THE IMPLICATIONS OF HETEROGENEITY AND INEQUALITY FOR ASSET PRICING

Stavros Panageas

The Implications of Heterogeneity and Inequality for Asset Pricing
Stavros Panageas
NBER Working Paper No. 26974
April 2020
JEL No. E21,G12

ABSTRACT

Does heterogeneity matter for asset pricing and in particular for risk premiums? Starting with an irrelevance result, I classify the literature into two groups of papers taking different routes to link investor heterogeneity and risk premiums. The first group contains models of investors who differ in terms of their preferences, beliefs, or access to markets. Despite their differences, these models have similar implications, and can be analyzed in a unified way. The second group of papers consists of models where investors experience uninsurable income shocks. The goal of this survey is to provide one unified framework to better understand this large literature, and especially to reconcile several of the seemingly inconsistent results found in some seminal papers.

Stavros Panageas
Anderson School of Management
University of California, Los Angeles
110 Westwood Plaza
Los Angeles, CA 90095-1481
and NBER
stavros.panageas@anderson.ucla.edu

1 Introduction and Summary

Heterogeneity is all around us. Besides the obvious dimensions of heterogeneity (income and wealth), people of similar wealth and income have different saving rates and attitudes towards risk. While it is hard to argue with the existence of heterogeneity, this survey asks a question with a less obvious answer: Does heterogeneity matter for asset pricing? Specifically, does it matter for the market price of risk?

Intuitively, it would seem that the answer should be a clear yes. The saving and portfolio choices of the median retiree and a rich entrepreneur are likely to be very different. As a result, the relative wealth share of these different groups of people should matter for both interest rates and risk premiums.

Yet, most leading asset pricing models tend to be "representative" agent models, i.e., models where the distribution of wealth, income, consumption and the associated dynamics of these quantities are all irrelevant for asset pricing.[1] One possible justification for abstracting from all distributional considerations was given in a seminal paper by Grossman and Shiller (1982). That paper revisited and substantially relaxed the assumptions of the Breeden (1979) aggregate consumption CAPM. Specifically, Grossman and Shiller (1982) showed that Breeden's aggregate-consumption-CAPM continues to determine risk premiums, even if risk sharing is imperfect, so that each consumer experiences a different consumption growth.

The argument is powerful and quite simple, since it essentially boils down to three equations. Specifically, Grossman and Shiller (1982) starts with the first order condition that if agent i is trading without frictions in some asset with (random) gross return $R_{t+\delta}$ between t and $t+\delta$ and also in the riskless asset with (certain) gross return $R_{t+\delta}^f$, then portfolio optimality requires that

$$E_t \left\{ u' \left(c_{t+\delta}^i \right) \left(R_{t+\delta} - R_{t+\delta}^f \right) \right\} = 0, \tag{1}$$

where $u' \left(c_{t+\delta}^i \right)$ is the marginal utility of agent i. Assuming that the agent trades frequently,

---

[1]Three leading such paradigms are Campbell and Cochrane (1999), Bansal and Yaron (2004), Barro (2006).

so that δ is small, and proceeding heuristically, a first order expansion of $u'\left(c_{t+\delta}^i\right)$ around c_t^i gives[2]

$$-\frac{u'\left(c_t^i\right)}{u''\left(c_t^i\right)} E_t\left\{R_{t+\delta} - R_{t+\delta}^f\right\} \approx E_t\left\{\left(c_{t+\delta}^i - c_t^i\right)\left(R_{t+\delta} - R_{t+\delta}^f\right)\right\}. \tag{2}$$

Defining aggregate consumption as $C_t \equiv \int_i c_t^i di$, and integrating both sides of (2) across i gives

$$
\begin{aligned}
E_t\left\{R_{t+\delta} - R_{t+\delta}^f\right\} &\simeq \left(-\int_i \frac{u'\left(c_t^i\right)}{u''\left(c_t^i\right)} di\right)^{-1} \times E_t\left\{\left(C_{t+\delta} - C_t\right)\left(R_{t+\delta} - R_{t+\delta}^f\right)\right\} \\
&= \left(-\int_i \frac{u'\left(c_t^i\right)}{c_t^i u''\left(c_t^i\right)} \frac{c_t^i}{C_t} di\right)^{-1} \times E_t\left\{\left(\frac{C_{t+\delta} - C_t}{C_t}\right)\left(R_{t+\delta} - R_{t+\delta}^f\right)\right\}. \tag{3}
\end{aligned}
$$

Equation (3) is remarkably similar to Breeden's aggregate consumption CAPM. It states that the excess return on an asset is given by the product of two terms. The first term is a consumption-weighted[3] "harmonic average" of the relative risk aversions $-\frac{c_t^i u''\left(c_t^i\right)}{u'\left(c_t^i\right)}$ of the different investors. The second term is the expected product of *aggregate consumption growth* with the excess return.

One immediate implication of (3) is that even in the presence of heterogeneity, the *conditional* aggregate consumption CAPM continues to hold. Other than affecting the weights in the harmonic average of relative risk aversion, heterogeneity in *individual consumption growth rates* – whatever the reason for this heterogeneity – does not impact the risk premium. In particular, income risks that can cause idiosyncratic consumption fluctuations due to imperfect risk sharing are irrelevant for risk premia. To exaggerate for the sake of clarity, the risk premium in an economy with and without uninsurable idiosyncratic risks will be the same as long as all agents have the same risk aversion.

The Grossman and Shiller (1982) result relies on one approximation step (equation 2). This approximation step is innocuous if asset prices and consumption growth are both dif-

---

[2]A first order Taylor expansion implies that $u'\left(c_{t+\delta}^i\right) \simeq u'\left(c_t^i\right) + u''\left(c_t^i\right)\left(c_{t+\delta}^i - c_t^i\right)$. Substituting this approximation inside (1) and re-arranging gives (2).

[3]The harmonic average of y_i with weights x_i is defined as $\left(\int_i x_i \left(y_i^{-1}\right) di\right)^{-1}$.

fusions. Indeed, in the continuous time limit that they consider, the approximation becomes exact, because of Ito's Lemma.

Even though the Grossman and Shiller (1982) result would appear to indicate a dead end, the almost forty years that followed its publication saw the development of a very active literature on the interactions between heterogeneity and asset pricing. Indeed, it seems that there is renewed interest in this question in recent years because of a broader trend in macroeconomics and finance to understand the economic implications of rising income and wealth inequality.

As it always happens with irrelevance results in economics (e.g., Modigliani Miller Theorem, Ricardian Equivalence, Revenue Equivalence, etc.), the Grossman and Shiller (1982) result is a useful pedagogical framework to explain how different papers in the literature "break" the irrelevance. With this in mind, in this survey I attempt a taxonomy of the different papers against the backdrop of Grossman and Shiller (1982). I classify the papers in two broad categories. The first broad category contains three strands depending on whether the source of heterogeneity is due to preferences, beliefs, or access to markets. The second broad category comprises papers with income heterogeneity and incomplete risk sharing. Inside this category there are again three main strands capturing models with (lack of) intra- or inter-cohort risk sharing and models that assume recursive preferences in a framework of imperfect risk sharing.

The first strand of the first broad category (section 2.1) contains models of risk aversion heterogeneity, typically in a framework where agents have expected utility preferences. These models are not departures from Grossman and Shiller (1982), since equation (3) continues to hold. Indeed, most of the papers in this literature assume that agents trade continuously and therefore equation (3) is exact, not approximate. The key feature of these models is the observation that if agents have different risk aversions then the consumption distribution evolves dynamically, favoring the bold in good times and the meek in bad times.

Specifically, the most interesting feature of these models is the countercyclicality of the market price of risk, or "Sharpe ratio": Because the relatively less risk averse agents choose

to be more exposed to aggregate shocks (as compared to the more risk averse agents), a positive aggregate shock increases their wealth and consumption weight, driving down the equilibrium Sharpe ratio. By contrast, a negative shock to aggregate consumption raises the Sharpe ratio. This negative correlation between aggregate shocks and changes in the market price of risk is a prediction that these models share with the Campbell and Cochrane (1999) representative agent model. In both types of models the aggregate consumption CAPM holds conditionally, but not unconditionally. The variations in "habits" that cause fluctuations in risk aversion in the Campbell and Cochrane (1999) model resemble the variations in the consumption weights of the heterogeneous agents in models of preference heterogeneity. As section 2.3 shows, the counter-cyclicality of the Sharpe ratio is a remarkably robust result in models of preference heterogeneity. These models also contain interesting implications on the determination of equilibrium interest rates and bond risk premia, which are also discussed in section 2.1.

The second strand of the first broad category (section 2.2) comprises models of belief heterogeneity. These models assume that some investors may have different beliefs than others, possibly not resulting from superior information, but from different priors or plain irrationality. In such models, equation (1) may not hold for some investors, as they use a different expectation operator. In terms of asset pricing implications, however, models of belief heterogeneity and preference heterogeneity are very closely related. In particular, the Sharpe ratio with heterogeneous beliefs is a consumption-weighted average of the Sharpe ratios that would obtain in homogeneous belief economies populated by only one of the constituent groups. Just as luck favors the bold in models of preference heterogeneity (in terms of increasing their wealth- consumption- share), luck favors the optimists in models of belief heterogeneity. Section 2.2 presents the formal connections between belief and preference heterogeneity.

The third strand of the first broad category (section 2.4) are models where equation (1) holds for some investors but not for others. For instance if investor i is not even participating in the market for the risky asset, then equation (1) does not apply to her. The implication

is that when aggregating across all agents to get from equation (2) to (3), one should only aggregate across the subset of agents that participate in the risky market. As section 2.4 shows, models of this type closely resemble models of heterogeneous preferences, where non-participants are viewed as investors with infinite risk aversion for the purposes of pricing the risky assets.

Overall, the three strands of the literature mentioned sofar belong to the same broad category, since they share more similarities than differences. In this survey I present and analyze them as part of a unified model in an effort to highlight these similarities. One could even argue that these models do not invalidate the core of the Grossman and Shiller (1982) result: While equation (1) may fail for some of the investors, at the end a conditional version of the aggregate consumption CAPM continues to hold, with the consumption weights acting as conditioning variables in the jargon used by finance econometricians.

The next broad category of papers can again be split into three strands and comprises models that assume identical investors experiencing idiosyncratic income and endowment shocks that cannot be insured due to some market failure. The papers in this group collide with a strong implication of Grossman and Shiller (1982), namely that in any model with diffusive (i.e., continuous) consumption and asset processes, heterogeneity should not matter in the continuous-time limit.

To be specific and give an example of this tension, one of the most influential papers in this literature is the paper by Constantinides and Duffie (1996). In a discrete time framework, the paper shows that a judicious specification of cross sectional income heterogeneity allows one to support any given stochastic discount factor (in a specific class) as an equilibrium outcome. Moreover, this stochastic discount factor may differ from the one implied by the aggregate consumption CAPM. Yet, the Grossman and Shiller (1982) result would seem to allow only the stochastic discount factor implied by the aggregate consumption CAPM as an equilibrium outcome, no matter what is assumed about income heterogeneity.

It would be natural to conjecture that the discrepancy between the two papers lies in the usage of discrete versus continuous time methods. If true, this would be a source of concern,

since it would indicate that if one were to shrink the assumed time interval in Constantinides and Duffie (1996) to zero, the results of the paper could be jeopardized.

Section 3.1, which discusses the first strand of the second broad category of the literature, reconciles the results of Constantinides and Duffie (1996) and Grossman and Shiller (1982). Using a minor modification of Constantinides and Duffie (1996) that allows consideration of the continuous time limit, this section shows that the key insight of Constantinides and Duffie (1996) is invariant to the assumed decision interval. The results of Grossman and Shiller (1982) do not apply because the continuous-time process is not a diffusion, but a process with discontinuous sample paths. Because of this, equation (2) does not hold, even in continuous time.

Reconciling the results of Constantinides and Duffie (1996) and Grossman and Shiller (1982) is not just a matter of resolving a mathematical conundrum. The discussion illuminates that for income heterogeneity to matter, it has to affect the covariance between higher order moments of individual consumption growth and asset returns. If consumption and asset price processes are diffusions, these higher order moments don't matter in the continuous time limit, since Ito's Lemma implies that the marginal utility of consumption behaves (locally) like a linear function. If consumption is a discontinuous function of time, then this locally linear relation fails and higher order moments start to matter.

Section 3.2 discusses the second strand of the second broad category, namely models where the risk sharing imperfection is not due to missing markets but rather due to missing market participants. Specifically, rather than assuming that existing cohorts cannot trade claims to their personal incomes with each other (which is the implicit market failure in models such as Constantinides and Duffie (1996)), the models in section 3.2 assume that it is impossible to trade with unborn agents. This results in a lack of inter-cohort risk sharing. While starting from different assumptions and setups, the lack of inter- and intra-cohort risk sharing imply identical stochastic discount factors.

The reason for the similarity between the two types of models is quite intuitive. The source of risk premia in models like Constantinides and Duffie (1996) is an agent's fear that

7

if there is a large redistribution amongst the existing cohorts of agents, she might end up being among the losers rather than the winners. Because of risk aversion, she overweights the possibility of being among the losers and demands a risk premium for assets with bad payoffs when redistribution is high. In models of imperfect inter-cohort risk sharing investors fear that an incoming cohort of agents might introduce the next big company that will displace the companies owned by current cohorts. Therefore any assets that are prone to this displacement command a risk premium.

Section 3.3 discusses the third strand of the second broad category. Models in this strand make make endowment heterogeneity matter by using recursive preferences rather than expected utility. In a seminal paper, Bansal and Yaron (2004) highlighted that if agents are not neutral to the timing of the resolution of uncertainty, then expected returns do not only reflect compensation for "short run" risk, but also for risks that are associated with consumption growth over the "long run". Mathematically, this means that equation (1) does not hold and the Grossman and Shiller (1982) argument fails at its origin.

Recursive preferences alone could be an irrelevant extension if individual consumption growth is i.i.d.. However, models with heterogeneous agents can easily lead to slow-moving predictable components in individual consumption growth, even if aggregate consumption growth is i.i.d. This is especially true in models where different birth cohorts experience different integrated consumption paths over their life time due to lack of inter-cohort risk sharing.

One interesting feature of recursive preferences is that they do not require a strong high-frequency co-movement between consumption inequality changes and asset returns to make income heterogeneity matter for the risk premium. Indeed over the short run there need not be any relation at all, and yet income heterogeneity can matter for asset returns.

The last section of the survey (section 4) addresses two further observations relating to equation (1). Section 4.1 discusses models, where for each risky asset, equation (1) applies only to a subset of agents. Section 4.2 discusses the validity of (1) in the presence of asymmetric information.

Section 4.1 discusses a multi-asset economy where all agents participate in some risky assets, but no agent participates in all asset classes. Models of this sort can feature equilibrium arbitrages. In such models it is natural to consider the incentives of agents to exploit the arbitrages by modeling pricing and participation decisions as joint outcomes. Interestingly, the presence of an arbitrage leads to a non-convexity in agent's optimization problems, which in turn leads to heterogeneous portfolio and participation decisions, even in the absence of any initial heterogeneity. In a sense, the extent of heterogeneity and equilibrium risk premia are jointly determined in response to endogenous participation decisions. Models of this sort are particularly well suited to study portfolio flows, leverage and asset price determination as joint outcomes.

Section 4.2 discusses the possibility that investors may have superior information compared to the econometrician. By itself, this is not a problem for the Grossman and Shiller (1982) argument, because the Euler equation (1) "conditions down" from the perspective of the econometrician. If, however, short selling constraints prevent equation (1) from holding for every investor and every asset, then agents may (endogenously) choose to not participate in certain asset classes. This means that the model features effectively heterogeneous stochastic discount factors. Models of this sort have important implications for portfolio biases, and performance evaluation.

In terms of presentation, this survey doesn't simply outline these models. All sections contain a simplified mathematical model that illustrates not only the economic ideas, but also the techniques that can be used to analyze these models. The reason for presenting these techniques is that models with heterogeneity can become intractable. One of the goals of this survey is to introduce the reader to some basic techniques to keep the mathematical strucuture tractable.

To be concrete, there are two difficulties when dealing with heterogeneous agents models. The first and obvious difficulty is that the wealth distribution becomes a state variable. And second, these models tend to be non-stationary, since the innate differences between the agents lead to different consumption growth rates, which end up driving the consumption

9

levels of different agents arbitrarily far apart in the long run.

I address these problems by placing the models within a "perpetual youth" framework. In such framework, new generations arrive constantly with new units of the aggregate endowment. This constant flow of wealth towards each group ensures their long term survival. Moreover, the solution of the model boils down to the solution of a system of differential equations. However, most of the insights and the analysis don't even require that one be able to solve these differential equations, so I relegate their formulation to the appendix.

After every section there is a literature review. Given the overwhelming size of the literature, the reader should view this literature review merely as providing some indicative pointers, not as an exhaustive list. It would be a mistake to presume that papers not included in the literature review are less important than the ones cited. The choice of which papers to cite was mostly dictated by the proximity of these papers to the ideas and techniques developed in each section.

2 Heterogeneous preferences, beliefs, and limited participation models

2.1 Heterogeneous preferences

The models of heterogeneous preferences, beliefs and limited participation can be presented in a unified framework. This section develops this framework in the context of heterogeneous preferences. The next two sections reuse this model with minor modifications to incorporate different beliefs and access to markets.

Time is continuous. This assumption is quite helpful to make the model tractable and obtain sharp results. To ensure stationarity, I assume the arrival of new agents who are endowed with new "trees". Per unit of time a mass of agents π is born. Existing agents face a constant hazard rate of death π. By the law of large numbers, the surviving population of agents who were born at time $s \leq t$ is $\pi e^{-\pi(t-s)}$ and total population is constant and equal to $\int_{-\infty}^{t} \pi e^{-\pi(t-s)} ds = 1$.

Arriving agents at time t are equally endowed with shares to a "tree" born at time t. Letting $s \leq t$ denote the time of "birth" of a tree, its time-t dividends are given by

$$D_{t,s} = \delta e^{-\delta(t-s)} D_t.$$

Accordingly, the total endowment of this economy is

$$\int_{-\infty}^{t} D_{t,s} ds = \left(\int_{-\infty}^{t} \delta e^{-\delta(t-s)} ds \right) \times D_t = D_t.$$

The aggregate endowment D_t follows a geometric brownian motion with some mean μ_D and some volatility σ_D

$$\frac{dD_t}{D_t} = \mu_D dt + \sigma_D dB_t, \tag{4}$$

where B_t is a standard brownian motion.

