Heterogeneity and Asset Prices: An Intergenerational Approach

Nicolae Gârleanu
Washington University in St. Louis and NBER

Stavros Panageas
UCLA Anderson School of Management and NBER*

March 2023

Abstract

In an overlapping-generations economy the consumption growth of a given cohort member (the "marginal agent") differs from the aggregate consumption growth. A cohort member is faced with long-run consumption uncertainty even in the absence of aggregate (and within-cohort) consumption risk. This uncertainty allows the model to account for several stylized asset-pricing facts (high market price of risk and volatility, return predictability, low and non-volatile interest rate) despite deterministic macroeconomic aggregates and inequality measures that are contemporaneously uncorrelated with asset returns. We devise and implement a methodology to measure the marginal agent's consumption growth and evaluate the model's quantitative implications.

Keywords: asset pricing, heterogeneity, imperfect risk sharing, overlapping generations
JEL Classification: G01, G12

*We are grateful for comments and suggestions from Valentin Haddad, Jon Ingersoll, Ralph Koijen, Lawrence Schmidt, Harald Uhlig, Pietro Veronesi, Motohiro Yogo, and two anonymous referees, as well as participants at WFA, NBER Summer Institute on Asset Pricing, Chicago Booth, Princeton, Insead, NYU Stern, Berkeley Haas-Stanford GSB joint seminar, UCLA Macro-Finance lunch, Washington University (Olin), Caltech, Richmond Fed, and University of Virginia.

1 Introduction

We develop a tractable asset-pricing framework characterized by imperfect risk sharing among cohorts, which experience different levels of endowments over the course of their lives. Because of the lack of inter-cohort risk sharing, these endowment fluctuations drive a wedge between aggregate per-capita consumption growth and the consumption growth that the representative member of a *fixed* cohort experiences over their lifetime. To illustrate this point, we focus on an economy where aggregate consumption growth is constant, and yet the long-run consumption growth of a fixed cohort member, who is the marginal agent in financial markets, exhibits long-run uncertainty. Paired with recursive preferences, this long-run consumption uncertainty translates into a risk premium. The model can reproduce several stylized asset-pricing facts (volatile asset prices, large and countercyclical Sharpe ratios, low and non-volatile interest rate), while all of the aggregate quantities (in particular, consumption and dividends) exhibit zero instantaneous covariance with asset returns, implying the failure of the conditional aggregate-consumption CAPM.[1] In line with the data, inequality is non-volatile, exhibits zero instantaneous correlation with the stock market, and has a substantially higher persistence than the price-dividend ratio.

In addition to highlighting the conceptual difference between aggregate consumption growth and the consumption growth of an investor who is "marginal" in financial markets we develop an empirical framework that allows us to measure and quantify the implications of this difference. This empirical framework is simple yet flexible enough to account for a multitude of factors, such as different cohort sizes, age-dependent life-cycle effects, shifts in the demographic pyramid, and different cohort productivities.

Specifically, the framework is a continuous-time, OLG economy. Agents arrive continuously, endowed with a claim to either a wage path ("workers") or a dividend stream ("entrepreneurs"). All shocks are exclusively distributional; they drive the income shares obtained by firms and workers born at different times, while aggregate labor and dividend income grow at the same constant rate. Moreover, to differentiate our results from the liter-

[1] The conditional Sharpe-Lintner-Mossin CAPM also fails, because the stock market is only a part of a consumer's total wealth in the model.

ature, which has predominantly focused on models with lack of intra-cohort risk sharing, we assume that intra-cohort risk sharing is perfect and the shares of labor or dividend income accruing to a given cohort of investors are locally deterministic processes,[2] albeit random over the long run.

In this setup we introduce a) imperfect inter-cohort risk sharing and b) recursive utility with a preference for early resolution of uncertainty. We utilize this framework to perform two exercises, one theoretical and one empirical.

Our theoretical exercise parallels the one in Constantinides and Duffie (1996). Specifically, we show the following "possibility" result: Share processes exist that support a broad class of given stationary processes for both the market price of risk and the price-dividend ratio. However, the endowment specifications, the source of market imperfection, and most importantly the implications for the joint movements of inequality and asset prices are different from Constantinides and Duffie (1996), as we explain in Section 1.1.

Our existence results are not abstract, but constructive; we use them to specify endowment-share processes that lead to closed-form expressions for asset-pricing quantities. The model is therefore highly tractable, despite the underlying heterogeneity.

The key distinguishing feature of our model relative to a representative-agent economy is that the aggregate (per capita) consumption growth does not coincide with the consumption growth of the marginal agent. In our OLG economy, marginal-agent consumption growth over a given interval is the consumption growth of cohorts that entered before the beginning of the interval, and consequently does not include the portion of aggregate consumption accruing to newly arriving cohorts. Over short time intervals the discrepancy between marginal and aggregate consumption growths is small but, due to its persistence, its cumulative effect at long horizons can be arbitrarily large. With recursive preferences the different long-run behavior of the two consumption processes has important asset pricing consequences.

We show how to use cross-sectional data to infer our notion of marginal agent's consumption growth. Essentially, we observe that the Euler equation implies a time, age, and cohort decomposition for log cohort consumption. This decomposition can be estimated with a

[2]Throughout the paper "locally deterministic" refers to a time-differentiable process. By definition, such a process has no diffusion component, but a possibly stochastic drift process.

simple regression from cross-sectional data, and the variation of the time effect corresponds to the consumption growth of a fixed cohort member (i.e., marginal-agent consumption growth) inside our model. Since detailed cross-sectional data are available only since the mid-eighties, we also develop an indirect inference approach to combine information contained in estimated cohort and age effects with market clearing to extend the sample farther back in time. Our methodology for measuring marginal-agent consumption growth relies on minimal assumptions and is valid irrespective of how one chooses to model production, the government, redistribution policies, demographics, aging effects, etc.[3]

We find that our measure of marginal-agent consumption growth exhibits more persistence and predictability than aggregate per capita consumption growth, and a stronger co-movement with the real expected interest rate over medium-run cycles. The reason is that distributional effects that are typically irrelevant for representative-agent asset pricing (e.g., the fact that the cohort of people born in 1940–1950 "did better" than the cohort born in the eighties) have an impact on interest rates and risk premiums in our framework.

To illustrate the quantitative implications of the model, we calibrate it to reproduce the time-series moments of our inferred measure of marginal-agent consumption growth. We show that the model produces realistic risk premiums, return predictability, interest rate levels, and volatility. We also show that the cross-sectional consumption variance has negligible volatility but follows a near-unit-root process, consistent with the data.

1.1 Relation to the literature

As mentioned above, the structure of our theoretical exercise is similar to Constantinides and Duffie (1996), but there are notable differences in outcomes. To start with, our construction can generate a persistent, rather than i.i.d., stochastic discount factor. The more important difference, however, pertains to the time-series implications for inequality measures, such as the cross-sectional variance of consumption. Constantinides and Duffie (1996) requires strong contemporaneous co-movement between stock market returns and the vari-

[3]Specifically, any two OLG economies calibrated to match the time series properties of our measure of marginal-agent consumption growth have the same asset-pricing implications.

ance of the idiosyncratic shocks to consumption growth. In our model, the contemporaneous correlation between these two quantities is zero, consistent with the low value of its empirical counterpart.[4] We nevertheless obtain a risk premium because our investors have recursive preferences. In that sense, our paper is similar in spirit to Schmidt (2015) and Constantinides and Ghosh (2017). Unlike Schmidt (2015) or Constantinides and Ghosh (2017), however, our model does not feature idiosyncratic risk. Further, in these papers idiosyncratic (labor income) risk has asset-pricing implications because of *missing markets* to hedge it, while in our framework the key friction is *missing market participants*, in the sense that the unborn investors cannot trade with the existing investors. This friction is sufficient to generate persistent, stochastic variation in the first moment, rather than higher moments, of the consumption growth of existing (marginal) agents, even though we assume that aggregate consumption growth is constant and we abstract from idiosyncratic labor-income risk. Finally, we note that an additional difference between Constantinides and Duffie (1996) and our paper is that we don't have to take a stance on whether higher cross-sectional moments of the consumption distribution are finite.[5]

For our theoretical results, we employ a stochastic, endowment version of the Blanchard (1985) model. An advantage of this framework is that it allows us to isolate the notion of imperfect risk sharing across cohorts while sidestepping the technical complications of more conventional OLG models.[6] Our empirical measure of marginal consumption growth, however, does not depend on the simplifying assumptions in Blanchard (1985), as we explain in Section 4.

Our paper also relates to the literature that studies the asset-pricing implications of OLG models,[7] to which we make two contributions. First, our paper highlights how to

[4]The strength of this co-movement becomes most apparent in a continuous-time version of Constantinides and Duffie (1996), which requires simultaneous jumps in consumption dispersion and stock prices. (For a continuous-time version of Constantinides and Duffie (1996), see Panageas (2020).)

[5]Toda and Walsh (2015) argues that higher moments of the cross sectional distribution of consumption may fail to exist, leading to erroneous conclusions in the Constantinides and Duffie (1996) model.

[6]Due to their tractability, in recent years perpetual-youth models have gained popularity in asset pricing. See, e.g., Campbell and Nosbusch (2007), Gârleanu and Panageas (2015), Ehling et al. (2018), Maurer (2017), Gomez (2017), Schneider (2017), or Farmer (2018) among others.

[7]Indicative examples of such papers include Constantinides et al. (2002), Gomes and Michaelides (2005), Storesletten et al. (2007), and Piazzesi and Schneider (2009). The literature on demographic shocks to asset prices, which we don't attempt to summarize here, is also (remotely) related to the present paper. Two

appropriately measure the consumption growth of a marginal agent — not just in this model but also in a broad family of OLG frameworks. Second, for many of the papers in this literature the risk-price effects due to the lack of intergenerational risk sharing would vanish if trading was allowed at a high rate, as all demography-related shocks are locally predictable. Just as in Grossman and Shiller (1982), only the risk-free rate would be affected. Our framework helps both formalize this criticism (Section 3.1) and address it by employing recursive preferences (Section 3.3).

There are several papers in this literature that rely on perpetual-youth models with stochastic fluctuations in the labor income or profit share of each cohort, and to which ours relates more closely. Gârleanu et al. (2012) employs a discrete-time, i.i.d. framework, and thus cannot generate time variation in asset-pricing moments.[8] Kogan et al. (2019) builds models of firm investment and heterogeneous rent allocation from technological progress. While present in that model, the lack of inter-generational risk sharing is not its centerpiece. In contrast, we consider a simpler endowment economy that allows a detailed theoretical analysis of the implications of the lack of inter-cohort risk sharing, and a methodology to evaluate the empirical connections between the lack of inter-cohort risk sharing and asset pricing.

Our paper also relates to the literature on long run risks, which was initiated by Bansal and Yaron (2004). The point we make in this paper is complementary to Bansal and Yaron (2004). We show (both theoretically and empirically) that inter-cohort risk sharing imperfections are an additional source of long-run risk, largely independent from the long-run risks in aggregate per-capita consumption growth. Further, by modeling dividend and labor income processes explicitly, we avoid the need for an intertemporal elasticity of substitution above one and we can model explicitly the source of non-cointegration between the dividends of existing stocks and marginal-agent consumption in general equilibrium.

There is a voluminous empirical literature that employs time, age, and cohort decompo-

indicative examples are Abel (2003) and Geanakoplos et al. (2004).

[8] From a technical perspective, Gârleanu et al. (2012) would involve discontinuous allocations to newly arriving cohorts, if one were to consider its continuous time limit. This means that a zero measure of cohorts would obtain an infinite per capita allocation, an outcome that the present paper avoids.

sitions in repeated cross sections. We do not attempt to summarize this literature here. We simply mention that a large number of empirical papers (especially in labor economics) also find a significant role for cohort effects on people's incomes, consumption, health, etc.[9] Our contribution is to show how to utilize the information contained in a time, age, and cohort decomposition to reconstruct the consumption evolution of the "marginal agent."

The study of inequality is outside the scope of this paper, except for its covariance with asset prices. While the model could be easily extended to account for income inequality at birth, it cannot account for the increase of inequality over the life cycle, which is due to lack of intra-cohort risk sharing coupled with idiosyncratic income shocks. There is a debate in the literature on whether intra-cohort inequality driven by idiosyncratic shocks has any implications for risk premiums.[10] For reasons of conceptual differentiation, we abstract from it in this paper without dismissing its importance or claiming that the two approaches to linking heterogeneity and asset prices are mutually exclusive.

The paper is organized as follows. Section 2 presents the model. Section 3 contains the possibility results. Section 4 develops the empirical implications of the model and uses them to measure the consumption and dividend share variations in the data. Section 5 calibrates the model to match the variation in consumption shares and derives the model's asset pricing applications. Section 6 concludes. The appendix contains proofs, model extensions, implementation details, and robustness exercises.

2 Model

We present the baseline model in two steps. In a first step we perform the analysis assuming that agents have expected utility, logarithmic preferences. In a second step we extend the analysis to recursive preferences.

[9]This literature is too large to summarize here. An indicative list includes Oyer (2006), Oyer (2008), Kahn (2010), von Wachter and Schwandt (2018), and Oreopoulos et al. (2012) amongst many others.

[10]Indicatively, Krueger and Lustig (2010) points out that conventional modeling approaches to idiosyncratic endowment risk result in cross-sectional consumption dynamics that don't matter for risk pricing. As mentioned earlier, Grossman and Shiller (1982) also reaches the conclusion that incomplete risk sharing amongst existing agents does not matter for risk pricing in a continuous time, Brownian setting.

2.1 Consumers

Time is continuous. Each agent faces a constant hazard rate of death $\lambda > 0$ throughout her life, so that a fraction λ of the population perishes at each instant. A new cohort of mass λ is born per unit of time, so that the total population remains at $\lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} ds = 1$. In Appendix B.1 we extend the model to accommodate a random birth rate.

Consumers maximize the utility they derive from their stream of consumption. In this section we illustrate our approach in the special case of logarithmic utility, i.e., consumers maximize

$$\mathrm{E}_s \left[\int_s^{\infty} e^{-\rho(t-s)} \log \left(c_{t,s} \right) dt \right], \tag{1}$$

where s is the time of their birth and t is calendar time. Consumers have no bequest (or gift) motives for simplicity.

2.2 Endowments

Following a long tradition in asset pricing, we consider an endowment economy. The total endowment of the economy is denoted by Y_t and evolves exogenously according to

$$\frac{\dot{Y}_t}{Y_t} \equiv g, \tag{2}$$

where $g > 0$. We intentionally model the aggregate endowment as a deterministic, constant-growth process in order to isolate the effect of distributional shocks. Proposition 3 in appendix B.1 allows for time-varying aggregate consumption growth.

This aggregate endowment accrues to the various agents populating the economy as follows. At birth, agents are of two types, to which we refer as "entrepreneurs" and "workers." They only differ with respect to their endowment.

Entrepreneurs join the economy at a rate of $\lambda\varepsilon$ per unit of time and constitute a fraction ε of the population. The entrepreneurs born at time s introduce a new cohort of firms into the market. The firms introduced at time s pay the following aggregate dividends at times

$t \geq s$:

$$D_{t,s} = \alpha Y_t \eta_s^d e^{-\int_s^t \eta_u^d du}. \tag{3}$$

The term $\alpha \in (0,1)$ in equation (3) is a constant, while $\eta_t^d \geq 0$ is assumed to follow a non-negative diffusion

$$d\eta_t^d = \mu_t^d dt + \sigma_t^d dB_t, \tag{4}$$

where B_t is a standard Brownian motion and μ_t^d and σ_t^d are processes that we specify later. In equation (3) we can interpret α as the fraction of output that is paid out as dividends, and $\eta_s^d e^{-\int_s^t \eta_u^d du} \geq 0$ as the fraction of dividends accruing to firms of vintage s, since $\int_{-\infty}^t \eta_s^d e^{-\int_s^t \eta_u^d du} ds = 1$ for any path of η_t^d.[11] Accordingly, aggregating across firms of all vintages gives $D_t^A \equiv \int_{-\infty}^t D_{t,s} ds \; = \alpha Y_t \int_{-\infty}^t \eta_s^d e^{-\int_s^t \eta_u^d du} ds = \alpha Y_t$.

An implication of the dividend specification (3) along with $\eta_t^d \geq 0$ is is that firms belonging to any given cohort s account for a smaller and smaller fraction of aggregate dividends as time t goes by. This feature of the model is empirically motivated.

We next turn to workers. The specification of workers' endowments mirrors the one for dividends and is a simple extension of the specification in Blanchard (1985). Specifically, per unit of time a mass $(1-\varepsilon)\lambda$ of workers are born. Accordingly, the time-t density of surviving workers who were born at time s is given by $l_{t,s} = \lambda (1-\varepsilon) e^{-\lambda(t-s)}$. The time-$t$ endowment $w_{t,s}$ of a worker who was born at time $s \leq t$ is given by

$$w_{t,s} \equiv \frac{(1-\alpha) Y_t \eta_s^l e^{-\int_s^t \eta_u^l du}}{l_{t,s}}, \tag{5}$$

where $\eta_t^l \geq 0$ is assumed to follow a non-negative diffusion

$$d\eta_t^l = \mu_t^l dt + \sigma_t^l dB_t,$$

[11] Specifically, this statement holds true for paths of η_t^d satisfying $\int_{-\infty}^t \eta_s^d ds = \infty$. For the type of stochastic processes that we consider for η_t^d this property holds almost surely.

where μ_t^l and σ_t^l are processes that we specify later. As with dividend income, the term $\eta_s^l e^{-\int_s^t \eta_u^l du}$ can be interpreted as the share of aggregate earnings that accrues to the cohort of workers born at time s. Repeating the observations we made earlier, aggregate wage earnings are $\int_{-\infty}^t w_{t,s} l_{t,s} ds = (1-\alpha) Y_t$.

2.3 Markets

Markets are dynamically complete. Investors can trade in instantaneously maturing riskless bonds in zero net supply, which pay an interest rate r_t. Consumers can also trade claims on all existing firms (normalized to unit supply). Following Blanchard (1985), investors can access a market for annuities through competitive insurance companies, allowing them to receive an income stream of $\lambda W_{t,s}$ per unit of time, where $W_{t,s}$ is the consumer's financial wealth. In exchange, the insurance company collects the consumer's financial wealth when she dies. Entering such a contract is optimal, given the absence of bequest motives.

A worker starts out with wealth $W_{t,t} = 0$ and faces the dynamic budget constraint

$$dW_{t,s} = (r_t + \lambda) W_{t,s} dt + (w_{t,s} - c_{t,s}) dt + \theta_{t,s} \left(dP_t + D_t^A dt - r_t P_t dt \right), \tag{6}$$

where P_t is the value of the market portfolio at time t and $\theta_{t,s}$ is the number of shares of the market portfolio. Without loss of generality, specification (6) assumes that the consumer trades only in shares of the market portfolio, rather than individual firms.[12]

An entrepreneur's dynamic budget constraint is identical to the worker's, except that the term $w_{t,s}$ is replaced by zero and the initial wealth $W_{t,t}$ is given by the value of the firm that the entrepreneur creates.

While there is no limitation on consumers' abilities to dynamically replicate any payoff while they are alive, the fact that financial markets at any point in time are missing some participants (the unborn consumers) makes this a model of incomplete market participation, and therefore limited inter-cohort risk sharing.

[12] This is without loss of generality in our setup because all existing firms have identical dividend growth rates, therefore the same price-to-dividend ratios and consequently the same return to avoid arbitrage.

