

w o r k i n g
p a p e r

16 34

Mobility

Daniel R. Carroll and Eric R. Young



FEDERAL RESERVE BANK OF CLEVELAND

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. The views stated herein are those of the authors and are not necessarily those of the Federal Reserve Bank of Cleveland or the Board of Governors of the Federal Reserve System.

Working papers are available on the Cleveland Fed's website:

<https://clevelandfed.org/wp>

Mobility

Daniel R. Carroll and Eric R. Young

This paper studies short-run wealth mobility in a heterogeneous agents, incomplete-markets model. Wealth mobility has a “hump-shaped” relationship with the persistence of the stochastic process governing labor income: low when shocks are close to i.i.d. or close to a random walk, and higher in between. The standard incomplete markets framework features less wealth mobility than found in the PSID wealth supplements. We include features commonly used in the literature to capture wealth inequality and find that they do little to improve the model’s performance for wealth mobility. Finally, we introduce state-contingent assets, which allow households to partially span the space of labor productivity. Moving toward a more “complete” market lowers wealth mobility unless the labor income process is very persistent.

Keywords: wealth mobility, inequality, incomplete markets.

JEL Codes: D52, D31, E21.

Suggested citation: Carroll, Daniel, and Eric Young, “Mobility,” Federal Reserve Bank of Cleveland, Working Paper no. 16-34.

Daniel R. Carroll is at the Federal Reserve Bank of Cleveland (daniel.carroll@clev.frb.org) and Eric R. Young (corresponding author) is at the University of Virginia (ey2d@virginia.edu). The authors thank Nicholas Hoffman for outstanding research assistant work.

1 Introduction

This paper examines wealth mobility in a simple dynamic stochastic general equilibrium model with incomplete markets in the spirit of Bewley (1986), Aiyagari (1994), and Huggett (1993). This model and its many variations has become the workhorse model of macroeconomics in great part because it generates an endogenous distribution of agents across income and wealth. This

endogenous distribution is ideal for studying the effects of policy on inequality. Very little is understood in this environment regarding wealth mobility – the frequency with which agents “switch places”.

Mobility is distinct from inequality. It should be obvious that inequality is a necessary condition for mobility – if everyone is the same, it makes no sense to talk about households switching places in any distribution. But inequality can arise in the absence of mobility – for example, without risk, inequality not only can be present but can be permanent, depending on how savings choices vary in the population, but mobility may be zero as agents are frozen at their relative place in the wealth distribution. Thus, inequality per se is not informative about the insurance opportunities that agents can access. Furthermore, given the generally-poor quality of the consumption panel data needed to directly characterize the incompleteness of asset markets, we think it useful to consider whether mobility data can help us characterize the asset markets used by households.

This question has policy implications. As discussed in Carroll and Young (2011), changes in the progressivity of income tax functions have qualitatively different effects depending on whether inequality is driven by uninsurable risk (as in Castaneda et al. (2003)) or time-invariant characteristics like preferences and average labor efficiency (as in Carroll and Young (2011)). Specifically, a more progressive income tax function leads to lower inequality under idiosyncratic risk, as it operates to reduce the volatility of the idiosyncratic component (labor productivity) and compresses the distribution of returns; in contrast, under no risk but permanent heterogeneity rich households actually increase their assets and poor households reduce them. It seems important therefore to get a clear picture of what drives inequality, and that requires an understanding of asset market opportunities. Since inequality is not sufficient, we turn to mobility. In this paper we will not completely answer the question of what drives mobility in the data; instead our goal is the more modest one of characterizing how mobility is determined with a given workhorse model and comparing it to the numbers found in the Panel Study of Income Dynamics Wealth Supplements.

To this end, we first present a battery of different measures of mobility found in the literature. The Shorrocks measure uses the trace of the Markov transition matrix only – a process has higher mobility if the trace is smaller, meaning that households are more likely to leave their

current quantile. In contrast, the Bartholomew measure uses a weighted average of transition probabilities, where the weights are the absolute number of quantiles that the agent 'passes through'; this measure views an economy as more mobile if agents move quickly, even if they simply bounce back and forth between two states. The 'second highest eigenvalue' measure is commonly used in macroeconomics; we show for two-state chains the autocorrelation of a chain equals the second highest eigenvalue.¹ Finally, the 'mean first passage time' calculates the expected number of periods before a household exiting an initial quantile reaches any other particular quantile for the first time.

We use these measures to interpret the wealth movements generated by our model. Comparing two environments in which the only difference is the persistence of the idiosyncratic shock, we find that mobility is 'hump-shaped' – mobility is low when shocks are close to iid and when they are close to a random walk, and higher at intermediate values. We decompose the change in mobility as the sum of three components, which we label luck, behavior, and structure. First, fix the behavior of all households at a given persistence value, and let one household draw a sample sequence from a process with a higher autocorrelation; all that changes is the particular realizations of the shock, which we label 'luck'. Second, now let this household realize that the persistence of her shock is different and reoptimize, leaving the behavior of all other households unchanged (including their persistence); we label this change 'behavior', since it captures the effect of different decision rules on mobility. Third, suppose all other households also face the new persistence coefficient, leading to changes in the distribution of wealth against which any particular household will be viewed; we call this the 'structural' effect.²

All four mobility measures return a similar decomposition pattern. Structure has a minimal effect on mobility, while behavior has a large negative effect and luck a large positive effect (given the change is to increase the autocorrelation). The negative effect of behavior results from the decreased sensitivity of saving to more persistent income shocks; this sensitivity is a reflection of consumption-smoothing, wherein shocks that are permanent are absorbed into consumption

¹This result is not general. We cannot prove anything for chains with more than two states, but a Monte Carlo experiment with random stochastic matrices shows a low correlation between the modulus of the second-highest eigenvalue and the autocorrelation of the chain.

²Technically, the behavior effect could also arise under the third change, since equilibrium prices differ. We ignore this distinction.

since there is no 'better future' to borrow from, while transitory shocks are smoothed away using a buffer-stock of assets. In contrast, increased persistence more often generates longer consecutive strings of high or low productivity draws (luck of the draw), and therefore generates more movement up and down the wealth ordering.

Using a reasonable calibration for the income process (taken from Floden and Lindé (2001)) with a high persistence of shocks, we find that the benchmark model delivers too little short-run mobility relative to the data (over five-year horizons). Specifically, we find that the model implies far too little mobility overall, but in particular fails to deliver the high mobility observed in the lowest and highest quintiles; in the model, households stay in these quintiles on average 38 and 63 years, in contrast to values in the data closer to 15 and 17 years, respectively. Furthermore, households in the model also stay in their initial quintile too frequently, and when they move they move only one quintile at a time; in the data wealth moves more rapidly, with significant numbers of households switching more than one quintile in either direction.

We then move to consider different environments, designed to illuminate what is causing the model to fail. We first look at a variety of changes used to match the extreme wealth inequality observed in the data. We examine the Krusell and Smith (1998) modification that introduces stochastic movements in discount factors that are highly persistent. The discount factor model improves a small amount by increasing mobility at the lower end of the wealth distribution, but actually reduces it at the high end; the reason the model gets high wealth concentration is that it nearly 'freezes' rich households in the top quintile, since high discount factor types will save a significant amount whether their income is high or not, and these discount factor states are very persistent.

We then examine a 'rockstar' model as in Castaneda et al. (2003), in which the earnings process has a rare and transitory state with very high income and a relatively high probability of dropping to the lowest state. The rockstar model works relatively well, as it increases mobility across the board and introduces some households that shift across more than one quintile; nevertheless, mobility at the high end is still substantially too low as households do not choose to let their wealth fall fast enough. This failure can be understood as the result of standard buffer-stock behavior combined with decreasing absolute risk aversion – with high temporary income, households save rapidly to move away from the borrowing constraint but dissave slowly.

We finally consider how the span of assets affects mobility – we permit households to purchase some contingent securities, but maintain a borrowing constraint. With contingent claims, wealth mobility can increase if the household takes extreme positions; if transitions are rare, then insurance against those transitions is cheap and permits very large portfolio positions, which are useful due to the borrowing constraint. On the other hand, as shown in Rampini and Viswanathan (2016), poor households hold portfolios that hedge against fewer states than rich households do, meaning that wealth mobility may fall. We find that the hump-shaped pattern from the one asset case holds in an environment with an incomplete set of state-contingent assets. Compared to the one-asset baseline economy, partially completing the market decreases wealth mobility when the underlying income shock persistence is not too high. When the persistence becomes sufficiently large, however, the partial insurance economy has greater wealth mobility, due to this portfolio-composition effect.

2 Model

As a starting point, we study the long run properties of Aiyagari (1994) with no borrowing.³ There is a unit measure of *ex ante* identical households. Every period, each household receives an idiosyncratic labor productivity shock, ε , from a finite set $\mathcal{E} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_J]$ with $\varepsilon_1 < \varepsilon_2 < \dots < \varepsilon_J$. The process for productivity shocks be Markov with stochastic transition matrix $\Pi = \Pr(\varepsilon_j | \varepsilon_i)$ for $j, i \in 1, \dots, J$. Every household supplies the same fixed number of hours, \bar{h} , and earns total labor income equal to $\omega \bar{h} \varepsilon$, where ω is a market-wide wage. Because the wage and hours supplied do not change across periods, labor productivity shocks are equivalent to random labor income endowments. As in the standard incomplete-markets model, there is only one asset, a , which is a claim to the capital stock K . Because no state contingent claims exist, households have a motive to self-insure through precautionary savings.

A stand-in firm combines capital and effective labor through a constant-returns-to-scale production technology $F : \mathfrak{R}^+ \times \mathfrak{R}^+ \rightarrow \mathfrak{R}^+$ to produce a final good which may be consumed or invested in capital for next period. The firm manages the capital stock from household's saving, pays an interest rate r on assets, hires labor, and invests in new capital. Capital depreciates at

³Because we are concerned with mobility in the stochastic steady state, we omit time subscripts.

a constant rate δ each period. We assume that the firm behaves competitively. Letting F be Cobb-Douglas, the optimal choice of the firm implies that each factor is paid its marginal product:

$$\omega = (1 - \alpha) \left(\frac{K}{N} \right)^\alpha$$

and

$$r = \alpha \left(\frac{K}{N} \right)^{\alpha-1} - \delta.$$

The state vector of the household has two elements: current wealth, a , and current labor productivity, ε . Let period utility be represented by a continuous, strictly concave function $u : \mathfrak{R}^+ \rightarrow \mathfrak{R}$, and assume that u is continuously differentiable as many times as necessary. The household problem in recursive form is

$$V(a, \varepsilon) = \max_{c, a'} \{u(c) + \beta E_{\varepsilon'|\varepsilon} [V(a', \varepsilon')]\}$$

subject to the budget constraint

$$c + a' \leq w\varepsilon + (1 + r)a$$

and lower bound constraints

$$c > 0; a' \geq \underline{a}.$$

Denote by $\Gamma(a, \varepsilon)$ the distribution of households over $\mathcal{A} \times \mathcal{E}$.

Definition 0.1 A *steady-state recursive competitive equilibrium* is a set of value functions $V(a, \varepsilon)$, policy functions $g_a(a, \varepsilon)$, $g_c(a, \varepsilon)$, pricing functions, r and w , and a distribution $\Gamma(a, \varepsilon)$ such that

1. Given prices, V , g_a and g_c solve the household's problem.
2. Firms maximize profits

$$\omega = (1 - \alpha) \left(\frac{K}{N} \right)^\alpha$$

and

$$r = \alpha \left(\frac{K}{N} \right)^{\alpha-1} - \delta.$$

3. Markets clear:

$$K = \sum_{j=1}^J \int a d\Gamma(a, \varepsilon_j)$$

$$N = \sum_{j=1}^J \int \bar{h}_{\varepsilon_j} d\Gamma(a, \varepsilon_j).$$

4. Γ is consistent with the saving decisions of households and the process for ε .

5. The joint distribution of wealth and productivity $\Gamma(a, \varepsilon)$ is stationary.

3 Measures of Mobility

The literature on measuring income mobility with transition matrices dates back to at least as early as Prais (1955) who examined transitions between occupational classes in England. There is no standardized measure in part because there are many aspects to mobility.⁴ In this paper, we are primarily interested in so-called *relative mobility*. Relative mobility measures how likely it is that a household in wealth quantile n_1 at time s will be in some other quantile n_2 at time $s + t$, where t is a fixed number of periods in the future.

Formally, represent by $x(\Gamma)$ the distribution over each of N wealth quantiles (i.e., $x = [\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}]$) and by $q(\Gamma)$ the wealth values defining the quantiles. That is,

$$q(\Gamma) = [q_1, q_2, \dots, q_N]$$

where $q_1 = \underline{a}$ and $q_i = a_i : \sum_{j=1}^J \int d\Gamma(a, \varepsilon_j) \mathbf{1}_{\{q_{i-1} \leq a < q_i\}} = \frac{1}{N}$, for $i = \{1, \dots, N\}$. The q_i values define the cutoff wealth values for entering the i^{th} quantile (the lowest wealth value in the quantile). Further, denote by $Q_i = \{a : a \in [q_i, q_{i+1})\}_{i=1, N-1}$ and $Q_N = \{a : a \geq q_N\}$; these sets define the wealth levels that constitute a given quantile. Finally, let $M_{N \times N}(\Gamma)$ be a regular transition matrix induced by Γ with the element m_{ij} indicating the probability that a household in quantile Q_i will be in quantile Q_j after some fixed number of periods.⁵

⁴For a broad overview of the literature, see Fields and Ok (1999).

⁵According to Theorem 4.1.2 in Kemeny and Snell (1976), a transition matrix is regular if and only if for some $t > 0$, M^t has no zero entries. Regularity guarantees that starting from any state in the Markov chain any other state can be visited in a finite amount of periods (that is, all states communicate). This condition is related to

We will consider four measures from the literature, discussed at length in Dardanoni (1993). In particular, we highlight how each measure captures somewhat different aspects of mobility (due to the loss of information generated by moving from a matrix to a scalar).

3.0.1 Shorrocks Measure

Shorrocks (1978) measure of mobility focuses on the probability weight along the diagonal of M . One interpretation of the measure is that it reports the 'stickiness' of initial conditions. Formally, Shorrocks' measure is

$$\mu_S(M) = \frac{N - \text{trace}(M)}{N - 1}.$$

The Shorrocks measure takes values between 0 and 1, with smaller values indicating a lower likelihood that a household will escape its initial quantile. Importantly, the measure is unaffected a reallocation of mass along off-diagonal elements. The Shorrocks measure makes no distinction between economies where households move immediately from rags to riches and those where the poor become only slightly less poor. That is, the two Markov processes

$$\Pi_A = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.25 & 0.5 & 0.25 \\ 0.0 & 0.5 & 0.5 \end{bmatrix}$$

and

$$\Pi_B = \begin{bmatrix} 0.5 & 0.0 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.5 & 0.0 & 0.5 \end{bmatrix}$$

would be regarded as equally mobile. Clearly, the second process moves 'faster', since households switch across multiple quintiles.

the 'monotone mixing condition' (see Hopenhayn and Prescott (1992)) used to prove the existence of a stationary distribution Γ , which Ríos-Rull (1998) labels 'the American Dream and the American Nightmare' condition. This condition is a long-run mobility requirement, whereas we are interested in short-run effects.