It should be noted here that in a typical perpetual youth, or overlapping generations model, the endowment of agents takes the form of labor income. The choice to endow agents with new trees rather than labor income is a (rather innocuous) shortcut to accelerate the presentation of the main results.[4]

Agents optimize their life-time consumption[5]

$$E_t \int_t^{\infty} e^{-\left(\rho^i + \pi\right)(u-t)} \frac{c_{u,t}^{1-\gamma^i}}{1-\gamma^i} du \tag{5}$$

where γ^i, ρ^i denotes the risk aversion (discount factor) of agent $i \in \{A, B\}$. Without loss of generality, $\gamma^A < \gamma^B$, so that agents of type A are the less risk averse agents. A fraction ν of agents is of type A and $1 - \nu$ of type B.

Once born, agents can trade with other agents (who are also alive at the same time) in

---

[4]The reader who wants to see how to modify the setup to allow for income is referred to Gârleanu and Panageas (2015).

[5]Note that the discount rate is $\rho^i + \pi$ to reflect the hazard rate of death π. See Gârleanu and Panageas (2015) on the specification of preferences in the presence of death.

complete markets. The associated stochastic discount factor in this economy is H_t. The goal of the analysis is to determine the dynamics of this stochastic discount factor.

2.1.1 Model Analysis

To start, note that the assumption of complete markets makes an agent's optimization problem simple to solve. Specifically, the consumers born at time t maximize (5) subject to the intertemporal budget constraint

$$E_t \int_t^\infty e^{-\pi(u-t)} \left(\frac{H_u}{H_t} \right) c_{u,t}^i du = \frac{1}{\pi} P_{t,t}, \tag{6}$$

where $\frac{1}{\pi} P_{t,t}$ denotes the (per-cohort member) value of the shares of the tree born at time t. Multiplying both sides of (6) by H_t and attaching a Lagrange multiplier λ to (6) gives

$$\max_{c_{u,t}} E_t \left\{ \int_t^\infty e^{-\pi(u-t)} \left(e^{-\rho^i(u-t)} \frac{c_{u,t}^{1-\gamma^i}}{1-\gamma^i} du - \lambda H_u c_{u,t} \right) du \right\} + \lambda \frac{1}{\pi} H_t P_{t,t},$$

which results in the first order condition

$$e^{-\rho^i(u-t)} c_{u,t}^{-\gamma^i} = \lambda H_u. \tag{7}$$

Equation (7) is the familiar requirement that the marginal utility of consumption be proportional to the stochastic discount factor.

Evaluating (7) at time t implies that $c_{t,t}^{-\gamma^i} = \lambda H_t$, and hence the consumption growth of a consumer of type i is

$$\frac{c_{u,t}^i}{c_{t,t}^i} = e^{-\frac{\rho^i}{\gamma^i}(u-t)} \left(\frac{H_u}{H_t} \right)^{-\frac{1}{\gamma^i}}. \tag{8}$$

Equation (8) is intuitive. Inspection of the right hand side shows that all consumers of type i that belong to the cohort born at time t experience the same consumption growth between times u and t. This reflects the assumption of a complete market, which allows

perfect risk sharing *within* a cohort.

In equilibrium, the consumption market needs to clear at each time t. To analyze the implications of market clearing, recall that a fraction let x_t^i denote the consumption share accruing to type-i agents

$$x_t^i \equiv \frac{\nu^i \int_{-\infty}^t \pi e^{-\pi(t-s)} c_{t,s}^i ds}{D_t}, \tag{9}$$

where $\nu^i = \nu$ if $i = A$ and $\nu^i = 1 - \nu$ if $i = B$. In other words, ν^i is the population weight of agents of type i, while x_t^i is their consumption weight. Market clearing requires that

$$x_t^A + x_t^B = 1. \tag{10}$$

Substituting (8) into (9) leads after some re-arranging to

$$x_t^i \equiv \frac{H_t^{-\frac{1}{\gamma^i}} \times \nu^i \int_{-\infty}^t \pi e^{-\pi(t-s)} c_{s,s}^i e^{-\frac{\rho^i}{\gamma^i}(t-s)} H_s^{\frac{1}{\gamma^i}} ds}{D_t}. \tag{11}$$

Equation (11) provides a key relationship between the consumption share of type i agents and the stochastic discount factor H_t. This relationship can be used to derive the joint dynamics of the consumption share, the interest rate and the Sharpe ratio.

2.1.2 Volatility of x_t^i and countercyclical Sharpe ratios

To start, is useful to recall a standard result in asset pricing, namely that the dynamics of the stochastic discount factor are given by

$$\frac{dH_t}{H_t} = -r_t dt - \kappa_t dB_t, \tag{12}$$

where r_t is the interest rate and κ_t the market price of risk (or Sharpe ratio). Applying Ito's Lemma to the right hand side of (11) and using (12) and (4) shows that x_t^i follows a diffusion

$dx_t^i = \mu_x^i dt + \sigma_x^i dB_t$, where

$$\sigma_x^i \equiv \left(\frac{\kappa_t}{\gamma^i} - \sigma_D \right) x_t^i. \tag{13}$$

The market clearing requirement (10) requires that $\sigma_x^A + \sigma_x^B = 0$.[6] This is intuitive, since any movement in the consumption share of type-A agents must be offset by an opposing movement in the consumption share of type-B agents. Combining (13) with (10) and the associated requirement $\sigma_x^A + \sigma_x^B = 0$ leads to

$$\kappa_t = \left(\sum_{i \in \{A,B\}} x_t^i \left(\gamma^i \right)^{-1} \right)^{-1} \times \sigma_D. \tag{14}$$

Equation (14) shows that the Sharpe ratio is proportional to the harmonic average of the risk aversions of the two agents times the volatility of the aggregate endowment (σ_D).

The weights of this harmonic mean are time-varying since x_t^i changes over time. Indeed, combining (13) with (14) leads to

$$\sigma_x^i \equiv \sigma_D \times \left(\frac{\left(\gamma^i \right)^{-1}}{\sum_{i \in \{A,B\}} x_t^i \left(\gamma^i \right)^{-1}} - 1 \right) x_t^i. \tag{15}$$

Equation (15) shows that x_t^A is procyclical and x_t^B is countercyclical. Given the normalization $\gamma^A < \gamma^B$, the term inside brackets in equation (15) is positive when $i = A$ and negative when $i = B$. Hence, a positive shock to the aggregate endowment increases the consumption share of type-A agents and reduces the consumption share of type-B agents.

This procyclicality of the consumption share of the less risk averse agents implies that the Sharpe ratio is countercyclical. Inspection of equation (14) shows that when the share of the less risk averse agents x_t^A increases (which happens in response to a positive aggregate shock), the Sharpe ratio declines. In that sense one can think of an economy with heterogenous agents as being equivalent to an economy with a representative agent but with a risk aversion that

---

[6]This can be proven by applying Ito's Lemma to both sides of (10).

declines in response to good aggregate shocks and increases in response to bad aggregate shocks.

The economic force behind this countercyclicality of aggregate risk aversion is simple. Less risk averse agents choose to bear more aggregate risk than less risk averse agents. Because of this higher exposure to aggregate risk, a positive aggregate shock leads to an increase in their consumption share. Being a harmonic weighted average of risk aversions, the market price of risk declines, reflecting the increased weight of the relatively more risk tolerant agents.

This basic mechanism that links the movements of the wealth distribution with the countercyclicality of the Sharpe ratio is a common feature in almost all models with heterogeneous agents, whether the source of heterogeneity stems from preferences, beliefs, or access to markets, as will become evident in the subsequent sections.

To complete our characterization of equilibrium, we next turn attention to the determination of the drift of x_t^i and the implications for the equilibrium interest rate and the stationarity of the consumption shares

2.1.3 Drift of x_t^i, equilibrium interest rate r_t, and stationarity of the consumption distribution

Applying Ito's Lemma to (11) and using (12) leads after some simplifications to the following expression for the drift of x_t^i :

$$\mu_x^i = \left(\frac{r_t - \rho^i}{\gamma^i} - \mu_D + \sigma_D^2 + \frac{1}{2} \left(\frac{\kappa_t}{\gamma^i} \right)^2 \left(\frac{1}{\gamma^i} + 1 \right) - \frac{\kappa_t \sigma_D}{\gamma^i} - \pi \right) x_t^i + \pi \nu^i \frac{c_{t,t}^i}{D_t}. \qquad (16)$$

In turn, market clearing implies that $\mu_x^A + \mu_x^B = 0$, since the two consumption shares have to always add up to one. Combining $\mu_x^A + \mu_x^B = 0$ with (10) and (16) gives the following

expression for the interest rate

$$
r_t = \underbrace{\left(\sum_{i \in \{A,B\}} \frac{x_t^i}{\gamma^i} \right)^{-1}}_{\text{inverse weighted average of IES}} \times
$$

(17)

$$
\times \left[\underbrace{\sum_{i \in \{A,B\}} x_t^i \frac{\rho^i}{\gamma^i}}_{\text{weighted discount rate}} + \mu_D - \sigma_D^2 - \underbrace{\sum_{i \in \{A,B\}} x_t^i \left(\frac{1}{2} \left(\frac{\kappa_t}{\gamma^i} \right)^2 \left(\frac{1}{\gamma^i} + 1 \right) - \frac{\kappa_t \sigma_D}{\gamma^i} \right)}_{\text{weighted precautionary savings}} + \underbrace{\pi \left(1 - \sum_{i \in \{A,B\}} \frac{\nu^i c_{t,t}^i}{D_t} \right)}_{\text{deaths-births}} \right].
$$

While the expression for the equilibrium interest rate is long, it has the same basic structure as the interest rate in a standard representative investor economy. Indeed if $x_t^A = 1$ and $\pi = 0$, the expression for the interest rate simplifies to $r = \rho^A + \frac{1}{\gamma^A} \left[\mu_D + \frac{1}{2} \sigma_D^2 \left(\frac{1}{\gamma^A} + 1 \right) \right]$, which is the interest rate that one would obtain in an economy populated by a single, infinitely-lived, representative agent of type A. A similar expression obtains when $x_t^B = 1$ and $\pi = 0$.

For intermediate values of x_t^i, the interest rate can be expressed as a product of two terms. The first term is the consumption-weighted average of the intertemporal elasticity of substitution (IES) of the two agents. (Recall that with CRRA preferences the IES is just the reciprocal of the risk aversion). The second term is the sum of a) a weighted sum of discount rates, b) the aggregate growth rate, c) a term reflecting a weighed average of the precautionary savings motives and d) a term that captures the effect of deaths and births.

There are a few things worth noting about the equilibrium interest rate:

First, since the interest rate is the product of a weighted IES and a summation term that also depends on consumption weights, the interest rate in a heterogenous-agents economy cannot be expressed as some weighted average of the interest rates in a pure type-A or pure type-B economy.

Second, the dependence of the interest rate on x_t^A is ambiguous and depends on several effects. As x_t^A increases, the weighted average of the IES increases. Assuming that the interest rate is positive, (i.e. assuming that the term inside square brackets in equation (17)

is positive), this effect pushes the interest rate up, reflecting the increased importance of agents who are more willing to substitute inter-temporally. However, as x_t^A increases, the precautionary savings term decreases, reflecting that the agents with the high IES are also the agents with the low risk aversion. The reduced importance of precautionary savings pushes the interest rate down, counteracting the effect of higher IES. The behavior of the term $\sum_{i\in\{A,B\}} x_t^i \frac{\rho^i}{\gamma^i}$ depends on whether one assumes that $\rho^A < \rho^B$ or the reverse. In short, the dependence of the interest rate on x_t^A is ambiguous.

Third, in a heterogeneous-agents economy, the interest rate fluctuates as the consumption weights change, even when there is no change in the conditional moments of aggregate consumption growth (i.e., when consumption growth is i.i.d). This would not be possible in a representative-agent economy, where the interest rate can only change if the conditional moments of aggregate consumption growth change. The fact that the interest rate exhibits variation even when aggregate consumption growth is i.i.d., could be one of the reasons for the very low estimates obtained from macro-level estimates of the IES. Motivated by representative agent models that ignore heterogeneity, researchers in this literature estimate the IES of the representative agent as the ratio of predictable variations in aggregate consumption growth to predictable variations in the (real) interest rate. In the economy analyzed above, this ratio would be zero even though any given agent has a strictly positive IES.

Fourth, even if all investors have the same preferences, the resulting interest rate would differ from the one in an economy populated by a single representative agent with the same preferences. There would be an additional term $\pi\left(1 - \frac{c_{t,t,}}{D_t}\right)$, which is due to the presence of deaths and births. In an economy with deaths and births and no altruistic linkages, the consumption growth rate of any fixed cohort of agents is not equal to aggregate consumption growth. The reason is that a fraction of consumption every period accrues to the newly born agents $\frac{c_{t,t}}{D_t}$, while at the same time a fraction π of agents dies, thus increasing the per capita consumption growth of surviving agents. Section 3.2 discusses this issue in further detail.

As a concluding observation, note that, despite allowing for heterogeneity, the consumption distribution is stationary. Specifically, note that equation (13) implies that $\lim_{x_t^i \to 0}$

$\sigma_x^i = 0$, while (16) implies that $\lim_{x_t^i \to 0} \mu_x^i = \pi \nu^i \frac{c_{t,t}^i}{D_t} > 0$. This means that as the consumption share of one type of agents approaches zero, the diffusion process starts resembling a deterministic process with positive growth. (Section 2.2.3 goes into some further details on the issue of stationarity). The economic intuition behind stationarity is that – irrespective of the past path of the economy – there are constantly new agents of type i arriving with new units of wealth. Indeed, note that as one group starts approaching extinction (i.e., as x_t^i approaches zero), it is exclusively the properties of the entering cohort of type i agents that matter for the dynamics of x_t^i (specifically, the mass $\pi \nu^i$ of the cohort times its consumption relative to aggregate consumption, $\frac{c_{t,t}^i}{D_t}$.)

2.1.4 Completing the construction of equilibrium

Taking stock, equations (14) and (15) provide the Sharpe ratio κ_t and the volatility of x_t^i as functions of x_t^i. Equations (16) and (17) provide the drift μ_x and the interest rate r_t as functions of x_t^i and the endogenous quantity $\frac{c_{t,t}^i}{D_t}$. The determination of $\frac{c_{t,t}^i}{D_t}$ requires the usage of the budget constraint of an agent who enters the economy at time t. Specifically,

$$E_t \int_t^\infty e^{-\pi(u-t)} \left(\frac{H_u}{H_t} \right) c_{u,t}^i du = \frac{1}{\pi} P_{t,t} = \frac{\delta}{\pi} E_t \int_t^\infty e^{-\delta(u-t)} \left(\frac{H_u}{H_t} \right) D_{u,t} du, \qquad (18)$$

which states that the present value of consumption should equal the value of the shares that an agent is endowed with, which in turn should equal the present value of the dividends accruing to the stock. Defining

$$g_t^i \equiv E_t \int_t^\infty e^{-\pi(u-t)} \left(\frac{H_u}{H_t} \right) \left(\frac{c_{u,t}^i}{c_{t,t}^i} \right) du, \text{ and } p_t \equiv E_t \int_t^\infty e^{-\delta(u-t)} \left(\frac{H_u}{H_t} \right) \left(\frac{D_{u,t}}{D_{t,t}} \right) du, \quad (19)$$

the intertemporal budget constraint (18) can be expressed compactly as

$$\frac{c_{t,t}^i}{D_t} = \frac{\delta}{\pi} \frac{p_t}{g_t^i}. \qquad (20)$$

18

The term p_t can be interpreted as the price-to-dividend ratio in this economy, while the term g_t^i is the wealth-to-consumption ratio for an agent of type i. The appendix shows that p_t and g_t^i are functions of x_t^i and solve a pair of differential equations, similar to the ones that one finds in the derivatives literature. This implies that $\frac{c_{i,t}^i}{D_t}$ is a function of x_t^i. Accordingly both σ_x^i and μ_x^i are functions of x_t^i, $i \in \{A, B\}$ which in turn means that x_t^i, is a Markov process.

2.1.5 Summary

If one had to isolate the main insight sofar, it is that a model of preference heterogeneity contains a natural mechanism to render the Sharpe ratio and the interest rate time-varying even when the aggregate endowment growth is i.i.d.. The time-variation in the Sharpe ratio is countercyclical, as positive shocks favor the consumption share of the bold agents, who require lower compensation for risk. The cyclicality of the riskless rate is ambiguous.

2.1.6 Literature review

The methodology presented in this section is based on Gârleanu and Panageas (2015), who consider a model with heterogeneous risk aversions and heterogeneous inter-temporal elasticity of substitution in a stationary, perpetual youth framework. Blanchard (1985) analyzes the perpetual youth framework in a deterministic framework without preference heterogeneity.

The literature on risk sharing with heterogeneous risk aversions dates back at least to Borch (1962). Dumas (1989) is an early contribution, characterizing the risk sharing problem in a dynamic framework. Wang (1996) considers the risk sharing problem in a dynamic, exchange economy and derives implications for the term structure of interest rates. Longstaff and Wang (2012) and Barro et al. (2017) derive implications for credit markets. Bhamra and Uppal (2013) uses Lagrange's Theorem to characterize the risk sharing problem with infinitely lived agents. Gârleanu and Pedersen (2011) considers a model of heterogeneous preferences with margin constraints.

Chan and Kogan (2002) and Xiouros and Zapatero (2010) consider an alternative approach to obtaining a stationary consumption distribution, which relies on specifying agents'

consumption preferences as being relative to aggregate consumption. Santos and Veronesi (2018) considers an alternative approach to stationarity that relies on a combination of habits and non-homothetic preferences.

Dumas et al. (2000) considers a planner's problem with heterogeneous recursive preferences. Schneider (2017) studies the term structure and bond risk premia in a model with heterogeneous risk aversion and IES. Recently several papers have used models of heterogeneous preferences to study issues related to inequality. Indicative examples include Gomez (2017), Pastor and Veronesi (2019) , Akira Toda and Walsh (2019). Kargar (2018), uses a model of heterogenous risk aversions and IES to study heterogeneous intermediation. Kondor and Vayanos (2019) studies the dynamics of arbitrage in a framework of heterogeneity. Alvarez and Atkeson (2018) consider a tractable model of random shifts in risk aversion. Calvet et al. (2019) study the empirical cross-section of household risk aversion and intertemporal elasticity of substitution.

