A Theory of Experience Effects*

Ulrike Malmendier† — *UC Berkeley, NBER, and CEPR*
Demian Pouzo‡ — *UC Berkeley*
Victoria Vanasco§ — *Stanford University*

October 27, 2016

Abstract

Financial crises and other macroeconomic shocks appear to have long-lasting effects on investor behavior and to alter aggregate dynamics in the long run. The theoretical foundations and dynamic implications of such behavior are still debated. Recent evidence suggests that individuals overweight personal experiences of macroeconomic shocks when forming beliefs about risky outcomes and making investment decisions. We propose a simple OLG model that formalizes 'experience-based learning.' Risk averse agents can invest in risky and risk-free assets. They form beliefs about the payoff of the risky asset (1) based on data observed during their lifetimes so far and (2) exhibiting recency bias, which are the two key components of experience effects. In equilibrium, prices depend on past dividends, but only those observed by the generations that are alive, and they are more sensitive to more recent dividends. Younger generations react more strongly to recent experiences than older generations. Hence, the young have higher demand for the risky asset than the old in good times, and lower demand in bad times. The model generates predictions for stock-market dynamics and trading volume. First, a recent crisis will increase the average age of agents holding stocks, while booms have the opposite effect. Second, the stronger the disagreement across generations (e.g. after a recent shock), the higher is the trade volume. We provide stylized facts from the *Survey of Consumer Finances* consistent with the model predictions.

*We thank workshop participants at LBS, LSE, NYU, Pompeu Fabra, Stanford, UC Berkeley, as well as the ASSA, NBER EFG Behavioral Macro, SITE (Psychology and Economics segment), SFB TR 15 (Special Research Program TR15, Tutzing, Germany) conferences for helpful comments. We also thank Felix Chopra, Leslie Shen, and Jonas Sobott for excellent research assistance.

†Department of Economics and Haas School of Business, University of California, 501 Evans Hall, Berkeley, CA 94720-3880, ulrike@berkeley.edu

‡Department of Economics, University of California, 501 Evans Hall, Berkeley, CA 94720-3880, dpouzo@berkeley.edu

§Stanford University, Graduate School of Business, 655 Knight Way, Stanford, CA 94305, vvanasco@stanford.edu

1 Introduction

Economists and policy-makers alike have long wrestled with the long-lasting effects of financial crises and other macroeconomic shocks. In the case of the Great Depression, Friedman and Schwartz (1963) argue that the experience of that time created a "mood of pessimism that for a long time affected markets." In the case of the recent financial crisis, Blanchard (2012) maintains that "the crisis has left deep scars, which will affect both supply and demand for many years to come." The notion that longer-lasting crisis effects alter the dynamics of markets is consistent with growing empirical evidence on *experience effects* that suggests that personal experiences of macroeconomic shocks leave an imprint on individuals' attitudes and willingness to take risk. For example, Malmendier and Nagel (2011) show that stock-market experiences predict future willingness to invest in the stock market, and Kaustia and Knüpfer (2008) argue the same for IPO experiences.[1]

On the theoretical side, the foundations of such behavior are still debated. Existing models point to altered investment behavior during recessions, causing "hysteresis effects" (Delong and Summers (2012)), or argue that we need to revise our understanding of the stochastic processes governing the economy to explain scaled down investments after a crisis, such as the "disasterization approach" proposed by Gabaix (2011, 2012). The interpretation of Friedman and Schwartz (1963) goes in a different direction. Their notion is that the experience of an economic crisis induces pessimism and alters expectations about the future, as also pointed out by Cogley and Sargent (2008). In a similar vein, Woodford (2013) has argued that it is time to step away from the rational-expectations hypothesis, as well as from ad-hoc variations thereof, in order to understand these and other stylized facts in macro-finance. Confirming this notion, much of the evidence on experience effects pertains directly to beliefs, e.g., expectations of future stock market performance in the UBS/Gallup data (Malmendier and

[1] Evidence of experience effects is also present in non-financial settings. For example, Oreopoulos, von Wachter, and Heisz (2012) show the long-term effects of graduating in a recession on labor market outcomes, and Alesina and Fuchs-Schundeln (2007) relate the personal experience of living in (communist) Eastern Germany to political attitudes post-reunification. See also Giuliano and Spilimbergo (2013) who relate the effects of growing up in a recession to redistribution preferences.

Nagel (2011)), inflation expectations in the Michigan Survey of Consumers (Malmendier and Nagel (2016)), or expectations of future unemployment rates and the outlook for durable consumption, also in the Michigan Survey of Consumers (Malmendier and Shen (2015)).

In this paper, we propose the first formal theoretical framework that captures both of the main two empirical features of experience effects: (1) over-weighing lifetime experiences and (2) recency bias. This approach builds closely on the psychology evidence on availability bias, initiated by Tversky and Kahneman (1974), and on the extensive evidence on the different effects of description versus experience.[2] Our framework is designed to study the long-term effects of personal experiences on the cross-section of stock-market participation and of portfolio decisions, as well as on the time series of financial market aggregates, such as equilibrium prices and trade volume. As such, it generates testable predictions of trading behavior and of the cross-sectional composition of stock-market investors, which relate to long-standing empirical puzzles such as the excess volatility puzzle (LeRoy and Porter (1981), Shiller (1981), LeRoy (2005)) or the predictive power of dividend-price ratios for future stock returns (Campbell and Shiller (1988)). We take the model predictions to the data, and find evidence on the cross-section of stock market participation, the cross-section of asset holdings, and trade volume that are consistent with our model. While more evidence on the exact process of household-level learning is needed to accompany the theoretical development (see the discussions in Campbell (2008) and Agarwal, Driscoll, Gabaix, and Laibson (2013)), the formal framework allows us to explore the aggregate dynamics of an economy with experience-based learners. It aims to lay the foundation for testing whether experience-based learning can provide a new modeling framework for expectation formation that would allow to capture the above mentioned stylized facts from macro-finance.