3.0.2 Bartholomew's Immobility Measure

In contrast to the Shorrocks measure, Bartholomew and Bartholomew (1967) deals exclusively with the off-diagonal elements:

$$\mu_B(M) = \frac{1}{N-1} \sum_{i=1}^N \sum_{j=1}^N m_{ij} |i-j|$$

is the expected number of quantiles a household would cross into each period. The measure puts positive weight only on the off-diagonal probabilities. The term $|i-j|$, the absolute number of quantiles crossed into, weights more heavily transitions that cross multiple quantiles; a transition matrix with more probability mass further from the diagonal has greater mobility (like Π_B in the previous subsection). Fields and Ok (1999) point out that Bartholomew's measure can be thought of as capturing total movement; economies in which households oscillate between being very rich and very poor would be measured as much more mobile than one where households transitioned more slowly through adjacent quantiles, even if the former involved fewer such transitions. To see how this measure works, consider the Markov processes

$$\Pi_A = \begin{bmatrix} 0.5 & 0.5 & 0.0 \\ 0.25 & 0.5 & 0.25 \\ 0.0 & 0.5 & 0.5 \end{bmatrix}$$

and

$$\Pi_B = \begin{bmatrix} 0.75 & 0.0 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.0 & 0.75 \end{bmatrix}$$

Under Bartholomew's measure, these chains are equally mobile:

$$\mu_B(\Pi_A) = 0.75$$

$$\mu_B(\Pi_B) = 0.75.$$

But clearly agents move 'more often' under A ; they move 'more' under B whenever they move, however.

3.0.3 Second Largest Eigenvalue

The second largest eigenvalue of a stochastic matrix governs the mixing rate of a Markov chain process, with a larger eigenvalue implying a slower mixing rate. Let $\lambda_i(M)$ be the i^{th} largest eigenvalue of M . A natural measure of mobility is $\mu_{2E}(M) = 1 - |\lambda_2(M)|$. Because M is regular $\lambda_1 = 1$, and $\lambda_i < 1$ for all $i > 1$. Sommers and Conlisk (1979) show that $\mu_{2E}(M)$ measures the total deviation of M from a matrix with perfect mobility.⁶

To understand why this measure captures mobility, we can show for a two-state Markov chain that the second highest eigenvalue is equal to the autocorrelation of the chain. Let the Markov chain transition matrix be

$$\Pi = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

which has invariant distribution

$$\pi^* = \left[\frac{1-q}{2-p-q}, 1 - \frac{1-q}{2-p-q} \right].$$

The autocorrelation is

$$\rho(z_t|z_{t-1}) = \frac{(z_1 - z_2)^2 (1-p)(1-q) \frac{p+q-1}{(2-p-q)^2}}{(z_1 - z_2)^2 (1-p) \frac{1-q}{(p+q-2)^2}} = p + q - 1$$

and the eigenvalues of Π are 1 and $p + q - 1$.

This result is not general, and we were unable to derive any analytical results for chains with more than 2 states. We therefore conducted a Monte Carlo exercise by drawing 5000 random stochastic matrices and computing the sample autocorrelation from a simulation of length 100,000; it turns out that this autocorrelation is only weakly correlated with the second-highest eigenvalue of the transition matrix.

3.0.4 Mean First Passage Time

The mean first passing matrix $T(M)$ is the expected number of periods until a household initially in quintile i first arrives in quintile j ; Meyer (1978) shows that

$$T = (I - K + J \text{diag}(K)) (\text{diag}(K))^{-1} + E$$

⁶Perfect mobility for a $N \times N$ matrix is one with all elements equal to $1/N$. This concept is related to 'origin independence.'

where

$$E = \begin{bmatrix} 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & \cdots & 0 & -1 \\ \vdots & \vdots & \cdots & 0 & -1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix},$$

J is a matrix of all ones,

$$K = \begin{bmatrix} U & 1^T \\ d^T & 1 \end{bmatrix}^{-1},$$

and

$$A = I - M = \begin{bmatrix} U & c \\ d^T & \alpha \end{bmatrix}.$$

Conlisk (1990) proposes using

$$MFP = x^T T x$$

as a measure of mobility; MFP is the expected number of periods before one household enters the quintile of another household when both are drawn at random from Γ . Because x has equal elements that sum to one (recall that x is a vector of quantiles), MFP is just the average value of the elements of T . For ease of comparison to the other measures, we define

$$\mu_{MFP}(M) = \frac{N}{MFP}.$$

If M is "perfectly mobile" $\mu_{MFP} = 1$. As the diagonal elements of M approach one, $\mu_{MFP} \rightarrow 0$.⁷

So far we have defined these measures generally for any set of evenly spaced quantiles. In the remainder of this paper, we will restrict attention to quintiles, that is $\dim(M) = 5$.

3.1 Structural vs. Exchange Mobility

We are concerned with how quickly and to what extent agents change their ordering within the stochastic stationary distribution of wealth (known as *relative mobility*). In the steady state, households' wealth positions change, but the wealth distribution itself is time-invariant. It would

⁷ μ_{MFP} cannot exceed 1 if M is *monotone* (i.e., each row is stochastically dominated by the one below it). Huggett (1993) proves the monotonicity of M in Bewley models with positively-autocorrelated shocks.

be intuitive to presume that relative mobility is just a simple function of the rates at which agents accumulate wealth and that greater relative mobility implies that households transition more quickly through quintiles, by rising and falling over a shorter time span. This *exchange (or pure) mobility*, however, is only one component of relative mobility. Differences in relative mobility can also arise from changes in the shape of the wealth distribution, even if individual savings behavior is the same. This concept is called *structural mobility*, and it can appear in the data when wealth inequality changes over time. In the stochastic steady state of a Bewley model, wealth inequality does not change over time. Nevertheless, structural mobility must still be taken into account when comparing the steady states from two models. Because of general equilibrium effects, changes in the model environment induce changes in the shape of the stationary distribution as well and are likely to alter the cutoffs defining wealth quintiles.

To illustrate, consider two distributions of wealth, Γ_1 and Γ_2 , and let Γ_2 be a shape-preserving spread of Γ_1 . Take a household from each distribution and label them according to their distribution of origin. Because there is more wealth in equality in Γ_2 than in Γ_1 , the cutoffs which define the quantiles will be spread more apart. Even if household 1 and 2 begin with the same initial wealth, have the same optimal saving policies, and experience identical realizations for labor productivity, household 2 will transition across quantiles less frequently over the same amount of time, and so our measures of mobility would rank Γ_2 as less mobile than Γ_1 . Figure 1 plots the cutoffs for entering each quintile as defined by the distribution of wealth from our experiments. Notice that there is not much change in the cutoffs until ρ exceeds 0.7. Beyond that, as the productivity process becomes more persistent, the distribution spreads out, and the cutoffs become further apart. In our numerical experiments, we will detail how we use the model to identify exchange mobility from structural mobility.

3.2 Exchange Mobility: Behavior vs. Luck

Once the movement of households through the distribution has been isolated from movements in the distribution itself, exchange mobility can be separated further into changes due to differences in productivity shock process and changes in household behavior. Consider two households A and

B with two Π matrices. Let $\rho_A = 0$ and $\rho_B = 0.5$ so

$$\Pi_A = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

and

$$\Pi_B = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}.$$

One might initially suppose that household A will have greater mobility than household B . After all, according to any one of the above measures, the earnings mobility of A is considerably greater than that of B . This fact however does not necessarily translate to greater wealth mobility. The reason is that randomness in the household earnings does not wholly determine a household's wealth. Because household utility is strictly concave, they try to smooth consumption over time. Since shocks for A are less persistent, the optimal response of household A to a switch in productivity is to adjust savings. The more persistent the shocks, the more closely earnings resemble permanent income and the less savings adjusts.

4 Data

Nearly all empirical studies of wealth mobility have focused on wealth changes across generations mainly due to the limited amount of panel data on wealth. Studies of intragenerational mobility have exclusively come from the PSID wealth supplement data. Castaneda et al. (2003) compute a 5-year measure using just the 1984 and 1989 PSID waves and report the following diagonal entries:

$$\begin{bmatrix} 0.67 & - & - & - & - \\ - & 0.47 & - & - & - \\ - & - & 0.45 & - & - \\ - & - & - & 0.50 & - \\ - & - & - & - & 0.71 \end{bmatrix}.$$

Hurst et al. (1998) examine the analyze the supplements on household family wealth in the PSID for 1984, 1989, and 1994 and report one 10-year and two 5-year wealth transition matrices by deciles.⁸ The implied 5-year quintile transition matrix for 1984 to 1989 is

⁸The rows of this matrix may not sum to 1 because of rounding.

$$\begin{bmatrix} 0.66 & 0.24 & 0.07 & 0.03 & 0.00 \\ 0.26 & 0.45 & 0.21 & 0.06 & 0.02 \\ 0.06 & 0.25 & 0.44 & 0.19 & 0.06 \\ 0.01 & 0.05 & 0.23 & 0.49 & 0.21 \\ 0.01 & 0.01 & 0.05 & 0.22 & 0.71 \end{bmatrix}$$

while that for 1989 to 1994 is

$$\begin{bmatrix} 0.60 & 0.27 & 0.08 & 0.03 & 0.01 \\ 0.27 & 0.45 & 0.17 & 0.09 & 0.03 \\ 0.08 & 0.22 & 0.44 & 0.21 & 0.06 \\ 0.04 & 0.05 & 0.26 & 0.44 & 0.21 \\ 0.01 & 0.03 & 0.05 & 0.24 & 0.68 \end{bmatrix} .$$

Díaz-Giménez et al. (2011) report a 6-year quintile matrix for head of households between ages 35-45 with positive earnings in the beginning and end of the sample

$$\begin{bmatrix} 0.65 & 0.28 & 0.05 & 0.01 & 0.01 \\ 0.23 & 0.47 & 0.25 & 0.04 & 0.01 \\ 0.07 & 0.16 & 0.49 & 0.20 & 0.08 \\ 0.01 & 0.07 & 0.17 & 0.54 & 0.21 \\ 0.03 & 0.02 & 0.05 & 0.21 & 0.70 \end{bmatrix} .$$

Table 1 reports mobility measures for these various matrices.

5 Numerical Experiments

5.1 Baseline

We choose fairly standard values for our structural parameters: we let utility be logarithmic, we choose $\beta = 0.99$ and $\delta = 0.025$ as roughly consistent with quarterly aggregates for the capital/output and investment/output ratios, and we set $\alpha = 0.36$ to match capital's share of income. We also choose a zero borrowing limit.

We follow Floden and Lindé (2001) who estimate an earnings process of $\rho = 0.92$ and $\sigma_\varepsilon = 0.21$ (annual) from the PSID. The resulting 5-year wealth transition matrix is

$$\begin{bmatrix} 0.87 & 0.14 & 0.00 & 0.00 & 0.00 \\ 0.13 & 0.73 & 0.14 & 0.0 & 0.00 \\ 0.00 & 0.14 & 0.74 & 0.12 & 0.00 \\ 0.00 & 0.00 & 0.12 & 0.81 & 0.07 \\ 0.00 & 0.00 & 0.00 & 0.08 & 0.92 \end{bmatrix}$$

which features far less wealth mobility than any of the transition matrices above. Because the underlying source of both inequality and mobility in this model is the stochastic earnings process, we examine how the transition matrix above responds to different assumptions about the Markov process.

5.1.1 Earnings Process

The fundamental force driving the distribution of wealth in the economy is the labor productivity process. We assume the Markov process above approximates

$$\log(\varepsilon') = \rho \log(\varepsilon) + \nu', \quad \nu' \sim N(0, \sigma^2).$$

We set J , the number of individual productivity states, to 2.⁹ Given this and the parameters ρ and σ , we use the Rouwenhorst method to construct the Markov chain process. Under the Rouwenhorst method, the Markov chain depends upon ρ and σ . The states are equally-space over the interval $[-\psi, \psi]$, where

$$\psi = \frac{\sqrt{(J-1)}}{\sqrt{(1-\rho^2)}}\sigma.$$

The transition matrix, Π , depends on two parameters, p and q . Following Kopecky and Suen (2010), we set

$$p = q = \frac{1 + \rho}{2};$$

⁹We have run our experiments with 7 productivity states as well. In general, the qualitative results do not change significantly. One issue that arises when there are more than 2 values for productivity is for very low values of ρ the transition matrix is no longer monotone (i.e., the conditional probability of moving from $\varepsilon = \varepsilon_i$ to $\varepsilon' = \varepsilon_j$, $j \neq i$, does not monotonically decrease as the distance between j and i increases). Since monotonicity of the transition matrix is important for understanding the mobility measures and this failure is simply an approximation error, we concentrate on the two-state case.

note that Π only depends upon the persistence parameter ρ .

A consequence of generating a Markov chain in this manner is that if one only varies ρ and keeps σ fixed, the vector of states will be different for each value of ρ . This dependence will cause the marginal distribution of effective labor to vary across experiments due solely to the approximation procedure, which could mess up our comparisons. To prevent this contamination, we make σ a function ρ . Given a baseline ρ_0 and σ_0 , we define

$$\sigma(\rho) = \sigma_0 \sqrt{\frac{1 - \rho^2}{1 - \rho_0^2}}.$$

This procedure guarantees that the ε state vector of productivity remains the same across ρ experiments and, because labor is supplied inelastically, so does N . Moreover, because Π depends solely on ρ , we can isolate changes to the transition probabilities without altering the states. In this way, ρ will increase the probability of earning the same (by construction) current labor income in the next period (it increases the weight along the diagonal of the transition matrix).

5.2 The three factors affecting wealth mobility as ρ changes

We conduct a series of computational experiments to identify the fundamental ingredients governing individual wealth mobility within the model. Specifically, we vary ρ , compute the stochastic steady state, approximate the quintile wealth transition matrix via simulation, and calculate mobility. Figure 2 plots the relationship between ρ and several measures of mobility. Mobility is hump-shaped across persistence with mobility being low when ρ is near 0 and when ρ is near 0.9, and reaches its peak for $\rho \in (0.75, 0.80)$. Because for each value of ρ the model is solved in general equilibrium, the market clearing interest rate and the wealth distribution itself will differ in each case. Thus, our results are the combination of changes in structure, behavior, and luck. In a later section, we describe our strategy for identifying the portion of mobility arising from each of these components, but first we will discuss persistence affects each in turn.