2.2 Heterogeneous beliefs

2.2.1 Model setup and analysis

For the purposes of this section, all investors have the same preferences ($\gamma^i = \gamma, \rho^i = \rho$), but different beliefs. For instance, suppose that agents of type A believe that the aggregate dividend grows at the rate μ^* rather than μ_D, while agents of type B have correct expectations.

The easiest way to capture this situation is to define $\eta^i \equiv \frac{\mu^i - \mu_D}{\sigma_D}$, where $\mu^A = \mu^*$ and $\mu^B = \mu_D$, and let

$$Z_t^i \equiv e^{-\frac{\left(\eta^i\right)^2}{2}t + \eta^i B_t} \tag{21}$$

denote the likelihood ratio between the correct probability measure and the probability measure perceived by agent i. Note that $\eta^B = 0$ and $Z_t^B = 1$, which reflects that agent B is rational. An implication of Girsanov's theorem is that we can write each agents'

maximization problem as

$$E_t \int_t^\infty e^{-(\rho+\pi)(u-t)} Z_{u,t}^i \frac{c_{u,t}^{1-\gamma}}{1-\gamma} du \tag{22}$$

The intertemporal constraint of each agent continues to be given by (6). Attaching a Lagrange multiplier and repeating the same steps as in the case with heterogeneous preferences leads to the following equation describing the evolution of optimal consumption

$$\frac{c_{u,t}^i}{c_{t,t}^i} = e^{-\frac{\rho}{\gamma}(u-t)} \left(\frac{Z_u}{Z_t}\right)^{\frac{1}{\gamma}} \left(\frac{H_u}{H_t}\right)^{-\frac{1}{\gamma}}. \tag{23}$$

2.2.2 Sharpe ratio and interest rate

Using the optimal consumption process (23) in the definition of x_t^i (equation 9), and applying Ito's Lemma (along with the fact that $\frac{dZ_t^i}{Z_t^i} = \eta^i dt$) gives the following expression for the volatility of x_t^i

$$\sigma_x^i \equiv x_t^i \left(\frac{\kappa_t + \eta^i}{\gamma} - \sigma_D\right). \tag{24}$$

In turn, using the market clearing requirement $\sum_{i \in \{A,B\}} \sigma_x^i = 0$ and $\sum_{i \in \{A,B\}} x_t^i = 1$ gives the following expression for the Sharpe ratio

$$\kappa_t = \gamma \sigma_D - \sum_{i \in \{A,B\}} x_t^i \eta^i. \tag{25}$$

Equation (25) shows that the Sharpe ratio in an economy where some investors have distorted beliefs is equal to the Sharpe ratio that would obtain in an economy without belief distortions $(\gamma \sigma_D)$ minus a term that represents a consumption-weighted belief distortion. Specifically, if $\eta^A > 0$ (meaning that the perceived growth rate of the aggregate endowment by agent A is higher than in reality), the Sharpe ratio is lower than $\gamma \sigma_D$, reflecting the optimism of a fraction of the agents. By contrast, when the economy is populated by some pessimistic agents $\left(\eta^A < 0\right)$, the Sharpe ratio is higher.

Substituting (25) into (24) and re-arranging gives

$$\sigma_x^i = \frac{x_t^i}{\gamma}\left(\eta^i - \sum_{i\in\{A,B\}} x_t^i\eta^i\right). \tag{26}$$

The sign of σ_x^A depends on the sign of η^A. If η^A is positive, then $\sigma_x^A \geq 0$, i.e., the consumption share of optimistic agents is pro-cyclical. Similarly, if η^A is negative, $\sigma_x^A \leq 0$. Since κ_t is declining in x_t^A when $\eta^A > 0$ and increasing in x_t^A when $\eta^A < 0$, it follows that the Sharpe ratio declines in response to a positive shock to the aggregate endowment irrespective of the sign of η^A. Therefore, as in the case of heterogenous preferences, the Sharpe ratio is unambiguously countercyclical, irrespective of whether the irrational agents are optimistic or pessimistic.

Proceeding as in the case with heterogeneous preferences, the drift of x_t^i is given by

$$\mu_x^i = x_t^i\left[\frac{r-\rho}{\gamma} - \mu_D + \sigma_D^2 - \frac{\kappa_t+\eta^i}{\gamma}\left(\eta^i + \sigma_D - \frac{1}{2}\left(\frac{\kappa_t+\eta^i}{\gamma}\right)\left(\frac{1}{\gamma}+1\right)\right) - \pi\right] + \pi\nu^i\frac{c_{t,t}^i}{D_t}, \tag{27}$$

and upon using $\sum_{i\in\{A,B\}}\mu_x^i = 0$ and $\sum_{i\in\{A,B\}} x_t^i = 1$ the interest rate is

$$r_t = \rho + \gamma\mu_D + \gamma\pi\left(1 - \sum_{i\in\{A,B\}}\nu^i\frac{c_{t,t}^i}{D_t}\right) + \gamma\sum_{i\in\{A,B\}} x_t^i\frac{\kappa_t+\eta^i}{\gamma}\left(\eta^i - \frac{1}{2}\left(\frac{\kappa_t+\eta^i}{\gamma}\right)\left(\frac{1}{\gamma}+1\right)\right)$$

Just as in the case of heterogeneous preferences, the interest rate cannot be expressed as a simple average of its value at $x_t^A = 0$ and $x_t^A = 1$. As π becomes small, one can however show that if $\eta^A > 0$ (optimism), then the interest rate is higher when $x_t^A = 1$ than when $x_t^A = 0$. Optimists believe in higher consumption growth than the rational agents and in an effort to smooth that perceived upward sloping consumption path, they end up dissaving, thus increasing the interest rate. (Vice versa in the case of pessimism $(\eta^A < 0)$).

As before, fully characterizing the interest rate r_t and the drift μ_x^i requires an expression for $\frac{c_{t,t}^i}{D_t}$. Obtaining such an expression requires one to impose the budget constraint at time 0 as in section 2.1.4. In the special case where both agents have logarithmic preferences

$(\gamma = 1)$, there is a simple expression for the initial consumption ratio $\frac{c^i_{t,t}}{D_t}$. In this case the ratio $\frac{c^i_{t,t}}{D_t}$ is constant and equal to $\frac{\delta}{\pi}$.

2.2.3 Long-term "survival" of irrational agents

As in the case of heterogeneous preferences, $\lim_{x^i_t \to 0} \sigma^i_x = 0$, while $\lim_{x^i_t \to 0} \mu^i_x = \pi \nu^i \frac{c^i_{t,t}}{D_t} > 0$, which means that there is a force to ensure stationarity and long term "survival" of all agent groups, even the ones that have wrong beliefs. To obtain some further insight into the role of births and deaths in guaranteeing survival of each agent type, it is useful to compute a "stochastic steady state value of x^B_t". This value is defined as the consumption share of rational agents $\overline{x}^B_t$ such that $\mu^B_x(\overline{x}^B_t) = 0$. In other words it is the value of the consumption share of rational agents, x^B_t, which has neither a tendency to grow or shrink in expectation. (Since $\mu^A_x + \mu^B_x = 0$, it follows that the consumption share of irrational agents is also drift-less.)

To simplify matters, it is instructive to focus on the case where all investors have logarithmic preferences ($\gamma = 1$). Substituting the expressions for the Sharpe ratio (25) and the interest rate (28) into (27) and solving for the value that sets $\mu^B_x = 0$ results in

$$\overline{x}^B_t = \nu^B \frac{\delta}{\delta - \left(1 - \overline{x}^B_t\right)^2 (\eta^A)^2} > \nu^B. \tag{28}$$

Equation (28) implies that any steady state value $\overline{x}^B_t$ that sets $\mu^B_x = 0$ must be such that the steady state consumption weight $\left(\overline{x}^B_t\right)$ of rational agents exceeds their population weight $\left(\nu^B\right)$. Since both $\overline{x}^A_t + \overline{x}^B_t = 1$ and $\nu^A + \nu^B = 1$ the implication is that the steady state consumption weight of irrational agents $\left(\overline{x}^A_t\right)$ is below their population weight ν^A. Accordingly, rational agents consume more per capita than irrational agents. Whether the irrationality takes the form of optimism or pessimism $\eta^A < 0$ or $\eta^A > 0$ is irrelevant, since only $(\eta^A)^2$ enters equation (28). What does matter is the magnitude of the belief distortion $(\eta^A)^2$.

What allows the survival of irrational agents is the fact that the incoming cohorts of

both types are endowed with new trees. Interestingly, the steady state value of $\overline{x}_t^B$ is not affected by the rate of death and birth π, but only by the endowment fraction that accrues to newly-borns per unit of time. This highlights an important aspect of the economic forces that ensure survival in this model, namely that incoming cohorts of both types are endowed with a fraction of the endowment irrespective of past history.

I conclude the discussion of survival with a technical remark. I argued earlier that since the volatility of $\sigma_x(x_t^i)$ goes to zero as x_t^i approaches zero, while $\mu_x(x_t^i)$ approaches a positive number, this model allows for a stationary distribution. While intuitive, this argument requires an extra step. The exact condition[7] to ensure stationarity is that

$$\lim_{x_t^A \to 0} \frac{\left(\sigma_x^A\left(x_t^A\right)\right)^2}{x_t^A} < 2\left|\mu_x^A(0)\right|. \tag{29}$$

Equation (26) implies that the left hand side of the above expression is zero, and hence positivity of $\mu_x^A(0)$ is sufficient for survival.[8]

2.3 Multiple types of agents

The assumption so far has been that there are two types of agents, in which case the distribution of consumption across the two groups is characterized by x_t^A, the consumption share of type A agents. (The consumption share of type B agents is simply $1 - x_t^A$). None of the conclusions of either the heterogeneous-preferences, or heterogeneous-beliefs model changes, if instead there are N type of agents having different preferences (or beliefs). All the equations derived above remain unchanged – other than the fact that all sums of the form $\sum_{i \in \{A,B\}}$ would have to be replaced with the sum $\sum_{i \in \{A,B,C...\}}$.

The central insight, namely that the Sharpe ratio is some weighted average of the Sharpe ratios that would obtain in the economies populated by type $A, B, C...$ etc agents continues to hold, since equation (14) (in the case of heterogeneous preferences) and equation (25) (in

---

[7]Gârleanu and Panageas (2020) shows the sufficiency of this condition.

[8]A similar argument shows that positivity of $\mu_x^i(0)$ is sufficient for stationarity in the case where investors have heterogeneous preferences.

the case of heterogeneous beliefs) continue to hold with the sum running over all groups.

Interestingly, one can extend the insight that the Sharpe ratio is countercyclical. To start, suppose that the groups $A, B, C, ..$, are defined in order of declining risk aversion (or declining optimism if investors have different beliefs).

Viewing the Sharpe ratio as a function of the vector $\overrightarrow{x}_t$ and applying Ito's Lemma to obtain the dynamics of $d\kappa\left(\overrightarrow{x}_t\right)$ shows that the diffusion component of $d\kappa\left(\overrightarrow{x}_t\right)$ is given by[9]

$$\sum_{i\in\{A,B,C...\}} \frac{\partial\kappa}{\partial x_i}\sigma_x^i = -\left(\sum_{i\in\{A,B,C...\}} x_t^i\left(\gamma^i\right)^{-1}\right)^{-2}\times\left(\sum_{i\in\{A,B,C...\}}\left(\gamma^i\right)^{-1}\sigma_x^i\right) < 0. \tag{30}$$

In other words, a positive shock to the aggregate endowment reduces the Sharpe ratio. Importantly, the Sharpe ratio is countercyclical no matter what assumptions one makes about the magnitude of the preference discrepancies, the fractions, the discount rates, the initial endowments etc. of each group. These assumptions may impact the quantitative magnitude, but not the qualitative nature of the result.

In the case of heterogeneous beliefs, the analog of equation (30) is[10]

$$\sum_{i\in\{A,B,C...\}} \frac{\partial\kappa}{\partial x_i}\sigma_x^i = -\sum_{i\in\{A,B,C...\}} \eta^i\sigma_x^i < 0,$$

Once again, the assumptions on the magnitude of the belief distortions, the fractions of each group, their preferences etc. are immaterial. All that matters is that there is some discrepancy in beliefs between different groups.

---

[9]To see that $\sum_{i\in\{A,B,C...\}} \frac{\partial\kappa}{\partial x_i}\sigma_x^i < 0$, note that $\sum_{i\in\{A,B,C...\}} x_t^i\left(\gamma^i\right)^{-1} > 0$. The second term on the right hand of (30) is also positive since it can be written as

$$\left(\sum_{i\in\{A,B,C...\}}\left(\gamma^i\right)^{-1}\right)\times\sum_{i\in\{A,B,C...\}}\frac{\left(\gamma^i\right)^{-1}}{\sum_{i\in\{A,B,C...\}}\left(\gamma^i\right)^{-1}}\sigma_x^i\left(x_t^i\right).$$

The expression immediately above is positive because a) $\sum_{i\in\{A,B,C...\}}\sigma_x^i\left(x_t^i\right) = 0$, b) the weights $\frac{\left(\gamma^i\right)^{-1}}{\sum_{i\in\{A,B,C...\}}\left(\gamma^i\right)^{-1}}$ are declining in i and c) σ_x^i is declining in i.

[10]As in the case of heterogenous preferences, the facts that a) $\sum_{i\in\{A,B,C...\}}\sigma_x^i = 0$, b) η^i is declining in i by construction, and c) σ_x^i is declining in i, imply that that $\sum_{i\in\{A,B,C...\}}\eta^i\sigma_x^i > 0$.

2.3.1 Literature review

Methodologically, the presentation in this section combined the approach of Kogan et al. (2006) with that of Gârleanu and Panageas (2015).

The literature on heterogeneous beliefs and the survival of irrational agents is voluminous. Scheinkman and Xiong (2003), Chen et al. (2012), Buraschi and Jiltsov (2006), David (2008), Dumas et al. (2016), Ehling et al. (2017) is a small, indicative sample of applications of such models in finance. The issue of survival of irrational traders is discussed in Sandroni (2000), Blume and Easley (1992), Kogan et al. (2006), Yan (2008) and Borovicka (2019).

2.4 Limited participation

Both in the case of heterogenous preferences and heterogenous beliefs agents were assumed to trade in complete markets. In this section, I show how to extend the analysis to a situation where some agents don't participate in the stock market. To simplify matters, suppose that there are two groups of investors who have the same discount rates and (rational) beliefs. Agents of type A have risk aversion γ^A and access to complete markets, while agents of type B have risk aversion $\gamma^B = 1$ (log preferences) and cannot enter into any contracts that are contingent on the realization of the brownian motion B_t; put simply agents of type B cannot dynamically trade stocks. The limited participation assumption of type B agents is encapsulated in the following assumption on the evolution of their wealth:

$$\frac{dW_t^B}{W_t^B} = \left(r_t + \pi - \frac{c_t^B}{W_t^B} \right) dt. \tag{31}$$

There are two observations about equation (31).

First, in an effort to save notation, equation (31) is in terms of c_t^B and W_t^B rather than $c_{t,s}^B$ and $W_{t,s}^B$. The reason is that since all agents of type B face the same hazard rate of death π and interest rate r_t, and have the same homogenous preferences, they choose the same consumption-to-wealth ratios irrespective of their cohort. Indeed, because $\gamma^B = 1$, it follows that $\frac{c_t^B}{W_t^B} = \rho + \pi$.

Second, the presence of the term π reflects the assumption of annuitization. Specifically, since agents have no bequest motives, they enter contracts that require them to surrender their wealth upon death in exchange for an income stream of πdt while alive. The (competitive) insurance company offering this contract breaks even as it collects a fraction π of aggregate wealth from the dying agents and distributes it as an income stream to the surviving agents.[11] Mathematically, equation (31) together with $\frac{c_t^B}{W_t^B} = \rho + \pi$ implies that

$$\frac{W_t^B}{W_s^B} = e^{\int_s^t (r_u + \pi - \rho - \pi) du} \Rightarrow \frac{c_t^B}{c_s^B} = e^{\int_s^t (r_u - \rho) du}. \tag{32}$$

Assuming momentarily the existence of a stochastic discount factor H_t, defining $x_t^A\ \left(x_t^B\right)$ as the consumption share of the type A (type B) agents, repeating the calculations of section 2.1 and using equation (32) for the consumption evolution of the type-B agent implies that

$$\sigma_x^A = \left(\frac{\kappa_t}{\gamma^A} - \sigma^D\right) x_t^A, \quad \text{and } \sigma_x^B = -\sigma_D x_t^B.$$

Imposing the market clearing conditions $\sigma_x^A + \sigma_x^B = 0$ and $x_t^A + x_t^B = 1$ results in

$$\kappa_t = \frac{1}{x_t^A} \gamma^A \sigma_D, \quad \text{and } \sigma_x^A = \left(1 - x_t^A\right) \sigma_D \tag{33}$$

Clearly, the Sharpe ratio satisfies $\kappa_t = \frac{\gamma^A \sigma_D}{x_t^A} \geq \gamma^A \sigma_D$. In other words, the Sharpe ratio exceeds the one that would obtain in an economy populated by a single representative agent with risk aversion γ^A. Morover, the Sharpe ratio is declining in x_t^A, and since the share of type A agents is procyclical ($\sigma_x^A = \sigma_D \left(1 - x_t^A\right) \geq 0$), the Sharpe ratio is countercyclical.

Note that setting $\gamma^B = \infty$ in equations (14) and (15) would lead to (33). In other words, for the purposes of the determination of the Sharpe ratio and the volatility of x_t^A, the limited participation model implies the same formulas as a model with heterogeneous preferences with the more risk averse agent having infinite risk aversion.