2.4 Equilibrium

The equilibrium definition is standard. We let $c^e_{t,s}$ (resp. $c^w_{t,s}$) denote the time-t consumption of an entrepreneur (resp. worker) born at time s and $\theta^e_{t,s}$ ($\theta^w_{t,s}$) her holding of stock. With $c_{t,s} = \varepsilon c^e_{t,s} + (1-\varepsilon)\, c^w_{t,s}$ the per-capita consumption of cohort s and, similarly, $\theta_{t,s} = \varepsilon\theta^e_{t,s} + (1-\varepsilon)\,\theta^w_{t,s}$, we look for consumption processes, asset allocations $\theta_{t,s}$, asset prices $P_{t,s}$, and an interest rate r_t such that a) consumers maximize objective (1) subject to constraint (6), b) the goods market clears, i.e., $\lambda\int_{-\infty}^{t} e^{-\lambda(t-s)}c_{t,s}ds = Y_t$, and c) assets markets clear, i.e., $\int_{-\infty}^{t} \lambda e^{-\lambda(t-s)}\theta_{t,s}ds = 1$ and $\int_{-\infty}^{t} \lambda e^{-\lambda(t-s)}(W_{t,s} - \theta_{t,s}P_t)ds = 0$.

3 Solution and Analysis

This section contains our theoretical results, which are in the spirit of the exercise in Constantinides and Duffie (1996). In particular, we establish the existence of share processes η^l_t and η^d_t that can support given processes for the asset-pricing quantities as equilibrium outcomes.

The section is divided into four subsections. In Section 3.1 we derive, under the logarithmic-preference assumption, a key relation linking the processes η^d_t and η^l_t to the dynamics of the price-dividend ratio, which we denote by q_t. We use this relation in Section 3.2 to establish the existence of processes η^d_t and η^l_t that can support any given process for q_t as an equilibrium outcome. In Section 3.3 we enrich the setup to allow for recursive preferences and show how to obtain any (joint) dynamics for the price-dividend ratio and the market price of risk (Sharpe ratio) in equilibrium.

Besides providing a comprehensive mapping from assumptions on η^l_t and η^d_t to the equilibrium processes for the price-dividend ratio and the market price of risk, the propositions of this section have a practical implication: They can help determine the functional forms that one needs to assume for the diffusions η^d_t and η^l_t in order to ensure a given (closed-form) expression for the price-dividend ratio and the Sharpe ratio. We illustrate this statement with two examples.

Section 3.4 contains a discussion of the implications of the model for the joint dynamics of inequality and asset prices and highlights the differences of our framework from the literature.

3.1 Logarithmic utility

We start by conjecturing that in this economy investors' consumption processes are locally deterministic. Given that agents have expected utility preferences, there are no risk premiums and the equilibrium stochastic discount factor m_t follows the dynamics

$$\frac{dm_t}{m_t} = -r_t dt, \tag{7}$$

for an interest rate process that is determined in equilibrium. We employ the following definition.

Definition 1 *Let $q^d_{t,s}$ denote the ratio of the present value of the dividend stream $D_{u,s}$ to the current dividend:*

$$q^d_{t,s} \equiv \frac{E_t \int_t^\infty \frac{m_u}{m_t} D_{u,s} du}{D_{t,s}}. \tag{8}$$

Similarly, we define the respective valuation ratio for earnings, $q^l_{t,s}$:

$$q^l_{t,s} \equiv \frac{E_t \int_t^\infty e^{-\lambda(u-t)} \frac{m_u}{m_t} w_{u,s} du}{w_{t,s}}. \tag{9}$$

Remark 1 *Both $q^d_{t,s}$ and $q^l_{t,s}$ are independent of s, since $\frac{D_{u,s}}{D_{t,s}}$ and $\frac{w_{u,s} l_{u,s}}{w_{t,s} l_{t,s}}$ are not functions of s. Accordingly, we shall write q^d_t and q^l_t instead of $q^d_{t,s}$, respectively $q^l_{t,s}$.*

Due to unitary IES, the consumption to (total) wealth ratio is constant and equal to $\beta \equiv \rho + \lambda$. The next lemma uses this fact to derive a simple affine relationship between q^d_t and q^l_t.

Lemma 1 *In any (bubble-free) equilibrium,*

$$\alpha q^d_t + (1-\alpha) q^l_t = \frac{1}{\beta}. \tag{10}$$

Equation (10) is intuitive. It states that the sum of the present values of all dividend income accruing to existing firms, $q_t^d \alpha Y_t$, and all earnings accruing to existing agents, $q_t^l (1-\alpha) Y_t$, equals the present value of the aggregate consumption of existing agents ($\frac{C_t}{\beta}$). Since $C_t = Y_t$ in equilibrium, equation (10) follows.

By Lemma 1, q_t^l can be expressed as a simple (affine) function of q_t^d. Therefore, from now on we can concentrate our efforts on determining q_t^d, the price-dividend ratio, and we'll simplify notation by writing q_t instead of q_t^d.

In the remainder of this section we determine the equilibrium interest rate r_t and the drift of the price-dividend ratio q_t as functions of the input variables η_t^d and η_t^l. In the next section, we use these results to construct a mapping from the dynamics of q_t to those of η_t^d and η_t^l.

Applying Ito's Lemma to (8) yields the drift of the diffusion process q_t as[13]

$$\mu_{q,t} \equiv \left(r_t - g + \eta_t^d\right) q_t - 1. \tag{11}$$

Equation (11) is an indifference relation between investing in stocks and bonds. After some re-arranging, it states that the expected percentage capital gain on stocks, $\frac{\mu_{q,t}}{q_t}$, plus the dividend yield, $\frac{1}{q_t}$, minus the depreciation (or appreciation) rate $\eta_t^d - g$, should equal the interest rate r_t.

Equation (11) expresses the drift of the price-dividend ratio, $\mu_{q,t}$, in terms of the equilibrium interest rate r_t. To determine this equilibrium interest rate, we proceed in three steps. First, the Euler equation for agents with log preferences implies that the consumption dynamics of any given agent are given by[14]

$$\frac{\dot{c}_{t,s}}{c_{t,s}} = -\left(\rho - r_t\right), \tag{12}$$

which also implies that $\frac{\dot{c}_{t,s}}{c_{t,s}}$ is independent of s.

[13] To see this, note that (8) implies that $m_t q_{t,s}^d D_{t,s} + \int_{-\infty}^{t} m_u D_{u,s} du$ must be a martingale, and hence the drift of this expression must be zero.

[14] For a derivation of the Euler equation in our perpetual youth model we refer to Gârleanu and Panageas (2015).

Second, as we show in the appendix, the market clearing condition for aggregate consumption implies that the consumption growth of an existing agent equals

$$\frac{\dot{c}_{t,s}}{c_{t,s}} = g + \lambda - \lambda \frac{c_{t,t}}{Y_t}, \tag{13}$$

which is intuitive: the consumption growth of a fixed cohort member consists of the growth in aggregate consumption (g), plus the consumption share that perishing agents do not consume (λ), minus the consumption shares accruing to newly born agents ($\lambda c_{t,t}/Y_t$).

Finally, the intertemporal budget constraint at the time of a consumer's birth leads to the following result.

Lemma 2 *Define* $\varphi_t \equiv \eta_t^d - \eta_t^l$ *and*

$$\nu_t \equiv (1 - \alpha\beta q_t)\eta_t^l + \alpha\beta q_t \eta_t^d = \eta_t^l + \alpha\beta q_t \varphi_t. \tag{14}$$

The arriving agents' consumption is given by

$$\lambda \frac{c_{t,t}}{Y_t} = \nu_t. \tag{15}$$

Equation (15) states that the per-capita consumption of an arriving cohort of agents is given by ν_t, which is a weighted average of η_t^l and η_t^d. To derive equation (15), we use the fact that an arriving cohort's initial consumption is the product of the consumption-to-wealth ratio (β) with the sum of the value of the new firms, $\alpha\eta_t^d q_t Y_t$, and the cohort's present value of labor income at birth, $(1-\alpha)\,\eta_t^l q_t^l Y_t$:

$$\lambda c_{t,t} = \beta \left(\alpha \eta_t^d q_t Y_t + (1-\alpha)\, \eta_t^l q_t^l Y_t \right). \tag{16}$$

Dividing both sides of the above equation by Y_t and using Lemma 1 and the definition of ν_t leads to (15).

Combining equations (13), (14), and (15) shows that the consumption drift of any given marginal agent is time varying, even though aggregate consumption growth is constant.

This difference between aggregate consumption growth per capita, g, and the consumption growth, $\frac{\dot{c}_{t,s}}{c_{t,s}}$, of any member of a fixed cohort s in equation (13) is the central feature of our model. In a representative agent economy, the consumption growth of a fixed cohort member and aggregate consumption per capita coincide, simply because the economy is populated by a single, two-sided altruistically linked family, which shares the cohort-specific endowment fluctuations of its members.[15] In our overlapping generations economy, $\frac{\dot{c}_{t,s}}{c_{t,s}}$, the consumption growth of the marginal agent, differs from g: consumers cannot hedge the uncertainty in the present values of their endowments at birth by trading while unborn. Consequently, these cohort-specific endowment shocks end up affecting the consumption distribution between existing and arriving agents.

A simple way to illustrate this point mathematically is to integrate (13) from s to t and use equation (15) to obtain the time-t consumption share of cohort s as

$$\frac{\lambda e^{-\lambda(t-s)} c_{t,s}}{Y_t} = \nu_s e^{-\int_s^t \nu_u du}. \tag{17}$$

Since ν_t is random, different cohorts experience different consumption growth rates over their lifetimes, despite constant aggregate (per-capita) consumption growth.

For asset-pricing purposes, only the existing (not the unborn) agents are marginal in financial markets. This is reflected in the fact that the Euler equation (12) relates the interest rate to the consumption growth of a given living cohort member ($\frac{\dot{c}_{t,s}}{c_{t,s}}$). Indeed, combining equations (12), (13), and (15) leads to the following result.

Lemma 3 *The equilibrium interest rate is given by*

$$r_t = \beta + g - \eta_t^l - \alpha\beta\varphi_t q_t. \tag{18}$$

Note that the interest rate in this economy differs from the constant interest rate that would

[15] The assumption that the set of families is non-expanding and that altruism is perfect and two-sided are all important for the validity of the “representative family” assumption. As Weil (1989) shows, even with perfect bequests, the mere arrival of new “dynasties” is enough to invalidate the assumptions of a representative agent model. Perfect and two-sided altruism is also important: parents are implicitly allowed to borrow against the endowments of their unborn descendants.

obtain in a representative agent economy, $\rho + g$.

Having solved for the equilibrium interest rate, we can now substitute (18) into (11) to obtain the following important result.

Lemma 4 *The drift of q_t is given by*

$$\mu_{q,t} = (\beta + \varphi_t)\, q_t - \beta\alpha\varphi_t q_t^2 - 1. \tag{19}$$

Equation (19) is central for our purposes, since it encapsulates all the equilibrium requirements that our model places on the drift of the price-dividend ratio.

3.2 Supporting a process for q_t as an equilibrium outcome

In this section we ask whether, taking two functions f and σ as given, one can specify a diffusion for $\varphi_t = \eta_t^d - \eta_t^l$ such that the equilibrium process for q_t is given by

$$dq_t = f(q_t)\, dt + \sigma(q_t)\, dB_t. \tag{20}$$

We leave some technical restrictions on f and σ to ensure that q_t is stationary and takes values in some bounded interval $[q^{\min}, q^{\max}]$ for the appendix, and present here the main argument, followed by an illustrative example and a general proposition.

Any process φ_t that supports (20) as an equilibrium price-dividend ratio must be such that $\mu_{q,t} = f(q_t)$. Using equation (19) allows us to explicitly solve for the process φ_t as a function of q_t:

$$\varphi_t = \varphi(q_t) = \frac{1 - \beta q_t + f(q_t)}{q_t (1 - \beta\alpha q_t)}. \tag{21}$$

We assume that φ thus defined is a strictly decreasing function of q_t, so that its inverse $\varphi^{-1}(\varphi_t)$ exists.[16] Combining equations (20) and (21), the dynamics of the process φ_t are

[16] We note that simple differentiation of (21) shows that φ decreases for $q \leq \frac{1}{2}\frac{1}{\alpha\beta}$ as long as f is decreasing. Hence, choosing the process q_t to have support in $[q^{\min}, q^{\max}]$ with $q^{\max} < \frac{1}{2}\frac{1}{\alpha\beta}$, or choosing a function f that has a sufficiently negative first derivative, is sufficient to ensure that φ is strictly decreasing.

easily obtained from Ito's Lemma as

$$d\varphi_t = \underbrace{\varphi'\left(\varphi^{-1}\left(\varphi_t\right)\right)\left(f\left(\varphi^{-1}\left(\varphi_t\right)\right)+\frac{1}{2}\sigma^2\left(\varphi^{-1}\left(\varphi_t\right)\right)\frac{\varphi''\left(\varphi^{-1}\left(\varphi_t\right)\right)}{\varphi'\left(\varphi^{-1}\left(\varphi_t\right)\right)}\right)dt}_{\equiv\mu_\varphi(\varphi_t)}$$
$$+\underbrace{\varphi'\left(\varphi^{-1}\left(\varphi_t\right)\right)\sigma\left(\varphi^{-1}\left(\varphi_t\right)\right)}_{\equiv\sigma_\varphi(\varphi_t)}dB_t. \tag{22}$$

Equation (22) provides the answer to the question that we posed at the outset. Specifically, if we started out with the primitive assumption that φ_t follows the Markov diffusion

$$d\varphi_t = \mu_\varphi(\varphi_t)dt + \sigma_\varphi(\varphi_t)dB_t, \tag{23}$$

with $\mu_\varphi(\varphi_t)$ and $\sigma_\varphi(\varphi_t)$ defined in equation (22), then — by construction — the equilibrium dynamics of the price-dividend ratio are given by (20).

Before stating a formal general result, we illustrate the above ideas with a concrete example.

Example 1 *Suppose that we fix a process x_t obeying the following dynamics*

$$dx_t = \left(-v_1 x_t + v_2\left(1-x_t\right)\right)dt - \sigma_x\sqrt{x_t\left(1-x_t\right)}dB_t \tag{24}$$

where a_1, a_2, v_1, v_2, and σ_x are positive constants. It is established in the literature (see, e.g., Karlin and Taylor (1981), p. 221) that x_t has a stationary (Beta) distribution with support in $[0,1]$ as long as $v_1+v_2>\frac{\sigma_x^2}{2}$.

Next suppose that we wish the equilibrium price-dividend ratio to be given by $q_t = a_1+a_2x_t$. Using (21), the (unique) dynamics of φ_t that support $q_t = a_1+a_2x_t$ as an equilibrium outcome are given as an explicit function of the Markov diffusion x_t:

$$\varphi_t = \frac{1-\beta(a_1+a_2x_t)+a_2v_2-a_2\left(v_1+v_2\right)x_t}{\left(a_1+a_2x_t\right)\left(1-\beta\alpha(a_1+a_2x_t)\right)}. \tag{25}$$

Assuming that the right hand side of the above equation is declining in x_t (which is guaranteed *if $a_1 + a_2 < \frac{1}{2\alpha\beta}$ or if $v_1 + v_2$ is sufficiently large), then φ_t can be expressed as a Markovian diffusion, since it is a monotone function of the Markov process x_t.*[17]

The above example illustrates a practical benefit of our analysis, namely how to guide the choice of a functional form specification for the dynamics of φ_t that leads to a closed form solution for the dynamics of q_t and avoids the need for numerical techniques or approximations. The following proposition provides the general result.

Proposition 1 *Suppose that technical Assumption 1 in the appendix is satisfied, and that the function $\varphi(\cdot)$ in equation (21) is decreasing. Then the equilibrium stochastic process for q_t is given by (20) if, and only if, φ_t follows the (Markovian) dynamics (22). Moreover, q_t is stationary and takes values in an interval $[q^{min}, q^{max}]$.*

We conclude with three remarks. First, the process φ_t that supports a given equilibrium stochastic process for q_t is unique. Second, the process q_t only determines $\varphi_t = \eta_t^d - \eta_t^l$. The individual processes η_t^d and η_t^l can be chosen freely as long as their difference obeys the dynamics (22) and the processes are non-negative. (For instance, one choice is to set $\eta_t^l = \eta^l = \varphi(q^{\max})$ and $\eta_t^d = \eta^l + \varphi_t$, which ensures that both processes are non-negative.) Finally, the requirement that $\varphi(\cdot)$ be decreasing is only useful for ensuring that the diffusion process $\varphi(\cdot)$ is Markovian; without this requirement, a stochastic process $\varphi(\cdot)$ supporting (20) would still exist, but would not be Markov.

3.3 Recursive preferences and risk premiums

With expected-utility preferences the model faces an important limitation: Any agent's consumption is locally deterministic and so is their marginal utility. Therefore the market price of risk in this economy is zero.

[17]The price-dividend ratio (i.e., the inverse function $q_t = \varphi^{-1}(\varphi_t)$) can be computed explicitly as

$$\varphi^{-1}(\varphi_t) = \frac{1}{2\beta\alpha}\left(\frac{\varphi_t + \beta + v_1 + v_2}{\varphi_t} - \sqrt{\left(\frac{\varphi_t + \beta + v_1 + v_2}{\varphi_t}\right)^2 - \frac{4\beta\alpha}{\varphi_t}\left(1 + a_2 v_2 + \alpha_1 (v_1 + v_2)\right)}\right).$$

Using this expression for $\varphi^{-1}(\varphi_t)$ inside (22) allows one to derive a stochastic differential equation for φ_t.

To introduce a non-zero market price of risk, in this section we allow for recursive preferences and show how to support any given dynamics for the price-dividend ratio and the market price of risk jointly. The construction of the appropriate processes η_t^d and η_t^l is conceptually similar to the construction in the previous section. Hence, in order to avoid repetition, we simply state the main results and refer the reader to the appendix for the derivations.

Specifically, we continue to assume that investors have unit intertemporal elasticity of substitution (IES), but allow for a risk aversion higher than one. In mathematical terms, the consumer maximizes

$$V_{t,s} = \mathrm{E}_t \left[\int_t^{\infty} f\left(c_{u,s}, V_{u,s}\right) du \right], \text{with } f\left(c_{t,s}, V_{t,s}\right) \equiv \beta\gamma V_{t,s} \left(\log\left(c_{t,s}\right) - \gamma^{-1} \log\left(\gamma V_{t,s}\right) \right). \quad (26)$$

Here, $V_{t,s}$ is a consumer's value function and the parameter $\gamma < 0$ controls risk aversion. Preferences of this sort are the continuous-time limit of the extensively used Epstein-Zin-Weil discrete-time preferences with unit IES.[18] We note that the parameter γ in the continuous-time specification maps to a risk aversion coefficient of $|\gamma| + 1$ in the discrete-time specification. Henceforth, whenever we refer to the risk aversion coefficient, we mean $|\gamma| + 1$.

With the recursive preferences in (26), equation (13) continues to hold and so do Lemmas 1 and 2. Since $\frac{\dot{c}_{t,s}}{c_{t,s}}$ is independent of s, we shall henceforth write $\frac{\dot{c}_t}{c_t}$.[19] The only object that changes when agents have recursive preferences is the stochastic discount factor, described by the following result.

Lemma 5 *Let the process Z_t be the solution to the backward stochastic differential equation*

$$Z_t \equiv E_t \int_t^{\infty} e^{-\beta(u-t)} \left(\gamma \left(\frac{\dot{c}_u}{c_u} \right) + \frac{1}{2} \sigma_{Z,u}^2 \right) du, \quad (27)$$

[18]See Duffie and Epstein (1992).

[19]Since preferences are homothetic, the hazard rate of death is constant, and the investment opportunities are the same for all existing agents, it follows that $\frac{\dot{c}_{t,s}}{c_{t,s}}$ continues to be independent of the cohort s to which the consumer belongs.

where $\sigma_{Z,t}$ is the volatility of Z_t. Then the stochastic discount factor evolves according to

$$\frac{dm_t}{m_t} = -r_t dt - \kappa_t dB_t, \tag{28}$$

where r_t, the interest rate in this economy, continues to be given by equation (18)*, while κ_t is the market price of risk in this economy and is given by $\kappa_t = -\sigma_{Z,t}$.*

Recursive preferences imply a risk premium because consumers demand compensation for uncertainty in their discounted "long-run" consumption growth as captured by Z_t.

We next ask a question similar to the one we asked in the previous subsection. Is it possible to choose diffusion processes for η_t^l and η_t^d to support given stock-market dynamics (q_t) and given dynamics of the Sharpe ratio (κ_t)?