Structure Figure 3 plots the steady state wealth distribution under different values of the persistence of the productivity process. There are two things to note about the distribution as ρ increases. First, the wealth becomes more unequally distributed as the right tail stretches out. Because there are only two productivity states, in equilibrium households with the high (low)

productivity are savers (dissavers). The closer ρ is to 1, the more likely households with high ε are to draw high ε' . As a consequence, some households will receive a very long string of good productivity shocks, allowing them to amass a considerable amount of wealth. In the same way, households that draw a low productivity will be more likely to draw low productivity in the future, leading to the second feature of a larger ρ : more households are borrowing constrained. These changes in the structure of the wealth distribution affect the boundaries between quintiles. Figure 1 plots these boundaries for different values of ρ . The cutoffs move apart gradually as ρ approaches 0.7. As the productivity process becomes more even persistent, however, the distribution spreads out rapidly, and the boundaries become further apart. When $\rho = 0.99$, the entire first quintile is at the borrowing limit.¹⁰

Behavior Optimal household behavior changes responds to the persistence of the shocks as well. The more sensitive is the saving policy to ε , the larger the wealth movements will be across periods, which in turn implies more rapid resorting. Here we state a proposition about the relationship between ρ and the saving policy function $g_a(a, \varepsilon)$ when the wealth distribution is fixed.

Proposition 1 *Consider two households, A and B, from the same steady state wealth distribution, and without loss of generality, let $\rho_A > \rho_B$. For $a > \underline{a}$, the distance between saving functions across productivity draws is larger for the household with a higher probability of switching productivity states,*

$$(i.e., |g_a^B(a, \varepsilon_2) - g_a^B(a, \varepsilon_1)| > |g_a^A(a, \varepsilon_2) - g_a^A(a, \varepsilon_1)|).$$

Proof. Consider two households in the same wealth distribution Denote by π_{ij} the conditional probability that $\varepsilon' = \varepsilon_j$ given $\varepsilon = \varepsilon_i$. The corresponding conditional probability that $\varepsilon' = \varepsilon_{-j}$ is $1 - \pi_j$. Because $\rho^A > \rho^B$, $\pi_{11}^A > \pi_{11}^B$, and $\pi_{21}^B > \pi_{21}^A$.

We will show that $g_a^B(a, \varepsilon_1) < g_a^A(a, \varepsilon_1) < g_a^A(a, \varepsilon_2) < g_a^B(a, \varepsilon_2)$. It follows from the conditions on u and on the compactness of the budget set that $g_a^i(a, \varepsilon)$ is strictly increasing both arguments, so the inner most inequality is immediate. Next we will prove that $g_a^B(a, \varepsilon_1) < g_a^A(a, \varepsilon_1)$.

¹⁰Under some measures, the narrowness of the first quintile can lead to 'spurious' mobility because households will very frequently transition between the first and second quintiles despite almost no change in wealth.

Assume not so $g_a^A(a, \varepsilon_1) \leq g_a^B(a, \varepsilon_1)$. Then by the budget constraint $c^B \leq c^A$, where c^i is consumption of household i . By the strict concavity of u ,

$$u'(c^A) \leq u'(c^B)$$

which from the Euler equation implies

$$\pi_{11}^A V_1^A(g_a^A(a, \varepsilon_1), \varepsilon_1) + (1 - \pi_{11}^A) V_1^A(g_a^A(a, \varepsilon_1), \varepsilon_2) \leq \pi_{11}^B V_1^B(g_a^B(a, \varepsilon_1), \varepsilon_1) + (1 - \pi_{11}^B) V_1^B(g_a^B(a, \varepsilon_1), \varepsilon_2)$$

where V_1 is the derivative of V with respect to wealth.

We can use Theorem 6.8 from Acemoglu (2009) to establish that V is strictly concave in a .

The strict concavity of V in a leads to a contradiction since

$$\begin{aligned} V_1^A(g_a^A(a, \varepsilon_1), \varepsilon_1) &< \pi_{11}^A V_1^A(g_a^A(a, \varepsilon_1), \varepsilon_1) + (1 - \pi_{11}^A) V_1^A(g_a^A(a, \varepsilon_1), \varepsilon_2) \\ &\leq \pi_{11}^B V_1^B(g_a^B(a, \varepsilon_1), \varepsilon_1) + (1 - \pi_{11}^B) V_1^B(g_a^B(a, \varepsilon_1), \varepsilon_2) \\ &< V_1^B(g_a^B(a, \varepsilon_1), \varepsilon_1) \end{aligned}$$

which implies

$$g_a^A(a, \varepsilon_1) > g_a^B(a, \varepsilon_1).$$

Finally, we will show that $g_a^A(a, \varepsilon_2) < g_a^B(a, \varepsilon_2)$. Once again, assume not. Then

$$\begin{aligned} g_a^B(a, \varepsilon_2) &\leq g_a^A(a, \varepsilon_2) \\ u'(c^B) &\leq u'(c^A) \end{aligned}$$

$$\begin{aligned} \pi_{21}^B V_1^B(g_a^B(a, \varepsilon_2), \varepsilon_1) + (1 - \pi_{21}^B) V_1^B(g_a^B(a, \varepsilon_2), \varepsilon_2) &\leq \pi_{21}^A V_1^A(g_a^A(a, \varepsilon_2), \varepsilon_1) + (1 - \pi_{21}^A) V_1^A(g_a^A(a, \varepsilon_2), \varepsilon_2) \\ V_1^B(g_a^B(a, \varepsilon_2), \varepsilon_1) &< V_1^A(g_a^A(a, \varepsilon_2), \varepsilon_2) \end{aligned}$$

Again by strict concavity of V in a ,

$$g_a^B(a, \varepsilon_2) > g_a^A(a, \varepsilon_2)$$

which is a contradiction. ■

Intuitively, Proposition 1 is the permanent income hypothesis. If household A and household B have the same assets today and each draws the good shock, but A believes that its shock comes from a more persistent process than B does, then A 's consumption will be more responsive and so A 's saving will move less than B 's will. The consequence is that, all else equal, mobility due to behavior should decrease as ρ increases.

Luck Finally, a household’s mobility will be affected by the particular sequence of productivity draws. Within a given measurement window, if a household, beginning from a low wealth level, happens by chance to get higher productivity than would be expected, then that household will have high wealth mobility. The effect on mobility of more persistence in good and bad luck is not monotone. Generally, mobility will be low when persistence is either very low or very high. At very high ρ , households that start with good fortune will tend to continue having good productivity, increasing their saving and moving further away from other less fortunate households. At very low ρ , mobility is low because households switch too frequently. If the household starts in a low quintile and receives a good shock, it saves and moves up a bit in the wealth ordering, but in order to move even further up and transition through multiple quintiles over time, the household needs to get *a string* of positive shocks that is well above average. The probability of getting such a string however increases in ρ . The result is that for low ρ households tend to move around only a small region of their initial wealth position. Luck will tend to push up mobility if ρ lies in some intermediate range. In that region, households will tend to get sufficiently long strings of positive shocks to transition across quintiles, but switch between states frequently enough to support mixing.

Total mobility With these three factors in mind, the inverted U -shape of mobility over ρ can now be understood more easily. As ρ increases, agents experience longer sequences of above (below) average productivity, leading to longer strings of saving (dissaving) and a wider distribution of wealth. The expansion of the distribution should reduce mobility since it increases the distance between quintile boundaries (with the possible exception of the one between the first and second quintiles). More autocorrelated shocks should increase mobility since it allows households to experience longer strings of movement in the same direction, whether up or down; however, this effect is somewhat offset by the reduction in the sensitivity of savings to the shocks. While at higher ρ , households move in the same direction longer, they in smaller steps.

The above proposition explains the hump-shape in mobility. At low ρ , a move from state (k, ε_1) to (k, ε_2) induces a large change in k' . In itself, this would increase mobility, but because ρ is low, the probability of returning to the lower $g_k(k, \varepsilon_1)$ rule is high. Thus, it is likely that such a household will not experience a long enough string of high productivities to accumulate a

lot more wealth and move up into other quintiles. By a similar logic, a household that just drew ε_1 after having been ε_2 is unlikely to move down quintiles. On average households in a low- ρ environment, are very unlikely to move far away from their initial wealth level, k , though they will move very frequently within a small neighborhood of k .

As ρ increases, the distance of between savings functions does not fall much but the likelihood of experiencing a long string of consecutive ε_2 productivities rises. This allows households to move greater distances within the wealth distribution over a fixed amount of time. At some point however, ρ becomes so large that households switch productivities very infrequently, and the distance between savings rules gets very small. A household that starts on the savings path implied by $g(k, \varepsilon_2)$ is likelihood to continue building up wealth for a long time but very slowly so that it takes many periods to transition between quintiles. In our numerical experiments, we find a ρ near 0.7 returns the highest measure of mobility over quintiles.

Figure 2 plots these mobility measures as functions of ρ (again where σ is normalized). While the levels of the mobility measures differ, the orderings are very similar. For instance, the correlations are nearly 1 as shown in Table 2.

5.2.1 Ghost households

In order to isolate the effects of structure, behavior and luck to mobility, we introduce 'ghost' households into the computed steady state wealth distributions. A ghost is single, zero-measure agent that differs from the other households in the economy in some way. Because a ghost is atomistic, its presence does not alter either equilibrium prices or the quintile boundaries of the wealth distribution. By changing the ghost's environment, policy rules, or labor productivity we can control for each of the other factors. In the first step toward constructing our decomposition, we introduce ghosts with different labor income processes into each of the steady wealth distributions found in the baseline. For exposition, we will draw a distinction between the ρ value of the process faced by normal households (that is, the value which gave rise that particular wealth distribution) and the ρ value of the ghost. Denote the first, ρ_{GE} , and the second, ρ_G . We then simulate and construct a 5×5 mobility matrix for each ghost. We will perform this exercise for two types of ghost households. The first ghosts understand that their process has a different autocorrelation than that of the other households around them. As a result, their

saving decision rules will differ from those of the standard households in the economy, as will the realization of their productivity shocks. The second type of ghosts, believe that they have the same process as the standard households but experience the productivity sequence of a household of with a different ρ . These households do not have different savings rules, only different shock realizations.

Informed ghosts The informed ghost understands the true value of its ρ . It takes prices as given and solves the household problem. The ghost differs from the standard households in its economy in both how it responds to shocks conditional on current wealth and the shock sequence it faces. We calculate the ghost's mobility matrix under the wealth distribution generated by $\rho_{GE} \neq \rho_G$ and compare it to the mobility matrix generated by the $\rho_{GE} = \rho_G$ economy and attribute the differences to structure. Figures 4-7 plot contours of the surface generated by the (ρ_{GE}, ρ_G) pairs. The 45 degree line running through the contour is general equilibrium mobility measures from our baseline experiments. Starting at a point on that line, mobility declines as we move along.

On Figure 4, we draw an example of the structural vs. exchange mobility calculation. Comparing mobility at point A to mobility at point B, our method first picks out point C where ρ_{GE} is the same as in B but ρ_G is equal to the persistence in A. Any differences in mobility between C and A must come from facing a different distribution of wealth (i.e., structure). Movement from A to C then is 'structure' and movement from C to B is 'exchange'.

Figure 8 plots the savings decision rules of three households with different when the economy-wide ρ is 0.73. First notice Proposition 1 at play. Ghosts with low ρ have savings decisions that are much more distant across ε realizations, while those with ε near 1 have policy rules near the 45 degree line. Agents with $\rho = 0.05$ will experience relative large and frequent changes in wealth across one period, while those with $\rho = 0.98$ will switch infrequently but their wealth will also change very little each period. Importantly, notice that the change in distance from $\rho = 0.05$ to $\rho = 0.73$ is much smaller than it is from $\rho = 0.73$ to $\rho = 0.98$. This is a key factor for the hump-shape in total mobility. Depending upon the measure used, the trade off between persistent shocks and smaller step sizes reaches maximum mobility value somewhere between $\rho = 0.7$ and $\rho = 0.8$. For values below 0.7, mobility is reduced because agents are switching from savers

to dissave too frequently. For values above 0.8, households are accumulating (decumulating) wealth too slowly.

We find an analogy to driving helpful for explaining how mobility works in this model. Think of the support of wealth as a highway that runs east and west. Take any location on that highway and call all locations to the west of it 'poorer' and all locations to the east 'richer'. 'Checkpoints' along the highway correspond to quintiles of wealth (also called 'class boundaries'). Household decision rules are lanes on a highway. Some lanes move east (toward higher wealth) and others move west (toward lower wealth); and the fastest lanes are on the outside of the highway. The fastest westbound lane corresponds to the lowest labor income value, and the fastest eastbound lane to the highest value. The further the saving decision is from the 45 degree line, the faster it moves. Changing ρ alters how likely one is to switch out of their current lane and into another one. In the two ε case, there is only one westbound and one eastbound lane. If ρ is high, then a household will likely stay in its lane continuing to move up or down in the wealth ordering. As Figure 8 shows however, the more persistent the Markov process the closer the decision rules are to the 45 degree line and so the more slowly will be the pace of the lane in our analogy. If ρ is low the lane speeds will be faster, but the households will switch directions frequently, moving up and the moving down the ordering. Maximum wealth mobility is achieved where lanes move quickly enough to allow for distant movement, but also where they are likely not to switch too often, allowing for a sufficiently long chain of movements in the same direction.

Uninformed ghosts To decompose exchange mobility from between behavior and luck, we run the same type of experiment as above, but now the ghost does not realize that its labor productivity process has a different autocorrelation. This ghost uses the same decision rules as the other households in the distribution, but it realizes a different sequence of shocks. Figure 9-12 plot mobility of these agents as a function of (ρ_G, ρ_{GE}) . As before with the informed ghosts, we draw path to highlight one of the three components, here being luck. We have a similar breakdown on figure 9. Moving from A to B is a combination of all three components, but movement between A and C is entirely due to luck because the ghosts in both cases reside in the same distribution and have the same decision rules. The only difference is that a ghost at C has a more persistent shock process (identical to the ghost at B).

The differences in the measures are also notable. The Bartholomew and Shorrocks measures show mobility increasing as the ghost's persistence parameter increases. For the mean first passage measure, the relationship has a similar hump-shaped pattern. Holding ρ_{GE} constant, mobility increases in ρ_G until it reaches a maximum somewhere between 0.70 and 0.80; then it declines rapidly. Oddly, the 2nd largest eigenvalue measure actually decreases in ρ_G .

Here we see the hump-shaped pattern in mobility. When the economy-wide ρ is low, the savings rules are far apart so non-phantom agents in a fast lane but change often. Mobility is low. The non-optimizing phantom agents with higher ρ share the same fast lanes but are much less likely to switch. They have longer chains of wealth accumulations and decumulations, and so their mobility is higher. One again, when ρ gets too high, the ghost agents remain in their lane for a very long time. They will move through the distribution but only very infrequently, and they will usually just 'pass through' one intermediate quintile. Those with low ε will spend a large number of periods in the bottom quintile before finally drawing a good shock and making a transition back through the distribution toward the top quintile where they will once again remain for a large number of periods.

5.2.2 Decomposing changes in mobility

We have identified three sources for the differences in mobility as the labor income process becomes more persistent. In order to disentangle the contributions of each source to the total change in steady state mobility, we will run several counterfactual experiments. Consider the steady states of two economies, one with $\rho = \rho_x$ and one with $\rho = \rho_y$; and without loss of generality, let $\rho_y > \rho_x$. Denote by $\mu_{[j,j,j]}$, the measured mobility induced by an agent acting in a distribution produced by agents with $\rho = \rho_j$, having optimal policy rules consistent with $\rho = \rho_j$, and experiencing a realized sequence of labor productivity shocks generated according to $\rho = \rho_j$. For ease of exposition, let $\mu_{[J,J,J]} = \mu_J$. Finally, let $\Delta\mu_{xy} = \mu_y - \mu_x$. $\Delta\mu_{xy}$ is the total change in mobility between the economy with a labor income persistence of ρ_x and ρ_y .