Specifically, we develop a stylized overlapping generations (OLG) general equilibrium model in which agents form their beliefs by overweighting their own experiences. Investors have CARA preferences and live for a finite number of periods. During their lifetimes, they

[2] See, for example, Weber, Böckenholt, Hilton, and Wallace (1993), Hertwig, Barron, Weber, and Erev (2004), and Simonsohn, Karlsson, Loewenstein, and Ariely (2008).

choose portfolios of a risky and a risk-free security to maximize their per-period payoff. The risky asset is in unit net supply and pays random dividends every period. The risk-less asset is in infinitely elastic supply and pays a fixed return. Investors do not know the true mean of the distribution of dividends, but they learn about it by observing the history of realized dividends.

The novel feature of the model is that investors are *experience-based learners*. That is, they over-weigh the outcomes they have experienced in their lives when forming beliefs about the mean of dividends. Specifically, we assume that when forming their beliefs agents (i) only use data observed during their lifetimes, and (ii) may over-weigh more recent observations. These two assumptions capture, in a simplified form, the psychology evidence on availability bias as first discussed by Tversky and Kahneman (1974).

Our stylized model allows us to fully isolate the forces introduced by the presence of experience-based learners. We begin by characterizing the benchmark economy in which agents know the true mean of dividends. In this setting, the model features constant equilibrium prices and zero trade volume. Prices are constant because agents' demands and the asset supply are constant over time; trade volume is zero due to the lack of disagreements among agents. Since all agents have the same demand, their holdings of the risky asset in equilibrium are independent of the mean of dividends and the risky asset is split in equal shares among all agents in the market. Any departure from this benchmark case can thus be cleanly attributed to experience-based learning.

By introducing experience-based learning into the model, we identify long-lasting effects of economic shocks on equilibrium prices, trade volume, and the cross section of asset holdings. We emphasize two channels. The first channel is through the belief formation process: shocks to dividends shape agents' beliefs about future dividends. Each cohort uses the dividends observed during their lifetimes so far to form their beliefs, and thus the aggregate demand for the risky asset depends on the weighted sum of cohort's beliefs about the payoff of this asset. As a result, the market-clearing price is a function of the history of dividends observed by at least one market participant. This implies that agent's demands for the risky asset also depend on the history of dividends observed by (at least some) market participants, since agents care about current, and future, prices. For example, a young investor knows that the presence in the market of an older cohort that has experienced recession times will depress the price of the asset. This reflects that agents in our model "agree to disagree" about the distribution of dividends.

The second channel through which experience-based learning matters for equilibrium outcomes is the cross-sectional heterogeneity in the population. Different lifetime experiences generate persistent belief heterogeneity among cohorts. Furthermore, even if different cohorts have experienced the same history and thus hold the same belief at some point, they will react differently to the same macroeconomic shock. This is because younger cohorts will react more strongly to a common shock than older cohorts as this new experience makes up a larger part of their lifetimes so far. A positive shock will induce younger cohorts to invest relatively more in the risky asset, while a negative shock will tilt the composition towards older cohorts. In fact, we show that periods of booms, interpreted as periods with sustained increases in dividends, result in younger generations holding a larger share of the risky asset than older generations; and vice-versa.

Relative to the benchmark of agents knowing the distribution of dividends, experienced-based learning introduces excess volatility and auto-correlation of prices, as well as return predictability. The extent of these features goes above and beyond the stochastic structure of the assumed dividend process. In addition, the model also generates implications for the time series of trade volume. We show that changes in the level of disagreement between cohorts lead to higher trade volume in equilibrium. The mechanism is intuitive: an increase (decrease) in dividends induces trade since young agents become more optimistic (pessimistic) than old agents, and disagreement generates gains from trade.

The model captures an interesting tension between the role of experience-based learning (that drives belief heterogeneity across cohorts with different experiences), and recency bias (that drives beliefs toward the most recent observations, reducing heterogeneity). When the recency bias is strong, all agents pay a lot of attention to the most recent dividend realization. Thus, agents reactions to a given shock are closer together: this increases the aggregate response to a shock, but reduces heterogeneity across cohorts. As a result, price volatility increases and price auto-correlation and trade volume decrease. The opposite holds when the recency bias is weak, and agents form their beliefs using their experienced history.