Repeating the steps of earlier sections, one can derive the drifts μ_x^i and the interest rate

---

[11]See Blanchard (1985) for further details.

r_t. The expression for μ_x^A continues to be given by (16). Using (32), the definition of x_t^B is now

$$x_t^B = \frac{\nu^B \int_{-\infty}^t \pi e^{-\pi(t-s)} c_{s,s}^B e^{\int_s^t (r_u - \rho) du} ds}{D_t},$$

and therefore Ito's Lemma implies that

$$\mu_x^B = \left[(r_t - \rho) - \mu_D + \sigma_D^2 - \pi \right] x_t^B + \pi \nu^B \frac{c_{t,t}^B}{D_t}. \tag{34}$$

Requiring that markets clear implies that $\mu_x^A + \mu_x^B = 0$, which leads to the equilibrium interest rate

$$r_t = \left(\frac{x_t^A}{\gamma^A} + x_t^B \right)^{-1} \times \tag{35}$$

$$\left[\rho \left(\frac{x_t^A}{\gamma^A} + x_t^B \right) + \mu_D \underbrace{- \sigma_D^2 - x_t^A \left(\frac{1}{2} \left(\frac{\kappa_t}{\gamma^A} \right)^2 \left(\frac{1}{\gamma^A} + 1 \right) - \frac{\kappa_t \sigma_D}{\gamma^A} \right)}_{\text{Impact of type } A\text{'s precautionary savings}} + \pi \left(1 - \sum_{i \in \{A,B\}} \frac{\nu^i c_{t,t}^i}{D_t} \right) \right].$$

Unlike the expression for the Sharpe ratio, the equilibrium interest rate in the presence of limited participation cannot be obtained by simply setting $\gamma^B = \infty$ in the heterogeneous-preferences model. There are, nonetheless, several similarities between the interest rate in the two models. In both models, the inverse of the weighted IES of the two agents,[12] $\frac{x_t^A}{\gamma^A} + x_t^B$, multiplies a term that is seemingly identical across the two models; the main difference is that in the limited participation model the term reflecting the effect of precautionary savings only captures the precautionary savings undertaken by agent A. This is intuitive , since only agent $A's$ consumption has a positive diffusion component.

I conclude this section with an important caveat. Equation (33) illustrates a potential problem of models with limited participation, namely that the Sharpe ratio may not be bounded. Indeed as x_t^A approaches zero, κ_t grows without bound. This is unlike the models

---

[12]Recall that the type B agents have an IES of one.

of heterogenous preferences or beliefs, where both the Sharpe ratio and the interest rate are bounded. Since the Sharpe ratio is unbounded, Novikov's condition does not automatically apply. For instance, as is shown by Hugonnier (2012) in the context of the Basak and Cuoco (1998) model, the marginal utility of the participating investor is not a mathematically valid stochastic discount factor and valuation bubbles are needed to support the equilibrium. A full discussion of this issue is beyond the scope of this chapter. (The literature review includes references for further reading). In that sense the reader should interpret the results on non-participation as purely formal, i.e. with an understanding that some further assumptions need to be made to ensure the existence of a stochastic discount factor. For instance one could assume that when the Sharpe ratio becomes sufficiently high, then some non-participants may switch to the group of participants, an approach taken by Khorrami (2019a).

2.4.1 Literature Review

Basak and Cuoco (1998) solves a model very similar to the one considered in this section, but utilizing different techniques. Specifically Basak and Cuoco solve their model by using the concept of a central planner with state dependent weights. This approach does not apply in a perpetual youth framework and therefore section 2.4 used a direct approach to solving for equilibrium. See also Chabakauri (2013) for a different approach to solving models with heterogeneous investors and portfolio constraints.

The idea that limited stock market participation may be to blame for the poor performance of asset pricing models goes back at least to Mankiw and Zeldes (1991). Vissing-Jorgensen (2002) estimates the intertemporal elasticity of substitution for different groups of investors, depending on the markets that they choose to participate in. Parker and Vissing-Jorgensen (2009) discusses who bears aggregate fluctuations.

Slight modifications of the techniques to solve limited participation models can be applied to a variety of constraints (leverage constraints, shorting constraints, etc.). Indicative examples include He and Modest (1995), Detemple and Murthy (1997) and Detemple and Serrat (2003).

Favilukis (2013) , Guvenen (2005) and Gomes and Michaelides (2008) use numerical methods to solve models of limited participation and discuss the joint implications of these models for the equity premium and participation decisions.

Hugonnier (2012) discusses the technical problems that arise in limited participation models. Khorrami (2019a) discusses entry in such models, which can help keep the Sharpe ratio bounded.

2.5 Recursive preferences and the Sharpe ratio

Using recursive preferences, Gârleanu and Panageas (2015) solve a model whereby investors are heterogeneous both in terms of their risk aversion, but also their IES. Specifically, type A investors have risk aversion γ^A and IES equal to $\left(1 - \alpha^A\right)^{-1}$ and investors of type B have risk aversion γ^B and IES equal to $\left(1 - \alpha^B\right)^{-1}$. Without loss of generality, $\gamma^A < \gamma^B$.

In the presence of recursive preferences[13] the optimal consumption of agent i is[14]

$$
\frac{c_{u,t}^i}{c_{t,t}^i} = e^{\frac{1}{\gamma^i} \int_t^u \left(\Xi_1^i \left(g_w^i\right)^{-1} + \Xi_2^i + \pi\right) dw} \left(\frac{g_u^i}{g_t^i}\right)^{\frac{\Xi_1^i}{\gamma^i}} \left(\frac{H_u}{H_t}\right)^{-\frac{1}{\gamma^i}}
\tag{36}
$$

where $\Xi_1^i \equiv -\frac{\alpha^i + \gamma^i - 1}{\alpha^i}, \Xi_2^i \equiv -\frac{\rho + \pi}{\alpha^i}\left(1 - \gamma^i\right),$ and g_t^i is defined in equation (19). In the special case where the inverse of the IES is equal to the risk aversion of the agent $\left(1 - \alpha^i = \gamma^i\right),$ $\Xi_1^i = 0$ and $\Xi_2^i = -\left(\rho + \pi\right)$ and hence equation (36) reduces to (8).

Using (36) inside the definition of x_t^i (equation (9)), applying Ito's Lemma to obtain $\sigma_x^i,$ and requiring that $\sum \sigma_x^i = 0$ leads to the following expression for the Sharpe ratio

$$
\kappa_t = \Gamma_t \sigma_D - \sum_{i \in \{A,B\}} \omega_t^i \left(\frac{1 - \gamma^i - \alpha^i}{\alpha^i}\right) \frac{g^{i\prime}(x_t^A)}{g^i(x_t^A)} \sigma_{x,t}^A,
\tag{37}
$$

---

[13] For a discussion of recursive preference in continuous time, see e.g., Duffie and Epstein (1992).

[14] Note that equation (36) differs from equation (A.21) in the appendix of Gârleanu and Panageas (2015), because in this survey I define $g_t^i(x_t^A)$ as the wealth-to-consumption ratio. Gârleanu and Panageas (2015) defines $g_t^i(x_t^A)$ as the consumption-to-wealth ratio. This explains some differences between the formulas in this section and Gârleanu and Panageas (2015).

where $\Gamma_t \equiv \left(\sum\limits_{i=\{A,B\}} x_t^i \left(\gamma^i\right)^{-1} \right)^{-1}$, $\omega_t^i \equiv \left(\gamma^i\right)^{-1} \Gamma_t x_t^i$, and $g^{i\prime}$ is the partial derivative of g^i with respect to x_t^A.

The Sharpe ratio is the sum of two terms. The first term $(\Gamma_t \sigma_D)$ is identical to equation (14). The second term is a weighted average of a term that measures the discrepancy from expected utility $\left(\frac{1-\gamma^i-\alpha^i}{\alpha^i} \right)$ times the volatility of the two agents' wealth-to-consumption ratios. When all agents are neutral to the resolution of uncertainty $(1 - \alpha^i = \gamma^i)$, equation (37) reduces to (14).

When, however, $1 - \gamma^i - \alpha^i \neq 0$ and $\frac{g^{i\prime}}{g^i} \sigma_{x,t}^A \neq 0$ for some agent i, then the Sharpe is no longer a simple harmonic average of the Sharpe ratios that would obtain in an economy populated exclusively by type A or exclusively by type B agents. (In such single-agent economies the Sharpe ratio would be given by $\gamma^i \sigma_D$, since the aggregate endowment growth is i.i.d.).

The reason for the additional term in equation (37) is that the equilibrium fluctuations of wealth and consumption shares lead to persistent components in individual consumption growth, even though the growth rate of the aggregate endowment is i.i.d.. As is well understood in the literature on "long run risks" these persistent components affect the Sharpe ratio whenever investors are not neutral to the resolution of uncertainty $1 - \gamma^i - \alpha^i \neq 0$.

The fact that the Sharpe ratio is no longer a simple average of the Sharpe ratios that would obtain in pure type-A or type-B economies, implies that the Sharpe ratio in an economy with heterogenous recursive preferences may even exceed the Sharpe ratio that would obtain in an economy populated exclusively by the most risk-averse agents. Gârleanu and Panageas (2015) provide one such example along with conditions that make the second term of equation (14) positive. By contrast, in economies with either heterogeneous preferences or beliefs, the Sharpe ratio is bounded between the Sharpe ratios that would obtain in the respective single-agent economies.
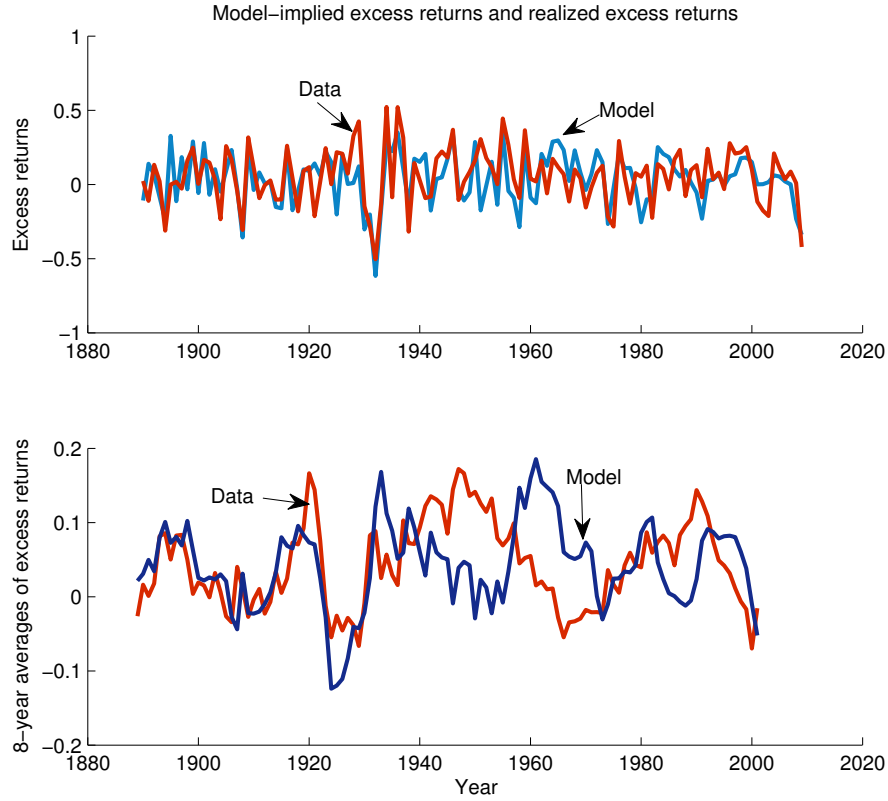
Figure 1: Year-over-Year and 8-year average excess returns in the Gârleanu and Panageas (2015) model plotted against the data.

2.6 Empirical Implications

The recurring theme of all the models presented sofar has been the counter-cyclicality of the Sharpe ratio. Campbell and Cochrane (1999) emphasized the importance of a counter-cyclical Sharpe ratio in terms of providing a unified explanation of asset pricing puzzles in a consumption-based framework.

As a result, it should not come as a surprise that calibrated heterogeneous-agent models can be successful in terms of explaining the data. Perhaps the most demanding test for an asset pricing model is to use the historical consumption path as an input and test whether the model-implied excess returns match up to the historical excess returns.

Figure 1 provides an illustration of such an exercise for the calibrated version of the

heterogeneous-preferences model of Gârleanu and Panageas (2015).[15] As is evident from Figure 1, the model manages to reproduce the historically observed path of the equity premium, both at the one- and eight-year horizons, simply by taking consumption data as an input. The main exception are the two decades after the second World War, where the model originally under- and then over-pedicts the realized excess returns. Outside of equities, Schneider (2017) discusses the empirical implications of this model for the term structure and bond risk premia.

One should not jump to the conclusion that any model featuring a countercyclical Sharpe ratio will automatically match the data. To provide a simple counter-example, if one were to take a model of heterogenous agents (say a limited participation model), but with all agents having unit IES, the price-to-aggregate-consumption ratio would be constant and return volatility would be counterfactually low, despite the Sharpe ratio being countercyclical. This happens because the same wealth shifts that cause the Sharpe ratio to decline (increase) also cause precautionary savings to decline (increase), making the interest rate move in the opposite direction of the Sharpe ratio, thus cancelling the overall impact of the wealth movements on the total discount rate. To match the asset pricing data, a heterogeneous-agents model must not only feature a sufficiently countercyclical Sharpe ratio, but also have some mechanism to keep the real interest rate stable in response to wealth shifts, which is a feature of both Campbell and Cochrane (1999) (in the context of a representative agent model) and Gârleanu and Panageas (2015) (in the context of a heterogeneous agents model).[16]

However, the most important distinguishing feature of heterogeneous agents model from their closest representative-agent relative (the Campbell and Cochrane (1999) model) is the fact that external habits are unobservable, in contrast to wealth distributions. Heterogeneous-agents models contain predictions on the joint movement of asset prices and the wealth distribution, which are absent in representative agent models. The added value of such models

---

[15]The reader is referred to the paper for details of the calibration

[16]Gârleanu and Panageas (2015) achieves a stable interest rate by assuming that investors with a low risk aversion also have high intertemporal elasticity of substitution. As a result, shifts in precautionary savings are offset by shifts in the consumption weighted intertemporal elasticity of substitution.

lies in testing these joint implications, which impart an additional layer of discipline on the theory. Recent empirical studies of these joint implications include Gomez (2017) (which studies the implications of these models for the tail of the wealth distribution), Akira Toda and Walsh (2019) (which focuses on the joint movements of the top wealth shares and equity prices), and Greenwald et al. (2019), which provides a framework for accounting for equity valuations in an environment of shifting factor shares.

3 Heterogeneous endowments

3.1 Lack of intra-cohort risk sharing

Cross-sectional data suggest that within any given birth cohort the cross-sectional heterogeneity of income is large and generally increasing as the cohort ages. A large literature in macroeconomics has investigated whether the increasing cross-sectional dispersion of income over the life-cycle of any given cohort also translates into an increasing consumption dispersion, suggesting incomplete risk sharing within each cohort.

The finance literature has generally taken the lack of risk sharing as given and asked whether the associated dispersion of consumption growth rates is relevant for risk premiums. At first pass, Grossman and Shiller (1982) would suggest that consumption growth dispersion caused exclusively by incomplete risk sharing should be irrelevant for risk premiums. Inspection of equation (3) shows that if all agents have the same preferences, then return differentials between different assets are still determined by the aggregate CAPM, no matter how cross-sectional consumption heterogeneity is specified.

In a seminal paper, Constantinides and Duffie (1996) reach a conclusion that seems to contradict the results of Grossman and Shiller (1982). Specifically, in a discrete time model Constantinides and Duffie (1996) showed that any stochastic discount factor *in a specific class* can be supported as the equilibrium stochastic discount factor, if the individual consumption growth process follows some appropriately specified dynamics. For instance, an implication of Constantinides and Duffie (1996) is that even if aggregate consumption

growth is constant, then the discount factor can still be stochastic (e.g., exhibit a positive risk premium), something that would be impossible in an economy where the aggregate consumption CAPM holds. In that sense Grossman and Shiller (1982) and Constantinides and Duffie (1996) seem to contradict each other.

This section reconciles the findings of the two papers. To do this, and compare the results of the two papers more effectively, it is essential to develop a minimal modification of the Constantinides and Duffie (1996) model that allows the time interval between decision to become arbitrarily small. This is important because the Grossman and Shiller (1982) argument becomes exact only as the time interval shrinks to zero.

Specifically, consider again an endowment economy and simplify things by assuming no births or deaths, but a continuum of infinitely lived agents (or more appropriately altruistically linked "dynasties"). This will jeopardize stationarity, but stationarity is irrelevant for the arguments of this section.

Assume next that each agent's income follows a process of the form

$$I_{i,t} = s_{i,t} \times C_t - D_t$$

where C_t is aggregate consumption, D_t are the dividends of the (representative) "Lucas tree" in this economy and $s_{i,t}$ is a share process with the property $\int_i s_{it} di = 1$ for all t.

The next assumption is key for the results. Suppose that $s_{i,t}$ obeys the following process:

$$s_{i,t+\Delta} = s_{i,t} \times \exp\left\{ -\frac{(\Lambda_{t+\Delta} - \Lambda_t)}{2} + \left(B^{(i)}_{\Lambda_{t+\Delta}} - B^{(i)}_{\Lambda_t} \right) \right\}, \tag{38}$$

where Λ_t is a *positive, increasing, discontinuous, random process* and $B^{(i)}_{\Lambda_t}$ is a Brownian motion that is specific to each agent i and independent of all other brownian motions $B^{(j)}_{\Lambda_t}$, $j \neq i$. Note that the variance of $B^{(i)}_{\Lambda_{t+\Delta}} - B^{(i)}_{\Lambda_t}$ between times $t + \Delta$ and t is not Δ, but is random and equal to $\Lambda_{t+\Delta} - \Lambda_t \geq 0$. The process Λ_t is common to all agents, and thus it controls the cross-sectional dispersion of the increments $B^{(i)}_{\Lambda_{t+\Delta}} - B^{(i)}_{\Lambda_t}$ between t and $t + \Delta$. Because the process Λ_t can only increase, but not decrease, it cannot be a diffusion. It is

instead a "jump process".

Assumption (38) implies that $s_{i,t}$ is a martingale. The law of the iterated expectation together with standard formulas for the expectation of a (conditionally) log-normal variable imply that $E_t \{s_{i,t+\Delta}\} = s_{i,t}$.[17] Initializing the process at $s_{i,0} = 1$, we have that for any time t, $E_t \{s_{i,t}\} = 1$ and a version of the law of large numbers implies that $\int_i s_{i,t} di = 1$ for all times t.

While the specification (38) is slightly different than the one employed in Constantinides and Duffie (1996), it preserves the key economic element of that specification, namely the presence of a shock $(\Lambda_{t+\Delta} - \Lambda_t)$ that controls the cross-sectional dispersion of $\frac{s_{i,t+\Delta}}{s_{i,t}}$. One advantage of (38) is that it allows a study of the limit as Δ becomes small, which allows a comparison with the results of Grossman and Shiller (1982). A further advantage is that $s_{i,t}$ can be chosen to be "infinitely divisible".[18] In practical terms, the assumption of infinite divisibility helps ensure that all risk premiums are the same (per unit of time) irrespective of the length of the time-interval Δ.