We decompose $\Delta\mu_{xy}$ in the following manner:

$$\begin{aligned}\Delta\mu_{xy} &= \Delta\textit{structure} + \Delta\textit{behavior} + \Delta\textit{luck} \\ &= (\mu_y - \mu_{[x,y,y]}) + (\mu_{[x,y,y]} - \mu_{[x,x,y]}) + (\mu_{[x,x,y]} - \mu_x).\end{aligned}$$

Each component removes one conflating factor in the relative mobility difference, starting with the structures of the ρ_x and ρ_y distributions, moving to differences in the decision rules (behavior) of the agents, and finishing with differences in the realized sequence of productivity shocks.

Figure 13 decomposes the total change in mobility as ρ rises into these three components. Across all four measures, the decomposition is qualitatively the same. Structure has a small negative effect on mobility, while behavior and luck make larger contributions, negative and positive, respectively. At low levels of ρ , mobility rises in the shock persistence because luck offsets behavior. Past a certain point, however, behavior becomes more powerful and pulls total mobility down.

5.2.3 Borrowing Limits and Elastic Labor

So far we have imposed a strict borrowing limit of zero. A large fraction of households can find this constraint binding, particularly when the labor income shocks are very persistent. As a result, the steady state wealth level separating the first and second quintiles can be very close to 0 so that even a small movement away from the borrowing limit can move a household into the second quintile. In this case, households in the first (second) quintile would appear to be very upwardly (downwardly) mobile. We have run cases with high persistence and exogenous borrowing limits near the natural borrowing limit and found that while it has little effect on our mobility measures. Therefore, we do not think that our assumption of no borrowing is restricting our findings.

In Carroll et al. (2016) we study an economy with elastic labor and fiscal policy and find that the quantitative problems carry over to that environment. For that reason we do not further explore adding elastic labor here.

6 Mobility and other features

6.1 Increased skewness in the wealth distribution

It is well-known that a Bewley model with idiosyncratic labor income risk alone does a poor job matching the high concentration of wealth in the right tail.¹¹ The fundamental issue is that

¹¹See Quadrini and Ríos-Rull (1997) and Carroll (1998) for discussion.

the sufficient amount of wealth to self-insure is low when agents are very patient and shocks are relatively small. Once a household can adequately smooth its consumption, it has no other incentive to continue saving, since interest rates are necessarily lower than time rates of preference. Several approaches have been used to generate longer right tails in the wealth distribution.¹²

Krusell and Smith (1998) replace the scalar household discount factor with a 3-state, highly persistent Markov chain. The three values are $[0.9763, 0.9812, 0.9861]$, and the transition matrix is

$$\begin{bmatrix} 0.99654 & 0.00346 & 0 \\ 0.00043 & 0.999135 & 0.00043 \\ 0 & 0.00346 & 0.99654 \end{bmatrix};$$

these choices deliver a Gini coefficient of wealth equal to 0.78. The invariant distribution of β is $[0.1, 0.8, 0.1]$ and the average duration in either extreme- β state is 200 quarters. The 5-year wealth mobility matrix for the stochastic- β environment is

$$\begin{bmatrix} 0.84 & 0.16 & 0.00 & 0.00 & 0.00 \\ 0.16 & 0.70 & 0.15 & 0.0 & 0.00 \\ 0.00 & 0.15 & 0.73 & 0.11 & 0.00 \\ 0.00 & 0.00 & 0.11 & 0.84 & 0.05 \\ 0.00 & 0.00 & 0.00 & 0.05 & 0.95 \end{bmatrix}$$

The stochastic- β model makes the mobility match worse – the top quintiles get even more persistent, since drawing a high discount factor leads even agents with temporarily low income to save, and discount factor shocks are very persistent. It is this immobility that delivers the high wealth concentration that was the goal of Krusell and Smith (1998), but it does not come for free.¹³

Castaneda et al. (2003) add a very high productivity state with relatively low persistence and a high probability of transitioning immediately to the lowest productivity. The transitory nature of this 'rockstar' state combined with the increased risk motivates households in this state to build up a substantial amount of precautionary savings. When a household draws the rockstar state, it takes advantage of its temporary good fortune by saving rapidly. This 'burst of saving'

¹²DeNardi et al. (2016) investigates the effect of introducing highly-leptokurtic income processes into the Bewley model; Guvenen et al. (2016) provides empirical support for such processes.

¹³Carroll (2001) shows that a permanent 'two- β ' model looks very much like the stochastic- β model, so the fact that the discount factors mean-revert does not seem important provided they do so slowly.

produces the matrix below which has considerably more upward mobility than the benchmark:

$$\begin{bmatrix} 0.73 & 0.17 & 0.05 & 0.04 & 0.00 \\ 0.24 & 0.52 & 0.19 & 0.05 & 0.00 \\ 0.00 & 0.34 & 0.52 & 0.14 & 0.00 \\ 0.00 & 0.00 & 0.24 & 0.59 & 0.17 \\ 0.00 & 0.00 & 0.00 & 0.17 & 0.83 \end{bmatrix} .$$

Nevertheless, the rockstar model still has too little downward mobility. The consumption-smoothing motive implies that while households save rapidly, they dissave slowly – staying away from the borrowing constraint is the reason they save, after all. Shocks which target wealth directly (such as medical expenditures, divorces, or business failures) could produce these large drops in wealth in principal; shocks to labor earnings, which are background risks, do not seem likely to work however.

Benhabib et al. (2015) use in a partial equilibrium OLG model with deterministic, heterogeneous earnings profiles and rates of return on saving to match aspects of inequality and mobility in the US wealth distribution (see also Hubmer et al. (2015)). They argue that three factors are critical for modeling wealth inequality and wealth mobility: stochastic earnings, capital income risk, and differential saving and bequest motives. We have already addressed the first and third features in our model.¹⁴ The extent to which *ex post* return heterogeneity can generate the high degree of wealth mobility in this model seems limited. The increase in income risk from stochastic returns would induce more precautionary saving, but not necessarily more rapid accumulation and decumulation of wealth.

7 Mobility and Market Incompleteness

We know that market incompleteness is a necessary condition for permanent mobility – mobility may be present along a transition path if agents have different preferences, but eventually it will disappear as the economy transitions to a steady state (see Caselli and Ventura (2000) and Carroll

¹⁴One can interpret households in our model as infinitely-lived dynasties in which the parent internalizes the child’s welfare with perfect altruism, so that discount factor shocks look like shocks to bequest motives. One complication would be how to handle the inter vivos transfers between family members; we defer this question to future work.

and Young (2011)). We now take up the question of how mobility is connected to incompleteness, in the sense of the spanning of assets. We consider a simple experiment – suppose there exists two assets, one of which pays off if $\varepsilon \geq E[\varepsilon]$ and one that pays off if $\varepsilon < E[\varepsilon]$. Asset markets are ‘more complete’, but mobility could easily go either way. Since the price of these assets is smaller than the price of a risk-free security, portfolios that ‘lever up’ in certain states can lead to large changes in wealth should those states realize; the results in Rampini and Viswanathan (2016) show that agents in our economy will in fact choose to endogenously hold a skewed portfolio if they are sufficiently poor.¹⁵

We compare the mobility results from this partial insurance case to one the baseline process. In each case, we set the number of productivity states to 7. For simplicity, denote the productivity states where $\varepsilon \geq E[\varepsilon]$ ‘good’ states, and the other ‘bad’ states.

Figure 14 plots the portfolio decisions of several informed ghost households. In each case, the ρ value of the underlying economy is 0.73. Each subplot shows the decisions of two ghosts with the same persistence value, one with $\varepsilon = \varepsilon_{\min}$ and one with $\varepsilon = \varepsilon_{\max}$. The solid lines represent the number of claims purchased which pay off if the next period’s productivity belongs to the same state as today’s productivity. The dashed lines are the claims which pay off in the opposite state from today’s. For example, for the $\varepsilon = \varepsilon_{\min}$ household, the solid line is the stock of claims that pay off if one of the bad states is realized next period, and the dashed line is those that pay off if the good state is realized instead.

First notice that the a household currently in the bad (good) state purchases contingent claims against the bad (good) state near the 45 degree line. In fact, the household’s decision rules in this regard are similar in appearance to those in the one asset case. Just as in the baseline case, these saving rules become closer as the probability of remaining in the same state increases. Again, households consume a larger fraction of income from more persistent shocks. This feature of the portfolio induces more mobility as it allows for long strings of consistent wealth accumulation and decumulation, as we illustrated in the section above.

The other side of the portfolio, that is the holding of claims which pay off only if the household’s

¹⁵It is straightforward conceptually to permit an arbitrary number of state-contingent claims, but the high autocorrelation of the states means that some of these assets will have essentially zero price; prices that are too low lead to instability in our solution algorithm, so we content ourselves with a simple two-asset case. We are investigating the effects of state-contingent asset limits as in Mendoza, Quadrini, and Ríos-Rull (Mendoza et al.).

state switches (from good to bad or bad to good) in the next period, is quite different, and it can have a big effect on mobility, particularly in the ghost household cases. Households currently in a good productivity state purchase considerably more claims against switching to a bad productivity state. These claims compensate both for the low labor income from a bad state and provide additional precautionary savings the likely recurrence of bad state shocks. Moreover, because the probability of switching between good and bad states is low (especially for an ε_{\max} or ε_{\min} household), this insurance is very cheap. Naturally then, as ρ increases the good state household's claims against bad states rises, causing the balance of the portfolio to tilt more and more.

The portfolio of household's currently in a bad productivity depends on their wealth level. At sufficiently high wealth, the portfolio looks like a mirror image of the good state household's portfolio. The purchase of claims against a bad state lie close to the 45 degree line, while the purchases of claims against the good state are much lower. At lower levels of current wealth, households would like to short the claim against good states, since consumption in the bad states is very valuable. Since this shorting is not allowed, these households simply do not participate in that asset market. With the exception of the wealth region where the non-negativity constraint binds, the response of any household portfolios can be generalized in the following way: as ρ increases, the demand for claims that pay off if the current state continues become less sensitive to income shocks, while the demand for assets that payoff if the state switches becomes more sensitive.

The consequence of this portfolio behavior for mobility across ρ is that as households become less and less likely to switch states, their wealth path is characterized by small, gradual movements interspersed with infrequent large shifts. Figure 15 plots mobility in the partial insurance case against the single asset baseline. Notice that mobility is lower in the partial insurance environment unless the labor income process is quite persistent. Regardless of the type of measure, mobility under partial insurance peaks at a higher ρ and may even reach a higher (absolute) level before quickly descending again as shocks approach being permanent.

Although the partial insurance environment features more wealth mobility at high values of ρ , there is still less mobility than in data for our chosen value of ρ . The five-year wealth transition

matrix is

$$\begin{bmatrix} 0.75 & 0.25 & 0.00 & 0.00 & 0.00 \\ 0.24 & 0.56 & 0.19 & 0.05 & 0.00 \\ 0.01 & 0.19 & 0.66 & 0.14 & 0.00 \\ 0.00 & 0.00 & 0.15 & 0.77 & 0.08 \\ 0.00 & 0.00 & 0.00 & 0.08 & 0.92 \end{bmatrix}.$$

8 Conclusion

We have studied wealth mobility in a Bewley model. In particular, we have shown how assumptions about the underlying process driving long run wealth inequality affect relative mobility. As labor income shocks become more persistent, relative mobility displays a hump-shape, starting low growing monotonically to a maximum around $\rho = 0.75$ and then declining sharply towards 0 as the process becomes closer to permanent. Using 'ghost' households, we run several counterfactuals in order to decompose the pattern in mobility into the change in the structure of the wealth distribution, the change in optimal savings behavior in the face of different income risk, and changes in sequence of labor income itself (i.e., luck). We find that the hump-shape is generally attributable to the mixture of behavior and luck. The first contributes negatively to mobility as household's saving is less sensitive to more persistent shocks. The second contributes positively by generating longer strings of low or high income allowing wealth to accumulate or decline for longer over a fixed amount of time.

We document that the baseline Bewley model generates a stationary wealth distribution with lower short-run wealth mobility than has been found empirically. In the data, a non-trivial fraction of households experience large movements across wealth quintiles, even over fairly short horizons, while these movements do not occur in the model. We extend the baseline model in several ways commonly used in the literature to better match wealth inequality. While the inclusion of a very high income state with low persistence as in Castaneda et al. (2003) improves the model's predictions for upward mobility somewhat, it does not match the observed downward mobility. In all versions of the model studied, households move down in wealth too slowly, a natural result of the precautionary saving motive present in the incomplete markets model.

Finally, we examine the relationship between market completeness and wealth mobility. We

find that replacing the non-contingent capital asset with two state-contingent claims (i.e., partial insurance) may reduce or increase mobility depending upon the underlying persistence of the income shock process. If ρ is sufficiently high, the more complete markets economy has higher mobility. Nevertheless, the model still fails to quantitatively match the observed mobility.

References

- Acemoglu, D. (2009). *Introduction to modern economic growth*. Princeton University Press.
- Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics* 109(3), 659–684.
- Bartholomew, D. J. and D. J. Bartholomew (1967). *Stochastic models for social processes*. Wiley London.
- Benhabib, J., A. Bisin, and M. Luo (2015). Wealth distribution and social mobility in the us: a quantitative approach. Technical report, National Bureau of Economic Research.
- Bewley, T. F. (1986). Stationary monetary equilibrium with a continuum of independently fluctuating consumers. In W. Hildenbrand and A. Mas-Collel (Eds.), *Contributions to mathematical economics in honor of Gerard Debreu*, pp. 79–102. North Holland.
- Carroll, C. D. (1998). Why do the rich save so much? Technical report, National Bureau of Economic Research.
- Carroll, C. D. (2001). A theory of the consumption function, with and without liquidity constraints. *The Journal of Economic Perspectives* 15(3), 23–45.
- Carroll, D. R., J. Dolmas, and E. R. Young (2016). The politics of flat taxes. Technical report.
- Carroll, D. R. and E. R. Young (2011). The long run effects of changes in tax progressivity. *Journal of Economic Dynamics and Control* 35(9), 1451–1473.
- Caselli, F. and J. Ventura (2000). A representative consumer theory of distribution. *American Economic Review*, 909–926.