We further explore the connection between demographics and the long-lasting effects of macroeconomic shocks, and study the effect of a (one time) demographic shock to the economy (e.g. a baby-boom or a war). We find that shocks to the demographic composition of markets have important implications for the response to aggregate shocks. For example, the price response to a shock is stronger when the generation of young agents is relatively large, e.g., due to a baby boom. Conversely, a given shock to dividends has a lower impact on prices if it occurs during periods where the generation of young agents is relatively small, e.g., due to a war. This highlights the interaction between aggregate shocks and the relative size of the generations that have experienced these shocks in an economy. Different lifetime experiences generate persistent belief heterogeneity among cohorts. Furthermore, even if different cohorts have experienced the same history and thus hold the same belief at some point, they will react differently to the same macroeconomic shock. This is because younger cohorts will react more strongly to a common shock than older cohorts as this new experience makes up a larger part of their lifetimes so far. A positive shock will induce younger cohorts to invest relatively more in the risky asset, while a negative shock will tilt the composition towards older cohorts. In fact, we show that periods of booms, interpreted as periods with sustained increases in dividends, result in younger generations holding a larger share of the risky asset than older generations; and vice-versa.

Relative to the benchmark of agents knowing the distribution of dividends, experienced-based learning introduces price volatility and auto-correlation, and return predictability. These features go above and beyond the stochastic structure of the assumed dividend process. The model also generates implications for the time series of trade volume. We show that changes in the level of disagreement between cohorts lead to higher trade volume in equilibrium. The mechanism is intuitive: an increase (decrease) in dividends induces trade since young agents become more optimistic (pessimistic) than old agents, and disagreement generates gains from trade.

The model captures an interesting tension between the role of experience-based learning (that drives belief heterogeneity across cohorts with different experiences), and recency bias (that drives beliefs toward the most recent observations, reducing heterogeneity). When the recency bias is strong, all agents pay a lot of attention to the most recent dividend realization. Thus, agents reactions to a given shock are closer together: this increases the aggregate response to a shock, but reduces heterogeneity across cohorts. As a result, price volatility increases and price auto-correlation and trade volume decrease. The opposite holds when the recency bias is weak, and agents form their beliefs using their experienced history.

We further explore the connection between demographics and the long-lasting effects of macroeconomic shocks, and study the effect of a (one time) demographic shock to the economy (e.g. a baby-boom or a war). We find that shocks to the demographic composition of markets have important implications for the response to aggregate shocks. For example, the price response to a shock is stronger when the generation of young agents is relatively large, e.g., due to a baby boom. Conversely, a given shock to dividends has a lower impact on prices if it occurs during periods where the generation of young agents is relatively small, e.g., due to a war. This highlights the interaction between aggregate shocks and the relative size of the generations that have experienced these shocks in an economy.

Our theoretical predictions are consistent with empirical stylized facts on portfolio decisions and trade volume. Using the representative sample of the *Survey of Consumer Finance*, CRSP, and historical data on stock-market performance, we show that cohorts differ both in their stock market participation (i.e., on the extensive margin) and in the fraction of liquid assets they invest in the stock market (i.e., on the intensive margin) in the same way they differ in their lifetime stock-market experiences. The cross-cohort differences vary over time as predicted by the time series of differences in experience. We also show that, in terms of abnormal trade volume, the de-trended turnover ratio is strongly correlated with differences in lifetime experiences of stock-market returns across cohorts.

Our theoretical predictions are consistent with empirical stylized facts on portfolio decisions and trade volume. Using the representative sample of the *Survey of Consumer Finance*, CRSP, and historical data on stock-market performance, we show that cohorts differ both in their stock market participation (i.e., on the extensive margin) and in the fraction of liquid assets they invest in the stock market (i.e., on the intensive margin) in the same way they differ in their lifetime stock-market experiences. The cross-cohort differences vary over time as predicted by the time series of differences in experience. We also show that, in terms of abnormal trade volume, the de-trended turnover ratio is strongly correlated with differences in lifetime experiences of stock-market returns across cohorts.

As final step, we investigate to what extent our results might be driven by the (stylized) myopic formulation of our model, i.e, how hedging concerns and lifetime horizon effects interact with experience-based learning. Towards that end, we extend the model of experience-based learning to a dynamic portfolio set-up where agents re-balance their portfolios every period to maximize their final period consumption.[3] Prior literature has shown that, in a rational expectations linear equilibrium, the agents' multi-period investment problem can be partitioned into a sequence of one-period ones (Vives (2010)). Under experience-based learning, however, such partitioning is no longer possible. Future beliefs and portfolio decisions of experience-based learners and, as a result, future prices depend on current dividends, making investors' wealth in the distant future correlated with next period's returns. By exploiting the CARA-Gaussian setup, however, we are able to show that the demand of experience-based learners coincides with the one in a static problem where dividends are drawn from a *modified* Gaussian distribution. That is, we can still partition the multi-period investment

[3] This form of modeling of dynamic portfolio choices is standard in the literature, see Vives (2010).

problem into a sequence of one-period problems, albeit with a probability distribution of dividends that differs from the original one. This latter result might also be of interest as an independent technical contribution in solving belief dependencies beyond the specific model proposed in this paper.

In this dynamic portfolio problem, we decompose agents' demands for the risky asset into a belief term, a hedging term, and a horizon term. The *belief term* is given by the demand of the myopic agents. Thus, all the forces that are present in our baseline model are captured by this term. The dynamics inherent to the multi-period problem generate the two additional demand motives. The *hedging term* captures that agents anticipate that they will learn about the risky asset from future dividends, and that this will in turn affect prices and future returns. In order to hedge their exposure to changes in beliefs, they distort their portfolio decisions relative to the static model. The *horizon term* captures that younger agents react less aggressively to a given change in beliefs due to their longer remaining investment horizon. In other words, their longer investment horizon makes them behave in a more risk-averse fashion. We focus on a two-period setting and show that the qualitative results presented in the baseline model, where agents maximize their per-period utility, pass through to the dynamic portfolio problem.