From this point onward, the construction of an equilibrium follows a "guess and verify" approach. Suppose that the representative investor maximizes an objective of the form

$$E_t \sum_{t_i=0,\Delta,2\Delta,\dots} e^{-\rho t_i} \frac{c_{t_i}^{1-\gamma}}{1-\gamma} \Delta. \tag{39}$$

Guess that in equilibrium there is no trade in financial markets, in which case each investor holds one share of the tree, no bonds, and consumes her endowment and the dividends, so that $c_{i,t} = s_{i,t} C_t$. Next, postulate that the (gross) interest rate and the stock market price

---

[17]

$$E_t \{s_{i,t+\Delta}\} = s_{i,t} \times E_t \left\{ E_t \left[\exp\left(-\frac{\Lambda_{t+\Delta} - \Lambda_t}{2} + \left(B_{\Lambda_{t+\Delta}}^{(i)} - B_{\Lambda_t}^{(i)} \right) \right) \middle| \Lambda_{t+\Delta} - \Lambda_t \right] \right\}$$
$$= s_{i,t} \times E_t \{1\} = s_{i,t}.$$

[18]One possible example is to choose Λ_t to have increments that are Gamma distributed. In this specific case, $\log(s_{i,t})$ follows a so-called variance-gamma process, which has a known (Laplace) distribution. This specification is quite tractable and has been used in other contexts in finance (e.g, option pricing) to model fat-tailed distributions.

P_t are given by

$$R^f_{t\to t+\Delta} = \frac{1}{E_t e^{-\rho\Delta}\left(\frac{c_{i,t+\Delta}}{c_{i,t}}\right)^{-\gamma}}, P_t = \Delta \times E_t \sum_{t_i=t+\Delta, t+2\Delta,...} e^{-\rho t_i}\left(\frac{c_{i,t_i}}{c_{i,t}}\right)^{-\gamma} D_{t_i}. \qquad (40)$$

Since the distribution of $\frac{c_{i,t+\Delta}}{c_{i,t}} = \frac{s_{i,t+\Delta}C_{t+\Delta}}{s_{i,t}C_t}$ is the same for all i at time t, it follows that both $R^f_{t\to t+\Delta}$ and P_t do not depend on i. This means that the postulated pair $R^f_{t\to t+\Delta}, P_t$ would make the Euler equations of all investors hold, if they chose to not trade in the bond market and simply hold one share of stock. Alternatively phrased, not trading is optimal. Moreover, if everyone holds a zero allocation in bonds and a unit allocation in stocks, both the bond market and the stock market clear. So does the consumption market, since $\int_i c_{i,t} di = \left(\int_i s_{i,t} di\right) \times C_t$.

An alternative way of writing equation (40) is to define the stochastic discount factor as the expected value (across all agents i) of the marginal utility of their consumption conditional on $\Lambda_{t+\Delta} - \Lambda_t$

$$\begin{aligned}\frac{H_{t+\Delta}}{H_t} &\equiv e^{-\rho\Delta}E_{(i)}\left\{\left(\frac{c_{i,t+\Delta}}{c_{i,t}}\right)^{-\gamma}|\Lambda_{t+\Delta} - \Lambda_t\right\}\\ &= \underbrace{e^{-\rho\Delta}\left(\frac{C_{t+\Delta}}{C_t}\right)^{-\gamma}}_{\text{Aggregate consumption CAPM}} \times \underbrace{e^{\frac{\gamma(\gamma+1)}{2}(\Lambda_{t+\Delta}-\Lambda_t)}}_{\text{Impact of heterogeneity}},\end{aligned} \qquad (41)$$

in which case the law of the iterated expectation implies that (40) can be put in the familiar form $R^f_{t\to t+\Delta} = \left[E_t\left(\frac{H_{t+\Delta}}{H_t}\right)\right]^{-1}$ and $P_t = \Delta \times E_t \sum_{t_i=t+\Delta,..} \frac{H_{t_i}}{H_t} D_{t_i}$.

An implication of (41) is the following: Suppose that one desired to choose some process Λ_t in a way that would lead to some desired stochastic discount factor H^*_t as an equilibrium outcome. Then (41) implies that one would simply have to define the process $\Lambda_{t+\Delta} - \Lambda_t$ as

$$\Lambda_{t+\Delta} - \Lambda_t = \frac{2}{\gamma(\gamma+1)}\left[\log H^*_{t+\Delta} - \log H^*_t + \rho\Delta + \gamma(\log C_{t+\Delta} - \log C_t)\right]. \qquad (42)$$

To ensure that Λ_t is an increasing process, one needs to assume that the right hand side of (42) is positive, irrespective of the realization of uncertainty. Therefore, $\log(H^*_t) - \log(e^{-\rho t}C_t^{-\gamma})$

needs to be an increasing process (assumption 6 in Constantinides and Duffie (1996)).

To summarize, the results of Constantinides and Duffie (1996) imply that heterogeneity matters for asset prices. Even if aggregate consumption growth is constant, the stochastic discount factor can be stochastic and claims whose value depend on the realization of the process Λ_t can exhibit risk premiums. It is important to note that the stochastic process H_t^* cannot be chosen arbitrarily. The discrepancy between $\log H_t^*$ and the log-SDF that would be implied by the aggregate consumption CAPM $\log(e^{-\rho(t+\Delta)}C_{t+\Delta}^{-\gamma})$ must be an increasing process.

But shouldn't the Grossman and Shiller (1982) result imply that the ability to support such a stochastic discount factor disappears as Δ approaches zero? The answer is no. The next example shows that even as the time interval shrinks to zero, the risk premium (per unit of time Δ) stays constant.

Example 1 *Suppose that $\frac{C_{t+\Delta}}{C_t} = e^{g\Delta}$, i.e., aggregate consumption grows deterministically. Define $x_{t+\Delta} \equiv (\Lambda_{t+\Delta} - \Lambda_t)$, and suppose that $x_{t+\Delta}$ is gamma distributed with density:*

$$f(x_{t+\Delta}) = \frac{\beta^{\alpha\Delta}}{\mathcal{G}(\alpha\Delta)}(x_{t+\Delta})^{\alpha\Delta-1}e^{-\beta x_{t+\Delta}}$$

for some parameters $\alpha > 0, \beta > \frac{\gamma(\gamma+1)}{2}$ and $\mathcal{G}(\alpha\Delta)$ the gamma function evaluated at $\alpha\Delta$. Consider a one-period state-dependent claim with cash flow $e^{-\eta x_{t+\Delta}}$ for some $\eta > 0$. This claim delivers lower payoffs the higher is the cross sectional dispersion $x_{t+\Delta}$. Letting $E(R_{t\to t+\Delta})$ denote the expected return of such a claim, and using the closed-form expressions for the moment-generating function of a gamma-distributed variable gives after some calculations:

$$\frac{\log\left(\frac{E(R_{t\to t+\Delta})}{R_{t\to t+\Delta}^f}\right)}{\Delta} = \frac{1}{\Delta}\log\left(\frac{E\left(e^{\frac{\gamma(\gamma+1)}{2}x_{t+\Delta}}\right)E\left(e^{-\eta x_{t+\Delta}}\right)}{E\left(e^{\left(\frac{\gamma(\gamma+1)}{2}-\eta\right)x_{t+\Delta}}\right)}\right) = -\alpha\log\left(\frac{\beta-\frac{\gamma(\gamma+1)}{2}}{\beta-\frac{\gamma(\gamma+1)}{2}+\eta}\frac{\beta+\eta}{\beta}\right) > 0.$$

$$(43)$$

There are several observations about (43). First, the expected excess return of the pos-

tulated claim (per unit of time Δ) is constant and does not depend on Δ. This is despite the fact that aggregate consumption growth is constant, which would seemingly imply (by the Grossman and Shiller (1982) argument) that the risk premium should approach zero as Δ shrinks to zero.

The reason why Grossman and Shiller (1982) does not apply is that the individual consumption process $s_{i,t}$ is not a diffusion in this example, but to a process with discontinuous sample paths (jumps). Intuitively, this happens because Λ_t is random and discontinuous, which makes the paths of B_{Λ_t} discontinuous, despite the fact that B_t itself is continuous in t. By contrast, for the Grossman and Shiller (1982) approximation to become exact as Δ becomes small, one has to assume that the underlying data-generating process for individual consumption growth is a diffusion, so as to apply Ito's Lemma and make equation (2) hold as an exact equation.

The above discussion shows that the Constantinides and Duffie (1996) argument does not depend on whether the model is set up in discrete or continuous time, but rather on whether the model is set up to make higher order moments of consumption growth relevant. In a diffusion setup, any non-linear function (such as the marginal utility of consumption) behaves like a linear function in the continuous time limit. Accordingly, the covariance of an individual's marginal utility of consumption with an excess return becomes proportional to the covariance of the individual's consumption growth with the same return. But when individual consumption growth doesn't have continuous sample paths, then one cannot simply approximate the marginal utility of consumption with a linear function, since consumption can experience large moves even as the time interval shrinks toward zero. In these situations, the covariance of higher moments of individual consumption growth with an asset's return become relevant, reflecting the fears of risk-averse agents that if re-distribution abruptly intensifies (a high jump of $\Lambda_{t+\Delta} - \Lambda_t$) they could find themselves among the losers of this redistribution at the same time as when the asset's returns are bad.

3.1.1 Literature review

While most of the discussion in this section focused on Constantinides and Duffie (1996), there are numerous papers in the literature that have explored the interactions of market incompleteness and asset pricing. A precursor to Constantinides and Duffie (1996) is Mankiw (1986). Krueger and Lustig (2010) discuss an irrelevance result in a discrete time model utilizing a different argument than Grossman and Shiller (1982), which applies to discrete time models. Lucas (1994), Heaton and Lucas (1996), Telmer (1993), Luttmer (1996), Angeletos and Calvet (2006), Kogan et al. (2019), Gârleanu and Panageas (2018) all discuss alternative setups of imperfect risk sharing with asset pricing implications. Storesletten et al. (2004), Cogley (2002), Brav et al. (2002), Kocherlakota and Pistaferri (2009) are empirical studies of the implications of incomplete risk sharing for asset pricing.

Several papers in the literature study models of investors facing liquidity or collateral constraints. In this literature, the investor's indebtedness is constrained by her ability to pledge some collateral whenever she wishes to borrow. The most well studied version of such models considers "liquidity" constraints, which allow an investor to borrow, but only against the value of the liquid assets in her portfolio. Importantly, borrowing against future labor income is not allowed.

The difference between discrete-time and continuous-time trading manifests itself in models of liquidity (or collateral) constraints as well, especially as it pertains to the implications of such models for risk premia. Indicatively, Alvarez and Jermann (2000) find that solvency constraints matter for the market price of risk in discrete time, while Detemple and Serrat (2003) prove that the related concept of liquidity constraints don't affect the conditional aggregate consumption CAPM in the continuous-time limit, implying that the quantitative results of discrete time models will tend to disappear as the trading interval shrinks.

One way to understand the different conclusions of continuous-time and discrete-time setups is to view them through the lens of the "Lagrange-Multiplier approach" of and Marcet and Marimon (1998). Both He and Pages (1993) and Marcet and Marimon (1998) observe that models with liquidity constraints (or collateral constraints more generally) lead to an

optimal individual consumption process that is the product of two terms: a) the consumption process that would result in a frictionless market and b) a non-decreasing, continuous process that is related to the cumulative sum of the Lagrange multipliers on the liquidity constraint. Because this second process is non-decreasing, it has bounded variation and zero quadratic variation in continuous time. The implication of zero quadratic variation is that the fluctuations caused by this process cannot command a risk price. Detemple and Serrat (2003) use this insight in general equilibrium to prove the absence of risk premia associated with liquidity constraints.

The economic intuition is that if asset price processes and consumption processes follow continuous-time diffusions, the continuity of the paths allows investors to essentially know (at least over the next small time interval) whether the liquidity constraint is likely to bind or not. As a result, there is no genuine "shock" or "risk" associated with liquidity constraints in the continuous time limit. This is unlike the setup of section 3.1, where the *discontinuity* of Λ_t faces the investor with "shocks" even over a small time-interval.

3.2 Lack of inter-cohort risk sharing

In the Constantinides and Duffie (1996) model the lack of risk sharing manifests itself within existing households. This section presents an alternative way to make heterogeneity matter, by re-introducing births and deaths and assuming that income risk is imperfectly shared across cohorts. The goal of the section is to show an equivalence between models featuring imperfect inter- and intra-cohort risk sharing.

Specifically, consider the following discrete time variant of the perpetual-youth, endowment economy of section 2.1, with the additional assumption that the depreciation (and arrival) rate of trees is random. Specifically, the dividend process for trees of cohort s, $D_{t,s}$, obeys the dynamics

$$\frac{D_{t+\Delta,s}}{D_{t,s}} = \left(\frac{D_{t+\Delta}}{D_t} \right) e^{-\delta_{t+\Delta}},$$

where $D_t = \sum_{s \leq t} D_{t,s}$ is the aggregate dividend process at time t and $\delta_{t+\Delta}$ is a non-negative, random, and i.i.d. shock across time. The initial dividends of the trees born at time t are given by

$$D_{t+\Delta,t+\Delta} = \left(1 - e^{-\delta_{t+\Delta}}\right) D_{t+\Delta}. \tag{44}$$

As in section 2.1, assume that the newly-born obtain the new trees. To highlight the differences from the results of that section, all agents have the same expected utility preferences and maximize (39).

The Euler equation for in this model is given by

$$E_t e^{-\rho\Delta} \left(\frac{c_{t+\Delta,s}}{c_{t,s}}\right)^{-\gamma} R_{t \to t+\Delta} = 1, \tag{45}$$

for the return of any asset. Note that the consumption growth entering equation (45) is the consumption growth $\frac{c_{t+\Delta,s}}{c_{t,s}}$ of a fixed cohort s, not aggregate consumption growth, $\frac{C_{t+\Delta}}{C_t}$. To relate the two notions of consumption growth, conjecture that $\frac{c_{t+\Delta,s}}{c_{t,s}}$ is independent of s, so that $\frac{c_{t+\Delta,s}}{c_{t,s}} = \frac{C_{t+\Delta}}{C_t}$. With this assumption, the consumption growth of a fixed cohort is given by[19]

$$\frac{c_{t+\Delta,s}}{c_{t,s}} = \frac{C_{t+\Delta}}{C_t} = \left(\frac{C_{t+\Delta}}{C_t} - \pi \frac{C_{t+\Delta,t+\Delta}}{C_t}\right) e^{\pi\Delta}. \tag{46}$$

The consumption growth of any fixed cohort is comprised of two multiplicative terms. The first term is equal to aggregate consumption, but reduced by the consumption fraction accruing to the newly-born agents. This term captures the idea that a fraction of aggregate consumption every period is the property of the arriving cohort and should be subtracted

---

[19]

$$
\begin{aligned}
C_{t+\Delta} &= \sum_{s \leq t+\Delta} \left(\pi e^{-\pi(t+\Delta-s)} c_{t+\Delta,s}\right) = \sum_{s < t+\Delta} \left(\pi e^{-\pi(t+\Delta-s)} c_{t,s} \frac{c_{t+\Delta,s}}{c_{t,s}}\right) + \pi c_{t+\Delta,t+\Delta} \\
&= \left(\frac{c_{t+\Delta,s}}{c_{t,s}}\right) e^{-\pi\Delta} \sum_{s \leq t} \left(\pi e^{-\pi(t-s)} c_{t,s}\right) + \pi c_{t+\Delta,t+\Delta} = \left(\frac{c_{t+\Delta,s}}{c_{t,s}}\right) e^{-\pi\Delta} C_t + \pi c_{t+\Delta,t+\Delta}
\end{aligned}
$$

from aggregate consumption growth in order to arrive at the consumption growth accruing to cohorts born before $t + \Delta$. The second term $\left(e^{\pi\Delta} > 1\right)$ is due to deaths. Since only a fraction $e^{-\pi\Delta}$ of agents survives from t to $t + \Delta$, per-capita consumption growth is increased by $e^{\pi\Delta}$.

Determining $\frac{c_{t+\Delta,t+\Delta}}{C_t}$ is particularly easy in this example because all agents have identical preferences. One can show that applying the budget constraint at birth and using equation (64) in the appendix leads to

$$\pi \frac{c_{t+\Delta,t+\Delta}}{C_t} = \left(1 - e^{-\delta_{t+\Delta}}\right) \frac{C_{t+\Delta}}{C_t}. \tag{47}$$

Combining (46) and (47) gives

$$\frac{c_{t+\Delta}}{c_t} = \frac{C_{t+\Delta}}{C_t} e^{-\delta_{t+\Delta}+\pi}$$

which implies that the stochastic discount factor is

$$\frac{H_{t+\Delta}}{H_t} = e^{-(\rho+\gamma\pi)\Delta} \left(\frac{C_{t+\Delta}}{C_t}\right)^{-\gamma} e^{\gamma\delta_{t+\Delta}}. \tag{48}$$

The stochastic discount factor of equation (48) is proportional to the SDF of the previous section (equation 41), as long as the increments of the process Λ_t are chosen as

$$\Lambda_{t+\Delta} - \Lambda_t = \frac{2}{\gamma + 1} \delta_{t+\Delta}.$$

The remarkable similarity in the SDFs of the two models is not just a coincidence. There are economic reasons for the correspondence. With lack of inter-generational risk sharing the existing cohorts of agents fear that a large value of $\delta_{t+\Delta}$ will lead to a large redistribution of dividends from existing trees toward the new vintages of trees. This redistribution benefits the (not yet born) entrepreneurs who introduce the new trees; however it hurts the owners of the existing trees, who understand that a large value of $\delta_{t+\Delta}$ shifts dividend income away from them ("displacement risk"). To achieve some risk mitigation, the owners of the existing

firms would have to trade with the not-yet born cohorts, who are also exposed to fluctuations in $\delta_{t+\Delta}$ but in the opposite direction (the arriving cohorts benefit from large values of $\delta_{t+\Delta}$). But since trade with the unborn is impossible, the risk $\delta_{t+\Delta}$ is not diversifiable for the existing cohort and hence becomes priced in the SDF.