- Castaneda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003). Accounting for the us earnings and wealth inequality. *Journal of Political Economy* 111(4), 818–857.
- Conlisk, J. (1990). Monotone mobility matrices. *Journal of Mathematical Sociology* 15(3-4), 173–191.
- Dardanoni, V. (1993). Measuring social mobility. *Journal of Economic Theory* 61(2), 372–394.
- DeNardi, M., G. Fella, and G. Paz-Pardo (2016). The implications of richer earnings dynamics for consumption and wealth. Technical report, National Bureau of Economic Research.
- Díaz-Giménez, J., A. Glover, and J.-V. Ríos-Rull (2011). Income and wealth in the united states: 2007 update. *Quarterly Review* 34(1).
- Fields, G. S. and E. A. Ok (1999). The measurement of income mobility: an introduction to the literature. In *Handbook of Income Inequality Measurement*, pp. 557–598. Springer.
- Floden, M. and J. Lindé (2001). Idiosyncratic risk in the united states and sweden: is there a role for government insurance? *Review of Economic Dynamics* 4(2), 406–437.
- Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2016). What does data on millions of u.s. workers reveal about life-cycle earnings dynamics? Technical report.
- Hopenhayn, H. A. and E. C. Prescott (1992). Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica*, 1387–1406.
- Hubmer, J., P. Krusell, and A. Smith (2015). The historical evolution of the wealth distribution: a quantitative-theoretic investigation. In *2015 Meeting Papers*, Number 1406. Society for Economic Dynamics.
- Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control* 17(5), 953–969.
- Hurst, E., M. C. Luoh, F. P. Stafford, and W. G. Gale (1998). The wealth dynamics of american families, 1984-94. *Brookings Papers on Economic Activity* 1998(1), 267–337.
- Kemeny, J. G. and J. L. Snell (1976). Finite markov chains, undergraduate texts in mathematics.

- Kopecky, K. A. and R. M. Suen (2010). Finite state markov-chain approximations to highly persistent processes. *Review of Economic Dynamics* 13(3), 701–714.
- Krusell, P. and A. A. Smith, Jr (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* 106(5), 867–896.
- Mendoza, E., V. Quadrini, and J.-V. Ríos-Rull. Financial integration, financial development, and global imbalances. *Journal of Political Economy* 117(3), 371–416.
- Meyer, C. D. (1978). An alternative expression for the mean first passage matrix. *Linear algebra and its applications* 22, 41–47.
- Prais, S. (1955). Measuring social mobility. *Journal of the Royal Statistical Society. Series A (General)* 118(1), 56–66.
- Quadrini, V. and J.-V. Ríos-Rull (1997). Understanding the us distribution of wealth. *Quarterly Review Federal Reserve Bank of Minneapolis* 21(2), 22.
- Rampini, A. A. and S. Viswanathan (2016). Household risk management. Technical report, National Bureau of Economic Research.
- Ríos-Rull, J.-V. (1998). Computing equilibria in models with heterogenous agents. *Computational Methods for the Study of Dynamic Economics*, 238–264.
- Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica*, 1013–1024.
- Sommers, P. M. and J. Conlisk (1979). Eigenvalue immobility measures for markov chains 1. *Journal of Mathematical Sociology* 6(2), 253–276.