In summary, our paper provides a simple formalization of experience effects. It generates testable implications for individual financial decision-making and the resulting stock-market dyamics, including the long-term effects of crisis experiences. The model, together with our empirical findings, suggest that a deeper understanding of the influence of past experiences is important not only to improve the micro-modeling of financial risk-taking, but also for our understanding of the aggregate dynamics of financial markets and the long-run effects of macro-shocks.

Related Literature. The above-cited empirical literature on "experience effects" explains the long-run effects of macroeconomic shocks by showing that personal experiences of macroeconomic shocks leave a lasting imprint and significantly affect individuals' decision-making over lifetimes. Our paper provides a theoretical foundation for such behavior. Closely related to our approach, Cogley and Sargent (2008) propose a model in which the representative consumer uses Bayes' theorem to update estimates of transition probabilities as realizations accrue. As in our paper, agents use less data than in a standard framework (less "than a rational-expectations-without-learning econometrician would give them," as the authors put it). The main difference to our paper is that, in our setup, agents are not Bayesian and live for a finite number of periods. Consequently, observations during the agents' lifetime have a non-negligible effect on their beliefs. We think that this feature provides an alternative modeling device to capture Friedman and Schwartz's idea that economic events, such as the Great Depression, shape the attitude of agents towards financial markets in the future.

Our paper also relates to the work on extrapolation by Barberis, Greenwood, Jin, and Shleifer (2015) and Barberis, Greenwood, Jin, and Shleifer (2016). Barberis et al. (2015) also depart from the Bayesian paradigm by considering a consumption-based asset pricing model populated by "rational" agents and "extrapolative" agents.[4] Extrapolative agents believe that positive changes in prices will be followed by positive changes. One main difference to our paper is the approach to modeling agents' beliefs. In our model, agents hold misspecified beliefs over the expectation of future dividends, but hold correct beliefs of the mapping between equilibrium prices and dividends. As a second, and perhaps more important difference, their model of infinitely-lived agents does not allow for cross-sectional heterogeneity. In our model, however, different generations (of finitely-lived agents) assign different weights to past dividends, and such weights change as the generations age. This feature allow us to explore the impact of cohort heterogeneity and of the demographic structure of an economy on equilibrium outcomes.

More generally, our paper relates to a large literature in asset pricing that departs from the correct-beliefs paradigm. For instance, Barsky and DeLong (1993), Timmermann (1993), Timmermann (1996), and Adam, Marcet, and Nicolini (2012) study the implications of learn-

[4] A similar approach is used in Barberis, Greenwood, Jin, and Shleifer (2016) where an extrapolative model of bubbles is presented.

ing for stock-return volatility and predictability. Cecchetti, Lam, and Mark (2000) construct a Lucas asset-pricing model with infinitely-lived agents where the representative agent's subjective beliefs about endowment growth are distorted. On a similar note, Jin (2015) rationalizes financial booms and busts in a model where agents learn about the probability of a crash, but hold incorrect beliefs about the underlying process of this risk.

At the same time, our approach is different to asset pricing models with asymmetric information, surveyed in Brunnermeier (2001). A key distinction between experienced-based learning and models where agents have private information is that, in the former, information is available to all agents, while in the latter agents want to learn the information their counter parties hold. Experience-based learners choose to down-weigh the observations they have not directly observed when forming their beliefs, even though such observations are available to them (and all other agents).

Finally, there are contemporaneous papers to ours exploring the role of learning in overlapping generations models (Collin-Dufresne, Johannes, and Lochstoer (2014), Schraeder (2015)). The paper most closely related to ours is Ehling, Graniero, and Heyerdahl-Larsen (2015), who explore the role of experience in portfolio decisions and asset prices in a complete markets setting. Differently from our paper, they do not aim to actually capture "experience effects" in the sense of the empirically observed pattern in Malmendier and Nagel (2011), which involves a declining weighting function and thus recency bias. Instead, they are interested in the pure effect of individuals restricting their use of data to their lifetimes. Similar to the *Bayesian Learners from Experience* in our analysis, agents in their paper start from a given prior (the truth) which they update only using lifetime observations. The authors use this setting to develop a theoretical underpinning for trend chasing and the negative relationship between beliefs about expected returns and realized future returns, as shown by Greenwood and Shleifer (2014). Instead, our incomplete markets setting allows us to focus on the cross-section of asset holdings and the relation between trade volume and price behavior in the presence of recency bias.

There is a large literature which proposes other mechanisms, such as borrowing constraints, as the link from demographics, or life-cycle considerations, to asset prices and other equilibrium quantities. We view these other mechanisms as complementary to our paper, and are omitted for the sake of tractability of the model.

2 Model Set-Up

2.1 Lucas Tree Economy

Consider an infinite-horizon economy with overlapping generations of a continuum of risk-averse agents. At each $t \in \mathbb{Z}$, a new generation is born and lives for q periods, with $q \in \{1, 2, 3, ...\}$. Hence, there are $q+1$ generations alive at any t. The generation born at time $t = n$ is called generation n. Each generation has a mass of q^{-1} identical agents.