To provide an answer to this question, we proceed as in the previous section. Specifically, we fix functions f_Z, σ_Z, f_q, and σ_q and intervals $[Z^{\min}, Z^{\max}]$ and $[q^{\min}, q^{\max}]$ and try to determine (η_t^l, η_t^d) such that the equilibrium process Z_t follows the dynamics

$$dZ_t = f_Z(Z_t)\, dt + \sigma_Z(Z_t)\, dB_t, \tag{29}$$

and has support in $[Z^{\min}, Z^{\max}]$, while the process for q_t has support in $[q^{\min}, q^{\max}]$ and follows the dynamics

$$dq_t = f_q(q_t)\, dt + \sigma_q(q_t)\, dB_t. \tag{30}$$

As we show in the appendix, the equilibrium dynamics of Z_t and q_t obey equations (29) and (30) when and only when the functions f_Z and f_q satisfy the relations

$$f_Z(Z_t) = \beta Z_t + \gamma \nu_t - \frac{1}{2}\sigma_Z^2(Z_t) - \gamma(\lambda + g) \tag{31}$$

$$f_q(q_t) = (\beta + \varphi_t)\, q_t - \beta\alpha\varphi_t q_t^2 - 1 - \sigma_Z(Z_t)\sigma_q(q_t). \tag{32}$$

Comparing the right-hand sides of (32) and (19) shows that the two expressions are identical, except for the last term in equation (32), which captures the presence of an equity premium.

Solving for ν_t from equation (31) and for φ_t from (32) leads to the following result.

Proposition 2 *Consider intervals* $[q^{min}, q^{max}] \subset \left(0, \frac{1}{\alpha\beta}\right)$ *and* $\left[Z^{\min}, Z^{\max}\right]$. *Then there exists a unique pair of stochastic processes satisfying*

$$\nu(Z_t) = \frac{1}{\gamma}\left(f_Z\left(Z_t\right) + \frac{1}{2}\sigma_Z^2(Z_t) - \beta Z_t\right) + \lambda + g \tag{33}$$

$$\varphi\left(q_t, Z_t\right) = \frac{1 - \beta q_t + f_q\left(q_t\right) + \sigma_Z\left(Z_t\right)\sigma_q\left(q_t\right)}{q_t\left(1 - \beta\alpha q_t\right)} \tag{34}$$

such that the equilibrium stochastic processes for Z_t *and* q_t *are given by the diffusions (29) with support* $\left[Z^{\min}, Z^{\max}\right]$ *and (30) with support* $[q^{min}, q^{max}]$*, respectively.*

Remark 2 *As we show in the appendix (Lemma 7), if* $\frac{d\nu}{dZ} > 0$ *for all* $Z \in \left[Z^{\min}, Z^{\max}\right]$ *and also* $\frac{\partial\varphi}{\partial q} < 0$ *for any* $Z \in \left[Z^{\min}, Z^{\max}\right]$ *and* $q \in \left[q^{\min}, q^{\max}\right]$*, the mapping* $(Z_t, q_t) \to (\nu_t, \varphi_t)$ *is invertible and the stochastic process* (ν_t, φ_t) *is a (joint) Markov diffusion.*

Equations (33) and (34) provide explicit solutions for $\nu(Z_t)$ and $\varphi_t = \eta_t^d - \eta_t^l$. The associated values of the primitive processes η_t^d and η_t^l follow easily as solutions to the linear two-by-two system constituted by $\varphi_t = \eta_t^d - \eta_t^l$ and equation (14):[20]

$$\eta_t^d = \nu_t + \left(1 - \alpha\beta q_t\right)\varphi_t \tag{35}$$

$$\eta_t^l = \nu_t - \alpha\beta q_t\varphi_t. \tag{36}$$

Example 2 *Suppose that* x_t *follows the process* (24) *and that we wish to obtain* $Z_t = b_1 + b_2 x_t$ *and* $q_t = a_1 + a_2 x_t$ *as equilibrium outcomes with* $b_1 = \frac{\gamma}{\beta}(\lambda + g)$ *and some constants* $a_1 > 0$, $a_2 > 0$*, and* $b_2 < 0$.

[20] We note that adding a constant to both η^d and η^l shifts Z_t by a constant, but leaves its dynamics (as well as the process q_t) the same. As a consequence, one can always construct positive processes η^d and η^l by starting from an arbitrary pair $(\hat{\eta}^d, \hat{\eta}^l)$ and letting $\eta^i = \hat{\eta}^i + k$, $i \in \{d, l\}$, for k large enough — in particular, $k \geq -\min_{Z\in[Z^{min},Z^{max}],q\in[q^{min},q^{max}]} \hat{\eta}^i(q, Z)$, where the function $\hat{\eta}^i(q, Z)$ defined by plugging (33)–(34) inside (35)–(36) is continuous.

In that case equations (33)–(34) imply that ν_t and φ_t are given by

$$\nu_t = -\frac{b_2}{\gamma}\left((v_1+v_2)\,x_t + \beta x_t - \frac{b_2\sigma_x^2}{2}x_t(1-x_t) - v_2\right). \tag{37}$$

$$\varphi_t = \frac{1-\beta(a_1+a_2x_t)+a_2v_2-a_2\,(v_1+v_2)\,x_t+a_2b_2\sigma_x^2x_t(1-x_t)}{(a_1+a_2x_t)\,(1-\beta\alpha(a_1+a_2x_t))}. \tag{38}$$

The interest rate is given by (18)*, the Sharpe ratio is* $|b_2|\sqrt{x_t(1-x_t)}$*.*

With some restrictions on the parameters, the conditions of Remark 2 are satisfied and ν_t and φ_t are Markovian diffusions. (For instance, a set of sufficient conditions is $v_1+v_2+\beta+\frac{b_2\sigma_x^2}{2}>0$, $v_1+v_2+b_2\sigma_x^2+\beta>0$, and $a_1+a_2<\frac{1}{2\alpha\beta}$.)

As a final remark, we note that we have assumed throughout that Z_t and q_t (and by implication η_t^d and η_t^l) are driven by the same Brownian motion. If Z_t and q_t are driven by two Brownian motions with correlation coefficient ρ^B, the only modification is to replace $\sigma_Z\sigma_q$ in equations (32) and (34) with $\rho^B\sigma_Z\sigma_q$.[21]

3.4 Discussion

Proposition 2 is a "possibility" result, similar to the one provided in Constantinides and Duffie (1996), but predicated on qualitatively different specifications of inequality dynamics and market incompleteness. It shows that the model is able to produce a wide range of dynamics for the price-dividend ratio and the Sharpe ratio despite constant aggregate consumption and dividend growth.

It is an empirical matter to estimate the share processes and establish whether they are quantitatively consistent with the observed asset-pricing moments. We address this question in Section 4. Here we discuss (i) how this model differs from Constantinides and Duffie (1996) and (ii) the extent to which the key insights are robust to various model extensions.

One obvious difference to Constantinides and Duffie (1996) is that we do not require independent innovations to the stochastic discount factor; instead we can accommodate a Markovian structure. However, the more important difference between the two models (and

[21] To ensure market completeness, one would also need to assume a zero net supply asset to "span" the second Brownian shock.

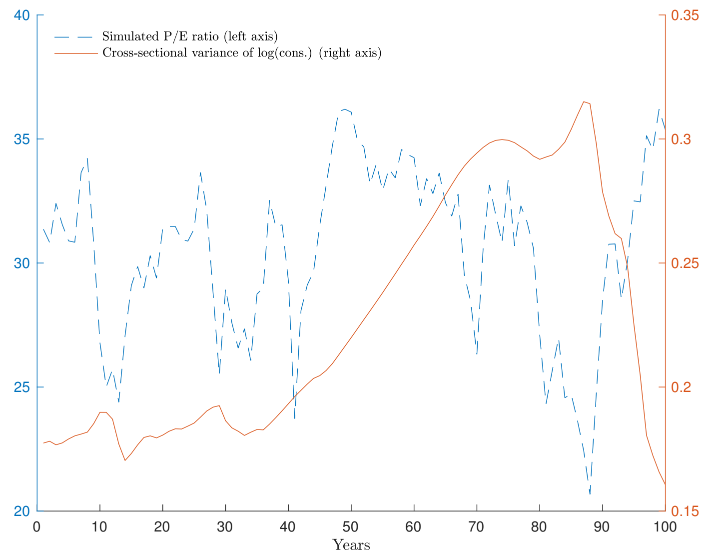

Figure 1: An indicative, model-implied path of the price-earnings ratio (left scale) and the cross-sectional standard deviation of log consumption (right scale).

indeed relative to many other models of heterogeneous agents) pertains to the dynamic behavior of inequality. To see this, it is useful to define the cross-sectional variance of log consumption as

$$\mathcal{V}_t = \lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} \left(\log\left(c_{t,s}\right) - \lambda \int_{-\infty}^{t} e^{-\lambda(t-u)} \log\left(c_{t,u}\right) du \right)^2 ds.$$

Time-differentiating $\mathcal{V}_t$ we obtain the following dynamics

$$d\mathcal{V}_t = -\lambda \mathcal{V}_t dt + \lambda \left(\log\left(c_{t,t}\right) - \lambda \int_{-\infty}^{t} e^{-\lambda(t-u)} \log\left(c_{t,u}\right) du \right)^2 dt \tag{39}$$

An immediate implication of the above equation is that $\mathcal{V}_t$ is a locally deterministic process, i.e., it has no diffusion component, and therefore essentially zero volatility over short-run time-intervals. However, the process is quite persistent, as it mean-reverts at the rate λ, the rate of population renewal. By contrast the price-dividend ratio follows the diffusion (30), which has positive volatility and may exhibit much smaller persistence.

Figure 1 provides an illustration by plotting an indicative path (of length similar to that of the post-war sample) of the price-dividend ratio and the cross-sectional variance of log consumption in the calibrated version of the model that we describe in Section 5. The weak comovement of the two series and the smooth but near unit root fluctuations in inequality are consistent with the data. We also note that this weak association between inequality and asset-price movements differentiates this model from other heterogeneous-agent models in which asset prices are driven by agents' different preferences, beliefs, etc.

Appendix B contains further discussion of the model and several possible extensions and modifications. Specifically, Appendix B.1 extends the results to time-varying aggregate growth rates and population growth rates. Appendix B.2 highlights the role played by the assumption that agents have multiple income sources. This assumption leads to a positive risk premium despite an IES no higher than one (in contrast to Bansal and Yaron, 2004). (Indeed, our analysis can be extended to an IES below one while preserving the positivity of the equity premium.) Appendix B.3 shows that allocating firms to arriving rather than existing agents is inconsequential for the results. Finally, Appendix B.4 presents an extension of the model in which only a fraction of the existing agents participate in financial markets.

4 Empirical Implications

We focused so far on the theoretical possibility of supporting given dynamics for the price-to-dividend ratio and the market price of risk as equilibrium outcomes. We turn now to the empirical measurement of the driver processes η_t^l and η_t^d. In Section 4.1 we develop a methodology to infer the consumption growth rate of a cohort member, $\frac{\dot{c}_{t,s}}{c_{t,s}}$, and we compare its properties to those of aggregate consumption growth per capita. In Section 4.2 we measure η_t^d. In Section 5 we calibrate the model to reproduce the properties of these time series.

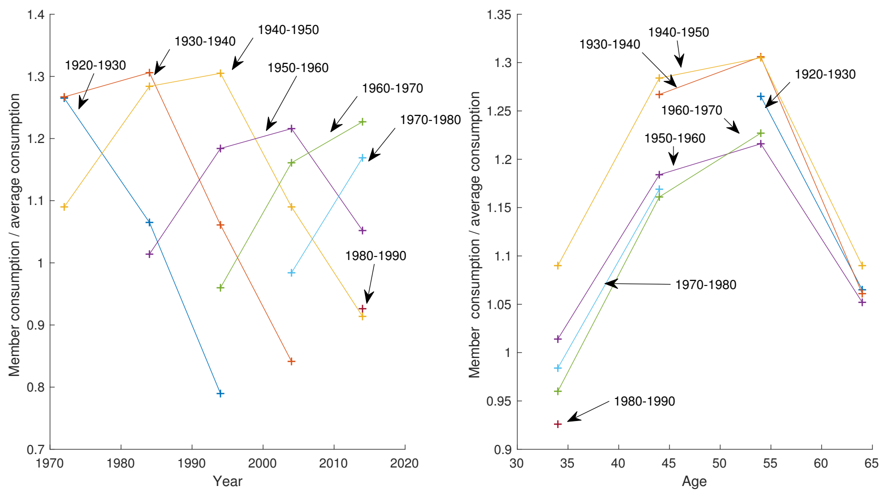

Figure 2: Average per-household consumption of cohorts born in different decades divided by average household consumption at the time of observation. The left plot depicts the information as a function of time, the right plot as a function of age.

4.1 Measuring cohort consumption growth

4.1.1. Data and preliminary observations

To motivate the results of this section, we present a figure that illustrate the differences in consumption allocation across different cohorts. We use the tables readily available on the website of the Consumer Expenditure Survey, which report household consumption by 10-year age groups[22] (25–34, 35–44, etc.) for the year 1972 and then annually from 1984 to 2018. We isolate the years 1972, 1984, 1994, 2004, and 2014, so that there is a ten year gap between cross sections (with the exception of the first cross section, after which the gap is 12 years). By creating a ten year gap between cross sections, we can follow the same cohort across time.[23] For each cohort and year of observation we divide by the average (across all age groups) household consumption in that year from the same CEX tables.

Figure 2 shows that every cohort exhibits a hump-shaped consumption pattern over the life cycle, which is similar across cohorts and across time. The right plot highlights that these hump-shaped consumption profiles are roughly parallel to each other, indicating the

[22]Age refers to the age of the head of the household.

[23]People born, say, between 1950 and 1960 that were in the 25–34 age group in the 1984 cross section will be in the 35–44 age group in the 1994 cross section.

presence of a permanent, cohort-specific effect on consumption. For example, compare the cohort born between 1940–1950 with the one born between 1960–1970. The first cohort consumed 1.1 of average consumption when aged 25–34 as opposed to 0.96 for the second cohort when aged 25–34. This gap continues to manifest itself at all stages of the two cohorts' life cycles.

These patterns are not special to CEX data. Figure 10 in Appendix D.1 shows that exactly the same patterns characterize the income data in tables P9 and P10 on the website of the Current Population Survey (CPS). The CPS data are at the person (rather than household) level, and any measurement errors are likely uncorrelated across the two surveys.[24]

The permanent differences between the consumption (and income) shares of different cohorts (controlling for life-cycle effects) is a central feature of our model. Indeed, with $s' < s$, equation (17) gives $\frac{c_{t,s'}}{c_{t,s}} = \frac{\nu_{s'}}{\nu_s} e^{-\int_{s'}^{s}(\nu_u - \lambda)du}$, which is independent of t and thus implies a constant (across time) consumption ratio between two distinct cohorts.

4.1.2. Estimation

Here we propose and implement a methodology to measure the marginal-agent consumption growth, $\frac{\dot{c}_{t,s}}{c_{t,s}}$. In performing this exercise, we pay particular attention to the facts that i) in the data cohorts have different sizes, ii) there are hump-shaped consumption profiles over the life cycle, and iii) the population mass at any given point in time may be an arbitrary function of t and s.

We start by remarking that equation (17) implies that $\log c_{t,s}$ can be decomposed as

$$\log c_{t,s} = A_s + L_t + G(t-s), \tag{40}$$

where $L_t \equiv -\int^t \nu_u du + \log C_t$ is a time effect, $A_s \equiv \log\left(\frac{\nu}{\lambda}\right) + \int^s \nu_u du$ is a cohort effect and $G(t-s) \equiv \lambda(t-s)$ is an age effect. The change in the time effects, $dL_t = d\log C_t - \nu_t dt$, captures the consumption growth of a fixed cohort member and therefore we will

[24]For both Figures 2 and 10 we used directly the summary tables on the websites of the CEX and the CPS (rather than the publicly available micro-data), which are based on the raw (non-top-coded) data available to these agencies. The patterns in Figure 10 remain the same when we examine median rather than mean income, and when we produce the graph by males and females separately.

henceforth refer to dL_t as the consumption growth of the marginal agent. Intuitively, the inclusion of cohort effects in the specification (40) means that dL_t can be understood as the consumption change *within* a given cohort. In addition, the inclusion of age effects removes the (deterministic) effects of aging on consumption growth. To allow for salient empirical features, in our estimation we do not impose the linear age effects implied by our baseline model, but rather estimate a general function G_{t-s} (which could be a reflection of age-dependent discount rates).[25]

Equation (40) suggests a straightforward approach to estimating marginal-agent consumption growth: regress $\log c_{t,s}$ on time, age, and cohort effects (dummy variables), and compute the first differences of the time effects. Such an approach, however, is limited by the lack of availability of long time series of cross-sectional consumption data. (Annual CEX cross sections start in the mid-eighties and therefore estimates of L_t are only available for the last three decades.) Quite remarkably, however, equation (40) allows an indirect approach to identify L_t even for times t for which cross-sectional data are not available.

Specifically, with $\Lambda_{t,s}$ the population of cohort s at time t, aggregating equation (40) yields

$$C_t = \int_{-\infty}^{t} \Lambda_{t,s} e^{A_s + L_t + G_{t-s}} ds. \tag{41}$$

[25]To see this more formally, suppose that we introduce age effects into our model by introducing an age-dependent discount factor ρ_{t-s}. Then the Euler equation becomes $\frac{\dot{c}_{t,s}}{c_{t,s}} = r_t - \rho_{t-s}$, which can be written in the regression form (40) with $L_t = \int^t r_u du$, $A_s = \log c_{s,s} - \int^s r_u du$, and $G_{t-s} = -\int_s^t \rho_{u-s} du$. This implies that the validity of an age, time, and cohort decomposition is unaffected by the presence of arbitrary age effects in consumption. The same goes for cohorts having different sizes, etc., which does not affect the validity of the Euler equation for each member of a cohort. In these model extensions, any common stochastic variation in the investment opportunity set that is relevant for asset pricing would be measured by dL_t, since the relation $L_t = \int^t r_u du$ holds irrespective of a constant, or age-dependent discount rate.

We should note, however, that while the validity of both a) a time, age, and cohort decomposition for log consumption and b) the relation between time effects and expected real rates are unaffected by age-dependent discount rates, this is not the case for all the aspects of the model. With age-dependent discount rates, one would need to replace $e^{-\beta(u-t)}$ in the definition of Z_t with $e^{-\int_t^u \beta_{z-t} dz}$, which makes Z_t and, in particular, the volatility σ_Z, depend on age. In that case, the Sharpe ratio is given by $-\int_{-\infty}^t w_{c,t-s} \sigma_{Z,t-s} ds$, where $w_{c,t-s}$ is the consumption share of cohort s. The consumption distribution $w_{c,t-s}$ therefore becomes an (infinite-dimensional) state variable. By using a constant discount rate β in our model, the volatility of σ_Z is age-independent, the Sharpe ratio does not depend on the wealth distribution, and an implication of equations (27) and (40) is that knowledge of the time-series dynamics of L_t is sufficient for the determination of the Sharpe ratio.

Defining

$$F_t \equiv \log\left(\int_{-\infty}^{t} \Lambda_{t,s} e^{A_s + G_{t-s}} ds\right) \tag{42}$$

and taking logarithms in equation (41) implies $\log(C_t) = L_t + F_t$ and therefore

$$dL_t = d\log(C_t) - dF_t. \tag{43}$$

Equation (43) presents an indirect way to infer the variation in dL_t. Computing F_t does not require a long time series of cross-sectional consumption data, since the only estimated quantities that enter the equation are the cohort effects (A_s) and the age effects (G_{t-s}) — thus, not the time effects (L_t). In principle, just two cross sections $T-1$ and T suffice to compute a long path of cohort and age effects, with more cross sections reducing the estimation error.

A common concern with regression (43) is that time, age, and cohort effects can only be identified up to an affine term.[26] The implication of this non-identifiability is that dF_t (and hence dL_t) can only be identified up to an additive constant.[27] However, the variation of dF_t (and hence dL_t) around its mean is uniquely identified, and this variation is the relevant quantity for asset-pricing purposes.