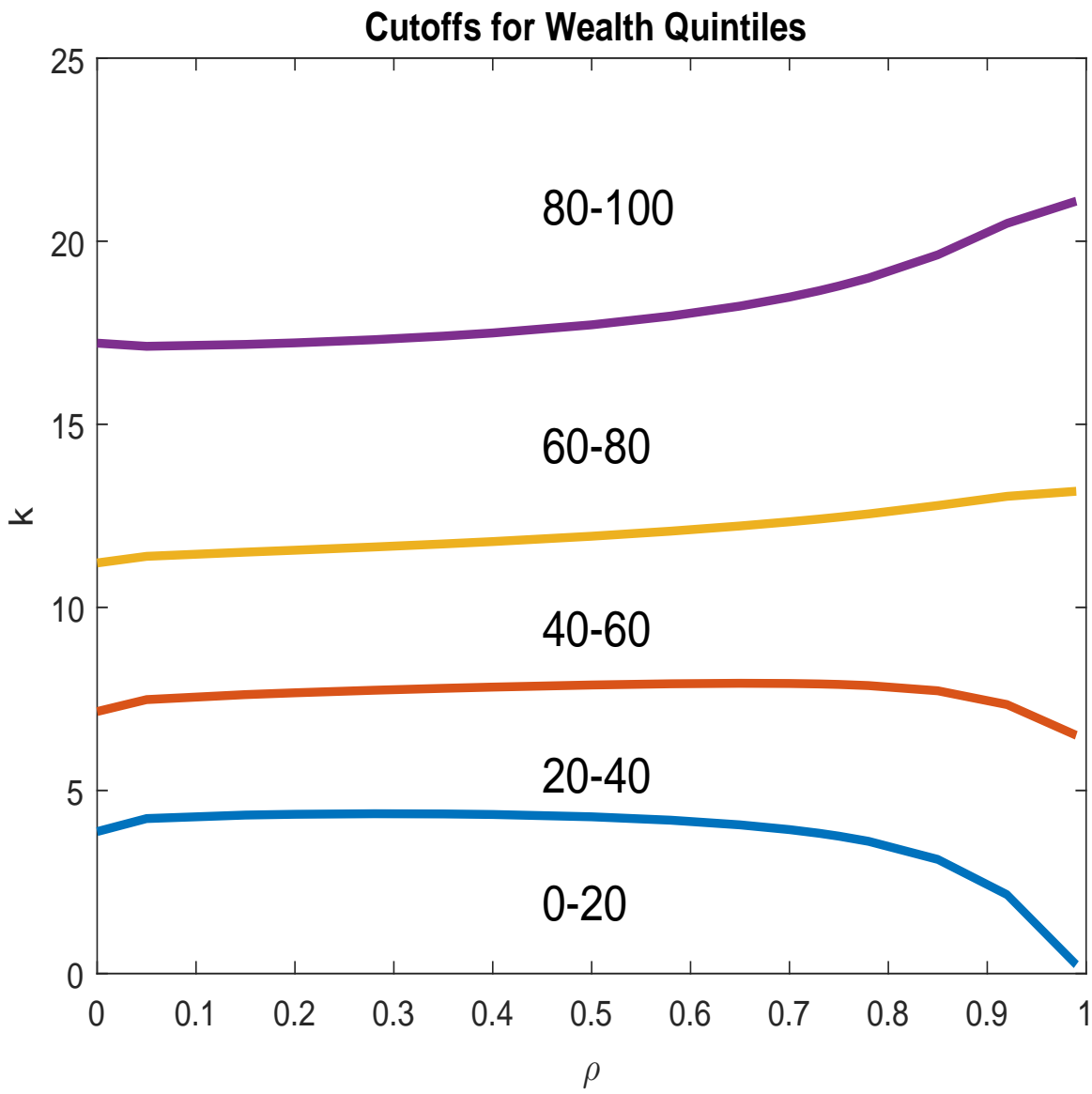


Figure 1: Cutoffs for wealth quintiles across persistence

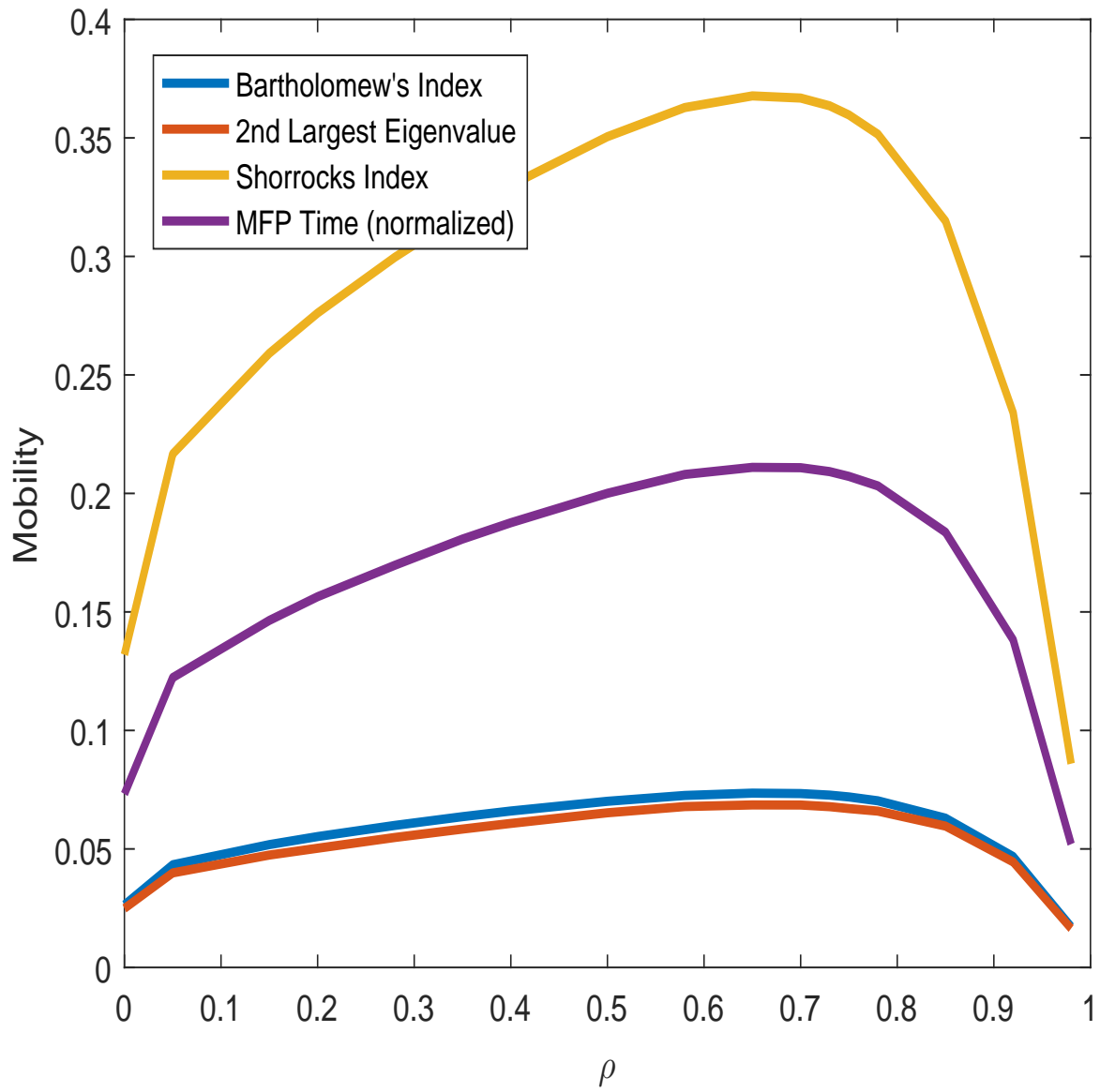


Figure 2: Mobility across persistence

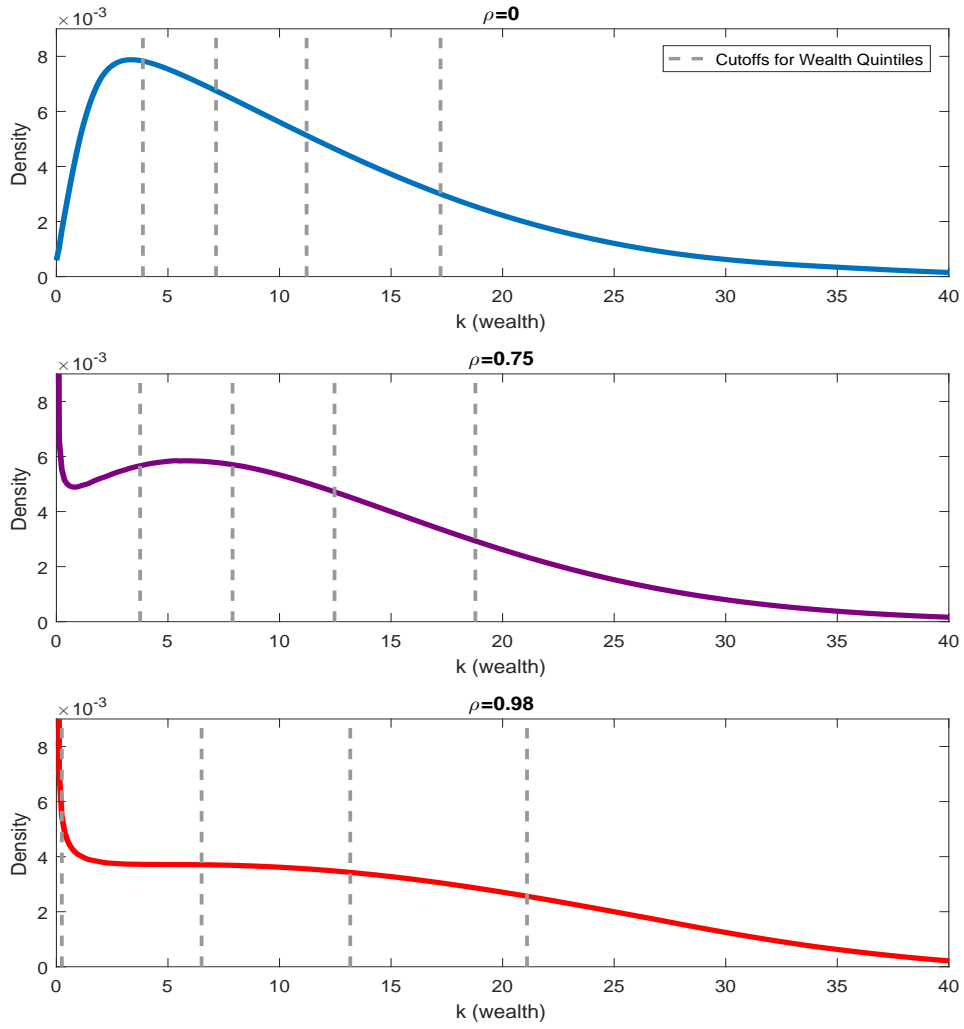


Figure 3: Boundaries between quintiles for different values of ρ

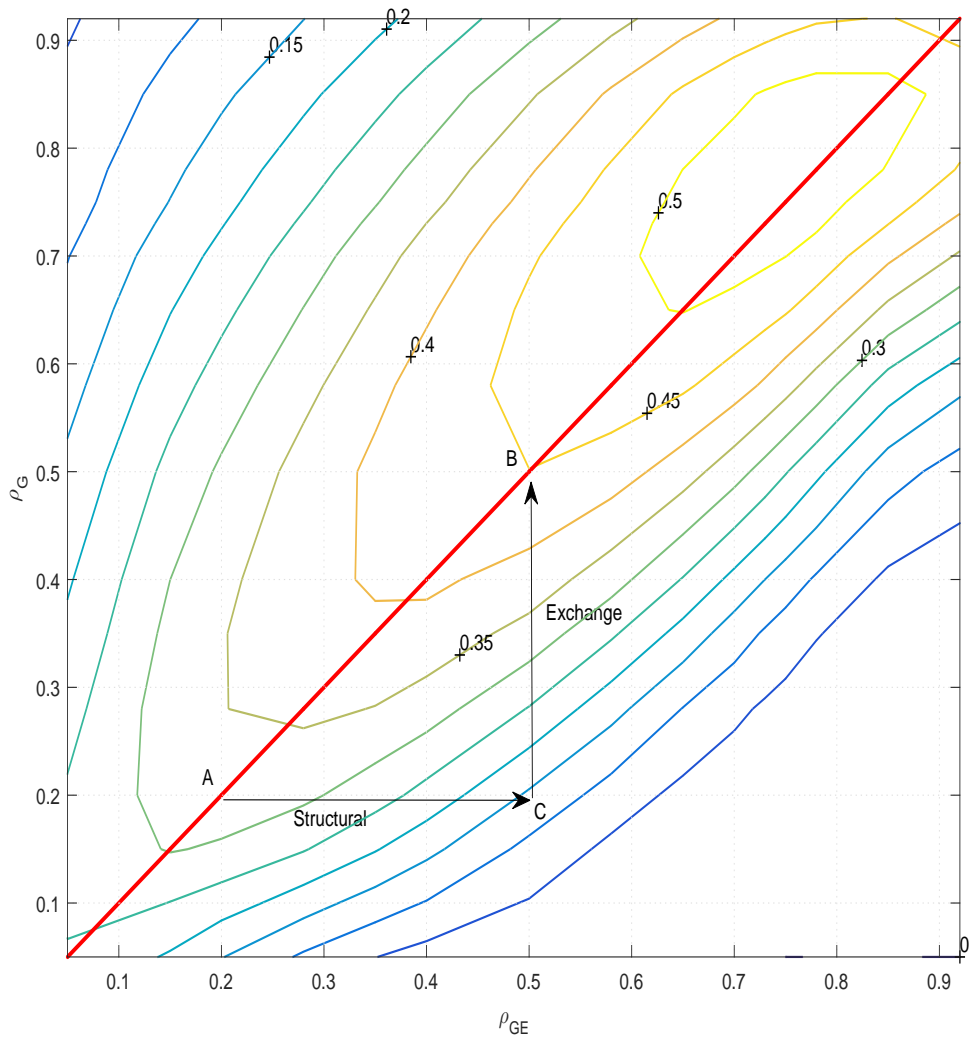


Figure 4: Mobility of optimizing ghost: μ_{MFP}

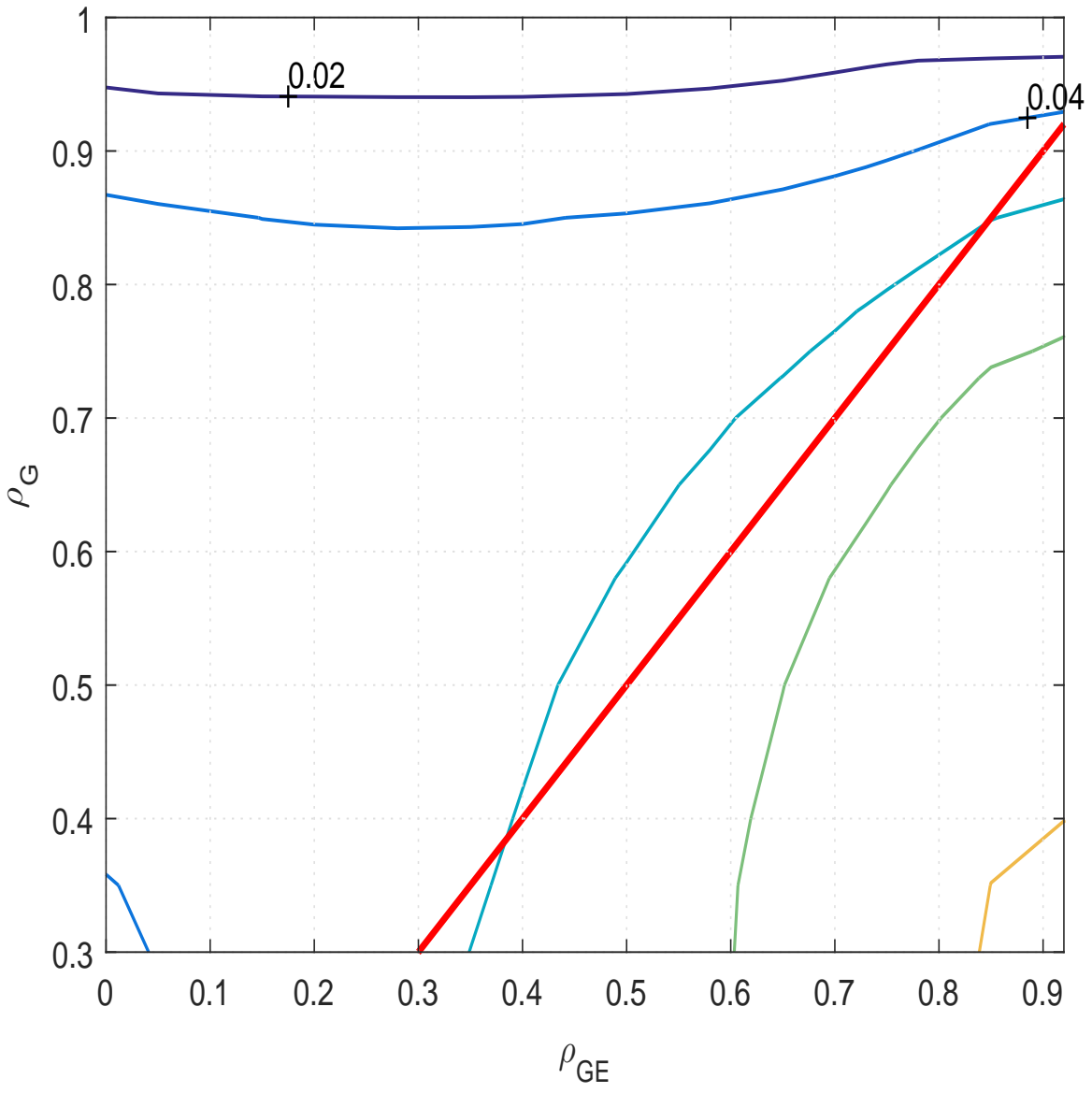


Figure 5: Mobility of optimizing ghost: μ_{2E}

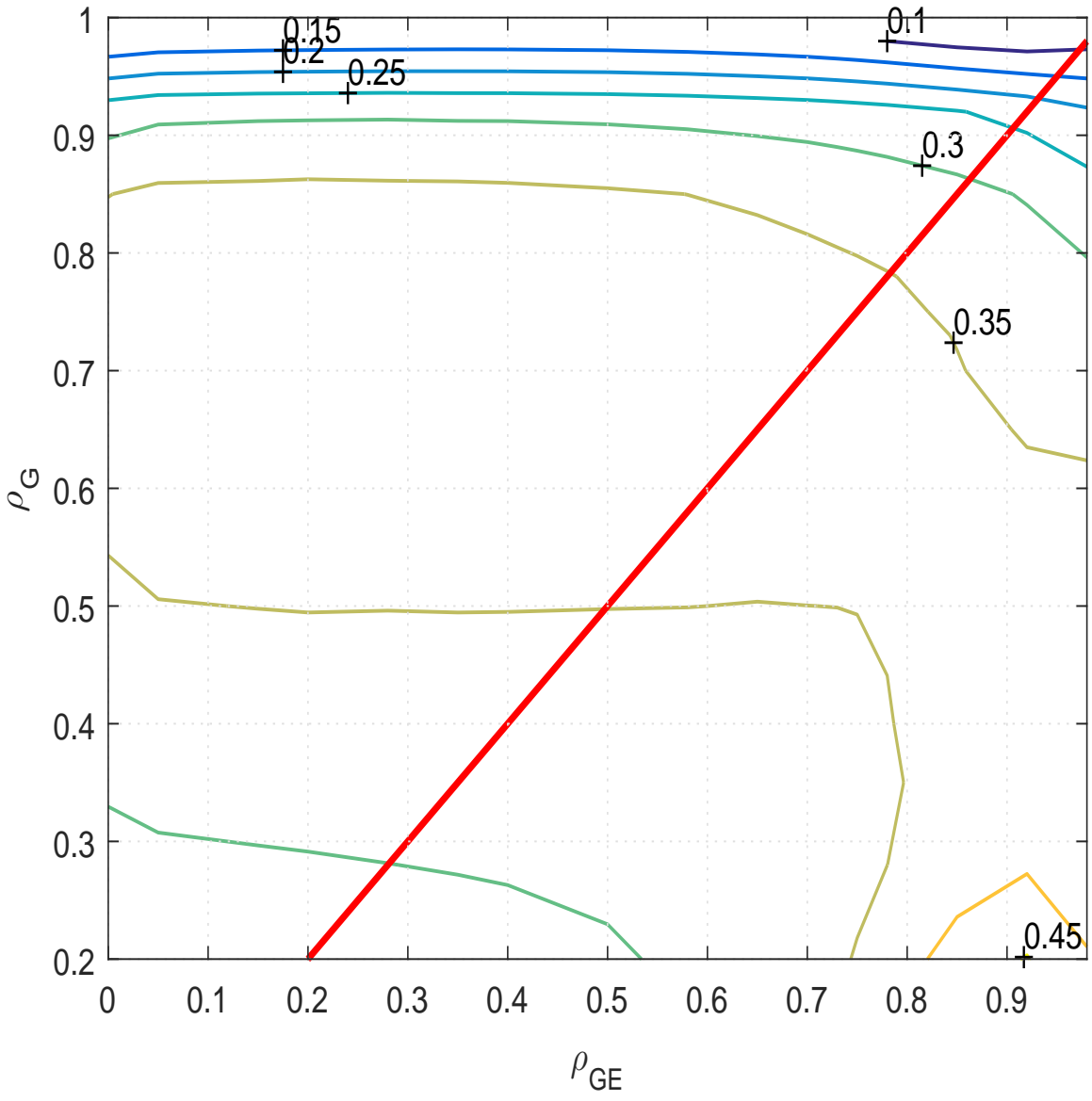


Figure 6: Mobility of optimizing ghost: μ_S

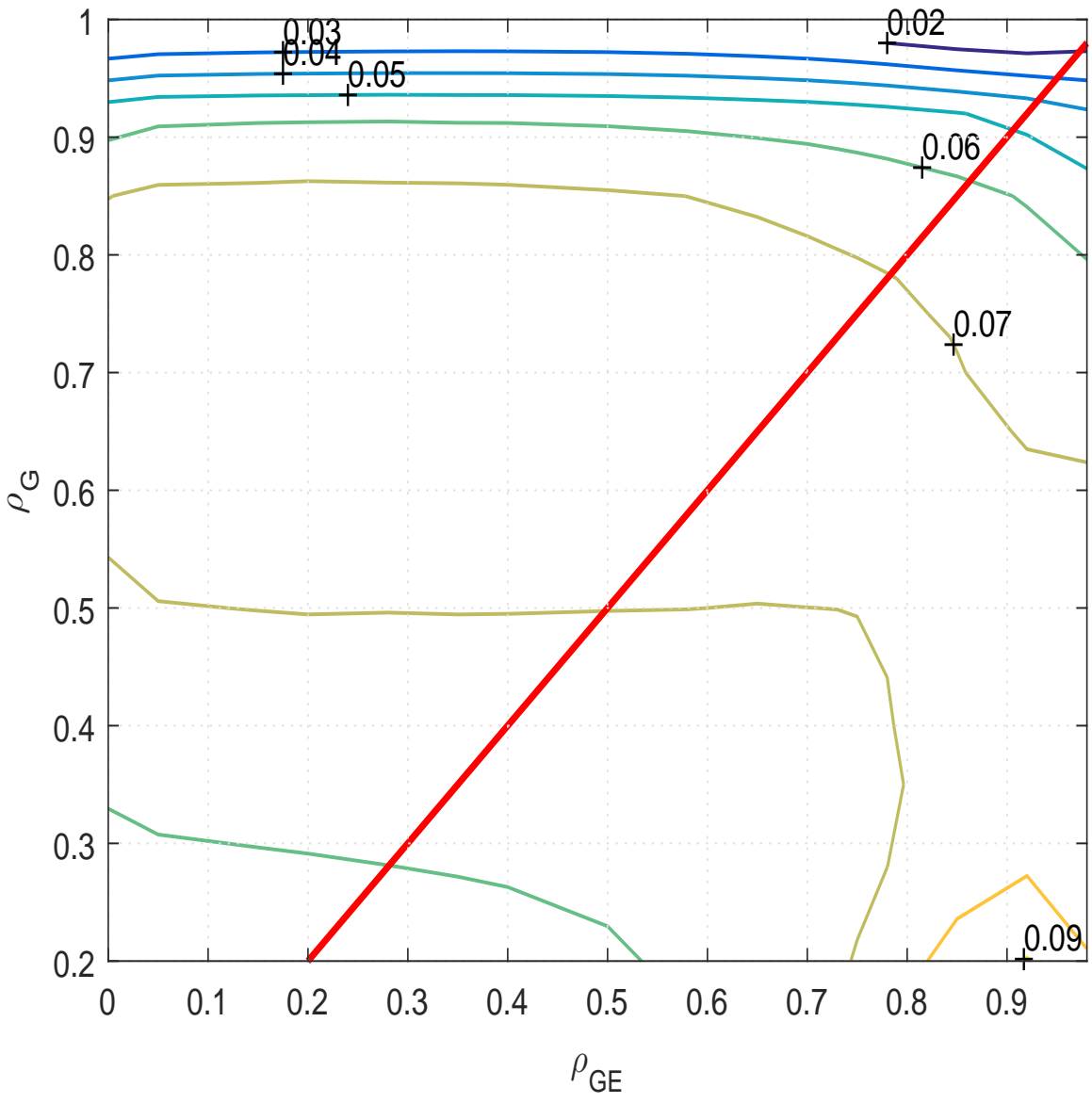


Figure 7: Mobility of optimizing ghost: μ_B