Agents have CARA preferences with risk aversion γ. They are born with no endowment and can transfer resources across time by investing in financial markets. Trading takes place at the beginning of each period. At the end of the last period of their lives, agents consume the wealth they have accumulated. We use n_q to indicate the last time at which generation n trades, $n_q = n + q - 1$. (If the generation is denoted by t we use t_q.) Figure 1 illustrates the timeline of this economy for two-period lived generations ($q = 2$).

There is a risk-less asset, which is in perfectly elastic supply and pays $R > 1$ at all times. There is a single risky asset (a Lucas Tree), which is in unit net supply and pays a random dividend $d_t \sim N\left(\theta, \sigma^2\right)$ at time t. To model uncertainty about fundamentals, we assume that agents do not know the true mean of dividends θ and use past observations to estimate the mean. To keep the model tractable, we assume that the variance of dividends σ^2 is known at all times.

For each generation $n \in \mathbb{Z}$ and any $t \in \{n, ..., n+q\}$, the budget constraint is given by

$$W_t^n = x_t^n p_t + a_t^n, \tag{1}$$

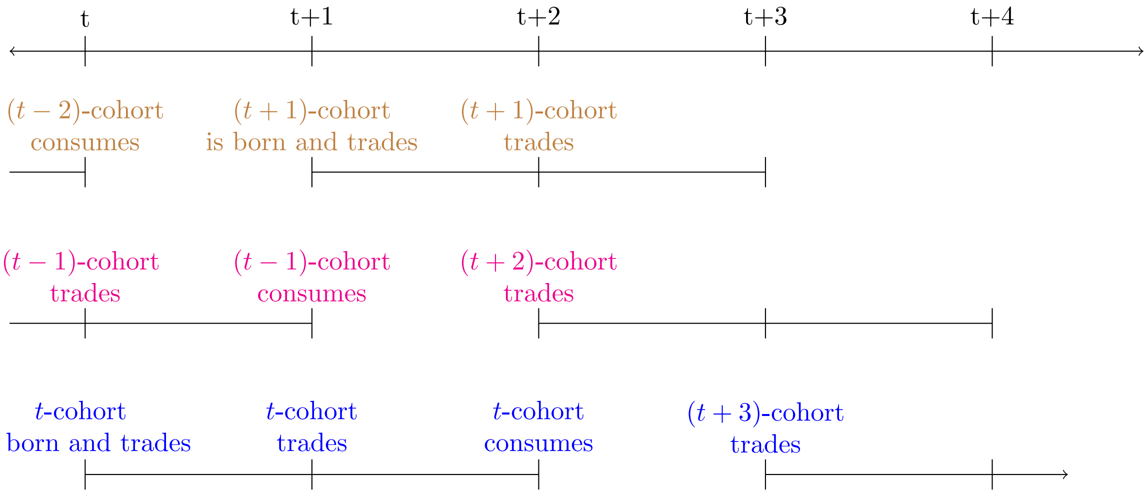


Figure 1: A timeline for an economy with two-period lived generations, $q = 2$.

where W_t^n denotes the wealth of generation n at time t, x_t^n is the investment in the risky asset (units of Lucas Tree output), a_t^n is the amount invested in the riskless asset, and p_t is the price of one unit of the risky asset at time t. As a result, wealth next period is

$$W_{t+1}^n = x_t^n(p_{t+1} + d_{t+1}) + a_t^n R = x_t^n(p_{t+1} + d_{t+1} - p_t R) + W_t^n R. \tag{2}$$

We denote the net (or excess) payoff received in $t+1$ from investing in one unit of the risky asset at time t as $s_{t+1} \equiv p_{t+1} + d_{t+1} - p_t R$. Note that $p_{t+1} + d_{t+1}$ is the payoff received at $t+1$ from investing in one unit of the risky asset at time t, and $p_t R$ is the (opportunity) cost of investing in one unit of the risky asset at time t. Using this notation, $W_{t+1}^n = x_t^n s_{t+1} + W_t^n R$.

In the baseline version of our model, we assume that agents are myopic and maximize their per-period utility. This assumption simplifies the maximization problem considerably, and highlights the main determinant of portfolio choice generated by experience-based learning. (In Section 7, we remove this assumption and show that the same mechanism is at work.)

For a given initial wealth level W_n^n, the myopic problem of a generation n at each time

$t \in \{n, ..., n_q\}$ is to choose x_t^n such that it solves $\max_{x \in \mathbb{R}} E_t^n[-\exp(-\gamma W_{t+1}^n)]$, and hence

$$x_t^n \in \arg\max_{x \in \mathbb{R}} E_t^n \left[-\exp(-\gamma x s_{t+1})\right]. \tag{3}$$

Given that agents only need to learn about the mean of dividends, $E_t^n[\cdot]$ is just the (subjective) expectation with respect to a Gaussian distribution with variance σ^2 and a mean denoted by θ_t^n. We call θ_t^n the subjective mean of dividends, and we define it below. Note that, when x_t^n is negative, generation n is short-selling units of the Lucas tree.

2.2 Experience-Based Learning

In this framework, experienced-based learning (EBL) means that agents overweight observations received during their lifetimes when forecasting dividends. For simplicity, we assume that agents *only* use observations realized during their lifetimes.[5] That is, even though they observe the entire history of dividends, EBL agents choose to disregard observations outside their lifetimes. Note that, in this full-information setting, prices do not add any additional information. While it is possible to add private information and learning from prices to our framework, adding these (realistic) feature would complicate matters without necessarily adding new intuition.