4.1.3. Data and implementation

To implement our empirical approach, we need data on annual aggregate consumption growth $d\log C_t$ (available from the BEA since 1929), the population $\Lambda_{t,s}$ of people alive at time t

[26]This means that the data cannot distinguish the model of equation (43) from the alternative model

$$\log c_{t,s} = \underbrace{A_s + \chi s}_{\text{modified cohort effect}} + \underbrace{L_t - \chi t}_{\text{modified time effect}} + \underbrace{G_{t-s} + \chi(t-s)}_{\text{modified age effect}},$$

for some arbitrary constant χ.

[27]This statement can be proven by using the modified age and cohort effects inside (42) to obtain

$$dF_t^{(\chi)} \equiv d\log\left(\int_{-\infty}^{t} \Lambda_{t,s} e^{A_s + \chi s + G_{t-s} + \chi(t-s)} ds\right) = d\log\left(e^{\chi t}\int_{-\infty}^{t} \Lambda_{t,s} e^{A_s + G_{t-s}} ds\right) = \chi dt + dF_t.$$

Hence, $dF_t^{(\chi)} = \chi dt + dF_t$.

that were born at time s (available from the Census and U.N. population statistics at annual frequency since 1910), and cross-sectional consumption data to estimate the cohort effects A_s and the age effects G_{t-s} (available from the Consumer Expenditure Survey) since 1990.

Since in the data we can only obtain estimates of cohort effects for people born from 1900 onwards, we extrapolate linearly cohort effects prior to 1900. In Appendix C we report results from alternative extrapolation or truncation methods. We show that, as long as we focus attention on post 1960 data, all methods give essentially the same results, since the population weight of pre-1900 cohorts is relatively small. That appendix also contains a detailed exposition of many other choices needed to obtain a measurement of F_t, along with a discussion of potential issues related to measurement error, discrepancies between NIPA and CEX data, etc.

In Appendix C, we also perform a validation exercise, whereby we compare the results of our indirect approach to estimating L_t with a direct estimation of L_t over the subsample where this is possible (since 1990). We show that both approaches lead to the same inference about the low-frequency movements of dL_t, which is the focus of our analysis. We relegate all measurement-related details to the appendix and continue with a presentation of the results.

4.1.4. Comparing aggregate and marginal-agent consumption growth

Letting Δ denote the first-difference operator (at annual frequency) and N_t denote the population at time t, Figure 3 plots the difference between marginal-agent consumption growth $\Delta L_t = \Delta \log C_t - \Delta F_t$ and the consumption growth typically used in representative-agent asset pricing, $\Delta \log C_t - \Delta \log N_t$. The resulting difference, $\Delta \log N_t - \Delta F_t$ is depicted in Figure 3. Since the mean of ΔF_t is not identified, we normalize the mean of $\Delta \log N_t - \Delta F_t$ to zero. The figure illustrates that the difference between marginal and per-capita consumption growth rates has small year-over-year volatility but is remarkably persistent, which is consistent with our model.[28,29]

[28] Equations (13) and (15) imply that the difference between marginal-agent consumption growth and per-capita consumption growth, $\frac{\dot{c}_{t,s}}{c_{t,s}} - g$, is equal to $\lambda - \nu_t$, which is a persistent process in the model.

[29] These low-frequency cycles can be traced to phenomena such as the baby boom and the relative economic weakness of cohorts born after the baby boomers. The figure shows that the difference between marginal consumption growth and aggregate per capita consumption growth peaks in 1980, hits a trough in 2000, and

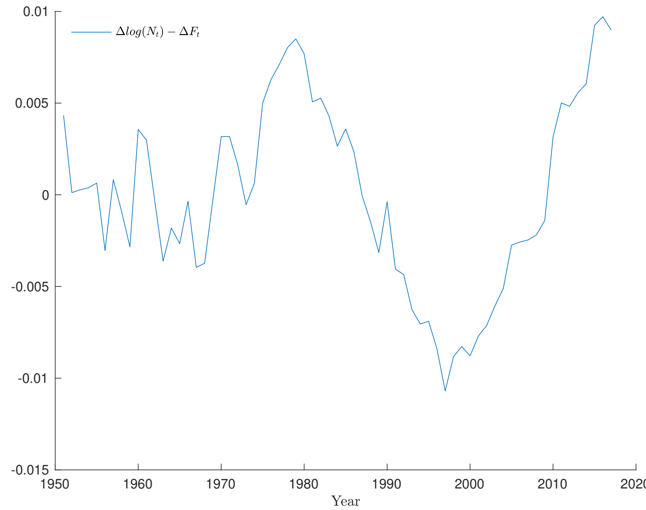

Figure 3: Difference between marginal-agent consumption growth and per capita consumption growth. The difference is normalized to have mean zero.

The slow moving economic and demographic forces that cause these persistent variations are largely independent from aggregate consumption growth per capita. Appendix D.2 contains an econometric VAR model of the joint time series properties of i) aggregate consumption growth per capita, $\Delta \log C_t - \Delta \log N_t$ and ii) the difference between marginal and aggregate consumption growth rates, $\Delta \log N_t - \Delta F_t$. Using this econometric framework, we show that the two series exhibit only a small low-frequency correlation (-0.11). This approximate orthogonality between the two series is one of the reasons why our model abstracts from fluctuations in aggregate consumption growth; the phenomena we wish to highlight are largely independent from the long-run risks in aggregate consumption growth per capita, which have been the focus of the long-run risks literature. As we discuss in Appendix D.2, an additional implication of this low correlation is that the the marginal-agent consumption growth, which is the sum of $\Delta \log C_t - \Delta \log N_t$ and $\Delta \log N_t - \Delta F_t$, has almost twice as high

then starts increasing again. The main driver of the slowdown after 1980 is that the fraction of aggregate consumption accruing to middle-aged and older groups (say, ages 45 and above) has been fairly stable over time (Figure 8 in Appendix C). As the populous baby boomers start becoming members of the middle-aged group in the eighties and early nineties, the implication is that the per-household consumption growth of the cohorts that are middle-aged in the mid-eighties experiences a slowdown when compared to aggregate per capita consumption growth. This effect reverses in early-to-mid 2000, as the cohorts that start joining the middle-aged population are both smaller and less economically successful.

a long-run standard deviation as aggregate consumption growth per capita.

4.1.5. Relating marginal-agent consumption growth to the interest rate

According to the Euler equation (12), there is a relation between the consumption growth of the marginal agent and the real interest rate. In this section we investigate this relation empirically.

The top left panel of Figure 4 plots the expected real interest rate[30] and marginal-agent consumption growth for the respective year. For comparison purposes, the top right panel of Figure 4 is similar to the the top left panel except that marginal-agent consumption growth is replaced with aggregate consumption growth per capita. To further aid the comparison between marginal-agent and aggregate consumption growth per capita, the bottom left panel plots the real expected rate and the difference between marginal-agent and aggregate consumption growth-per capita. Because of the unobserved discount rate, we add appropriate constants to the two consumption-growth measures to match the level of the real expected interest rate.[31]

Clearly, realized marginal-agent consumption growth differs from the real expected rate, a prime reason being that our model abstracts from shocks to aggregate consumption. However, if the real interest rate reflects the expected (rather than the realized) marginal-agent consumption growth, then we should find that the co-movement between the two series increases as we time aggregate them over longer horizons. The bottom right panel of Figure 4 shows that the the 12-year moving average of the real expected rate and the respective 12-year moving average of marginal-agent consumption growth co-move very closely (correlation approximately 0.8).

[30]To measure the expected real interest rate, we use the short term nominal interest rate from Robert Shiller's online data set "long term stock, bond, interest rate and consumption data" minus the (ex ante) expected inflation rate as formed in December of the preceding year. (Source: Philadelphia Fed Livingston inflation expectations survey — "Growth of median forecast for the Levels of Survey Variable (CPI).")

[31]Specifically, for the top two panels the constants are chosen so that the the consumption series start at the same level as the average real expected rate. Similarly, in the lower left plot, we add a constant to the difference between the two measures of consumption growth, so that the average value of this difference coincides with the average value of the real expected rate. For the bottom right figure, where we plot all three series together, the additive constants are chosen so that all three series start at the same value, to aid the visual comparison.

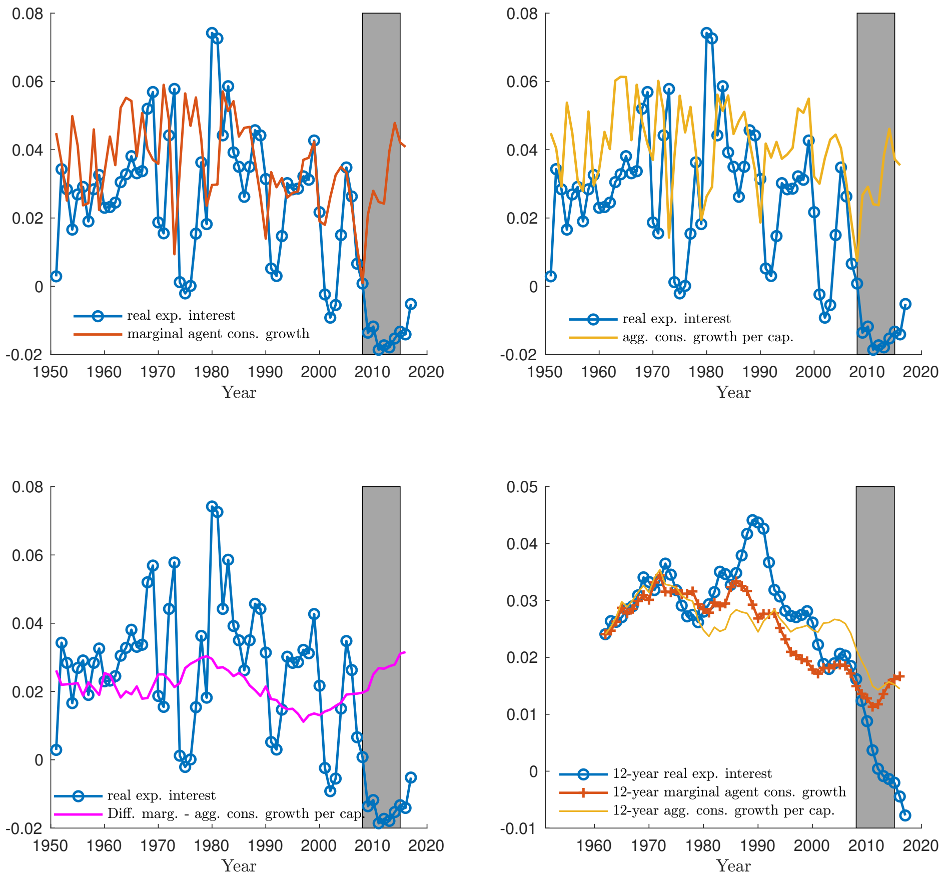

Figure 4: Top left plot: Expected real interest rate at the beginning of each year and marginal-agent consumption growth (plus constant) over the year. Top right plot: Expected real interest rate at the beginning of each year and aggregate consumption growth per capita (plus constant) over the year. Bottom left plot: Expected real interest rate at the beginning of each year and difference between marginal-agent and aggregate consumption growth per capita (plus constant). Bottom right: Twelve-year moving averages of expected real interest rate, marginal-agent consumption growth (plus constant) and aggregate consumption growth per capita (plus constant). For the choice of the additive constants, see footnote 31.

To see whether marginal-agent consumption growth can account for the fluctuations in the real expected rate better than the aggregate consumption growth per capita, it is useful to consider first the subsample that excludes the Great Financial Crisis (GFC) and its aftermath. (In Figure 4 we grey out the period 2008-2016, which encompasses the GFC and the zero-lower bound period). Prior to the GFC, aggregate consumption growth per capita shows no particular downward secular trend from 1982 onward, as the real expected rate does. By contrast, the top left panel of Figure 4 shows that marginal-agent consumption growth starts declining in 1980 and follows the decline in the expected real rate until the onset of the GFC. During the GFC both marginal-agent and aggregate consumption-growth per capita drop initially and rebound thereafter, with marginal-agent consumption growth rebounding more strongly than aggregate consumption growth per capita.

In Appendix D (Table 7) we provide a more formal empirical analysis of these visual impressions. Using GMM, we estimate the moment equation

$$\mathrm{E}_t\left[r_{t+1} - \left(\alpha + \beta_{\text{agg pc}}\Delta c_{t+1}^{\text{agg pc}} + \beta_{\text{marg}}\Delta c_{t+1}^{\text{marg}}\right)\right] = 0, \tag{44}$$

where r_{t+1} is the expected real rate, $\Delta c_{t+1}^{\text{agg pc}} \equiv \Delta \log C_t - \Delta N_t$ is aggregate consumption growth per capita and $\Delta c_{t+1}^{\text{agg pc}} \equiv \Delta \log C_t - \Delta F_t$ is marginal-agent consumption growth. Excluding 2008-2016, only $\Delta c_{t+1}^{\text{marg}}$ is significant (and positive). $\Delta c_{t+1}^{\text{agg pc}}$ is wrongly signed and insignificant, when both $\Delta c_{t+1}^{\text{marg}}$ and $\Delta c_{t+1}^{\text{agg pc}}$ are included in the estimation. (The same conclusions hold when we look at other subsamples that exclude the GFC). Including the GFC, both $\Delta c_{t+1}^{\text{marg}}$ and $\Delta c_{t+1}^{\text{agg pc}}$ are significant when considered in isolation. However, when both are included in the estimation, one cannot ascertain which consumption growth measure dominates; the standard errors become large resulting in $\beta_{\text{agg pc}}$ and β_{marg} being individually insignificant. The reason why $\Delta c_{t+1}^{\text{agg pc}}$ performs better when the GFC is included is that $\Delta c_{t+1}^{\text{agg pc}}$ exhibits a noticeable decline at the onset of the GFC and the years thereafter. Since the decline of the real rate accelerates during that period, $\Delta c_{t+1}^{\text{agg pc}}$ starts to show some signs of explanatory power.

Our interpretation of these findings is as follows. Marginal-agent consumption growth can better account (compared to aggregate consumption growth per capita) for the onset of the secular decline in the real expected rate in the early eighties. When the sample excludes the most significant business cycle (the GFC), the secular components of all three time series ($r_{t+1}, \Delta c_{t+1}^{\text{agg pc}}, \Delta c_{t+1}^{\text{marg}}$) become more important and this makes marginal-agent consumption growth outperform aggregate consumption growth per capita. When the major business

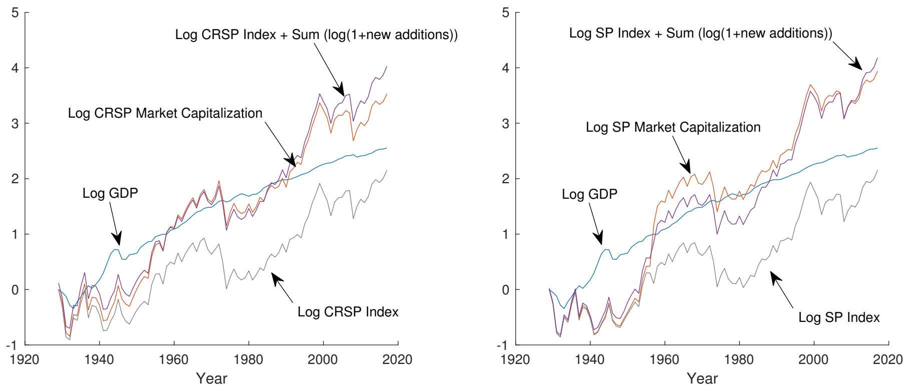

Figure 5: Left plot: Logarithm of GDP, logarithm of the CRSP market index (obtained by cumulating ex-dividend CRSP gross returns), log market capitalization, and log market capitalization plus the cumulative sum of the logarithm of 1+addition rate to the market. The addition rate is defined as the market 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.

cycle of the sample is included, then the cyclical components of all three time series become relatively more important. Since the two measures of consumption growth only differ in their "low frequency" components, it becomes harder to tell which consumption measure performs better in accounting for the interest rate.

We should also note that our approach to inferring marginal-agent consumption growth relies implicitly on the validity of a time, age, and cohort decomposition. The validity of such a decomposition and the relation between the estimated time-effects and the real expected rate are all premised on the validity of Euler equations. If credit markets are impeded (as they probably were during the GFC), then the discrepancy between marginal-agent and aggregate consumption growth per-capita need not be associated with the real expected rate and may just reflect the relative endowment changes of the different cohorts. In short, the GFC is a problematic period for testing the model on a priori theoretical grounds.

4.2 The measurement of η_t^d

The model implies the following result.

Lemma 6 *Let* $P_t^A = \int_{-\infty}^{t} P_{t,s}\, ds$ *denote aggregate market capitalization and let* $\pi_s = \frac{P_{t,s}}{P_t^A}$

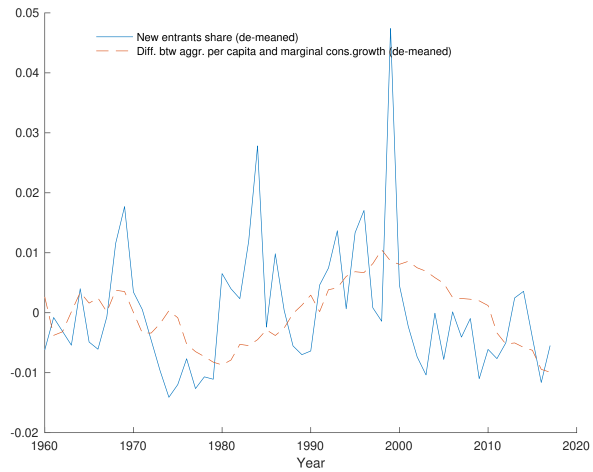

Figure 6: Market share of new entrants and difference between aggregate per capita and marginal-agent consumption growth. Since the level of marginal-agent consumption growth is not identified, we de-mean both series.

denote the market-capitalization weight of firms of vintage s. Then

$$\eta_t^d = \overbrace{\underbrace{\frac{P_{t,t}}{P_t^A}}_{\substack{\textit{Total market}\\ \textit{capitalization}}}}^{\substack{\textit{Market capitalization}\\ \textit{of firms added in year } t}} = \underbrace{\frac{\dot{P}_t^A}{P_t^A}}_{\substack{\textit{Aggregate market}\\ \textit{capitalization growth}}} - \underbrace{\int_{-\infty}^{t} \pi_s \frac{\dot{P}_{t,s}}{P_{t,s}} ds}_{\substack{\textit{Market capitalization growth of}\\ \textit{firms already in the market portfolio}}}. \tag{45}$$

The first equality in (45) provides a straightforward empirical proxy for η_t^d: the ratio of the market value of additions to the market index to the total market value of the index. We use this measure in our calibration. According to the model, this ratio also equals the discrepancy between aggregate market capitalization growth and the market capitalization growth of firms already in the market portfolio, i.e., the (ex-dividend) return on the index.

Figure 5 illustrates equation (45) in the data. The solid line in the figure depicts the (log) level of the market index. The figure also depicts the aggregate gross domestic product (GDP) series and the total stock-market capitalization. The figure shows that the log-level of the index (which is identical to the cumulative sum of the log gross ex-dividend returns) follows a markedly slower growth than the aggregate stock market capitalization. Interestingly, the discrepancy between these two comes down almost entirely to the value of additions to the index: adding the cumulative sum of $\log\left(1+\frac{P_{t,t}}{P_t^A}\right)$ to the log index results in a series that tracks the market capitalization growth closely, consistent with (45).

Figure 6 compares the path of the discrepancy between aggregate per capita and marginal-agent consumption growth that we obtained in the previous subsection with the path of the new entrant share, η_t^d.[32] Since our empirical approach can only identify marginal-agent consumption growth up to an additive constant, the figure removes the means from both series and depicts fluctuations around the respective means. Note that inside our model the difference between aggregate per capita and marginal-agent consumption growth is equal to ν_t (up to an additive constant[33]) and therefore the graph shows the empirical counterparts of the (de-meaned) processes η_t^d and ν_t. While the two series differ on a year-to-year basis, they appear to share the same low-frequency cycle.