Savings Decisions Across Persistence, $\rho_{\text{economy}} = 0.73$

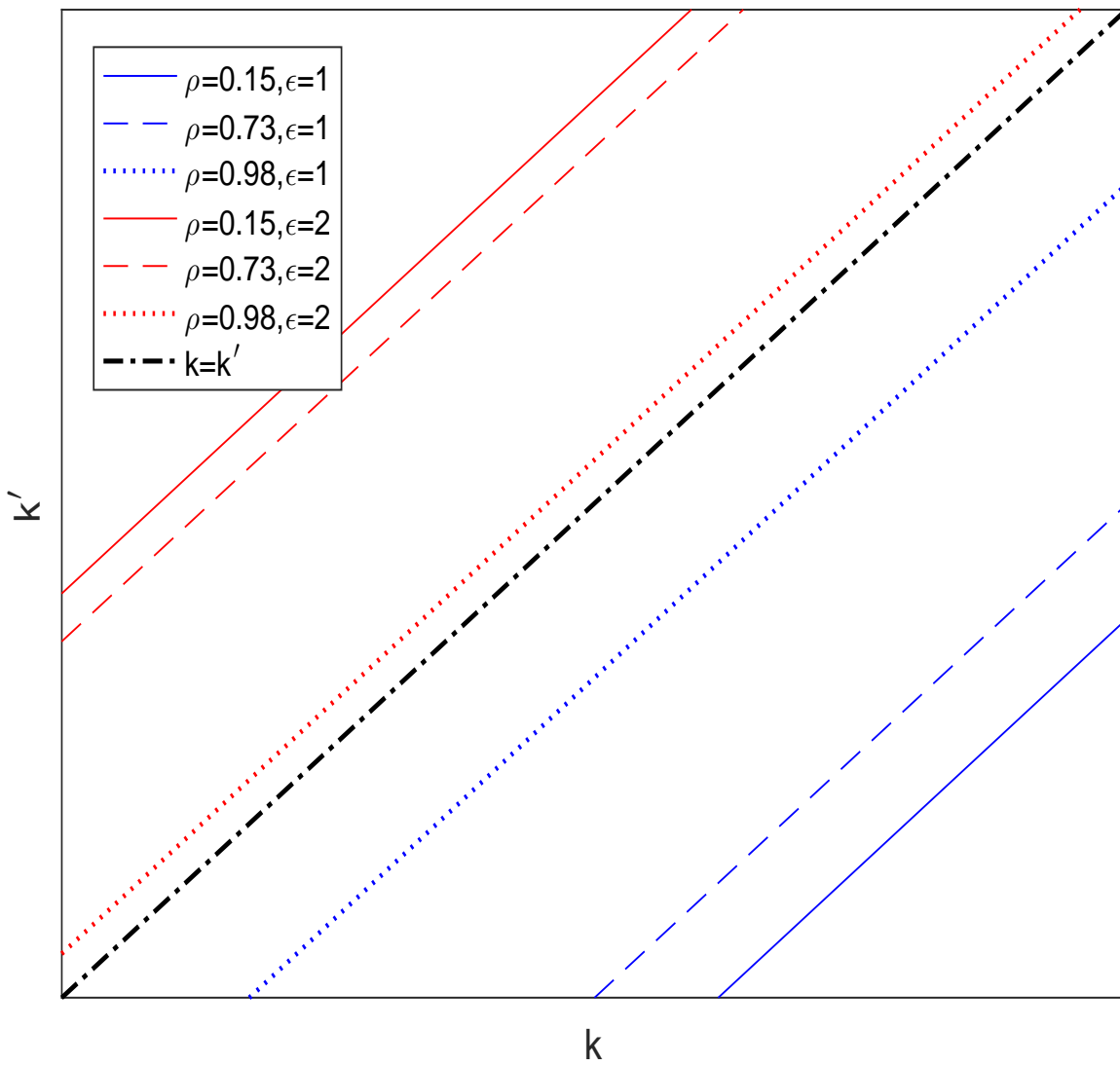


Figure 8: Saving Decisions across persistence

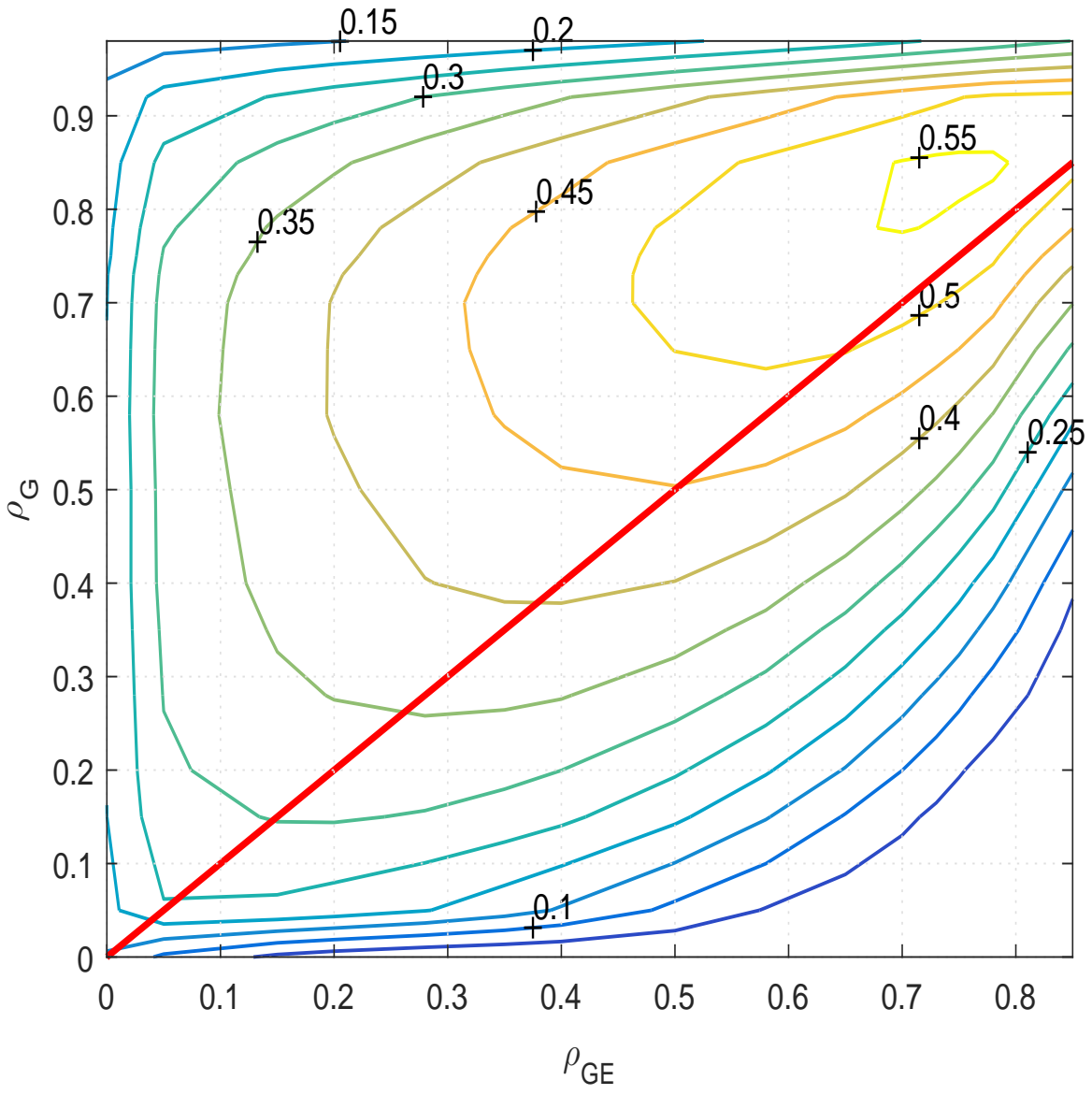


Figure 9: Mobility of non-optimizing ghost: μ_{MFP}

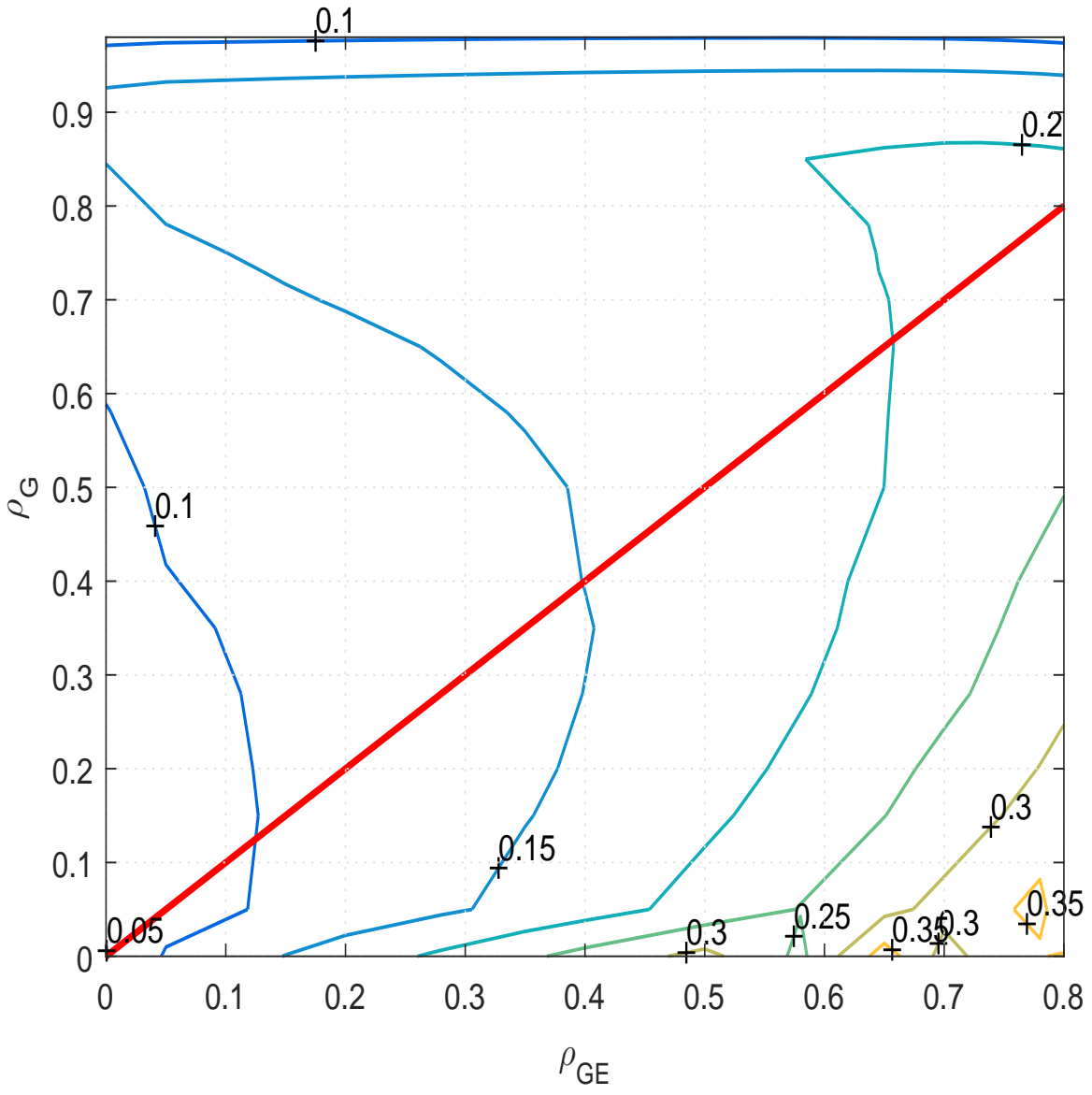


Figure 10: Mobility of non-optimizing ghost: μ_{2E}

Shorrocks Index: No Re-optimization

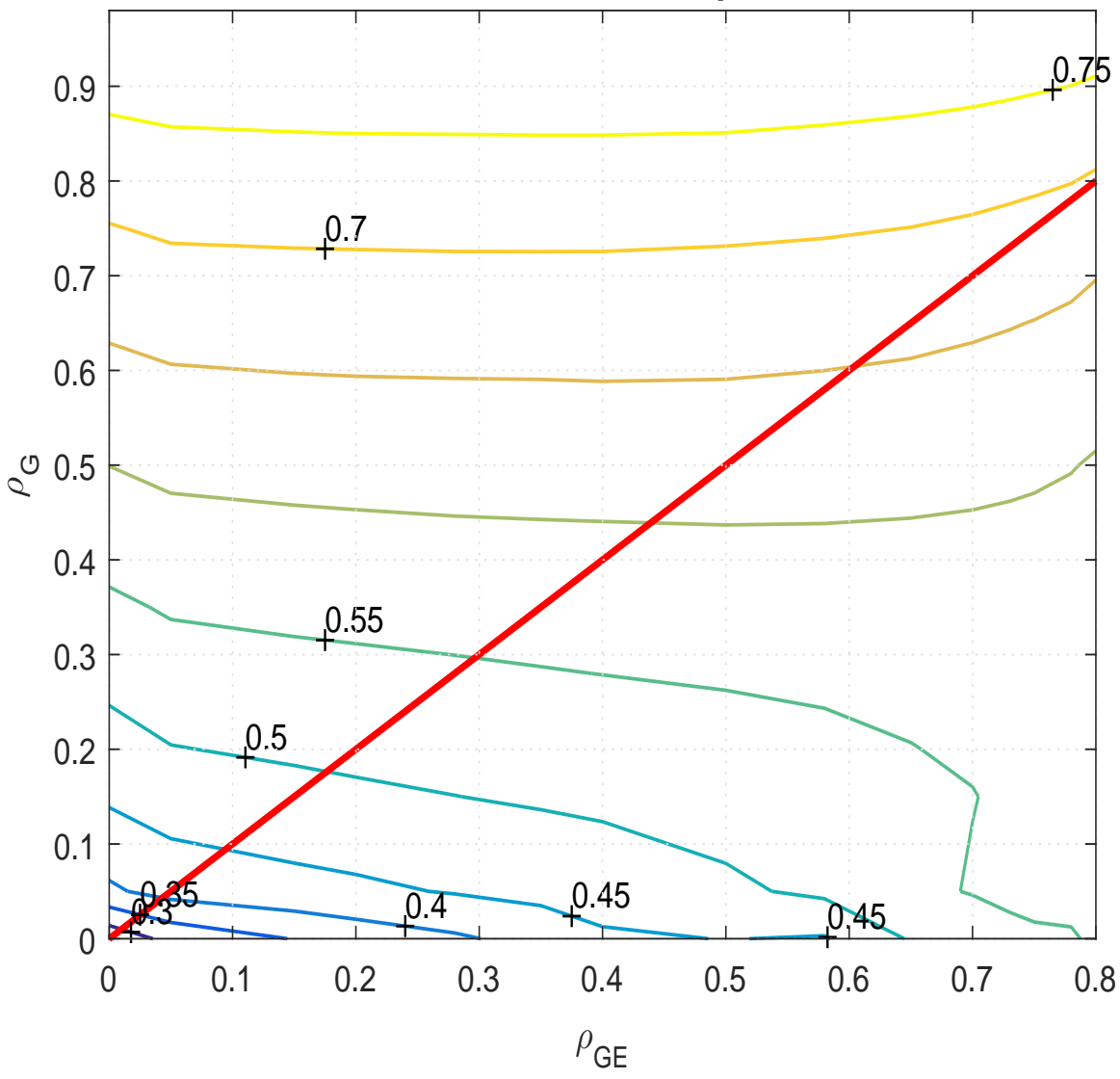


Figure 11: Mobility of non-optimizing ghost: μ_{sm}

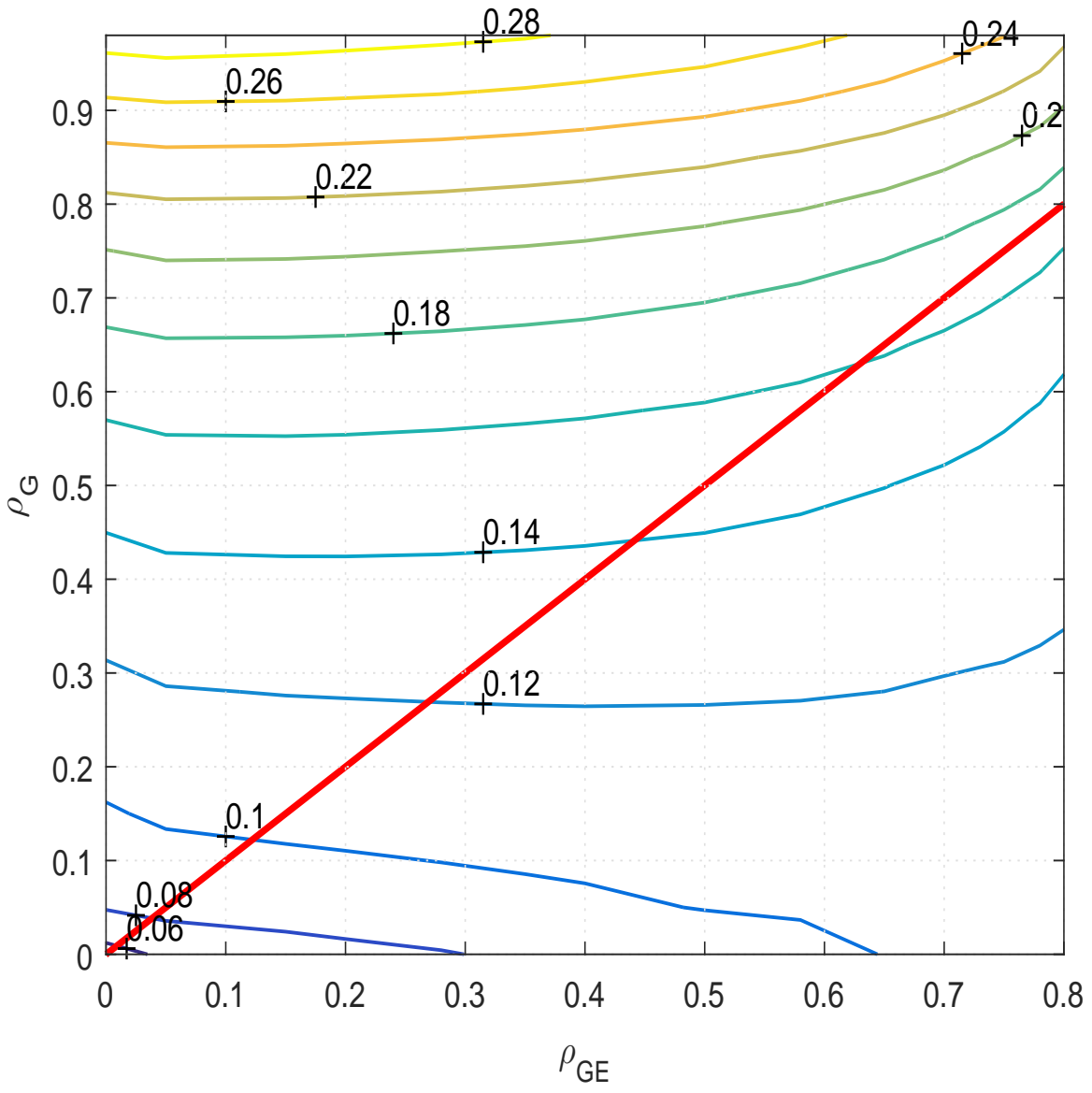


Figure 12: Mobility of non-optimizing ghost: μ_B

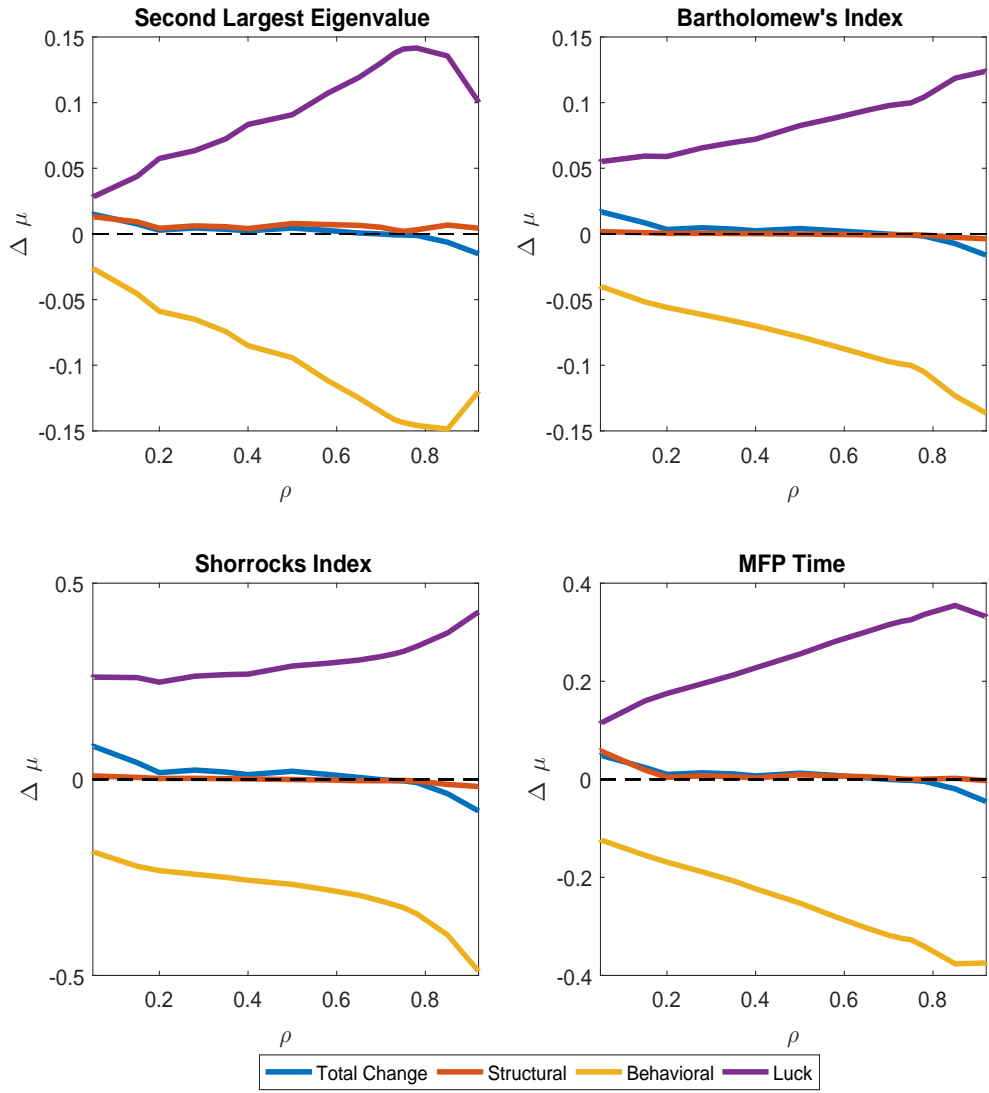


Figure 13: Decomposition of change in mobility as ρ increases