EBL differs from reinforcement learning-type models in two ways. First, as already discussed, EBL agents understand the model and know all the primitives, except the mean of the dividend process. Hence, they do not learn *about* the equilibrium, they learn *in* equilibrium. Second, EBL features a passive learning problem in the sense that actions of the players do not affect the information they receive. This would be different if we had, say, a participation decision that would link an action (participate or not) to the type of data obtained and to learning. We consider this to be an interesting line to explore in the future.

We construct the subjective mean of dividends of generation n at time t following Mal-

[5] All we need for our results to hold is that agents discount their pre-lifetime history relative to their experienced history when forming beliefs.

mendier and Nagel (2010):

$$\theta_t^n \equiv \sum_{k=0}^{age} w(k, \lambda, age) d_{t-k}, \tag{4}$$

where $age = t - n$, and where, for all $k \leq age$,

$$w(k, \lambda, age) = \frac{(age + 1 - k)^\lambda}{\sum_{k'=0}^{age} (age + 1 - k')^\lambda} \tag{5}$$

denotes the weight an agent aged age assigns to the dividend observed k periods earlier, and λ parameterizes this weighting function. That is, agents put weight $\frac{(age+1)^\lambda}{\sum_{k'=0}^{age}(age+1-k')^\lambda}$ on the most recent observation, $\frac{(age+1-1)^\lambda}{\sum_{k'=0}^{age}(age+1-k')^\lambda}$ on the previous one, and so forth for all observations experienced during their lifetimes so far. The sum of all weights an agent applies to lifetime experiences is always equal to one, $\sum_{k=0}^{age} w(k, \lambda, age) = 1, \forall\, age \in \{0, 1, ..., q\}$. For example, if $q = 2$, the old generation that has lived for one period uses weights $w(0, \lambda, 1) = \frac{2^\lambda}{1+2^\lambda}$ on the current realization, d_t, and $w(1, \lambda, 1) = \frac{1}{1+2^\lambda}$ on the previous one, d_{t-1}, while the young generation born in t places full weight, $w(0, \lambda, 0) = 1$, on the current observation d_t. Note that the denominator in (5) is a normalizing constant that depends only on age and λ.

The parameter λ regulates the relative weights of earlier and later observations. For $\lambda > 0$, more recent observations receive relatively more weight, whereas for $\lambda < 0$ the opposite holds.

Here are three examples of weighing schemes:

Example 2.1 (Linearly Declining Weights, $\lambda = 1$)**.** *For $\lambda = 1$, weights decay linearly, i.e., for any $0 \leq k, k + j \leq age$,*

$$w(k + j, 1, age) - w(k, 1, age) = -\frac{j}{\sum_{k'=0}^{age}(age + 1 - k')} = -\frac{2j}{(age + 1)(age + 2)}.$$

Example 2.2 (Equal Weights, $\lambda = 0$)**.** *For $\lambda = 0$, lifetime observations are equal-weighted, i.e., for any $0 \leq k \leq age$,*

$$w(k, 0, age) = \frac{1}{age + 1}.$$

Example 2.3. *For $\lambda \to \infty$, the weight assigned to the most recent observation converges to 1, and all other weights converge to 0, i.e., for any $0 \leq k \leq age$*

$$w(k, \lambda, age) \to 1_{\{k=0\}}.$$

Remark 2.1. *We note that the subjective mean θ_t^n in equation (4) can itself be thought of as an expectation with respect to the probability measure implied by the weights $w(k, \lambda, age)$, i.e.,*

$$\mathbb{P}_t^n(d) = \sum_{k=0}^{age} 1_{\{d_{t-k}\}}(d) w(k, \lambda, age), \ \forall d \in \mathbb{R} \tag{6}$$

That is, given a realization of past dividends $(d_\tau)_{\tau=-\infty}^t$, $\mathbb{P}_t^n(d)$ constitutes the experience-based empirical probability measure *of generation n at $t \in \{n, ..., n+q\}$.*[6]

Observe that by construction, $\theta_t^n \sim N(\theta, \sigma^2 \sum_{k=0}^{age}(w(k, \lambda, age))^2)$. Hence, θ_t^n does not necessarily converge to the truth as $t \to \infty$; it depends on whether $\sum_{k=0}^{age}(w(k, \lambda, age))^2 \to 0$. This in turn depends how fast the weights for "old" observations decay to zero (i.e., how small λ is). When agents have finite lives, convergence will not occur. In addition, since separate cohorts weight different realizations differently, we should expect belief heterogeneity, driven by different experiences, at any point in time.

We conclude this section by showing a useful property of the weights, which is used in the characterization results below.

Lemma 2.1. *[Single-Crossing Property] Let $age' < age$ and $\lambda > 0$. Then the function $w(\cdot, \lambda, age) - w(\cdot, \lambda, age')$ changes signs (from negative to positive) exactly once over $\{0, ..., age' + 1\}$.*[7]

Proof. See Appendix A. $\square$

[6] The function $x \mapsto 1_A(x)$ takes value 1 if $x \in A$, and 0 otherwise.

[7] Here and throughout the rest of the paper we set $w(k, \lambda, age) \equiv 0$ for all $k > age$.