5 Calibration

We start by choosing functional forms for the dynamics of η_t^l and η_t^d. We choose these parametric forms judiciously so as to support closed-form solutions for the dynamics of the price-dividend ratio q_t and the Sharpe ratio κ_t. Second, we choose the parameters governing the dynamics of η_t^l and η_t^d to match the empirical moments of η_t^d and ν_t in the data. (Note that by equations (35)–(36) there is a one-to-one correspondence between the pairs (η_t^l, η_t^d) and (η_t^d, ν_t).) Then we examine the resulting moments of asset-price dynamics and compare them to the data.

We modify Example 2 to allow for an imperfect instantaneous correlation between ν_t and φ_t and a log-linear specification for q_t. To achieve this, we start with two (ancillary) processes x_t and y_t given by $dx_t = (-v_1 x_t + v_2(1-x_t))dt + \sigma_x x_t(1-x_t)dB_t^{(1)}$ and $dy_t = (-v_3 y_t + v_4(1-y_t))dt + \sigma_y\sqrt{y_t(1-y_t)}dB_t^{(2)}$, where the Brownian motions $B_t^{(1)}$ and $B_t^{(2)}$ have

[32] The graph treats 1963 and 1972 (years when AMEX, respectively NASDAQ, were added to CRSP) as outliers. To preserve continuity of the series, we just replace the value in 1963 by the average of the values in 1962 and 1964 and similarly for 1972).

[33] As we note in footnote 28, in the model $g - \frac{\dot{c}_{t,s}}{c_{t,s}} = \nu_t - \lambda$.

v_1	0.011	β	0.02	γ	-8
v_2	0.078	α	0.4	g	0.025
v_3	0.014	σ_x	0.25	σ_y	0.132
v_4	0.06	a_1	0.5	a_2	3.2
b_2	-11				

Table 1: Parameters used in the model calibration. The value $\gamma = -8$ maps into a risk aversion coefficient of $1 - \gamma = 9$.

instantaneous correlation ρ^B. We then specify ν_t and φ_t as

$$\nu_t = -\frac{b_2}{\gamma}\left[(v_3 + v_4)\, y_t + \beta y_t - \frac{b_2\sigma_y^2}{2} y_t(1 - y_t) - v_4\right] \tag{46}$$

$$\varphi_t = \frac{e^{-a_1 - a_2 x_t} - \beta + a_2(v_2 - x_t(v_1 + v_2)) + \frac{a_2^2}{2}\sigma_x^2 x_t^2(1 - x_t)^2 - a_2\rho^B\kappa_t\sigma_x x_t(1 - x_t)}{1 - \alpha\beta e^{a_1 + a_2 x_t}}, \tag{47}$$

where $\kappa_t = |b_2|\sigma_y\sqrt{y_t(1 - y_t)}$. With these specifications for ν_t and φ_t, the formulae of Section 3.3 imply that the equilibrium price-dividend ratio is log-linear in x_t, $\log(q_t) = a_1 + a_2 x_t$, while the Sharpe ratio is given by κ_t.[34]

We fix preference parameters to $\beta = 0.02$ and $\gamma = -8$, which implies a risk-aversion coefficient of $|\gamma| + 1 = 9$. We set α to a level that reflects the share of capital income in output (0.4). The aggregate growth rate (g) is set to 0.025, similar to the data.[35] We choose the parameters that govern the dynamics of ν_t and φ_t in equations (46)–(47) — that is, v_1, v_2, v_3, v_4, σ_x, σ_y, b_2, a_1, a_2, and ρ^B — to (approximately) match the autocorrelation coefficients of ν_t and η_t^d, and minimize the distance between the empirical and model-implied stationary distributions of ν_t and η_t^d (as measured by the Kolmogorov-Smirnov distance).[36] For the

[34]We allow for slightly different volatility specifications for x_t and y_t, as it helps better match the stationary distributions of ν_t and φ_t in the data.

[35]The model does not feature population growth (for simplicity), while the data does. Aggregate (real) consumption growth per capita is roughly 2% in the data, whereas aggregate (real) consumption growth is roughly 3%. We set g to 0.025 to match the average of these two numbers. We note that the choice of g is not important for the results we highlight.

[36]We measure the autocorrelation of ν_t and η_t^d by fitting a univariate ARMA(1,1) for both series and coefficient. We choose this approach because the series for η_t^d in the data has a negative MA(1) coefficient, presumably due to the fact that some market additions tend to cluster around "hot markets," which tend to revert the following year. The MA(1) term in the estimation is meant to control for such effects. To be consistent, we estimate the same ARMA(1,1) inside our model and compare the autoregressive coefficients in the model and the data.

	Data	Model
Median arrival rate of new firms	1.71%	2.29% (1.16%)
Standard deviation of the arrival rate of new firms	1.81%	1.74% (1.07%)
Autocorrelation of the arrival rate of new firms	0.87	0.83 (0.0933)
Median value of ν_t		2.61% (1.16%)
Standard deviation of of ν_t	0.53%	0.57% (0.38%)
Autocorrelation of imputed ν_t	0.93	0.83 (0.1087)
Correlation of imputed ν_t and new firm arrival rate	0.44	0.47 (0.3774)

Table 2: Targeted moments, model and data. We simulate 1000 independent paths of similar length as the data, and compute each of the six moments for every path. We then report the mean and standard deviation (across the 1000 paths) for each moment. The term "arrival rate of new firms" refers to the ratio of the value of the market value of additions to the market portfolio to the total value of the market portfolio. The correlation between ν_t and η_t^d is the correlation of 5-year moving averages of ν_t and η_t^d, so as to better capture the low frequency co-movement of the series.

chosen parameters, the Kolmogorov-Smirnov test cannot reject that the model-implied stationary distributions of ν_t and η_t^d are the same as the respective stationary distributions in the data (p-values of 0.21 and 0.23). Figure 11 in the appendix provides a visual comparison between the model-implied and empirical distributions of ν_t and η_t^d in the model and the data. We note that since the mean value of ν_t cannot be identified with the time, age, and cohort decomposition method of the previous section, we target instead $\frac{c_{t,t}}{Y_t} \approx 1$ motivated by the evidence in Figure 2.[37] Finally, the parameter ρ^B is chosen to match the low-frequency correlation between ν_t and η_t^d illustrated in Figure 6. Specifically, we choose ρ^B so that the correlation of five-year moving averages of ν_t and η_t^d in the model and the data coincide.[38]

[37]This implies a mean value of ν_t close to the sum of the death, birth, and (net) immigration rate, which is about 2.5% in our sample.

[38]In the data the correlation between ν_t and η_t^d increases as we time-average the series over longer horizons (from 0.25 at the one-year horizon to 0.45 at the five-year horizon). We note, however, that even without considering moving averages of the two series, the empirical value of the correlation between ν_t and η_t^d is well within two standard deviations of its model-implied counterpart, when when we repeatedly sample 60-year-long samples from the model.

	Data	Model
Sharpe ratio	0.29	0.26
Stock market volatility	18.2%	14.19%
Equity premium	5.2%	3.86%
Average interest rate	2.8%	1.72%
Standard deviation of real interest rate	0.92%	0.85%
Average (log) PE ratio	2.65	2.77
Standard deviation of (log) PE ratio	0.34	0.21
Autocorrelation of (log) PE ratio	0.86	0.91

Table 3: Unconditional moments for the data and the model. The data for the average equity premium, the volatility of returns, and the level of the interest rate are from Gârleanu and Panageas (2015) (Table 1), and are sourced from the long historical sample available from the website of R. Shiller (http://www.econ.yale.edu/ shiller/data.htm), as are the data for the price-earnings ratio. The volatility of the real rate is inferred from the yields of five-year constant maturity TIPS. We note that the volatility of the real expected rate in Figure 4 is about 2%, which is higher than the volatility of the real rate implied by TIPS (0.92%). (This statement applies also to the smaller subsample over which TIPS data are available.) Because inflation expectation surveys may contain error (due to sampling of forecasters) we choose to target the lower real rate volatility implied by TIPS.

Table 2 shows that these parameter choices allow us to broadly reproduce the targeted empirical moments within our model. To account for estimation error, we do not only report average values of the targeted moments inside the model, but also the standard deviation for these values from simulations of the model over a similar number of years as in the data. The table shows that the moments in the data are within two standard deviations of their simulated means inside the model. Figure 11 in the appendix provides an alternative, graphical illustration of Table 2 by comparing the empirical and the simulated distributions of ν_t and η_t^d.

Having determined the parameters to match the moments of the share processes, we next examine what these parameter choices imply for asset-pricing moments. Table 3 provides a comparison between the model-implied unconditional moments and the respective moments in the data. In reporting the results we follow the approach of Barro (2006) to relate the results of our model (which produces implications for an all-equity financed firm) to the data (where equity is levered). Specifically, we use the well known Modigliani-Miller formula,

Year	β (Data)	β (Model)	R^2 (Data)	R^2 (Model)
1	-0.130	-0.080	0.040	0.029
		[-0.204 0.014]		[0.000 0.098]
3	-0.350	-0.212	0.090	0.073
		[-0.525 0.039]		[0.000 0.227]
5	-0.600	-0.315	0.180	0.107
		[-0.727 0.078]		[0.000 0.314]
7	-0.750	-0.399	0.230	0.134
		[-0.893 0.106]		[0.001 0.380]

Table 4: Long-horizon regressions of excess returns on the log P/D ratio. The simulated data are based on 1000 independent simulations of 100-year long samples. For each of these 100-year long simulated samples, we run predictive regressions of the form $\log R^e_{t\to t+h} = \alpha + \beta \log(P_t/D_t)$, where $\log R^e_{t\to t+h}$ denotes the time-t gross excess return over the next h years. We report the mean values for the coefficient β and the R^2 of these regressions, along with the respective [0.025, 0.975] percentiles. The columns labeled "Data" are from Gârleanu and Panageas (2015), Table 2.

according to which the levered equity return is equal to the unlevered equity return times 1.7 (the leverage ratio in the data — see, e.g., Barro, 2006).[39] We note in passing that another popular choice in the literature is to set the leverage parameter to three (e.g., Bansal and Yaron, 2004). To compare our results to these papers, one would have to multiply our equity premium and return volatility by $\frac{3}{1.7} = 1.76$.

Table 3 shows that the model accounts for a non-trivial fraction of all asset pricing moments. To put these numbers in the proper relation to the literature, it is worth highlighting that aggregate consumption and dividend growth are constant in this model. The numbers should therefore be interpreted as the asset-pricing moments that would obtain in an economy that abstracts from all aggregate sources of uncertainty and examines the impact of the share processes in isolation.[40]

[39] Another implication of leverage is that the price-to-earnings ratio of a levered firm is $\frac{1}{1.7}$ times the price-to-earnings ratio of an unlevered firm, since equity accounts for $\frac{1}{1.7}$ of the total value of the firm. To be consistent, in this section we report the log-price-earnings ratio of a levered firm, which is lower than the price-earnings ratio of an unlevered firm by the constant log(1.7).

[40] Some additional notes on Table 3: The reported Sharpe ratio in Table 3 is the ratio of the excess return of the stock market to its volatility. The volatility of the stochastic discount factor (or equivalently the Sharpe ratio of the growth-optimal portfolio) in our model is 0.52. Since the return on the growth-optimal portfolio is not observable, we focus on the Sharpe ratio of the stock market. Also, unlike in the model,

Table 3 only pertains to unconditional moments. In terms of conditional moments, Table 4 shows that the model-simulated data exhibit the same predictability patterns as empirical data.

Figure 12 in the appendix performs a different exercise to evaluate the model's pathwise properties. Using the paths of η_t^d and ν_t in the data, we plot the $\log(q_t)$ path that would obtain in the model and compare it to the data. Section D.5 provides the details.

6 Conclusion

Aggregate per capita consumption growth and the consumption growth of a fixed birth-cohort member can differ in an economy characterized by imperfect, inter-cohort risk sharing. The Euler equation applies at the level of a given cohort member, but not at the aggregate level. We illustrate these notions by using an overlapping generations model whereby random endowment shocks over a cohort's lifetime affect the consumption of its members, even though aggregate consumption evolves deterministically. When coupled with recursive preferences, the long-run uncertainty in a representative cohort member's consumption path requires risk compensation. This risk compensation is not explained by the conditional or unconditional versions of the aggregate-consumption CAPM, or the Sharpe-Lintner-Mossin CAPM, which don't hold inside the model. Compared to earlier approaches linking heterogeneous consumption processes and asset prices, such as Constantinides and Duffie (1996), our framework implies an inequality process that is non-volatile, weakly related to asset-price fluctuations at high frequencies, and substantially more persistent than the price-to-dividend ratio.

We develop an empirical strategy to infer the wedge between the consumption growth at the aggregate and cohort-member levels by utilizing a time, age, and cohort decomposition of cross-sectional consumption data and imposing market clearing. We evaluate the model quantitatively and show that the consumption growth of a fixed cohort member con-

where earnings are a continuous flow, in the data earnings are a flow over an interval and we need to adopt a convention as to whether they accrue at the beginning or the end of the period ("beginning of period" or "end of period" convention). We choose to average the two earnings measures surrounding the stock price at time t, since this quantity is more reflective of the (unobserved) instantaneous earnings flow at time t.

tains noticeable persistent components that are largely orthogonal to aggregate consumption growth, thus resulting in extra risk compensation for long-run risk. Consistent with Euler-equation reasoning, a fixed cohort's consumption growth exhibits similar secular trends to the expected real interest rate.

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.

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.

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

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

Constantinides, G., J. B. Donaldson, and R. Mehra (2002). Junior can't borrow: A new perspective on the equity premium puzzle. *The Quarterly Journal of Economics 117*(1), 269–296.

Constantinides, G. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*, 219–240.

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

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

Ehling, P., A. Graniero, and C. Heyerdahl-Larsen (2018). 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.

Gârleanu, N., L. Kogan, and S. Panageas (2012). The demographics of innovation 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.

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 (2005). Optimal life-cycle asset allocation: Understanding the empirical evidence. *Journal of Finance 60*(2), 869–904.

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

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.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–54.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303 – 316.

Karlin, S. and H. Taylor (1981). *A Second Course in Stochastic Processes*. Academic Press.

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

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

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jørgensen (2009). Long-run stockholder consumption risk and asset returns. *The Journal of Finance 64*(6), 2427–2479.

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

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

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of Monetary Economics 15*(2), 145–161.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The short- and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Oyer, P. (2006). Initial labor market conditions and long-term outcomes for economists. *Journal of Economic Perspectives 20*(3), 143–160.

Oyer, P. (2008). The making of an investment banker: Stock market shocks, career choice, and lifetime income. *The Journal of Finance 63*(6), 2601–2628.

Panageas, S. (2020). The implications of heterogeneity and inequality for asset pricing. Foundations and Trends in Finance, forthcoming.

Piazzesi, M. and M. Schneider (2009). Inflation and the price of real assets. Federal Reserve Bank of Minneapolis, Staff Report: 423.

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

Toda, A. A. and K. Walsh (2015). The Double Power Law in Consumption and Implications for Testing Euler Equations. *Journal of Political Economy 123*(5), 1177–1200.

von Wachter, T. and H. Schwandt (2018). Unlucky cohorts: Earnings, income, and mortality effects from entering the labor market in a recession. forthcoming, Journal of Labor Economics.

Weil, P. (1989). Overlapping families of infinitely-lived agents. *Journal of Public Economics 38*(2), 183–198.

Online Appendix

A Proofs and additional theoretical results

Proof of Lemma 1. The absence of bubbles together with the assumption of a unit elasticity of substitution implies that aggregate consumption is given by $C_t = \beta\left(\bar{W}_t + \bar{H}_t\right)$, where

$$\bar{W}_t = \int_{-\infty}^{t} q_{t,s}^{d} D_{t,s} ds = \alpha q_t^d Y_t \tag{48}$$

is the present value of all dividends to be paid by existing firms and

$$\bar{H}_t = \int_{-\infty}^{t} q_{t,s}^{l} l_{t,s} w_{t,s} ds = (1-\alpha)\, q_t^l Y_t. \tag{49}$$

is the total value of the human capital of existing agents. Combining goods market clearing ($C_t = Y_t$) with (48) and (49) and re-arranging leads to (10). ■

Proof of Lemma 2. The present value of all newly-born workers' wages is given by $(1-\alpha)\,\eta_t^l q_t^l Y_t$, while the present value of all newly created firms is $\alpha \eta_t^d q_t^d Y_t$. The sum of these quantities gives the total wealth of newly born agents. Given that the consumption-to-wealth ratio for investors is β, the per-capita consumption of the newly born, as a proportion of total consumption, is given by (15). ■

Proof of Lemma 3. The only step of the proof not made completely explicit in the proof is the one yielding equation (13). To show this relation, time-differentiate aggregate consumption $C_t = \lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} c_{t,s} ds$ to get

$$\dot{C}_t = -\lambda C_t + \lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} \dot{c}_{t,s} ds + \lambda c_{t,t} = -\lambda C_t + \frac{\dot{c}_{t,s}}{c_{t,s}} \left(\lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} c_{t,s} ds\right) + \lambda c_{t,t}, \tag{50}$$

where we used Leibniz's rule and the fact that $\frac{\dot{c}_{t,s}}{c_{t,s}}$ is independent of s. Dividing both sides of (50) by C_t, using $C_t = \lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} c_{t,s} ds = Y_t$ and $\frac{\dot{C}_t}{C_t} = g$ leads to (13). ■

The following technical restrictions on f and σ ensure existence of a stationary q.

Assumption 1 *The functions f and σ are Lipschitz continuous on the bounded interval $[q^{min}, q^{max}] \subset (0, \frac{1}{\alpha\beta})$. Moreover, f is twice differentiable, monotonically decreasing, and satisfies $f\left(q^{min}\right) > 0$ and $f\left(q^{max}\right) < 0$. Finally, $\sigma(q) \geq 0$, $\sigma\left(q^{min}\right) = \sigma\left(q^{max}\right) = 0$, and*

$$\lim_{q \to q^{max}} \frac{\sigma^2(q)}{q^{max} - q} < 2\left|f\left(q^{max}\right)\right|, \quad \lim_{q \to q^{min}} \frac{\sigma^2(q)}{q - q^{min}} < 2\left|f\left(q^{min}\right)\right|. \tag{51}$$

Proof of Proposition 1. Let $\hat{q}_t$ be a solution to the stochastic differential equation (20) with support in $[q^{\min}, q^{\max}]$. By construction, the process $\varphi(\hat{q}_t)$ solves the SDE (22). Since $\hat{q}_t$ is bounded, so is φ_t, and we can construct two positive processes η_t^l and η_t^d such that $\eta_t^d - \eta_t^l = \varphi_t$.

Taking these two processes as given, posit that the equilibrium price dividend ratio is $q_t^d = q_t = \hat{q}_t$ and q_t^l is given by Lemma 1. Further, conjecture that the interest rate r_t is given by equation (18). We next confirm that these postulates for q_t^d, q_t^l and r_t constitute an equilibrium.

Given the dynamics of q_t, the definition of $\varphi(\cdot)$, and the definition of r_t, pricing equation (11) is satisfied. Further, using also the definition of q_t^l, which implies $\alpha dq_t^d + (1-\alpha)dq_t^l = 0$, we obtain the analogous pricing equation for q_t^l:

$$\mathrm{E}[dq_t^l] = \left(r_t - g + \eta_t^l\right) q_t^l - 1. \tag{52}$$

Agents' consumption optimality require $c_{t,t} = \beta W_{t,t}$, yielding Lemma 2, as well as the Euler equation (12). Starting with equation (12), then applying Lemma 2 and equation (18) in succession, we obtain

$$\begin{aligned} C_t &= \int_{-\infty}^{t} \lambda e^{-\lambda(t-s)} e^{\int_s^t (r_u - \rho) du} c_{s,s} ds && (53)\\ &= \int_{-\infty}^{t} e^{\int_s^t (r_u - \beta) du} \left(\eta_s^l + \alpha\beta q_s \varphi_s\right) Y_s ds && (54)\\ &= \int_{-\infty}^{t} e^{\int_s^t (r_u - \beta - g) du} \left(\beta + g - r_s\right) Y_t ds && (55)\\ &= Y_t, && (56) \end{aligned}$$

given that r_t is bounded above away from $\beta + g$. Proposed consumption processes are therefore optimal and clear the consumption market, given the interest rate.