There is a close analogy with models featuring lack of intra-cohort risk sharing, such as Constantinides and Duffie (1996). In such models a large increment in $\Lambda_{t+\Delta}$ causes a large dispersion in income and consumption growth. While this redistribution of the endowment could turn out to be either good or bad for a given agent, risk aversion makes them more concerned with the possibility of being amongst the losers of this redistribution rather than the winners. By overweighting the possibility of losing, their decisions resemble those of agents in models with imperfect inter-cohort risk sharing, who are certain that a large value of $\delta_{t+\Delta}$ is associated with a higher displacement for them. As a result, in both setups agents want a risk premium for assets that perform badly in states of the world where there is more re-distribution (either within or across cohorts).

These re-distributional fears may offer a potential explanation for the emergence of risk premia that seem somewhat unrelated to aggregate outcomes, but may affect the distribution of profits across different firms. For instance, increased "creative destruction", which redistributes profits both within existing and across existing and arriving firms can lead to the emergence of a "value premium". Firms that are more exposed to displacement by entrants ("value firms") must deliver a risk premium compared to firms that are less vulnerable to discplacement ("growth firms").[20]

3.2.1 Literature Review

The discussion of lack of inter-generational risk sharing is based on Gârleanu et al. (2012). Gârleanu and Panageas (2018) shows the connections between lack- of inter- and intra- cohort risk sharing in a framework where existing investors cannot share the benefits of innovation neither within, nor across cohorts. Kogan et al. (2019) also present a model where innovation

---

[20]See, Gârleanu et al. (2012) and Kogan et al. (2019), Gârleanu and Panageas (2018).

has redistribution implications across and within investor cohorts.

A recent set of papers (Gomez (2019), Zheng (2019)) investigate the importance of displacement effects for the dynamics of top wealth shares. The key issue in these papers is to disentangle the dynamics of top wealth shares into an intensive margin (wealth growth of the existing rich) and a displacement effect (entrants who displace the existing rich from the top wealth shares).

Indicative examples of papers that use overlapping generations models for asset pricing are Constantinides et al. (2002), Abel (2003), Geanakoplos et al. (2004), Krueger and Kubler (2006), Storesletten et al. (2007), Campbell and Nosbusch (2007), Gomes and Michaelides (2008), Maurer (2017), Farmer (2018). A large fraction of these papers employ discrete-time setups, with households choosing portfolios and consumption only at a few discrete points in their life-time (sometimes only once, or twice over the course of their lives). This makes it hard for many of these models to deliver predictions for returns as the trading frequency shortens, which can be a source of concern in light of the Grossman and Shiller (1982) result. For this reason, in this survey I adopted a perpetual youth framework, which allows trading to be modeled over any frequency without jeopardizing tractability.[21]

3.3 Recursive preferences

The increased popularity of long run risk models has led several researchers in recent years to propose alternative mechanisms in order to make heterogeneity matter, by exploiting recursive preferences. With recursive preferences, the Grossman and Shiller (1982) argument ceases to apply even if the underlying asset price and individual consumption processes are all diffusions in the continuous time limit. The reason is that equation (1) no longer holds. The stochastic discount factor "prices" not only the immediate impact of consumption shocks, but requires compensation also for consumption risks that affect consumption over the long run.

---

[21]It should be noted, however, that for some papers infrequent trading is a central feature that they intentionally employ, highlight, and analyze. See, e.g., Abel et al. (2007), Abel et al. (2013), Chien et al. (2012).

Of course, by themselves, recursive preferences might make no difference. For instance, if individual consumption growth is i.i.d. (as in the models of sections 3.1 and 3.2), then recursive preferences make no difference. However, if the interaction of income heterogeneity and imperfect risk sharing leads to persistent components in the drift or volatility of individual consumption growth, then heterogeneity does matter even if all underlying processes are diffusions.

To illustrate these ideas, I revisit the baseline continuous-time setup of section 2.1 , but assuming that a) all investors have identical, recursive preferences with an intertemporal elasticity of substitution equal to $(1 - \alpha)^{-1}$ and a risk aversion $\gamma > 1$ and b) the depreciation rate is stochastic, so that the dividend process of a tree born at time s is

$$D_{t,s} = \left(\delta_s e^{-\int_s^t \delta_u du} \right) D_t, \tag{49}$$

where D_t is the aggregate dividend. (By construction $\int_{-\infty}^t D_{t,s} ds = D_t$ for any positive depreciation process δ_t). Assume furthermore that δ_t follow some positive, Markov, stationary diffusion processes

$$d\delta_t = \mu_\delta (\delta_t) dt + \sigma_\delta (\delta_t) dB_t.$$

for some functions $\mu_\delta (\delta_t), \sigma_\delta (\delta_t)$. To sharpen results, and draw a clear distinction from Grossman and Shiller (1982) assume that the aggregate dividend is deterministic

$$\frac{\dot{D}_t}{D_t} = \phi, \tag{50}$$

i.e., there is no short run risk in this economy, so that the aggregate consumption CAPM would imply zero risk premia.

To analyze this model, note first that aggregate consumption growth and consumption growth for any fixed cohort s are related by[22]

---

[22]To arrive at this equation, time differentiate aggregate consumption $C_t = \int_{-\infty}^t \pi e^{-\pi(t-s)} c_{t,s} ds$ and use the fact that $\frac{\dot{c}_{t,s}}{c_{t,s}}$ is independent of s.
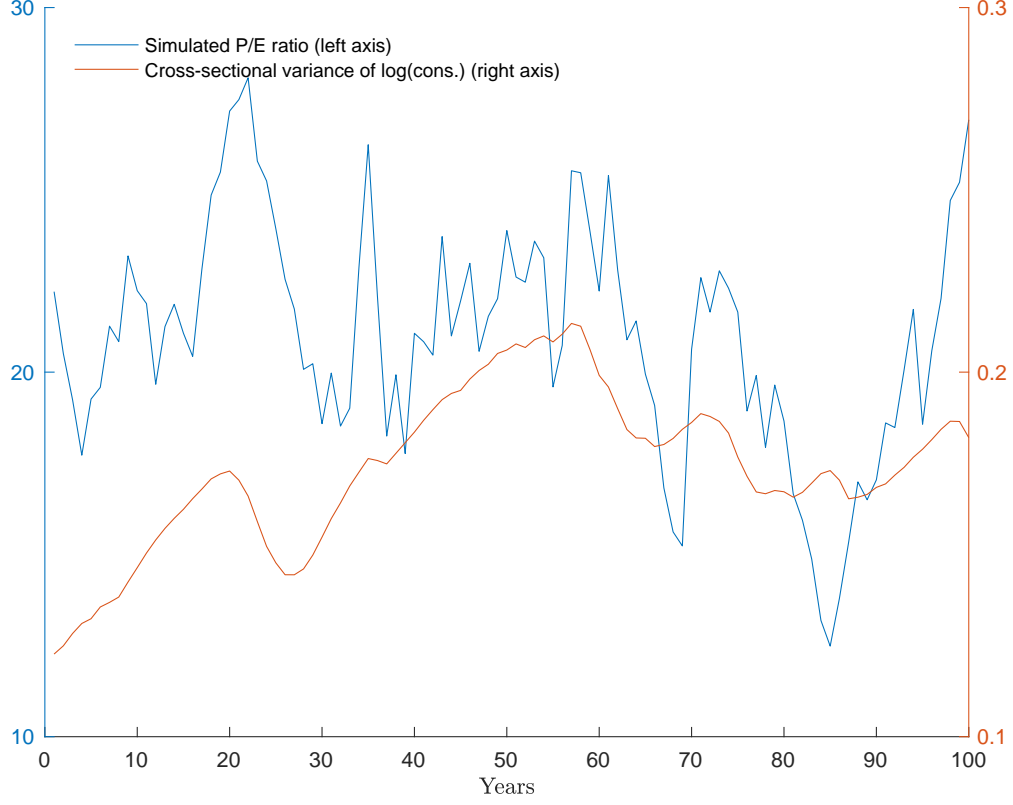
Figure 2: An indicative, model-implied path of the price-earnings ratio (left scale) and the cross- sectional standard deviation of log consumption (right scale) in the Gârleanu and Panageas (2019) model.

$$\frac{\dot{c}_{t,s}}{c_{t,s}} = \frac{\dot{C}_t}{C_t} - \pi \frac{c_{t,t}}{C_t} + \pi. \tag{51}$$

Equation (51) is the continuous time analog of equation (46). It states that the consumption growth of a fixed cohort is equal to aggregate growth minus the fraction of consumption accruing to the newly borns $\left(\pi \frac{c_{t,t}}{C_t}\right)$ plus the death rate π.

Using the intertemporal budget constraint at birth and equation (64) in the appendix implies that $g_t = p_t$ and therefore

$$\pi \frac{c_{t,t}}{C_t} = \delta_t, \tag{52}$$

Substituting (52) into (51) gives

$$\frac{\dot{c}_{t,s}}{c_{t,s}} = \phi + \pi - \delta_t. \tag{53}$$

This implies that the consumption growth of any fixed cohort has a stochastic drift (driven by the persistent component δ_t), even though aggregate consumption is deterministic. To determine the equilibrium risk premium, apply Ito's Lemma to both sides of equation (36) and match the diffusion components to arrive at

$$\kappa_t = -\frac{\alpha + \gamma - 1}{\alpha} \frac{g'(\delta_t)}{g(\delta_t)} \sigma_\delta(\delta_t), \tag{54}$$

where $g_t = g(\delta_t)$ is the wealth-to-consumption ratio defined in (19). Equation (54) is the Sharpe ratio that one would obtain in a Bansal and Yaron (2004) model without any short run risks. Indeed, if the IES is above one ($\alpha > 0$), an increase in the depreciation rate δ_t lowers expected consumption growth of any fixed cohort s, and hence the wealth-to-consumption ratio declines $g'(\delta_t) < 0$. If investors have preference for early resolution of uncertainty $\alpha + \gamma > 1$, this means that negative news for the drift rate of existing agents' consumption (and the dividends of existing firms) going forward, require risk compensation.

The takeaway from the above discussion is that imperfect risk sharing across cohorts leads to a model that resembles a representative-agent long run risks model. The major difference is that long-run risk in the Bansal and Yaron (2004) model is driven by persistent and random components in aggregate consumption growth, whereas here long-run risk is driven by persistent, random components in the consumption growth of a fixed cohort. Positive innovations to δ_t are negative news for the long run consumption growth of existing cohorts. However, δ_t is a purely re-distributional risk between existing and arriving cohorts. Aggregate consumption growth is constant.

3.3.1 Literature Review

The literature that uses heterogeneous-agents models together with recursive preferences is recent and includes Schmidt (2015), Constantinides and Ghosh (2017), (who utilize recursive preferences in a context of imperfect intra-cohort risk sharing) and Gârleanu and Panageas (2019) (which utilizes recursive preferences in a context of imperfect inter-cohort risk sharing).

3.4 Summary and Empirical Implications

Models of imperfect risk sharing are qualitatively different from models that feature heterogeneous preferences, beliefs, or access to markets. In models of imperfect risk sharing all agents have the same preferences, beliefs and access the same markets. However, either some key markets are missing (i.e., the markets for trading idiosyncratic income shocks), or some key market participants are missing (the unborn who cannot trade before their birth).

In models of imperfect risk sharing the aggregate consumption CAPM fails both conditionally and unconditionally. In these types of models it is possible to have positive risk premia even if aggregate consumption is deterministic.

The unifying mechanism behind all the models in section 3 is to drive a wedge between individual and aggregate consumption growth that is priced. The model of section 3.1 achieves this by exploiting higher-order movements in consumption growth, the model of section 3.2 achieves this by creating a random redistribution between young and old that resembles a jump-risk from the perspective of the old, while the model of section 3.3 makes the displacement activity "smoother" over time and introduces recursive preferences to make the representative agent unsure about its long term impact.

Ultimately, whether these models are plausible candidates to resolve asset pricing puzzles boils down to whether these models make empirically plausible predictions about the joint movement of inequality and asset returns.

Models that utilize expected utilities require a negative, short-run correlation between innovations to the cross-sectional consumption dispersion and asset returns. Especially as
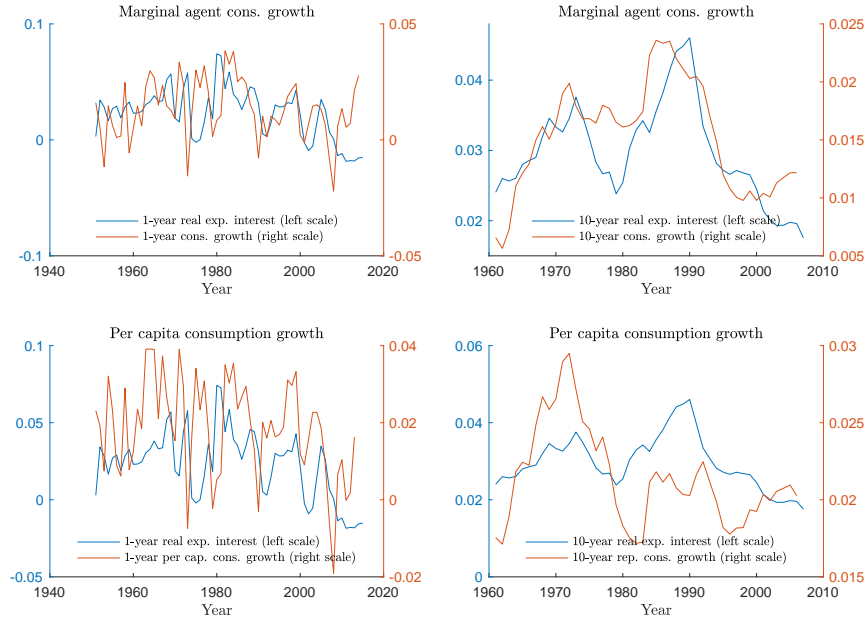
Figure 3: Top left plot: Expected real interest rate at the beginning of each year and "marginal-agent" consumption growth over the year. (For the computation of the expected real interest rate and "marginal-agent" consumption growth, see Gârleanu and Panageas (2019)). Top right plot: 10-year moving averages yearly marginal agent consumption growth and 10-year moving average of expected real interest rate. Bottom left and right plots: Identical to the top plots, except that marginal-agent consumption growth is replaced with aggregate consumption growth per capita.

one shrinks the time-interval to zero, this becomes more evident: for heterogeneity to matter, there have to be *joint* discontinuous jumps in both cross-sectional variance and risky asset returns.

One obvious response is that people don't trade very frequently in reality. Still, the issue remains that the asset return movements and shocks to inequality have to coincide over the course of a decision interval.

The model of recursive preferences in section 3.3 is more forgiving on this issue. Even with the first warning signs that she might be losing from redistribution, an investor wants risk compensation for assets with bad payoffs if redistributive activity intensifies unfavorably. The empirical implication is that innovations to inequality and asset returns may look completely

unrelated in the short run. Indeed in the model of section 3.3 , the cross-sectional variance of consumption follows a locally deterministic (time-differentiable) process, while asset prices follow a diffusion with positive volatility. This means that over short time intervals, there is no correlation between the two series. Figure 2 provides an illustration of this fact. The figure is from Gârleanu and Panageas (2019), which analyzes a model with similar properties to the one in section 3.3 .

Before concluding this section, it is worth noting that models of imperfect risk sharing (be it imperfect inter- or intra-cohort) have interesting implications not just for risk premiums but also for real interest rates. An implication of equation (41) is that a location shift of the increments $\Lambda_{t+1} - \Lambda_t$ to the right lowers the interest rate. The reason is that increased uncertainty about individual-level consumption growth boosts precautionary savings.

The same tendency (namely to obtain low real rates) is also a typical feature of overlapping generations models. As Gârleanu and Panageas (2015) show in a model that features both capital and labor income, the interest rate in an overlapping generations economy is generally lower than in an economy with infinitely lived investors, whenever the life cycle income profile declines with age. The reason is that agents in an overlapping generations economy cannot count on the income of their children to smooth their consumption and need to save more for the years when their income drops.

But leaving the level of interest rates aside, overlapping generations economies have different implications for the consumption growth $d\log(c_{t,s})$ of a fixed cohort versus aggregate consumption growth $d\log(C_t)$, with only the former being relevant for asset pricing. Gârleanu and Panageas (2019) propose an econometric technique to infer $d\log(c_{t,s})$, ("marginal agent consumption growth") utilizing a time-, age-, and cohort- decomposition of log consumption in cross sectional data.

According to the Euler equation, the drift of $d\log(c_{t,s})$ should reflect the movements of the real interest rate. Figure 3 shows that the inferred $d\log(c_{t,s})$ co-moves with the real interest rate better than aggregate consumption growth (per capita). Interestingly, marginal agent consumption growth starts declining steadily at the onset of the secular decline of the

real interest rate, which started in the mid-eighties.

I conclude this section with a digression on the co-movement between dividends and consumption.

3.4.1 Digression: The co-movement between the dividends of the market portfolio and consumption

Throughout this survey, the focus has been on the properties of the stochastic discount factor. For an equity premium to arise one has to also establish a link between the stochastic discount factor and the dividends of the market portfolio.

For instance in the models of sections 3.2 and 3.3, aggregate consumption is deterministic as are aggregate dividends. But then how can there be an equity premium?[23]

The short answer to this question is that even though aggregate dividends are deterministic, the dividends of the market portfolio are stochastic and not co-integrated with aggregate dividends. There is a subtle distinction between the dividends of the market portfolio and aggregate dividends that pertains not just to this survey, but all asset pricing models.

In the real world the number of companies that are included in the market portfolio is constantly expanding. While obvious, it is worth stressing that the market value of all companies in existence at time t equals the present value of the dividends that these companies (and these companies *alone*) will produce from t onwards. This is not the same as the present value of the aggregate dividends from t onwards, since some of the dividends in the future will be produced by future companies, which are not part of the market portfolio today.

This idea is reflected in the definition of the price-to-dividend ratio in equation (19), namely

$$p_t = E_t \int_t^\infty e^{-\delta(u-t)} \left(\frac{H_u}{H_t}\right) \left(\frac{D_{u,t}}{D_{t,t}}\right) du.$$

---

[23]Because the stock market is a long-lived claim, it could be that there is an equity premium that is a pure term premium. So the right question is: How can there be an equity premium above and beyond any term premium?