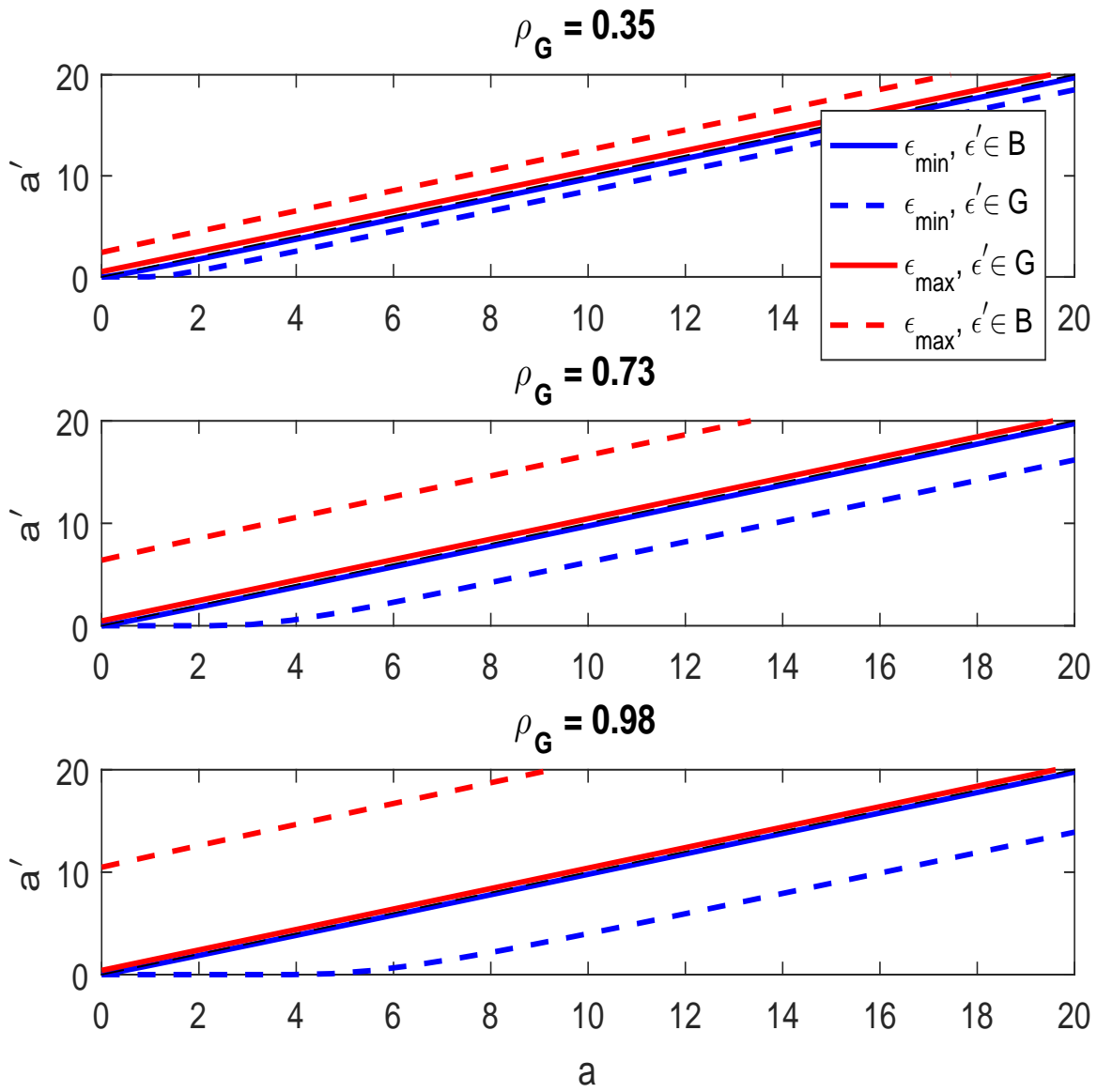


Figure 14: Portfolio decisions

Table 1: Wealth mobility in data

Source	PSID Years	μ_S	μ_{2E}	μ_B	μ_{MFP}
Hurst et al. (1998)	PSID 1984-1989	0.563	0.203	0.140	0.47
Hurst et al. (1998)	PSID 1989-1994	0.597	0.242	0.155	0.53
Díaz-Giménez et al. (2011)	PSID 2001-2007	0.608	0.237	0.158	0.53
Castaneda et al. (2003)	PSID 1984-1989	0.505	-	-	-

Table 2: Correlation between mobility measures

Correlation Coefficients			
	μ_B	μ_{2E}	μ_S
μ_{MFP}	0.9991	0.9997	0.9991
μ_S	1.0000	0.9991	—
μ_{2E}	0.9991	—	—

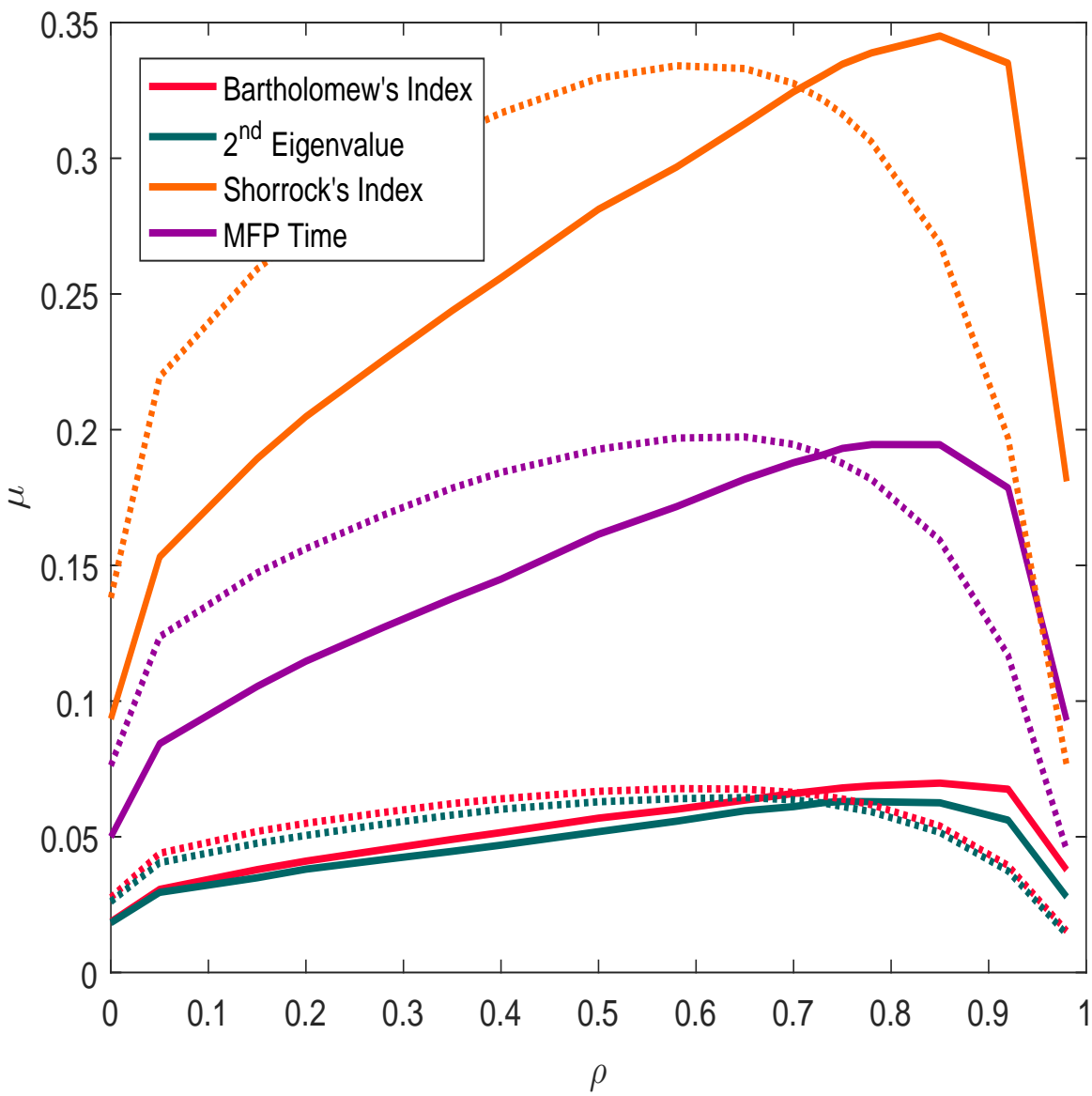


Figure 15: Mobility across ρ ; Incomplete markets vs. partial insurance

Solid - 2 assets; Dashed - 1 asset