Finally, with q_t^d and q_t^l the valuation ratios, the total wealth in the economy is

$$\frac{1}{\beta} C_t = \frac{1}{\beta} Y_t = \alpha q_t^d Y_t + (1-\alpha) q_t^l Y_t. \tag{57}$$

To see that asset markets clear, note that integrating forward the budget constraint (6) of all agents born at time s and alive at time t and taking expectations gives

$$\lambda e^{-\lambda(t-s)} W_{t,s} = \lambda e^{-\lambda(t-s)} E_t \left[\int_t^\infty \frac{m_u}{m_t} e^{-\lambda(u-t)} (c_{u,s} - (1-\epsilon) w_{u,s}) du\right],$$

where m_t is given by (7). Aggregating across all cohorts, and using the same arguments as in Lemma 1 shows that

$$\lambda \int_{-\infty}^{t} e^{-\lambda(t-s)} W_{t,s} = \int_{-\infty}^{t} P_{t,s} ds,$$

i.e., the stock market is clearing. Consumption market clearing and stock market clearing implies bond market clearing by Walras' law. Uniqueness of the process φ_t is a direct consequence of the analysis in the text, in particular equation (19).

We end the proof with a technical detail — a sketch of an argument that shows that q_t is stationary. We make use of results in Karlin and Taylor (1981). Specifically, we start by defining

$$s(q) \equiv \exp\left\{-\int^q \frac{2f(\xi)}{\sigma^2(\xi)} d\xi\right\},$$

noting that by assumption (51) there exists $\bar{v} > 1$ such that, for ε small enough and $q \in (q^{\max} - \varepsilon, q^{\max})$ we have

$$\frac{s(q)}{s(q^{\max} - \varepsilon)} = \exp\left\{ - \int_{q^{\max} - \varepsilon}^{q} \frac{2f(\xi)}{\sigma^2(\xi)} d\xi \right\} < \exp\left\{ - \int_{q^{\max} - \varepsilon}^{q} \frac{\bar{v}}{q^{\max} - \xi} d\xi \right\} = \left(\frac{q}{q^{\max} - \varepsilon} \right)^{-\bar{v}}.$$

Hence, for q "close" to $q^{\max}$ the function $s(q)$ (and accordingly the speed measure $S(q) = \int^{q} s(\eta)\, d\eta$) behaves as in Example 5 on page 221 in Karlin and Taylor (1981). (A similar argument applies to the boundary $q = q^{\min}$.) It then follows that the boundaries $q^{\min}$ and $q^{\max}$ are entrance boundaries whenever condition (51) holds and a stationary distribution exists. ■

Proof of Lemma 5. The fact that m_t is a (spanned) stochastic discount factor (SDF) means

$$d\log(m_t) = -r_t dt - \frac{\kappa_t^2}{2} dt - \kappa_t dB_t, \tag{58}$$

where κ_t is the market price of risk (the maximal Sharpe ratio). In the special case when preferences are specified by (26), and given the existence of annuities, optimality implies that the process $\log(m_t)$ satisfies[41]

$$d\log(m_t) = \beta\left(\gamma \log(c_t) - \log(\gamma V_t)\right) dt - \rho dt + d\log(\gamma V_t) - d\log(c_t). \tag{59}$$

An agent's value function V is homogeneous of degree γ in the her total wealth $\hat{W}_t$, which is the sum of her financial wealth and the present value of her future earnings. We consequently write

$$V_t(\hat{W}_t) = \frac{\hat{W}_t^{\gamma}}{\gamma} e^{\tilde{Z}_t} \tag{60}$$

for an appropriate process $\tilde{Z}_t$. Furthermore, from the envelope condition we have

$$\frac{\gamma}{\hat{W}_t} V_t = \frac{\partial V_t}{\partial \hat{W}_t} = f_c = \frac{\beta \gamma V_t}{c_t},$$

giving $c_t = \beta \hat{W}_t$.

For any $s < t$, the definition of V_t implies

$$V_t + \int_s^t \beta \gamma V_u \left(\log(c_u) - \frac{1}{\gamma} \log(\gamma V_u) \right) du = E_t \int_s^{\infty} \beta \gamma V_u \left(\log(c_u) - \frac{1}{\gamma} \log(\gamma V_u) \right) du.$$

Since the right-hand side is a martingale, the drift of the left-hand side equals zero, implying that $dV_t + \left(\beta \gamma V_t \left(\log(c_t) - \gamma^{-1} \log(\gamma V_t) \right) \right) dt$ is a martingale increment and therefore, upon applying Ito's Lemma we obtain

$$d\log(\gamma V_t) = -\beta\gamma \left(\log(c_t) - \gamma^{-1} \log(\gamma V_t) \right) dt - \frac{1}{2} \frac{\sigma_{V,t}^2}{V_t^2} dt + \frac{\sigma_{V,t}}{V_t} dB_t, \tag{61}$$

[41]See Duffie and Epstein (1992) for details.

where $\sigma_{V,t}$ denotes the instantaneous volatility of V_t at time t. Combining (61) and (59), we obtain

$$d\log(m_t) = -\beta dt - \frac{1}{2}\frac{\sigma_{V,t}^2}{V_t^2}dt - d\log(c_t) + \frac{\sigma_{V,t}}{V_t}dB_t.$$

Comparison with (58), along with the fact that aggregate consumption growth is deterministic and as a result individual consumption growth needs to be locally deterministic implies

$$\frac{\sigma_{V,t}}{V_t} = -\kappa_t \tag{62}$$

$$\dot{c}_t = (r_t - \rho)c_t. \tag{63}$$

Consumption c_t is therefore locally deterministic, and so is $\hat{W}_t = \beta^{-1}c_t$, which, upon using equation (60), leads to

$$\frac{\sigma_{V,t}}{V_t} = \sigma_{\tilde{Z}_t} = -\kappa_t.$$

Combining $c_t = \beta\hat{W}_t$ and (60) leads to

$$d\log(\gamma V_t) = d\tilde{Z}_t + \gamma d\log c_t = -\beta\left(\gamma\log\beta - \tilde{Z}_t\right)dt - \frac{1}{2}\frac{\sigma_{V,t}^2}{V_t^2}dt + \frac{\sigma_{V,t}}{V_t}dB_t, \tag{64}$$

where the second equality follows from (61). Letting $Z_t \equiv \tilde{Z}_t - \gamma\log(\beta)$ and noting that $\frac{\sigma_{V,t}}{V_t} = \sigma_{\tilde{Z},t} = \sigma_{Z,t}$ leads to

$$dZ_t = -\gamma d\log c_t + \beta Z_t dt - \frac{1}{2}\sigma_{Z,t}^2 dt + \sigma_{Z,t}dB_t. \tag{65}$$

Integrating (65), and noting that σ_Z is bounded, gives equation (27).

■

Lemma 7 *Suppose that $\frac{d\nu}{dZ} > 0$ for all $Z \in \left[Z^{\min}, Z^{\max}\right]$ and also $\frac{\partial\varphi}{\partial q} < 0$ for any $Z \in \left[Z^{\min}, Z^{\max}\right]$ and $q \in \left[q^{\min}, q^{\max}\right]$. Then the mapping (33)–(34) is invertible.*

Proof of Lemma 7. Since the right hand side of (33) depends only on Z_t, it is immediate that strict monotonicity is equivalent to invertibility. Fixing Z_t and therefore ν_t, $\frac{\partial\varphi(q_t,Z_t)}{\partial q_t} < 0$ implies that there is a unique $q_t = \varphi^{-1}(\varphi_t, \nu_t)$. ■

Proof of Proposition 2. The proof of the proposition follows the same logic as that of Proposition 1. In the interest of completeness, we start by invoking Ito's Lemma to write down the SDE for ν:

$$\begin{aligned} d\nu_t = {} & \nu'\left(\nu^{-1}(\nu_t)\right)\left(f_Z\left(\nu^{-1}(\nu_t)\right) + \frac{\sigma_Z^2\left(\nu^{-1}(\nu_t)\right)}{2}\frac{\nu''\left(\nu^{-1}(\nu_t)\right)}{\nu'\left(\nu^{-1}(\nu_t)\right)}\right)dt \\ & + \nu'\left(\nu^{-1}(\nu_t)\right)\sigma_Z\left(\nu^{-1}(\nu_t)\right)dB_t \end{aligned} \tag{66}$$

Similiarly, one can write the dynamics of

$$\varphi_t = \varphi(q_t, Z_t) \tag{67}$$

based on the dynamics of q_t and Z_t, and then plug in $q_t = \varphi^{-1}(\varphi_t, \nu_t)$ and $Z_t = \nu^{-1}(\nu_t)$.

The existence of the inverse functions ν^{-1} and φ^{-1} is ensured by Lemma 7. To avoid repetition, we only justify two key statements in the text, namely equations (31) and (32).

As before, the definition of q_t implies that

$$m_t q_t D_{t,s} + \int_s^t m_t D_{t,s} = E_t \int_s^\infty m_u D_{u,s} du. \tag{68}$$

is a martingale. Using Ito's Lemma and $\kappa_t = -\sigma_{Z,t}$ yields equation (32).

From equations (65), (13), and the definition of ν_t, the drift of Z_t equals

$$\beta Z_t - \gamma \frac{\dot{c}_t}{c_t} - \frac{1}{2}\sigma_Z^2(Z_t) = \beta Z_t - \gamma(\lambda + \rho - \nu_t) - \frac{1}{2}\sigma_Z^2(Z_t), \tag{69}$$

which is equated to $f_Z(Z_t)$ to yield equation (31). ■

Proof of Lemma 6. The first equality follows from $\eta_t^d = \frac{D_{t,t}}{D_t^A} = \frac{q_t D_{t,t}}{q_t D_t^A} = \frac{P_{t,t}}{P_t^A}$. The second equality follows upon time-differentiating $P_t^A = \int_{-\infty}^t P_{t,s} ds$ to obtain $dP_t^A = \int_{-\infty}^t d_t P_{t,s} ds + P_{t,t}$ and then dividing by P_t^A. ■

B Model extensions and modifications

In this section of the appendix we consider several extensions of the baseline model.

B.1 Time-varying population and aggregate consumption growth

In the text we assume that both population and aggregate consumption growth per capita are both constant. Suppose instead that the population size N_t and the per capita output growth $\frac{Y_t}{N_t}$ are stochastic and given by

$$dN_t = (-\lambda + b_t) N_t dt \tag{70}$$

$$d\left(\frac{Y_t}{N_t}\right) = g_t \frac{Y_t}{N_t} dt, \tag{71}$$

for some birth process b_t and some per-capita output — and therefore consumption — growth process g_t. We have the following result.

Proposition 3 *Suppose that the share processes η_t^d and η_t^l support a given set of dynamics for the price-dividend ratio and the Sharpe ratio in the baseline economy, i.e., with $g_t = g$ and $b_t = \lambda$. Then, the modified share processes $\hat{\eta}_t^d = \eta_t^d - g + g_t + b_t - \lambda$ and $\hat{\eta}_t^l = \eta_t^l - g + g_t + b_t - \lambda$ support the same dynamics for the price-divided ratio, the interest rate, and the Sharpe ratio in an economy described by* (70)–(71).

Proof of Proposition 3. The adjustments to the share processes, $\eta_t^d \mapsto \hat{\eta}_t^d$ and $\eta_t^l \mapsto \hat{\eta}_t^l$, ensure that $\frac{\dot{w}_{t,s}}{w_{t,s}}$ and $\frac{\dot{D}_{t,s}}{D_{t,s}}$ remain unaffected:

$$\frac{\dot{w}_{t,s}}{w_{t,s}} = (g_t + b_t - \lambda) - \hat{\eta}_t^l + \lambda = g + \lambda - \eta_t^l \tag{72}$$

$$\frac{\dot{D}_{t,s}}{D_{t,s}} = (g_t + b_t - \lambda) - \hat{\eta}_t^d = g - \eta_t^d. \tag{73}$$

It follows that the same consumption growth $\frac{\dot{c}_{t,s}}{c_{t,s}}$ and valuation ratio q_t as before satisfy the equilibrium conditions. In other words, the real interest rate, the Sharpe ratio, and the price-dividend ratio are identical in the two economies. ■

The intuition behind Proposition 3 is straightforward. The adjustments to the share processes ensure that the entire additional growth accrues to the new cohort, leaving marginal agents' consumption growth, $\frac{\dot{c}_{t,s}}{c_{t,s}}$, the same as in the baseline model. Accordingly, the model's asset pricing implications remain unaffected.

Proposition 3 illustrates a more general point. The asset-pricing implications of the model depend only on the properties of the marginal-agent consumption growth. Modifications or extensions of the model (e.g., production, government, redistribution policies) that imply the same stochastic process for marginal-agent consumption growth as our simple fluctuating endowment-share economy will have identical asset pricing implications.

B.2 The role of multiple sources of income

Here we discuss the role of multiple sources of income. To start, it is useful to consider the limit where there are no workers ($\epsilon = 1$) and all income is capital income ($\alpha = 1$).

In that case Lemma 1 implies that the $q_t^d = \frac{1}{\beta}$. Therefore q_t^d is constant, and since the dividends of any given firm are locally deterministic, the stock market return has no diffusion component. As a result, stocks cannot command a risk premium, or else there would be an arbitrage between stocks and bonds. This result holds irrespective of whether consumers have logarithmic or recursive preferences.

This result is due to the assumption of a unitary IES. With unitary IES and a single source of income, the present value of existing agents' consumption (which is equal to $\frac{C_t}{\beta}$) must equal the value of the stock market. In turn, market clearing requires $C_t = D_t$ and therefore the value of the stock market is just $\frac{D_t}{\beta}$, or equivalently the price-dividend ratio is $\frac{1}{\beta}$.

By assuming multiple sources of income, the present value of consumption does not equal the value of the stock market. This allows us to obtain a positive equity premium despite a unitary IES. Indeed, even for values of the IES smaller than one, the equity premium would be positive (as long as preferences are non-trivially recursive).

With an IES above one, the model generates a positive equity premium even with a single source of income.[42] Besides not having to take a stance on whether the IES should be above or below one, the main reason for assuming multiple sources of income is that in the data ν_t and η_t^d are distinct processes, whereas with a single source of income they would coincide. (With a single source of income, $\beta q_t = 1$ and $\alpha = 1$; therefore (14) implies $\nu_t = \eta_t^d$.)

[42] Panageas (2020) shows this in a model with a single source of income.

B.3 Assigning new firms to existing agents

In the baseline model new firms are assigned to arriving cohorts. This is not essential for our results. In this section we modify the model to assign the shares of the new firms to existing agents. We show that an alternative specification of the process η_t^l supports the same processes ν_t and η_t^d as in the baseline model (and by implication has the same implications for asset pricing).

Specifically, rather than being allocated to arriving agents, new firms are allocated to all of the agents, as we describe below. In addition, for simplicity, suppose that we remove the distinction between entrepreneurs and workers ($\varepsilon = 0$); the only difference among members of a given cohort is the realization of their random death time.

The time-t endowment of cohort s consists of two components: the amount $w_{t,s}l_{t,s}$ in wages, and an allotment $\phi_{t,s}l_{t,s}$ of shares of the firms introduced in the market at time t. Specifically, at time t each (surviving) member of cohort s receives a fraction

$$\phi_{t,s} = \frac{\eta^\phi e^{-\int_s^t \eta^\phi du}}{l_{t,s}} = \frac{\eta^\phi}{\lambda} e^{-(\eta^\phi - \lambda)(t-s)} \tag{74}$$

of the shares of the newly arriving firms, for a constant $\eta^\phi > 0$. The sum of all shares alloted to all cohorts at time t satisfies $\int_{-\infty}^t \phi_{t,s}l_{t,s}ds = 1$. Markets remain dynamically complete, in the sense that all asset payoffs can be replicated by dynamically trading the stock and the bond.

From this point onward one can repeat the analysis of the paper. Specifically, we define the valuation ratio as $q_{t,s}^\phi$

$$q_{t,s}^\phi \equiv \frac{E_t \int_t^\infty e^{-\lambda(u-t)} \frac{m_u}{m_t} \phi_{u,s} q_u^d D_{u,u} du}{\phi_{t,s} q_t^d D_{t,t}}. \tag{75}$$

The numerator of $q_{t,s}^\phi$ is the expected present value of the shares that will be received by a member of cohort s and the denominator is the valuation of the batch of shares that the member just received. Similar to the valuation ratios $q_{t,s}^d$ and $q_{t,s}^l$, $q_{t,s}^\phi$ is independent of s.

Repeating the analysis of Section 3 for this modified setup, one can show that the equation stated in Lemma 1 now becomes

$$\frac{1}{\beta} = \alpha q_t^d \left(1 + \eta_t^d q_t^\phi\right) + (1-\alpha)\, q_t^l, \tag{76}$$

equations (12), (13), and (14) remain unchanged, and the definition of ν_t in Lemma 2 becomes

$$\nu_t = \frac{\lambda c_{t,t}}{Y_t} = \beta \left[(1-\alpha)\, \eta_t^l q_t^l + \alpha \eta_t^d q_t^d \times \eta^\phi q_t^\phi\right]. \tag{77}$$

Combining (76) with (77) gives

$$\begin{aligned} \nu_t &= \beta \left[\eta_t^l \left(\beta^{-1} - \alpha q_t^d \left(1 + \eta_t^d q_t^\phi\right)\right) + \alpha \eta_t^d q_t^d \eta^\phi q_t^\phi\right] \\ &= \eta_t^l \left(1 - \alpha\beta q_t^d\right) + \alpha\beta \left(\eta^\phi - \eta_t^l\right) q_t^\phi \eta_t^d q_t^d. \end{aligned} \tag{78}$$

Now fix a specific equilibrium set of processes ν_t^*, $\eta_t^{d^*}$, $q_t^{d^*}$, and m_t^* that correspond (respectively) to the consumption share of arriving cohorts, the market value share of arriving firms, the price-dividend ratio, and the stochastic discount factor in the baseline version of the model. Setting $\nu_t = \nu_t^*$, $\eta_t^d = \eta_t^{d^*}$, and $q_t = q_t^{d^*}$, letting $q_t^{\phi^*}$ be the corresponding process defined by (75), and solving equation (78) for η_t^l gives

$$\eta_t^l = \frac{\nu_t^* - \alpha\beta\eta^\phi \eta_t^{d^*} q_t^{d^*} q_t^{\phi^*}}{1 - \alpha\beta q_t^{d^*}\left(1 + \eta_t^{d^*} q_t^{\phi^*}\right)}. \tag{79}$$

By specifying the process η_t^l according to equation (79) and setting $\eta_t^d = \eta_t^{d^*}$, the modified model and the model contained in the paper imply exactly the same processes for the consumption share of the marginal agent $c_{t,s}$, the same stochastic discount factor, price-dividend ratio, dividend growth of the market portfolio, etc.[43]

The broader point of this section (and of Section B.1) is that any model modification has the same asset-pricing implications as the baseline model as long as the consumption shares (ν_t) and the dividend shares (η_t^d) of the baseline model remain unchanged. Additionally, the empirical measurement of ν_t in the data does not depend on the assumptions that we make about the allocation of the new company shares in the population. The same is true for η_t^d, whose measurement reflects the fraction of stock market value that accrues to new firms, but does not rely on who obtains these shares.

B.4 Limited stock market participation

Here we extend the model to allow for the possibility that a fraction ϕ of workers are "hand-to-mouth" consumers. We show that under plausible assumptions, the variability of consumption growth of the consumers who do participate in financial markets is higher than implied by the baseline model. Since these are the consumers that are relevant for asset pricing, this higher volatility would translate into a higher equity premium (ceteris paribus). Appendix section D.6 provides empirical evidence on the higher variability of stock market participants' consumption growth.