This definition says that the dividend growth in the definition of the price-to-dividend ratio is the dividend growth of a fixed firm, not aggregate dividend growth $\frac{\dot{D}_u}{D_t}$. The two notions of dividend growth are related as follows: Using the definition of aggregate dividend growth $D_t = \int_{-\infty}^{t} D_{t,s} ds$, implies the decomposition

$$\underbrace{\frac{\dot{D}_{t,s}}{D_{t,s}}}_{\text{Dividend growth of existing firms}} = \underbrace{\frac{\dot{D}_t}{D_t}}_{\text{Aggregate Dividend Growth}} - \underbrace{\frac{D_{t,t}}{D_t}}_{\text{Dividends of the new entrants}} . \quad (55)$$

The dividend-growth decomposition (55) is just an accounting identity that applies to any model with an expanding set of firms. A very similar decomposition applies to the capital gains of the market portfolio, since the ratio of prices to dividends is independent of s. Indeed, letting $P_{t,s}$ denote the market value of firms born into cohort s, and $P_t = \int_{-\infty}^{t} P_{t,s} ds$ denote the market capitalization of these firms gives

$$\underbrace{\frac{dP_{t,s}}{P_{t,s}}}_{\text{Capital Gains of existing firms}} = \underbrace{\frac{dP_t}{P_t}}_{\text{Growth in market capitalization}} - \underbrace{\frac{P_{t,t}}{P_t} dt}_{\text{Market Value of new entrants}} . \quad (56)$$

By the definition of a self-financing strategy, the return on any portfolio with weights $w_{t,s} ds$ summing to one is defined as $\int_{-\infty}^{t} w_{t,s} dR_{t,s}$, where $dR_{t,s} = \frac{dP_{t,s}}{P_{t,s}} + \frac{D_{t,s}}{P_{t,s}} dt$. Using the market weights $w_{t,s} = \frac{P_{t,s}}{P_t} ds$ implies that

$$dR_t = \int_{-\infty}^{t} w_{t,s} dR_{t,s} = \frac{dP_t}{P_t} + \frac{D_t}{P_t} dt - \frac{P_{t,t}}{P_t} dt. \quad (57)$$

An equivalent way to arrive at equation (57) is to define a "share" or "divisor" process $S_t = e^{\int^{t} \frac{P_{s,s}}{P_s} ds}$ and treat the entire stock market as a single company with value equal to the market capitalization P_t, but a time-varying number of shares S_t. The return "per-share" of the market portfolio is then defined as the change in the price of each share plus the

aggregate dividends per share

$$dR_t = \frac{d\left(\frac{P_t}{S_t}\right) + \frac{D_t}{S_t}dt}{\frac{P_t}{S_t}}. \tag{58}$$

Mathematically, (58) coincides with (57). Ecomomically, (58) shows that the dividends that are pertinent for the computation of the market portfolio are the dividends per share $\frac{D_t}{S_t}$.

An equivalent way of thinking about the dividends of the market portfolio in a world where the mass of productive units ("trees") is expanding, is to treat the stock market as consisting of a "representative" firm whose yearly profits are equal to aggregate profits D_t, but the firm invests $P_{t,t}$ to purchase the newly arriving productive entities. Accordingly, the payoff to investors is equal to profits minus investment $(D_t - P_{t,t})$ and the firm's market value can be expressed as

$$P_t = E_t\left\{\int_t^\infty \frac{H_u}{H_t}\left(D_u - P_{u,u}\right)du\right\} = E_t \int_t^\infty \frac{H_u}{H_t}\left(\frac{S_t}{S_u}\right)D_u du \tag{59}$$

where the last equation follows from the definition of S_t.[24]

Figure 4 plots the relation between the value of an index $\frac{P_t}{S_t}$ and aggregate market capitalization, while Figure 5 gives a visual depiction of the relation between aggregate dividends

---

[24]To derive (59) multiply both sides by H_t and add $-\int_0^t H_u P_{u,u}du + \int_0^t H_u D_u du$ to obtain

$$P_t H_t - \int_0^t H_u P_{u,u}du + \int_0^t H_u D_u du = E_t \int_0^\infty H_u D_u du.$$

The right hand side of the above equation is a martingale. Using the definition of $S_t = e^{\int^t \frac{P_{s,s}}{P_s}ds}$, one can write the above equation in differential form:

$$d\left(P_t H_t\right) - H_t P_t \left(\frac{dS_t}{S_t}\right) = -H_t D_t + dM_t,$$

where dM_t is a martingale increment. Noting that $d\left(P_t H_t\right) - H_t P_t \frac{dS_t}{S_t} = d\left(\frac{P_t H_t}{S_t}\right) \times S_t$ implies that the above equation can be written as

$$d\left(\frac{P_t H_t}{S_t}\right) = -\frac{H_t D_t}{S_t} + \frac{dM_t}{S_t}.$$

Integrating both sides and taking expectations (noting that dM_t is a martingale increment) gives (59).
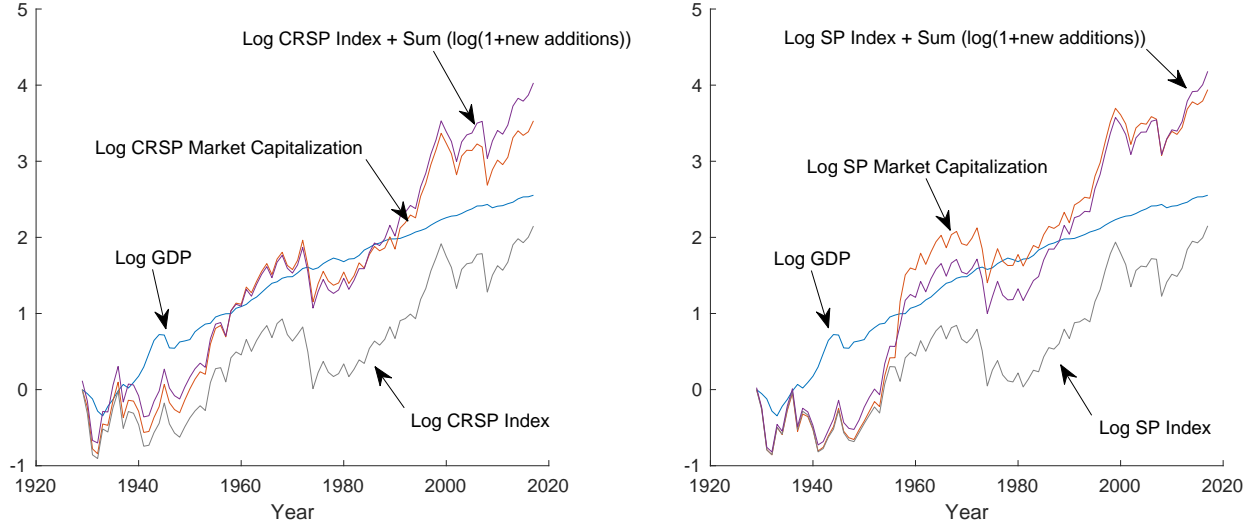
Figure 4: Left plot: Logarithm of GDP, logarithm of the CRSP market index (obtained by cumulating ex-dividend CRSP gross returns), log market capitalization, and log market capitalization plus the cumulative sum of the logarithm of 1+addition rate to the market. The addition rate is defined as the market value of additions to the index (valued at the end of each year) divided by the total value of the index at the end of each year. All series are deflated by subtracting the logarithm of the CPI. Right plot: Same as left plot, except that the market index is the S&P 500.

D_t and the dividends per share of the market portfolio $\frac{D_t}{S_t}$. These figures use the time series of price per share and dividends per share that are used in index construction. In the data, the number of shares of the index gets adjusted for a multitude of reasons not captured here (typically corporate payouts or issuances of existing firms). However, as the figure shows over long horizons, the dominant force in accounting for the discepancy between the (log) price-per-share of an index and (log) aggregate stock market capitalization is the cumulative sum of the addition rate $\frac{P_{t,t}}{P_t}$. The same applies to the discrepancy between log dividends-per-share and log aggregate dividends.

To return to the implications of these accounting identities for the model of section 3.3 , a positive shock to the depreciation rate δ_t acts as a permanent shock to both the consumption of a fixed cohort $c_{t,s}$ and the dividends of existing firms $D_{t,s}$. The investor realizes that because δ_t is persistent, this shock has a permanent effect on her long run consumption. Therefore she demands compensation for shocks to δ_t. The value of the market portfolio, being a long-
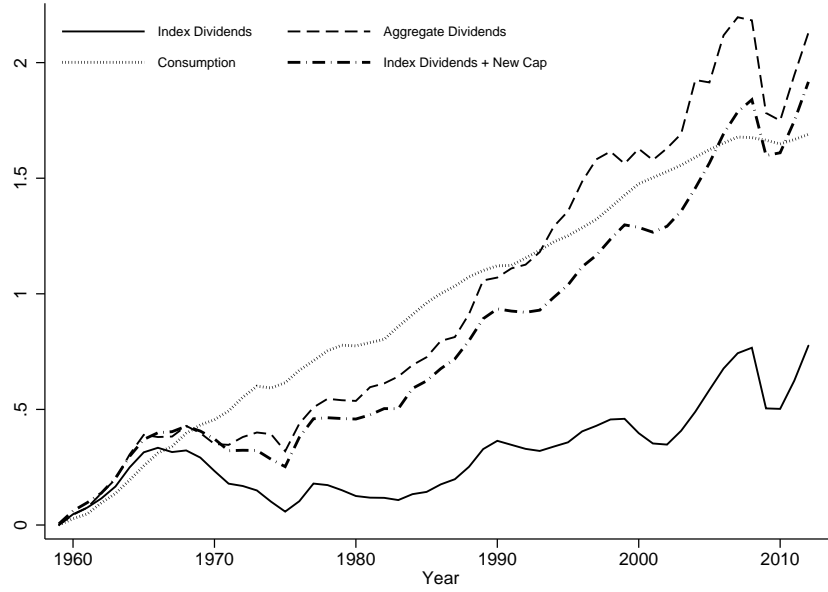
Figure 5: Real S&P 500 dividends per share, aggregate consumption and aggregate dividends in logarithms. The line "Index Dividends + New Cap" is equal to real log-dividends per share plus the progressive sum of the logarithm of the cumulative change in the number of shares of the index that is due to the introduction of new companies.

term claim to dividends-per-share, also drops in response to a shock to δ_t, and therefore an equity premium emerges.

4 Further applications: Equilibrium Arbitrage, heterogeneous stochastic discount factors

4.1 Equilibrium arbitrage

In models of market incompleteness, either a market, or some key market participants are missing. However, typically one can price *all* assets at time t by considering the consumption growth of a fixed agent at time t.

One possibility, though, is that no agent participates in all asset markets. In these situations, the existence of a stochastic discount factor is not guaranteed, idiosyncratic risk
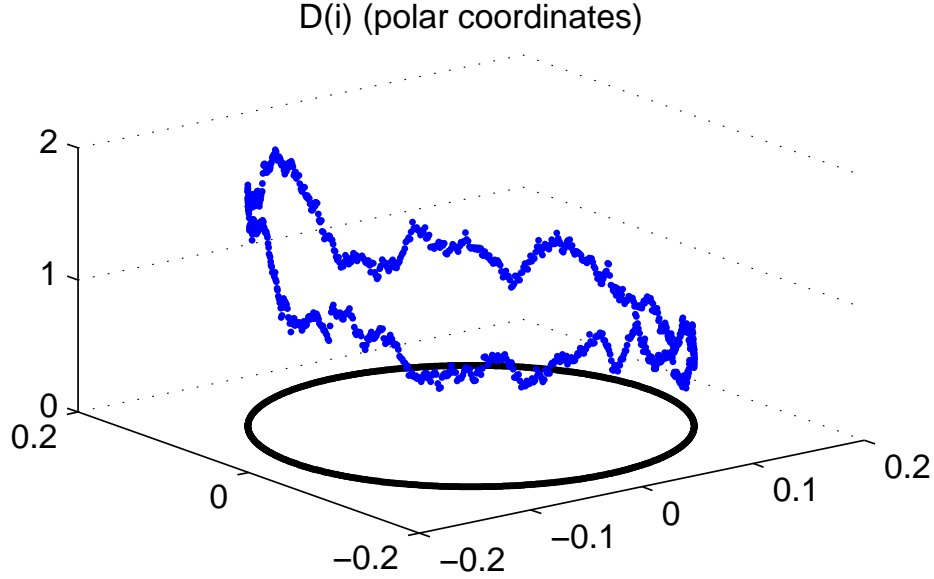
Figure 6: A sample path of the dividend process postulated in equation (60).

may be priced, and there could be an equilibrium arbitrage. Models of this sort become especially interesting if one introduces a participation decision, so as to study the incentives of agents to take advantage of the arbitrages, thus eliminating them.

To illustrate the issues that arise in models of endogenous participation and limited equilibrium arbitrage, this section utilizes the structure proposed in Gârleanu et al. (2015). The model is a one-period model, whereby investors are aligned on a circle with circumference normalized to one. An investor in location $i \in [0, 1)$ is endowed with a tree that pays a random, location-specific dividend. The distribution of this dividend is constructed to ensure that the distribution of the dividend is location-invariant, and that all risk is idiosyncratic. Specifically, letting B_i denote a brownian motion for $i \in [0, 1)$ with $B_0 = 0$, define

$$Z_i \equiv B_i - iB_1.$$

The process Z_i is called a "Brownian Bridge"'. While random for all $i \in (0, 1)$, the process

Z_i satisfies $Z_0 = 0, Z_1 = 0$, i.e., the process starts and ends at zero. Using this definition of Z_i, the time-1 dividend in location i is

$$D_i = 1 + Z_i - \int_0^1 Z_i di. \tag{60}$$

By construction, $\int_0^1 D_i = 1$, i.e., all location-specific risk is diversifiable. Moreover, the distribution of D_i is normal, independent of i, and the correlation between i and j is a declining function of the distance between i and j on the circle.[25] Figure 6 provides an illustration of this construction by plotting a sample path of of D_i (in polar coordinates). Given the normality of dividends and in order to obtain closed form solutions, investors have constant absolute risk aversion utilities (CARA).

In this framework investors have a clear incentive to diversify. In the absence of frictions, each investor in location i would sell her own tree, and purchase a portfolio of all other trees, which would be riskless. To make matters interesting, participation in some market $j \neq i$ is an increasing function of the distance between i and j. Gârleanu et al. (2015) assumes that participation costs start approaching infinity as an investor attempts to invest in all locations on the circle, so that in equilibrium no investor participates in all markets. (The only market where all investors participate is the zero-net-supply bond market.) Because no investor can fully diversify the location-specific asset risk, that risk becomes priced. This leads to an (unexploitable) arbitrage opportunity: the payoffs of an equally weighted portfolio of all risky assets (which has a certain payoff of one) costs less than a zero coupon bond with unit face value

Gârleanu et al. (2015) shows that in this economy an agent's joint participation-investment allocation decision is generally non-concave. The non-concavity of the optimization problem stems from the complementarity between leverage and participation. As an investor contemplates leveraging her portfolio, she has an incentive to also increase her participation in more markets in order to further diversify risk. Because idiosyncratic risk is priced, the lower variance of the portfolio translates into a higher Sharpe ratio, which in turn further

---

[25]See Gârleanu et al. (2015) for a proof.

increases the incentive to leverage. In equilibrium, the investor is faced with two maxima, one associated with a small participation arc and a positive allocation to bonds, while the other involving a large participation arc and a leveraged portfolio.

As in Aumann (1966), markets clear by determining the fractions of agents that choose the first and the second maximum. Specifically, a group of investors participate in a small arc and choose a positive allocation to bonds, while another mass of agents in the same location choose to participate in a large arc and borrow. The first group of investors accept a lower Sharpe ratio and pay less fees, while the latter group of investors obtain a higher Sharpe ratio and pay more fees for it.

Unlike the other models presented in this survey, in Gârleanu et al. (2015) all investors in a location are ex-ante identical . Yet, they are heterogeneous ex-post, since the interaction of participation and leverage decisions makes an investor's optimization non-convex, leading ex-ante identical investors to make different portfolio and leverage decisions. Specifically, the endogeneity of the participation (and leverage) decision separates investors into high-Sharpe ratio, high-leverage, high-participation strategies and low-Sharpe ratio, low participation, positive-bond-holding strategies.

Models of equilibrium arbitrages are useful for linking portfolio flows into high- and low- Sharpe ratio strategies with equilibrium risk premiums. An additional aspect of these models is that they can lead to multiple, Pareto ranked equilibrium if one imposes leverage constraints.

4.1.1   Literature Review

The literature on constrained arbitrage is voluminous and any attempt to summarize it within the confines of this survey would not do it justice. Some indicative papers (chosen to exemplify the varied approaches taken in this literature) include Basak and Croitoru (2000), Gromb and Vayanos (2018), Gârleanu and Pedersen (2011), Kondor (2009), Davilla and Korinek (2017), Liu and Longstaff (2003) among many others.

To highlight some recent contributions, Hebert (2019), Davilla and Parliatore (2019) pro-

pose frameworks to measure the welfare costs of equilibrium arbitrages. Zentefis (2019b), Zentefis (2019a), Khorrami (2019b) utilize models that feature a tradeoff between participation costs and diversification.

4.2 Asymmetric information and heterogeneous stochastic discount factors

The purpose of this section is not to survey the literature on asymmetric information, which is voluminous and outside the scope of this survey. Mostly, for reasons of completeness, this section simply mentions a high-level observation contained in Grossman and Shiller (1982), when the source of heterogeneity among investors is asymmetric information.

Specifically, the Grossman and Shiller (1982) result applies to cases where the econometrician has a coarser information set than the agents. Indeed, the law of the iterated expectation implies that

$$E_t^{(I)} \left[E_t^{(i)} \left\{ u' \left(c_{t+\delta}^i \right) \left(R_{t+\delta} - R_{t+\delta}^f \right) \right\} \right] = E_t^{(I)} \left[u' \left(c_{t+\delta}^i \right) \left(R_{t+\delta} - R_{t+\delta}^f \right) \right] = 0$$

where superscript I in the expectations operator denotes the econometrician's information set and i denotes the investor's information set. Accordingly, it would seem that if agents are better informed than the econometrician, then different information sets do not affect the validity of the Grossman and Shiller (1982) result.