2.3 Comparison to Bayesian Learners

To better understand the experience-effect mechanism, we compare EBL agents to agents who update their beliefs using Bayes rule. We consider two sub-cases: the standard case of Bayesian learning, wherein agents use all the available observations to form their beliefs; and an alternative case where agents "learn from experience" in a Bayesian manner, in the sense that they only use data realized during their lifetimes, but update their beliefs using Bayes rule. We call the former case *Full Bayesian Learning* (FBL), and in the latter case *Bayesian Learning from Experience* (BLE).

Full Bayesian Learners. Bayesian learners use all the available observations since the "beginning of time" to form their beliefs. Formally speaking, there is no "beginning of time" in our economy since we are analyzing an economy that has been running forever. Hence, loosely speaking, the beliefs of Bayesian agents who use all available information would have converged to θ at any point in time t, i.e., Bayesian agents would behave like agents that know the true mean.

In order to still illustrate the comparison to Bayesian agents learning from a common sample, we start the economy at an initial time $t = 0$ (for this analysis), and assume that all generations of FBL agents consider all observations since time 0 to form their belief. We denote the prior of FBL agents as $N(m, \tau^2)$. For simplicity, all generations have the same prior, though the analysis can easily be extended to heterogeneous Gaussian priors across generations.[8] The posterior mean of any generation alive at time t, γ_t, is given by

$$\gamma_t = \frac{\tau^{-2}}{\tau^{-2} + \sigma^{-2}t} m + \frac{\sigma^{-2}t}{\tau^{-2} + \sigma^{-2}t} \left(\frac{1}{t} \sum_{k=0}^{t} d_k \right)$$

That is, the belief of an FBL agent is a convex combination of the prior m and the average of all observations d_k available to date. The key difference to EBL agents is that FBL agents do not differ in their beliefs. All generations alive in any given period have the same belief

[8] The assumption of Gaussianity is also not needed but simplifies the exposition greatly.

about mean of dividends; different past experiences do not play a role, and hence there is no heterogeneity in posterior beliefs.

Beliefs are non-stationary (depend on the time period). As $t \to \infty$, the posterior mean eventually converges (almost surely) to the true mean. Hence, with FBL the implications of learning vanish as time goes to infinity. With EBL, this is not true. Since agents learn from their own experiences, our model generates learning dynamics even as time diverges.

Bayesian Learners from Experience. For BLE agents, the situation is different. We assume again that each generation has a prior $N(m, \tau^2)$ when they are born. Here, the posterior mean of generation n at period $t = n + age$, β_t^n, is given by

$$\beta_t^n = \frac{\tau^{-2}}{\tau^{-2} + \sigma^{-2}(age+1)} m + \frac{(age+1)\sigma^{-2}}{\tau^{-2} + \sigma^{-2}(age+1)} \left(\frac{1}{age+1} \sum_{k=n}^{t} d_k \right).$$

That is, the belief of a BLE generation is a convex combination of the prior m and the average of (only) the lifetime observations d_k available to date; in turn this average coincides with the belief of our learners from experience θ_t^n with $\lambda = 0$. Thus, the posterior mean of BLE agents and the belief of EBL agents with $\lambda = 0$ differ in the weight on the prior mean and in EBL agents not employing a prior. If the prior of the FBL agent is diffuse, i.e., $\tau \to \infty$, then β_{n+a}^n coincides with the θ_{n+a}^n of EBL agents for $\lambda = 0$. Additionally, as age increases, β_{n+a}^n gets closer to θ_{n+a}^n (and both gets, almost surely, closer to the true value θ). For lower values of age, instead, the prior will introduce a wedge between the BLE and EBL.

These two benchmark comparisons to Bayesian learning illustrate the role of experience-based learning, i.e., of the assumption that agents only use data observed during their lifetimes, in generating heterogeneity in beliefs. Under FBL, beliefs do not differ across agents and, eventually, will converge to the truth. Under BLE and our main approach, EBL, this is not true. Cohorts differ in their beliefs. However, BLE does not allow for the empirically documented recency bias. As such BLE is akin to over-extrapolation from one's lifetime.

An important difference between EBL, on the one hand, and both types of Bayesian agents, on the other hand, is that we assume that our EBL agents do not understand that their estimate for the mean of dividends is a random variable. A Bayesian agent would acknowledge that the perceived mean of dividends is random and, hence, her expected payoff will consist of two expectations – one with respect to dividends and one with respect to theta. An EBL agent only has the former expectation and not the latter. That is, we assume that EBL agents, who form their beliefs about the mean as described in equation (4), make decisions as if this was the true mean of the dividends.

2.4 Equilibrium Definition

We now proceed to define the equilibrium of the economy with EBL agents.