We proceed by modifying the model in a straightforward manner. Specifically, we assume that a fraction ϕ of workers do not participate at all in financial markets and instead consume their labor income. We sketch how this modification changes the model. To start with, the statement of Lemma 1 becomes

$$\alpha q_t^d + (1-\alpha)(1-\phi) q_t^l = \frac{1}{\beta}\left(1 - (1-\alpha)\phi\right). \tag{80}$$

The consumption of agents who participate in markets is $C_t^p = (1 - (1-\alpha)\phi)C_t$. The share of

[43] In equation (79), $\eta_t^l > 0$ as long as η^ϕ is sufficiently large. Indeed, by the definition of ν in Lemma 2, and letting η^{l^*} the process in the model in the paper, we have $\lim_{\eta^\phi\to\infty} \eta_t^l = \frac{\nu_t^* - \alpha\beta\eta_t^{d^*} q_t^{d^*}}{1-\alpha\beta q_t^{d^*}} = \eta_t^{l^*} > 0$, because $\lim_{\eta^\phi\to\infty} \eta^\phi q_t^{\phi^*} = 1$ and $\lim_{\eta^\phi\to\infty} q_t^{\phi^*} = 0$. In words, as η^ϕ approaches infinity, effectively all shares get assigned to the entering cohort and the equilibrium approaches the baseline model.

C_t^p accruing to new participating agents is

$$\begin{aligned}\nu_t^p &= \frac{\lambda c_{t,t}^p}{C_t^p} = \frac{\beta}{1-(1-\alpha)\phi}\left(\alpha\eta_t^d q_t^d + (1-\alpha)(1-\phi)\,\eta_t^l q_t^l\right) \\ &= (1-\omega_t)\,\eta_t^l + \omega_t\eta_t^d,\end{aligned}$$

where

$$\omega_t \equiv \frac{\alpha\beta q_t^d}{1-(1-\alpha)\phi}.$$

Equation (80) implies that $\omega_t \in (0,1)$. As in the baseline model, the consumption growth of participating agents is

$$\frac{\dot{c}_{t,s}^p}{c_{t,s}^p} = g + \lambda - \nu_t^p. \tag{81}$$

In this version of the model, the consumption growth of all agents (irrespective of whether they are participating in financial markets or not) is[44]

$$\frac{\lambda\int e^{-\lambda(t-s)}\dot{c}_{t,s}ds}{C_t} = g + \lambda - \nu_t, \tag{82}$$

where

$$\nu_t = (1-\widehat{\omega}_t)\,\eta_t^l + \widehat{\omega}_t\eta_t^d \tag{83}$$
$$\widehat{\omega}_t \equiv \alpha\beta q_t^d. \tag{84}$$

Note that $\widehat{\omega}_t < \omega_t$. The consumption growth of all existing and participating agents — equation (81) — and that of all existing agents — equation (82) — are given by similar expressions. The main difference is due to the fact that ν_t places a higher weight (compared to ν_t^p) on the share process for labor $\left(\eta_t^l\right)$.

Indeed, we have the relation

$$\begin{aligned}\nu_t^p &= \eta_t^l + \omega_t\left(\eta_t^d - \eta_t^l\right) = \eta_t^l + \frac{\omega_t}{\widehat{\omega}_t}\widehat{\omega}_t\left(\eta_t^d - \eta_t^l\right) \\ &= \eta_t^l + \frac{1}{1-(1-\alpha)\phi}\left(\nu_t - \eta_t^l\right) = \frac{1}{1-(1-\alpha)\phi}\nu_t - \frac{(1-\alpha)\phi}{1-(1-\alpha)\phi}\eta_t^l.\end{aligned} \tag{85}$$

[44]This follows from the computation

$$\begin{aligned}\lambda\int_{-\infty}^{t} e^{-\lambda(t-s)}\dot{c}_{t,s}ds &= \frac{\dot{C}_t}{C_t} + \lambda - \lambda\left(\frac{c_{t,t}^p}{C_t^p}\left(\frac{C_t^p}{C_t}\right) + \frac{c_{t,t}^{np}}{C_t^{np}}\left(\frac{C_t^{np}}{C_t}\right)\right) \\ &= g + \lambda - \lambda\left(\frac{c_{t,t}^p}{C_t^p}\left(1-(1-\alpha)\phi\right) + \frac{c_{t,t}^{np}}{C_t^{np}}(1-\alpha)\phi\right) \\ &= g + \lambda - \nu_t^p\left(1-(1-\alpha)\phi\right) + \eta_t^l(1-\alpha)\phi = g + \lambda - \nu_t.\end{aligned}$$

Given equation (85), and fixing the standard deviations σ_ν and σ_{η^l} of ν_t, respectively η_t^l, the variance σ_{ν^p} of ν_t^p is minimized when ν_t^p and η_t^l are perfectly correlated. We consequently have

$$\begin{aligned}\sigma_{\nu^p} &\geq \frac{\sigma_\nu - (1-\alpha)\,\phi\sigma_{\eta^l}}{1-(1-\alpha)\,\phi} \\ &= \sigma_\nu \frac{1-(1-\alpha)\,\phi\frac{\sigma_{\eta^l}}{\sigma_\nu}}{1-(1-\alpha)\,\phi}.\end{aligned}$$

As long as $\sigma_{\eta^l} \leq \sigma_\nu$, we obtain that $\sigma_{\nu^p}^2 > \sigma_\nu^2$. In our empirical estimation we obtain that indeed $\sigma_{\eta^l} \leq \sigma_\nu$.

C Details of the computation of marginal-agent consumption growth

In this appendix, we provide more details on the construction of our empirical measures for the marginal-agent consumption growth, dL_t, and its deviation from the aggregate per capita consumption growth, $d\log N_t - dF_t$. We also discuss issues related to measurement error and discrepancies between CEX and NIPA data and perform a direct validation exercise for the part of the sample where data availability allows it.

For the construction of our measure of marginal consumption growth, we need to choose: a) the measure of aggregate consumption C_t, b) how to estimate cohort and age effects in the data, c) the demographic table $\Lambda(t,s)$, d) how to extrapolate back in time cohort effects that we can't measure, and finally e) where to set the age cutoff at which agents enter the population and become marginal.

Concerning choice a), the aggregate consumption growth $d\log C_t$, we use NIPA aggregate consumption expenditure of goods and services deflated by the respective PCE deflator (available from the Bureau of economic Analysis since 1929). To construct aggregate per capita consumption growth we subtract total population growth $d\log N_t$ (available from the Census) from aggregate consumption growth. Whether we use NIPA aggregate consumption of goods and services, total NIPA consumption expenditure, or even the aggregate expenditure as measured by the CEX makes no difference to the calculation of the discrepancy between marginal agent and aggregate per capita consumption growth, which is the measure we use for our calibration. The reason is that aggregate consumption growth cancels in the computation of the difference $d\log N_t - dF_t$. As a robustness check, we also recomputed our measure of the discrepancy between marginal agent and aggregate per capita consumption growth using the population growth of the adult rather than the general population (ages greater or equal to 20). We report the results in column A of Table 5. We note that using either the adult population growth in the US (column A) or the growth in the number of households in the US (column B of Table 5) gives similar results.[45] The measure of $d\log N_t - dF_t$ in column B is more volatile than our baseline measure, which would imply larger Sharpe ratios in our calibration. However, the reason for this extra volatility is that the measure of households in the data contains a few abrupt year-over-year changes, presumably due to data revisions. More importantly, the demographic breakdown of households by age is sparse (households are binned in 10-year age-groups), and we need more granular information on the demographic pyramid (see

[45]The time series of US households is available since 1947.

point c below).

For point b), the estimation of cohort and age effects, there are two sets of issues to address. First, whether to aggregate at the household level or the cohort level, and second, how to adjust for different cohort sizes. Concerning the first issue, we choose to use the public use micro data (PUMD) and the computer programs provided by the CEX to aggregate consumption expenditure for an entire cohort s in year t and then divide by the number of households in year t whose head of household was born in year s, using the weighting variable provided by the CEX. This gives us a set of 1,900 observation of $c_{t,s}$, which can then be regressed on time, age, and cohort dummies. The advantage of this approach is that the CEX code isolates expenditure within a given calendar year and performs a "months in scope" adjustment. Moreover, it becomes possible to compare results with the published CEX tables.

An alternative approach to the estimation of time, age, and cohort effects is to aggregate expenditure at the level of a household (across all four interviews) and then regress that quantity on the appropriate dummies. The main disadvantage of this method is that, since households start and end their interviews at different times, part of a household's consumption will span two different years. As a robustness check, we used this alternative approach to the estimation of time, age, and cohort effects using the convention that year t corresponds to the calendar year of the earliest interview for each household. Using this convention, we reestimated the cohort and age effects, and recomputed our measure of the difference between marginal and aggregate per capita consumption growth. The correlation between our baseline measure and this alternative measure is reported in column C of Table 5 and is quite high (0.94).

Our results also don't depend on how we adjust for household size. Robustness check D recomputes our baseline measure by estimating time, age, and cohort effects but using only households with at least two members (to make sure that trends in single-member households don't affect our results). Robustness check E recomputes our baseline measure by aggregating consumption at the household level (as outlined in the above paragraph) and then including a control for log family size along with the dummies. The correlations with the baseline measure are above 0.9 in both cases.

For the time-t population born in year s — point c) — we use the demographic tables available from the Census, which go back to 1910 at annual frequency. Older ages are binned together past a cutoff for some census tables (typically past age 90). In these cases, we used the age-appropriate survival rates of the census closest to the calendar date (survival rates are available decennially) to split up the binned population, so as to make sure that for all years we have the population going to age 100. This procedure is inconsequential; simply truncating the sum F_t at age 90 yields (essentially) identical results.

Ideally, it would be best to use the age distribution of heads of households. Unfortunately, this time series is available only for ten-year binned age groups (25–34, 35–44, etc.) and only post 1960. We were able to confirm, however, that, the age distribution of heads-of-household largely mirrors the age distribution of adult persons when we bin the population of adult persons by ten-year age-groups and compare to the respective age distribution of heads of households (for the years where this information is available).

Concerning extending the estimates backwards — point d) — we start by noting that in our baseline construction we only have information on cohorts no older than 1900 only. For earlier cohorts, we extrapolate the cohort estimates A_s linearly.[46] We perform twoz robustness checks. In

[46]Specifically, for $s < 1900$, we set $A_s = A_{1900} + \chi(s - 1900)$ and used χ to correspond to the average value of ΔA_s for $s \in [1900, 1930]$.

	Baseline	A	B	C	D	E	F	G	H
Correlation (1930-2016)	1.00	0.75	0.79	0.94	0.94	0.95	0.90	0.99	0.96
Correlation (1960-2016)	1.00	0.91	0.69	0.97	0.97	0.97	0.99	1.00	0.96
Standard Deviation (1930-2016)	0.64	0.76	1.11	0.68	0.68	0.65	0.56	0.70	0.62
Standard Deviation (1960-2016)	0.54	0.78	1.03	0.60	0.61	0.58	0.56	0.61	0.57

Table 5: The first two rows display the correlations between the measures of the discrepancy between marginal agent growth and per-capita consumption growth $d\log N_t - dF_t$ computed as in the baseline specification, respectively according to alternative specifications. The last two rows display the standard deviations of these measures, expressed in percent per annum. Each of the columns A through H records results for variants that are explained in the text.

column F of Table 5, we use constant instead of linear extrapolation (i.e., we set $A_s = A_{1900}$ for $s < 1900$). In column G we recompute our measure introducing an upper cutoff age of 75 when calculating the moving sum in F_t. Any of these modifications makes little difference to our baseline measure for the entire sample. In the post-1960 subsample, the three measures produce practically identical results.

Before explaining how we choose the age cutoff for entering the population, we note that this cutoff age is immaterial as long as the same cutoff is used in the computation of the integrals in equations (41) and (42), which give C_t and F_t, respectively. Figure 7 illustrates this point in the data: Starting with 1984, the CEX tables allow us to compute aggregate consumption for age groups above given cutoffs. The different lines in the figure depict the right-hand side of equation (43) with different minimal age cutoffs (i.e., summing across age groups above 25, 35, etc., in the computation of both C_t and F_t). Since we are interested in low frequency co-movements between the series, we illustrate the co-movements of 5-year moving averages, which are essentially the same no matter which cutoff is used.

For our baseline results, we chose the minimal age cutoff to be 45. There are three reasons for this: First, for our purposes "birth" refers to the age at which an agent joins the financial market and her Euler equation starts holding. Accordingly, we want to ensure that agents have reached an age where borrowing constraints, which may invalidate equation (40) for younger cohorts, are likely to be irrelevant. We therefore aim for a cutoff where the consumption-age profile starts reaching its peak. Second, immigration (which is prevalent in younger cohorts and acts similarly to birth in our framework) is unlikely to be important past age 45. Third and most importantly, the aggregate consumption of people in the included age groups should be a relatively stable fraction of aggregate NIPA consumption. Indeed, Figure 8 shows that a cut-off of 45 meets this goal. (The ratio for the 1972 cross section is essentially the same as for the 1984 cross section and all the subsequent cross sections). We note that such stationarity does not characterize the ratio of aggregate consumption in the CEX data to NIPA aggregate consumption, which is about 0.8 in the 1972 cross section and in the 1984 cross section, then exhibits an almost linear decline to around 0.6 until 2003 and fluctuates around that level thereafter.

The fact that the ratio of aggregate consumption in the CEX to aggregate consumption in NIPA is below one is not a particular concern for our purposes.[47] If the two series were in a (con-

[47] This discrepancy is most likely due to different definitions, weighting schemes, and under-reporting of certain consumption categories in the CEX.

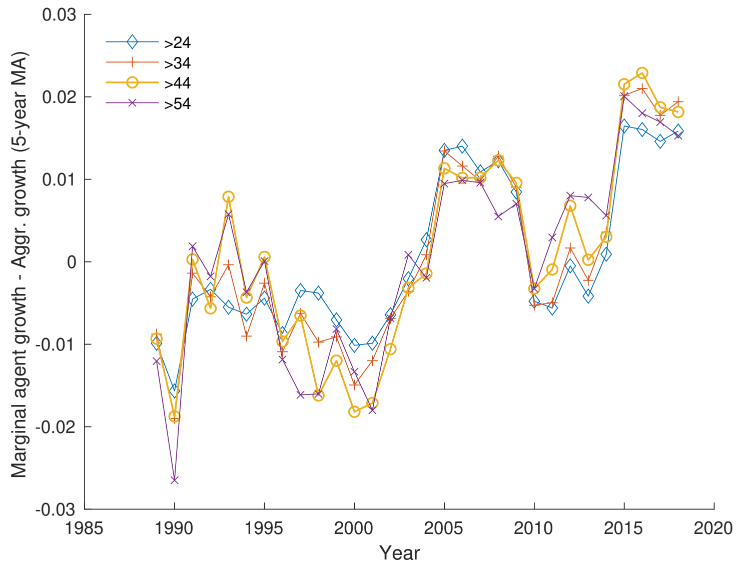

Figure 7: Difference between the consumption growth rate of the marginal agent and aggregate consumption growth using different minimal age cutoffs. All series have been de-meaned. We depict five-year moving averages to illustrate the low-frequency movements of the series, which is the focus of our analysis.

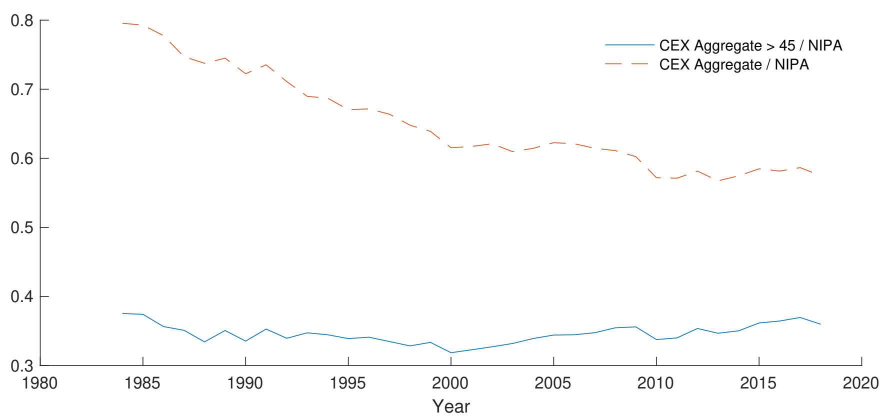

Figure 8: Ratio of aggregate consumption accruing to households 45 years or older according to the CEX divided by NIPA aggregate consumption expenditure. We also report the ratio of aggregate consumption according to the CEX divided by NIPA aggregate consumption expenditure.

stant) proportion to each other, this would not impact the computation of $d \log C_t$. What is more disconcerting is the period between 1984 and 2003, when aggregate consumption according to CEX exhibits a downward trend relative to its NIPA counterpart. While the aggregate consumption of people above our age cutoff has remained a roughly constant fraction of aggregate NIPA consumption, as a robustness check in column H of Table 5, we rerun our results using only the CEX cross sections after 2003 for the estimation of age and cohort effects that we use in the computation of

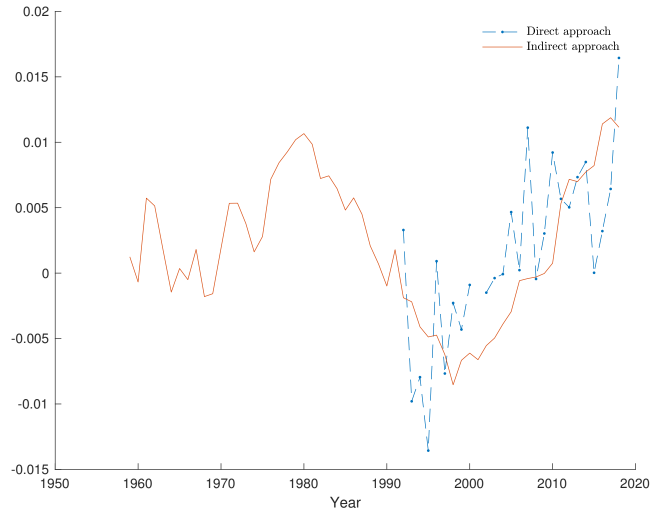

Figure 9: Comparison of $d \log N_t - dF_t$ (indirect approach) with $dL_t - (d \log C_t - d \log N_t)$ using regression estimates for dL_t (direct approach). The year 2001 is an outlier for the direct approach (the value is –0.034), due to a change in the CEX-PUMD data coverage. For this reason we omit the value for 2001.

F_t. For these cross sections, the ratio between CEX and NIPA aggregates has remained relatively stable. As column H of Table 5 shows, our results are unchanged.

In the context of measurement error, we would also like to note that the left plots of Figures 2 and 10 imply that the same patterns describe both the consumption and income of different cohorts. Income data are based on substantially more observations and are not subject to the same type of measurement errors as consumption surveys. Hence, whatever the source of measurement error in the CEX, it is unlikely to be impacting the measurement of age and cohort effects, which enter the computation of F_t.

We conclude this section by comparing our indirect approach to the direct approach of estimating dL_t with a regression. Figure 9 plots our indirect measure of $d \log N_t - dF_t$ and $dL_t - (d \log C_t - d \log N_t)$, where dL_t is the first difference of the regression estimates in equation (40). This comparison can only be performed for the last three decades.[48] The figure shows two patterns. First, the low-frequency movements of the two series appear identical, so both approaches imply the same persistence of the consumption growth of the marginal agent. Second, the direct approach exhibits more high frequency variation. We note that even though the two approaches would coincide in an infinite sample, due to sampling error they differ in finite samples.

[48]The variables required to perform the months-in-scope adjustments that we use to isolate calendar-year consumption in the CEX are available only in the post 1990 CEX crosssections.