One requirement, however, for the this argument to hold is that all investors choose interior portfolios in all assets. If assets cannot be shorted, then some investors may choose zero allocations to individual assets that their information set identifies as problematic, thus leading to endogenous non-participation. When the information asymmetry interacts with the participation decision, then the Grossman and Shiller (1982) fails. Gârleanu et al. (2019) provides a model of such a failure of Grossman and Shiller (1982) akin to the failure of the Euler equation that would obtain in an economy where agents face heterogeneous returns when investing in the same asset class. As a result, information heterogeneity matters for

equilibrium prices and the Grossman and Shiller (1982) fails.[26]

---

[26]Besides pricing implications, models of asymmetric information have natural applications to performance evaluation. Gârleanu et al. (2019) show that some classical results on performance evaluation cease to hold in a model where information asymmetries and portfolio constraints interact. For instance, the classical results of Mayers and Rice (1979) and Dybvig and Ross (1985), who show that the return of an informed investor regressed against the return of any benchmark portfolio (efficient or inefficient) has positive alpha in the presence of pure "selection" ability, ceases to hold when portfolio decisions are not interior.

A   Appendix

This appendix derives the differential equations characterizing g_t^i and p_t.

Using (8) inside (19) implies

$$g_t^i = E_t \int_t^\infty e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)(u-t)} \left(\frac{H_u}{H_t}\right)^{1-\frac{1}{\gamma^i}} du. \tag{61}$$

Equation (61) shows that g_t^i is essentially a present value relation. Using some basic arguments, this present value relationship satisfies an ordinary differential equation. In order to derive this ordinary differential equation for g_t^i, start by multiplying both sides of (61) by $e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)t} H_t^{1-\frac{1}{\gamma^i}}$ and add $\int_s^t e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)u} (H_u)^{1-\frac{1}{\gamma^i}} du$ to both sides, which results in

$$e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)t} H_t^{1-\frac{1}{\gamma^i}} g_t^i + \int_s^t e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)u} (H_u)^{1-\frac{1}{\gamma^i}} du = E_t \int_s^\infty e^{-\left(\pi + \frac{\rho^i}{\gamma^i}\right)u} (H_u)^{1-\frac{1}{\gamma^i}} du. \tag{62}$$

Observe that the right hand side of equation (62) is a conditional expectation and hence a martingale. This means that the left hand side must be a martingale as well.

To proceed, conjecture that the equilibrium is Markovian in x_t^A, which means that $r_t, \kappa_t, \mu_x^i, \sigma_x^i, g_t^i$ and p_t are exclusively functions of x_t^A. (This conjecture is verified at the end of the proof).

This conjecture implies that $\frac{c_{t,t}^i}{D_t} = \frac{\delta}{\pi} \frac{p_t}{g_t^i}$ is a function of x_t^A. The implication is that r_t and μ_x^i can be written exclusively as functions of x_t^A, (and so can κ_t and σ_x^i) since $x_t^B = 1 - x_t^A$.

Applying Ito's Lemma to compute the drift of the left-hand side of (62) and setting the resulting expression to zero leads after some simplifications to the following two differential equations for $i = A, B$:

$$\frac{\sigma_x^2 \left(x_t^A\right)}{2} \left(g^i\right)'' + \left(\mu_x^A \left(x_t^A\right) - \left(1 - \frac{1}{\gamma^i}\right) \kappa \left(x_t^A\right) \times \sigma_x \left(x_t^A\right)\right) \left(g^i\right)' - \left\{ \left(\pi + \frac{\rho^i}{\gamma^i}\right) + \left(1 - \frac{1}{\gamma^i}\right) \left[r\left(x_t^A\right) + \frac{1}{\gamma^i} \frac{\kappa^2 \left(x_t^A\right)}{2}\right] \right\} g^i + 1 = 0. \tag{63}$$

The final step is to obtain an expression for p_t, which enters the pair of differential equations (63) through the dependence of $r\left(x_t^A\right)$, $\mu_x^A \left(x_t^A\right)$ on $\frac{c_{t,t}^i}{D_t} = \frac{\delta}{\pi} \frac{p_t}{g_t^i}$. Asset market clearing implies

that aggregate wealth of the two agent groups needs to equal the value of the stock market, that is

$$W_t^A + W_t^B = P_t,$$

where W_t^i is each group's total wealth and P_t is the value of the stock market. Moreover, each group's budget constraint implies that $W_t^i = g_t^i \left(x_t^A \right) x_t^i C_t$. Accordingly

$$\sum_{i \in \{A,B\}} x_t^i g_t^i \left(x_t^A \right) = \frac{P_t}{C_t} = p \left(x_t^A \right). \tag{64}$$

Substituting this expression for $p \left(x_t^A \right)$ into the expressions for $r \left(x_t \right), \mu_x^i \left(x_t \right)$ implies that equations (63) are fully specified ODEs, i.e., can be written in the form $F \left(x_t^A, g^i \left(x_t^A \right), g' \left(x_t^A \right), g'' \left(x_t^A \right) \right) = 0$. Hence if $g^i \left(x_t^A \right), i \in \{A, B\}$ solves this pair of differential equations, then the conjecture that there exists a Markovian equilibrium (i.e., an equilibrium where $r_t, \kappa_t, \mu_x^i, \sigma_x^i, g_t^i$ and p_t are exclusively functions of x_t^A) is confirmed.

References

Abel, A. B. (2003). The effects of a baby boom on stock prices and capital accumulation in the presence of social security. *Econometrica 71*(2), 551–578.

Abel, A. B., J. C. Eberly, and S. Panageas (2007). Optimal inattention to the stock market. *American Economic Review 97*(2), 244–249.

Abel, A. B., J. C. Eberly, and S. Panageas (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica 81*(4), 1455–1481.

Akira Toda, A. and K. Walsh (2019). Asset Pricing and the One Percent. Forthcoming, Review of Financial Studies.

Alvarez, F. and A. Atkeson (2018). Random risk aversion and liquidity: a model of asset pricing and trade volumes. Working Paper, University of Chicago and UCLA.

Alvarez, F. and U. J. Jermann (2000). Efficiency, equilibrium, and asset pricing with risk of default. *Econometrica 68*(4), 775–797.

Angeletos, G.-M. and L. Calvet (2006). Idiosyncratic production risk, growth and the business cycle. *Journal of Monetary Economics 53*(6), 1095–1115.

Aumann, R. J. (1966). Existence of competitive equilibria in markets with a continuum of traders. *Econometrica 34*(1), 1–17.

Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance 59*(4), 1481–1509.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics 121*(3), 823–66.

Barro, R. J., J. Fernández-Villaverde, O. Levintal, and A. Mollerus (2017). Safe assets. Working paper.

Basak, S. and B. Croitoru (2000). Equilibrium mispricing in a capital market with portfolio constraints. *The Review of Financial Studies 13*(3), 715–748.

Basak, S. and D. Cuoco (1998). An equilibrium model with restricted stock market participation. *Review of Financial Studies 11*(2), 309–41.

Bhamra, H. S. and R. Uppal (2013). Asset Prices with Heterogeneity in Preferences and Beliefs. *The Review of Financial Studies 27*(2), 519–580.

Blanchard, O. J. (1985). Debt, deficits, and finite horizons. *Journal of Political Economy 93*(2), 223–247.

Blume, L. and D. Easley (1992). Evolution and market behavior. *Journal of Economic Theory 58*(1), 9 – 40.

Borch, K. (1962). Equilibrium in a reinsurance market. *Econometrica 30*(3), 424–444.

Borovicka (2019). Survival and Long-run Dynamics with Heterogeneous Beliefs Under Recursive Preferences. Forthcoming, Journal of Political Economy.

Brav, A., G. M. Constantinides, and C. C. Geczy (2002). Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence. *Journal of Political Economy 110*(4), 793–824.

Breeden, D. T. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*(3), 265 – 296.

Buraschi, A. and A. Jiltsov (2006). Model uncertainty and option markets with heterogeneous beliefs. *The Journal of Finance 61*(6), 2841–2897.

Calvet, L. E., J. Y. Campbell, F. J. Gomes, and P. Sodini (2019). The cross-section of household preferences. Working Paper, Harvard University.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

Campbell, J. Y. and Y. Nosbusch (2007). Intergenerational risksharing and equilibrium asset prices. *Journal of Monetary Economics 54*(8), 2251–2268.

Chabakauri, G. (2013). Dynamic Equilibrium with Two Stocks, Heterogeneous Investors, and Portfolio Constraints. *The Review of Financial Studies 26*(12), 3104–3141.

Chan, Y. L. and L. Kogan (2002). Catching up with the joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy 110*(6), 1255–1285.

Chen, H., S. Joslin, and N.-K. Tran (2012). Rare Disasters and Risk Sharing with Heterogeneous Beliefs. *The Review of Financial Studies 25*(7), 2189–2224.

Chien, Y., H. Cole, and H. Lustig (2012). Is the volatility of the market price of risk due to intermittent portfolio rebalancing? *American Economic Review 102*(6), 2859–96.

Cogley, T. (2002). Idiosyncratic risk and the equity premium: evidence from the consumer expenditure survey. *Journal of Monetary Economics 49*(2), 309–334.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2002). Junior can't borrow: A new perspective of the equity premium puzzle. *Quarterly Journal of Economics 117*(1), 269–96.

Constantinides, G. M. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*, 219–40.

Constantinides, G. M. and A. Ghosh (2017). Asset pricing with countercyclical household consumption risk. *The Journal of Finance 72*(1), 415–460.

David, A. (2008). Heterogeneous beliefs, speculation, and the equity premium. *The Journal of Finance 63*(1), 41–83.

Davilla, E. and A. Korinek (2017). Pecuniary Externalities in Economies with Financial Frictions. *The Review of Economic Studies 85*(1), 352–395.

Davilla, E. and C. Parliatore (2019). Trading costs and informational efficiency. Working paper, Yale and New York University.

Detemple, J. and S. Murthy (1997). Equilibrium Asset Prices and No-Arbitrage with Portfolio Constraints. *The Review of Financial Studies 10*(4), 1133–1174.

Detemple, J. and A. Serrat (2003). Dynamic Equilibrium with Liquidity Constraints. *The Review of Financial Studies 16*(2), 597–629.

Duffie, D. and L. G. Epstein (1992). Stochastic differential utility. *Econometrica 60*(2), 353–394.

Dumas, B. (1989). Two-person dynamic equilibrium in the capital market. *Review of Financial Studies 2*(2), 157–188.

Dumas, B., K. K. Lewis, and E. Osambela (2016). Differences of Opinion and International Equity Markets. *The Review of Financial Studies 30*(3), 750–800.

Dumas, B., R. Uppal, and T. Wang (2000). Efficient intertemporal allocations with recursive utility. *Journal of Economic Theory 93*(2), 240 – 259.

Dybvig, P. H. and S. A. Ross (1985). Differential information and performace measurement using a security market line. *Journal of Finance 40*(2), 383–399.

Ehling, P., A. Graniero, and C. Heyerdahl-Larsen (2017). Asset Prices and Portfolio Choice with Learning from Experience. *The Review of Economic Studies 85*(3), 1752–1780.

Farmer, R. E. (2018). Pricing assets in a perpetual youth model. *Review of Economic Dynamics 30*, 106 – 124.

Favilukis, J. (2013). Inequality, stock market participation, and the equity premium. *Journal of Financial Economics 107*(3), 740 – 759.

Gârleanu, N., L. Kogan, and S. Panageas (2012). Displacement risk and asset returns. *Journal of Financial Economics 105*, 491–510.

Gârleanu, N. and S. Panageas (2015). Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing. *Journal of Political Economy 123*, 670–685.

Gârleanu, N. and S. Panageas (2018). Finance in a time of gisruptive growth. Working paper, UC Berkeley and UCLA.

Gârleanu, N. and S. Panageas (2019). Heterogeneity and asset prices: A different approach. Working paper, UC Berkeley and UCLA.

Gârleanu, N. and S. Panageas (2020). What to expect when everyone is expecting: Self-fulfilling expectations and asset-pricing puzzles. Forthcoming, Journal of Financial Economics.

Gârleanu, N., S. Panageas, and J. Yu (2015). Financial entanglement: A theory of incomplete integration, leverage, crashes, and contagion. *American Economic Review 105*(7), 1979–2010.

Gârleanu, N., S. Panageas, and J. Yu (2019). Impediments to financial trade: Theory and applications. Forthcoming, Review of Financial Studies.

Gârleanu, N. and L. H. Pedersen (2011). Margin-based Asset Pricing and Deviations from the Law of One Price. *The Review of Financial Studies 24*(6), 1980–2022.

Geanakoplos, J., M. Magill, and M. Quinzii (2004). Demography and the long-run predictability of the stock market. *Brookings Papers on Economic Activity 2004*(1), 241–307.

Gomes, F. and A. Michaelides (2008). Asset pricing with limited risk sharing and heterogeneous agents. *Review of Financial Studies 21*(1), 415 – 449.

Gomez, M. (2017). Asset prices and wealth inequality. Working Paper, Columbia University.

Gomez, M. (2019). Displacement and the rise in top wealth inequality. Working paper, Columbia University.

Greenwald, D. L., M. Lettau, and S. Ludvigson (2019). How the wealth was won: Factors shares as market fundamentals. NBER Working paper, 25769.

Gromb, D. and D. Vayanos (2018). The dynamics of financially constrained arbitrage. *The Journal of Finance 73*(4), 1713–1750.

Grossman, S. J. and R. J. Shiller (1982). Consumption correlatedness and risk measurement in economies with non-traded assets and heterogeneous information. *Journal of Financial Economics 10*(2), 195 – 210.

Guvenen, F. (2005). A parsimonious macroeconomic model for asset pricing: Habit formation or cross-sectional heterogeneity. mimeo, University of Texas at Austin.

He, H. and D. M. Modest (1995). Market frictions and consumption-based asset pricing. *Journal of Political Economy 103*(1), 94–117.

He, H. and H. F. Pages (1993). Labor income, borrowing constraints, and equilibrium asset prices. *Economic Theory 3*(4), 663–696.

Heaton, J. and D. J. Lucas (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy 104*(3), 443–487.

Hebert, B. (2019). Externalities as arbitrage. Working Paper, Stanford Graduate School of Business.

Hugonnier, J. (2012). Rational asset pricing bubbles and portfolio constraints. *Journal of Economic Theory 147*(6), 2260 – 2302.

Kargar, M. (2018). Heterogeneous Intermediary Asset Pricing. Unpublished Working Paper, UCLA.

Khorrami, P. (2019a). Entry and slow-moving capital: using asset markets to infer the costs of risk concentration. unpublished manuscript, University of Chicago.

Khorrami, P. (2019b). The risk of risk-sharing: Diversification and boom-bust cycles. Working paper, University of Chicago.

Kocherlakota, N. R. and L. Pistaferri (2009). Asset pricing implications of pareto optimality with private information. *Journal of Political Economy 117*(3), 555–590.

Kogan, L., D. Papanikolaou, and N. Stoffman (2019). Left behind: Creative destruction, inequality, and the stock market. Journal of Political Economy, forthcoming.

Kogan, L., S. A. Ross, J. Wang, and M. M. Westerfield (2006). The Price Impact and Survival of Irrational Traders. *The Journal of Finance 61*(1), 195–229.

Kondor, P. (2009). Risk in dynamic arbitrage: The price effects of convergence trading. *The Journal of Finance 64*(2), 631–655.

Kondor, P. and D. Vayanos (2019). Liquidity risk and the dynamics of arbitrage capital. *The Journal of Finance 74*(3), 1139–1173.

Krueger, D. and F. Kubler (2006). Pareto-improving social security reform when financial markets are incomplete!? *American Economic Review 96*(3), 737–755.

Krueger, D. and H. Lustig (2010). When is market incompleteness irrelevant for the price of aggregate risk (and when is it not)?. *Journal of Economic Theory 145*(1), 1 – 41.

Liu, J. and F. A. Longstaff (2003). Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities. *The Review of Financial Studies 17*(3), 611–641.

Longstaff, F. A. and J. Wang (2012). Asset Pricing and the Credit Market. *The Review of Financial Studies 25*(11), 3169–3215.

Lucas, D. (1994). Asset pricing with undiversifiable income risk and short sales constraints: Deepening the equity premium puzzle. *Journal of Monetary Economics 34*(3), 325–341.

Luttmer, E. (1996). Asset pricing in economies with frictions. *Econometrica 64*(6), 1439–67.

Mankiw, N. (1986). The equity premium and the concentration of aggregate shocks. *Journal of Financial Economics 17*(1), 211 – 219.

Mankiw, N. and S. P. Zeldes (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics 29*(1), 97 – 112.

Marcet, A. and R. Marimon (1998). Recursive contracts. UPF Working Paper.

Maurer, T. (2017). Asset pricing implications of demographic change. Working Paper, Olin School of Business.

Mayers, D. and E. Rice (1979). Measuring portfolio performance and the empirical content of asset pricing models. *Journal of Financial Economics 7*, 3–28.

Parker, J. A. and A. Vissing-Jorgensen (2009). Who bears aggregate fluctuations and how? *American Economic Review 99*(2), 399–405.

Pastor, L. and P. Veronesi (2019). Inequality Aversion, Populism, and the Backlash Against Globalization. unpublished manuscript.

Sandroni, A. (2000). Do markets favor agents able to make accurate predictions? *Econometrica 68*(6), 1303–1341.

Santos, T. and P. Veronesi (2018). Leverage. NBER Working Paper 22905.

Scheinkman, J. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*(6), 1183–1220.

Schmidt, L. D. W. (2015). Climbing and falling off the ladder: Asset pricing implications of labor market event risk. Unpublished working paper, MIT Sloan.

Schneider, A. (2017). Risk sharing and the term structure of interest rates. Unpublished working paper, Federal Reserve Board of Governors.

Storesletten, K., C. Telmer, and A. Yaron (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy 112*(3), 695–717.

Storesletten, K., C. I. Telmer, and A. Yaron (2007). Asset prcing with idiosyncratic risk and overlapping generations. *Review of Economic Dynamics 10*(4), 519–548.

Telmer, C. I. (1993). Asset-pricing puzzles and incomplete markets. *The Journal of Finance 48*(5), 1803–1832.

Vissing-Jorgensen, A. (2002). Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy 110*(4), 825 – 853.

Wang, J. (1996). The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics 41*(1), 75–110.

Xiouros, C. and F. Zapatero (2010). The Representative Agent of an Economy with External Habit Formation and Heterogeneous Risk Aversion. *The Review of Financial Studies 23*(8), 3017–3047.

Yan, H. (2008). Natural selection in financial markets: Does it work? *Management Science 54*(11), 1935–1950.

Zentefis, A. (2019a). Bank net worth and frustrated monetary policy. Working Paper, Yale School of Management.

Zentefis, A. (2019b). Self-fulfilling asset prices. Working Paper, Yale School of Management.

Zheng, G. (2019). Wealth shares in the long run. Working Paper, UCLA.