wealth. *Mathematical Social Sciences 81*, 66–78.

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 (2021, Forthcoming). High marginal tax rates on the top 1%? Lessons from a life cycle model with idiosyncratic income risk. *American Economic Journal: Macroeconomics*.

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*.

Voth, J. (2021). Persistence - myth and mystery. In A. Bisin and G. Federico (Eds.), *Handbook of Historical Economics*. Amsterdam: Elsevier North Holland.

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 2.2, 2.3, and 2.4 and Theorem 2.5.

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 $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_N$. Following the econometric procedure described by Fernholz (2017), we can estimate the α_k parameters in (2.6) and map them to the corresponding α_k parameters from the rank-based model (2.1) so that condition (2.2) is satisfied and the rank-based approximation of the Standard model is stationary. □

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 $i = 1, \ldots, N$,

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

where the second equality follows from (2.11). □

Proof of Proposition 2.4. 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 $\{1, \ldots, N\}$ by Ψ_N, where $p(k) \in \{1, \ldots, N\}$ denotes the k-th element of the permutation $p \in \Psi_N$. Each permutation p describes a potential wealth ranking of the households $i = 1, \ldots, N$, with $p(k)$ denoting the index of the k-th ranked household. According to Corollary 4 from Ichiba et al. (2011), for all $i, k = 1, \ldots N$, 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 $k = 1, \dots, N-1$,

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

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 $k = 1, \ldots, N$. 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

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

which implies that

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

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

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

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

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

is increasing in k. $\square$