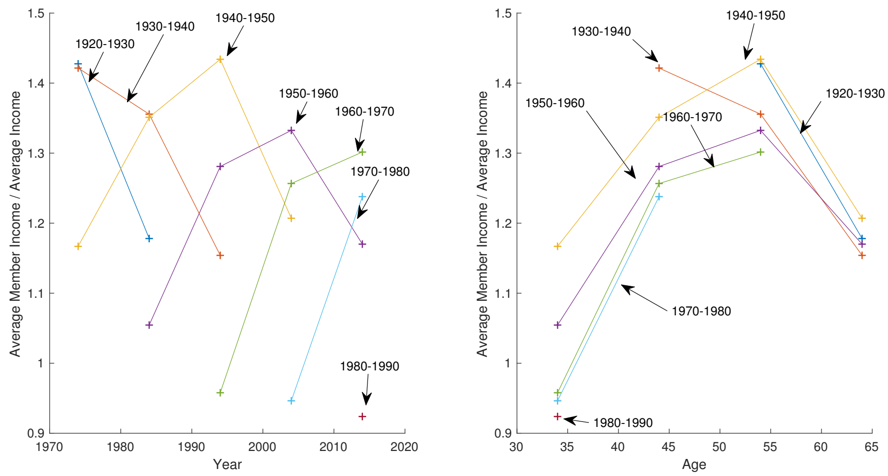

Figure 10: Average per-person income of cohorts born in different decades divided by average income at the time of observation. The left plot depicts the information as a function of time, the right plot as a function of age. For cohorts born during the decades 1920–1930, 1930–1940, and 1940–1950, we cannot provide a direct counterpart for the last observation in Figure 2, due to reporting differences in the CPS and CEX tables.

D Additional empirical and calibration results

D.1 Income cohorts

Figure 10 provides the counterpart of Figure 2 in the text, but for income rather than consumption. The figure is based on tables P9 and P10 of the CPS. The unit is a person rather than a household.

D.2 Joint time-series tests

Table 6 provides a formal econometric framework to model the joint time-series properties of i) aggregate consumption growth per capita and ii) the difference between marginal and per capita consumption growth rate as a bivariate first order vector autoregression

$$\begin{bmatrix} \Delta \log C_t - \Delta \log N_t \\ \Delta \log N_t - \Delta F_t \end{bmatrix} = B \begin{bmatrix} \Delta \log C_{t-1} - \Delta \log N_{t-1} \\ \Delta \log N_{t-1} - \Delta F_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}. \tag{86}$$

Using the estimates for B in the above equation and the covariance matrix of the residuals Σ thus obtained, we compute the long-run covariance matrix of the two time series,

$$\Omega \equiv \left[I + B + B^2 + ...\right] \Sigma \left[I + B + B^2 + ...\right]' = \frac{1}{100} \times \begin{bmatrix} 0.0444 & -0.0097 \\ -0.0097 & 0.1882 \end{bmatrix}. \tag{87}$$

The two time series exhibit low correlation at frequency zero (the correlation implied by Ω is around -0.11). Hence, the redistributional risk that arises from imperfect risk sharing is a separate source of long-run consumption uncertainty, and fluctuations in aggregate consumption growth per capita don't seem to offset them. This small correlation between the two series at low frequencies

	$\Delta \log C_t - \Delta \log N_t$	$\Delta \log N_t - \Delta F_t$
Lag $\Delta \log C_t^A - \Delta \log N_t$	0.4936	0.0255
	(0.1178)	(0.2752)
Lag $\Delta \log N_t - \Delta F_t$	-0.0037	0.9560
	(0.0207)	(0.0483)
σ_ϵ	0.0108	0.0019

Table 6: Bivariate vector autoregression of i) per capita consumption growth ($\Delta \log C_t - \Delta \log N_t$) and ii) the difference between marginal and per capita consumption growth ($\Delta \log N_t - \Delta F_t$).

is one of the reasons why we chose to abstract from fluctuations in aggregate consumption growth in our model, since they appear to be largely orthogonal to the consumption share variations that we wish to highlight.

An additional implication of this low correlation is that the the marginal-agent consumption growth, which is the sum of $\Delta \log C_t - \Delta \log N_t$ and $\Delta \log N_t - \Delta F_t$), has about twice as high a long-run standard deviation (the sum of all elements of Ω) as the long-run standard deviation of aggregate consumption growth rate per capita (the top left element of Ω).

D.3 Real interest rate, marginal-agent consumption growth, and aggregate consumption growth rate per capita

Table 7 reports results from estimating the moment condition (44) using the generalized method of moments of Hansen (1982), lagged values of r_t, $\Delta c_t^{\text{agg pc}}$, and Δc_t^{marg} as instruments, and a Newey and West (1987) HAC weighting matrix with three lags.[49] Columns (1) and (2) are full-sample "univariate" versions, where we include only $\Delta c_t^{\text{agg pc}}$ (resp. Δc_t^{marg}) in the regression. Column (3) includes both $\Delta c_t^{\text{agg pc}}$ and Δc_t^{marg}. Columns (4)–(6) repeat the same estimations, but this time excluding the Great Financial Crisis (GFC) and the period of (essentially) zero nominal rates that followed. Columns (7)–(9) report results for other subsamples with various arbitrary cutoffs prior to 2010. The results for these subsamples (columns 7–9) resemble more the results of column (6) rather than the results of the full-sample column (3), which suggests that the period 2008–2016 are unusual and exert a strong influence on the estimation results that one would obtain from other subsamples. Standard errors are in parentheses. The p-value of the test of over-identifying restrictions is reported in the last row.

D.4 Alternative risk aversion coefficients

In the paper, the parameter γ is not calibrated. We simply set it so that relative risk aversion is nine. This is motivated by the Mehra and Prescott (1985) assertion that values below ten are "reasonable." It is useful to consider what happens when we lower that number, which we report in Table 8. To consider the impact of risk aversion while keeping the rest of the model unchanged, we perform two exercises with results reported in the top and bottom panel of Table 8. In the first exercise we change b_2 along with γ so that that the ratio $\frac{b_2}{\gamma}$ preserves its value. An implication

[49] Newey and West (1987) require that the lag length parameter grow slower than $T^{\frac{1}{4}}$, so a common choice in the literature is to simply set the lag length to $T^{\frac{1}{4}}$.

	Full Sample			Excluding 2008-2016			Pre-2010	Pre-2005	Pre-2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\beta_{\text{agg. p .c.}}$	3.662**		2.919	1.415		-0.188	-1.656	-1.804	-1.751
	(1.421)		(2.252)	(1.228)		(2.073)	(1.557)	(1.377)	(1.163)
$\beta_{\text{marg.}}$		2.161**	0.866		1.821*	1.883*	2.215**	2.197**	1.925*
		(0.781)	(1.592)		(0.776)	(0.915)	(0.811)	(0.840)	(0.832)
N	65	65	65	56	56	56	56	51	46
J_p	0.654	0.284	0.505	0.0986	0.592	0.298	0.286	0.305	0.521

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: GMM estimation of Equation (44) for various subsamples of the data. Standard errors are based on a Newey-West (Bartlett kernel) weight matrix with three lags. The variable J_p is the p-value of Hansen's overidentification (J) test.

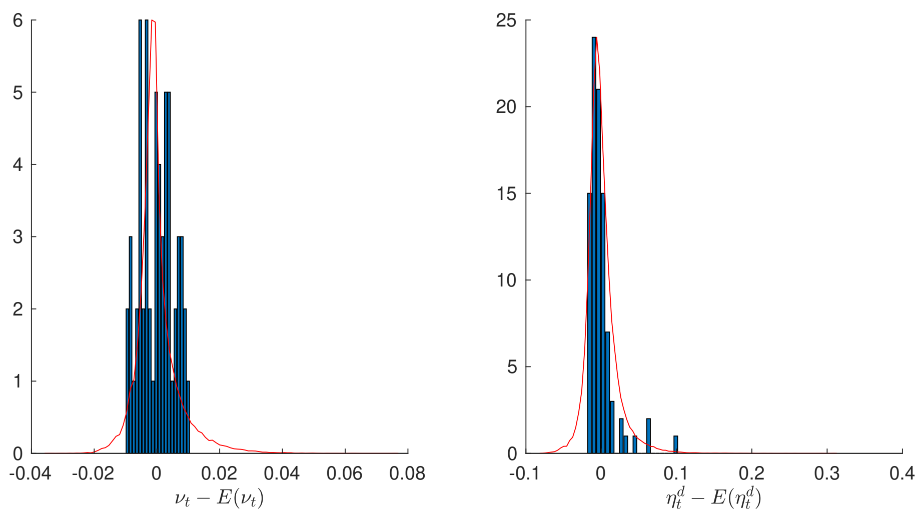

Figure 11: Left plot: Histogram of (de-meaned) inferred values of ν_t and kernel density of the respective quantity inside the model. Right plot: Histogram of the market capitalization of new index additions over the existing market capitalization of the index (de-meaned) and kernel density of the respective quantity inside the model. To obtain the model-implied quantities, we simulate 1000 paths of length identical to the length of the data sample, and de-mean the simulated data separately on each sample path, to account for sampling error in the computation of the means. We then compare the distribution of the de-meaned simulated data to the de-meaned empirical data.

of equation (46) is that, since $\frac{b_2}{\gamma}$ remains unchanged, the consumption-share process ν_t remains approximately unchanged.[50]

Comparing results across the various columns of the top panel of Table 8, we see that the Sharpe ratio and the equity premium change in proportion to the risk aversion, which is hardly

[50]To be precise, lowering b_2 to keep the ratio $\frac{\gamma}{b_2}$ constant has the effect of slightly increasing the average value of ν_t (by a few basis points) due to the third term inside square brackets in (46).

	Data	Risk aversion $\lvert\gamma\rvert + 1$		
		5	7	9
Only γ and b_2 recalibrated				
Sharpe ratio	0.29	0.14	0.20	0.27
Stock market volatility	18.2%	14.19%	14.19%	14.19%
Equity premium	5.2%	1.93%	2.90%	3.86%
Average interest rate	2.8%	0.82%	1.27%	1.72%
Standard deviation of real interest rate	0.92%	1.23%	1.03%	0.85%
Average (log) PE ratio	2.65	2.77	2.77	2.77
Standard deviation of (log) PE ratio	0.34	0.21	0.21	0.21
Autocorrelation of (log) PE ratio	0.86	0.91	0.91	0.91
Parameters a_1 and a_2 also recalibrated				
Sharpe ratio	0.29	0.14	0.20	0.27
Stock market volatility	18.2%	12.42%	13.75%	14.19%
Equity premium	5.2%	1.69%	2.81%	3.86%
Average interest rate	2.8%	0.82%	1.27%	1.72%
Standard deviation of real interest rate	0.92%	1.23%	1.03%	0.85%
Average (log) PE ratio	2.65	3.12	2.88	2.77
Standard deviation of (log) PE ratio	0.34	0.19	0.21	0.21
Autocorrelation of (log) PE ratio	0.86	0.91	0.91	0.91

Table 8: Unconditional moments for the data and the model for various levels of the risk aversion. The headers to the last three columns represent the effective risk aversion coefficient, given by $\lvert\gamma\rvert+1$. As we lower risk aversion, we also lower the value of $\lvert b_2\rvert$ so as to keep the ratio $\frac{b_2}{\gamma}$ unchanged.

surprising. To keep this exercise transparent we choose to only modify γ and b_2 as we move across columns, so as to keep ν_t approximately unchanged. Of course, the change in b_2 has an indirect impact on η_t^d as well, since b_2 affects the Sharpe ratio, κ_t, which appears on the right-hand side of (47). To address this issue, the bottom panel of the Table adjusts the parameters a_1 and a_2 as we change b_2 and γ, so that the moments of both ν_t and η_t^d remain approximately the same across the columns.[51] Once again, as one might expect, the main quantity that is affected across the various columns is the Sharpe ratio and by implication the equity premium.

D.5 Price-earnings ratio: data and model

The left plot of Figure 12 depicts $\log(q_t)$ in the model and the data (correlation 0.29). In our model there is no distinction between earnings and dividends, and hence between the price-earnings and the price-dividend ratio. We choose to focus on the price-earnings ratio, to mitigate the impact of the growing popularity of share repurchases on the price-dividend ratio over this period.

[51] Specifically, the parameters for column "5" include $a_1 = 1.2$ and $a_2 = 2.8$, while for column "7" $a_1 = 0.7$ and $a_2 = 3.1$.

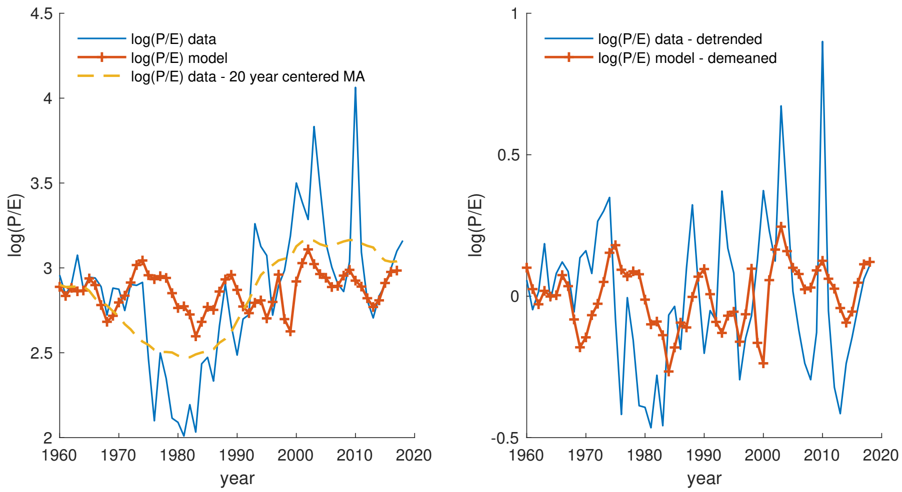

Figure 12: Left plot: Historical (log) price-to-earnings ratio and data-implied (log) price-to-earnings ratio. The line labeled "20-year centered moving average" corresponds to a two-sided (centered) 20-year moving average of the log price-to-earnings ratio in the data. Right plot: The line "log(P/E) data – detrended" graphs the difference between the historical log price-to-earnings ratio and the two-sided (centered) 20-year moving average of the log price-to-earnings ratio in the data. The line labeled "log(P/E) model – demeaned" is the model-implied (log) price-to-earnings ratio minus its mean over the sample period.

The left plot of Figure 12 also depicts a 20-year centered moving average of the price-earnings ratio.[52] This slow-moving moving average constitutes a simple and transparent way to capture the very low frequency variation of the log price-earnings ratio, which is responsible for making this ratio appear almost like a nonstationary process, especially since 1980. (We propose one potential explanation for this near nonstationarity in the next paragraph.) The right plot of Figure 12 subtracts this moving average from the historically observed price-earnings ratio to compute a quantity more easily comparable with the price-earnings ratio in the model.

Perhaps the main reason why the price-earnings ratio is a challenging time series to match is that we intentionally abstracted from aggregate fluctuations of any sort. For instance, the model's failure to capture the very low frequency cycle in the log price-earnings ratio may be due to the rise in corporate profits as a fraction of GDP. To illustrate this, Figure 13 shows the log price-earnings ratio plotted against the ratio of corporate profits to GDP four years ahead.[53] The figure suggests that one of the reasons for the secular increase in the Price-to-earnings ratio may be that market participants anticipated a rise in the aggregate profit share, which the model abstracts from. The same can be said for all other aggregate (cyclical or secular) shocks that may have impacted the price-earnings ratio in the data and are absent from the model.

In summary, the right way to interpret Figure 12 is as an illustration of how much of the variation of the price-earnings ratio could be attributed purely to the two redistributional shocks

[52]To compute this moving average, we use the function "movmean" in Matlab, which — close to both ends of the sample — uses as many lead and lag observations as are available to compute the moving average.

[53]Because the price-earnings ratio is a forward looking variable, it makes sense that market participants may have anticipated some of the future evolution of corporate profits.

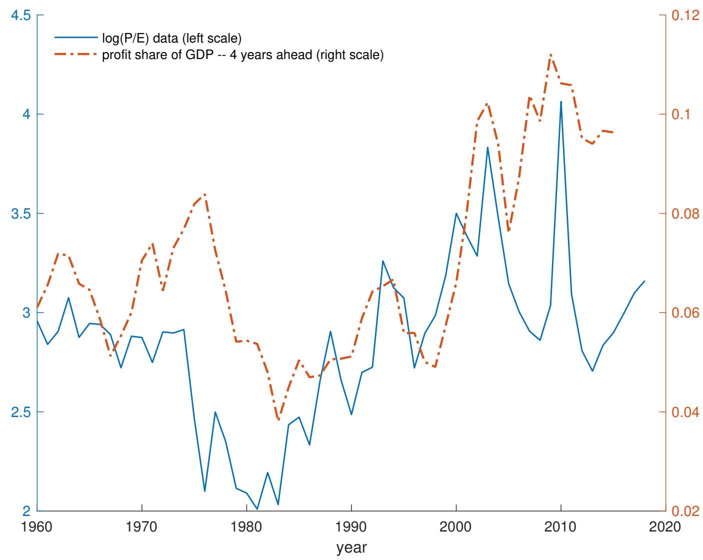

Figure 13: Left Plot: Historical (log) Price-to-earnings ratio (left scale) and the four-years-ahead historical ratio of corporate profits to GDP (right scale).

η_t^d and ν_t that enter the model.

D.6 Estimating ν_t for stock market participants

In this section we measure the discrepancy between marginal and aggregate per-capita consumption growth for stockholders. We exploit the fact that the CEX asks participants about the value of their stock holdings. This question was asked in the same way between 1990 and 2013. (From 2013 the CEX changed the way it asks questions related to stock ownership, which resulted in a change in the response rates to this question and made it hard to compare the results pre- and post-2013.) We reestimate the model, restricting attention only to the sample of people declaring positive stockholdings. While this restriction achieves the goal of focusing on households that hold stocks, it also reduces the sample since only about 5-10% of households provide any response, with the remainder leaving the answer blank. Because of this reason, we also consider a different subsample: for each year, we select the households belonging to the top 10% of the income distribution based on the variable "income before taxes" in the CEX. While this selection criterion does not get at the question of stock ownership directly, high-income people are more likely to be stock market participants.

Figure 14 shows the results of our analysis. We compare the results for the entire sample with those for the two subsamples mentioned in the above paragraph. For each of the three groups we repeat our empirical measurement of $dL_t + d\log N_t - d\log C_t$ (the direct approach, as in Figure 9).

The figure shows that whether we use the entire sample or the two subsamples (selected based on stock market participation, or on income) the low-frequency movements of the series are similar. In particular, all three series show a slow increase over this time period. The time series for stockholders is noticeably more volatile than the other two series. This is consistent with the theoretical findings of Section B.4 and the empirical literature, e.g., Mankiw and Zeldes (1991) and Malloy

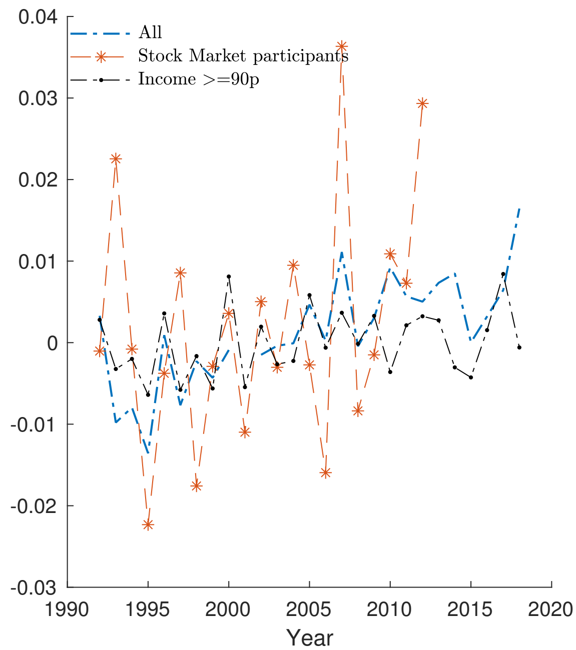

Figure 14: Inferred discrepancy between marginal-agent and aggregate consumption growth per capita for three different groups: a) Entire sample, b) Agents who report positive stock holdings, c) agents belonging to the top 10% of the income distribution. The CEX changed the stock ownership questions post 2013, so we focus on the 1990–2013 period for the group of stockholders.

et al. (2009).