Definition 2.1 (Equilibrium)**.** *An equilibrium is a demand profile for the risky asset* $\{x_t^n\}$*, a demand profile for the riskless asset* $\{a_t^n\}$*, and a price schedule* $\{p_t\}$ *such that:*

1. *given the price schedule,* $\{(a_t^n, x_t^n) : t \in \{n, ..., n_q\}\}$ *solve the generation-n problem, and*

2. *the market clears in all periods, i.e.,*

$$1 = (q)^{-1} \sum_{n=t-q+1}^{t} x_t^n, \ \forall t \in \mathbb{Z}. \tag{7}$$

We will focus the analysis on the class of equilibria with *affine prices.*

Definition 2.2 (Linear Equilibrium)**.** *A linear equilibrium is an equilibrium wherein prices are an affine function of dividends. That is, there exists a* $K \in \mathbb{N}$*,* $\alpha \in \mathbb{R}$*, and* $\beta_k \in \mathbb{R}$ *for all* $k \in \{0, ..., K\}$ *such that:*

$$p_t = \alpha + \sum_{k=0}^{K} \beta_k d_{t-k}. \tag{8}$$

Thus, the price p_t *is a linear functions of the current and the last* K *dividends.*

Benchmark with known mean of dividends. For the sake of benchmarking our results for EBL agents, we characterize the equilibrium of the economy where the mean of dividends,

θ, is known by all agents, i.e., $E_t^n = \theta \; \forall \, n, t$. In this scenario, there are no disagreements across cohorts, and the demand of any cohort trading at time t is given by:

$$x_t^n \in \arg\max_{x \in \mathbb{R}} E\left[-\exp(-\gamma x s_{t+1})\right] \tag{9}$$

The solution to this problem is standard, and given by:

$$x_t^n = \frac{E\left[s_{t+1}\right] - r p_t}{\gamma V[s_{t+1}]} \tag{10}$$

for all $n \in \{t - q + 1, ..., t\}$, and zero otherwise. To solve the problem, we guess that the solution is $p_t = P$ constant. We verify this guess with the market clearing condition (7) and obtain $P = \frac{\gamma\sigma^2 - \theta}{1-r}$. Furthermore, there is no heterogeneity in cohort's portfolios, and thus, in equilibrium, $x_t^n = 1$ for all $n \in \{t - q + 1, ..., t\}$, and zero otherwise.

3 Illustration: Toy Model

To illustrate the mechanics of the model and to highlight the main results of the paper, we first solve the model for $q = 2$. We will generalize this toy model to any $q > 1$ in the next section (and solve the non-myopic case in Section 7).

When $q = 2$, there are three cohorts alive at each point in time: a young cohort, which enters the market for the first time; a middle-age cohort, which is participating in the market for the second time; and an old cohort, whose agents simply consume the payoffs from their lifetime investments. Since the old cohort has no impact on equilibrium prices or quantities, we focus our analysis on the behavior of the young and middle-aged agents. At time t, the problem of generations $n \in \{t, t-1\}$ is given by (3), and thus demands for the risky asset are

$$x_n^t = \frac{E_t^n\left[s_{t+1}\right]}{\gamma V_t^n[s_{t+1}]}$$

We will show in Section 4 that prices depend on the history of dividends observed by the oldest generation trading in the market, i.e., that $K = q - 1$ in equation (8). Hence, for $q = 2$ we have $K = 1$ and thus

$$p_t = \alpha + \beta_0 d_t + \beta_1 d_{t-1}.$$

Given this functional form for prices, demands can be re-written as

$$\begin{aligned} x_t^t &= \frac{\alpha + (1+\beta_0) E_t^t \left[d_{t+1}\right] + \beta_1 d_t - p_t R}{\gamma \left(1+\beta_0\right)^2 \sigma^2} \\ x_t^{t-1} &= \frac{\alpha + (1+\beta_0) E_t^{t-1} \left[d_{t+1}\right] + \beta_1 d_t - p_t R}{\gamma \left(1+\beta_0\right)^2 \sigma^2}, \end{aligned}$$

noting that $s_{t+1} = p_{t+1} + d_{t+1} - p_t R$. We see that the only difference between the different cohorts trading in the market is their beliefs about future dividends, $E_t^t[d_{t+1}]$ and $E_t^{t-1}[d_{t+1}]$. Since our agents are EBL, each cohort's beliefs about future dividend d_{t+1} are given by:

$$\begin{aligned} E_t^t \left[d_{t+1}\right] = {} & \theta_t^t = d_t \\ E_t^{t-1} \left[d_{t+1}\right] = \theta_t^{t-1} & = \underbrace{\left(\frac{2^\lambda}{1+2^\lambda}\right)}_{w(0,\lambda,1)} d_t + \underbrace{\left(\frac{1}{1+2^\lambda}\right)}_{w(1,\lambda,1)} d_{t-1} \end{aligned}$$

The younger generation has only experienced the dividend d_t and expects the dividends to be identical in the next period. The older generation, having more experience, incorporates the previous dividend in its weighting scheme. Furthermore, belief heterogeneity is increasing in the change in dividends, $|d_t - d_{t-1}|$, and decreasing in the recency bias, λ.

Given these demands, we impose the market clearing condition, $\frac{1}{2}(x_t^t + x_t^{t-1}) = 1$, to derive the equilibrium price:

$$1 = \frac{\alpha + \frac{1}{2}(1+\beta_0)\left[d_t + \frac{2^\lambda}{1+2^\lambda} d_t + \frac{1}{1+2^\lambda} d_{t-1}\right] + \beta_1 d_t - R(\alpha + \beta_0 d_t + \beta_1 d_{t-1})}{\gamma \left(1+\beta_0\right)^2 \sigma^2}$$

We use the method of undetermined coefficients to solve for $\{\alpha, \beta_0, \beta_1\}$. By setting the constants and the terms that multiply d_t and d_{t-1} to zero, we obtain the following conditions:

$$
\begin{aligned}
R\alpha &= \alpha - \gamma \left(1 + \beta_0\right)^2 \sigma^2 \\
R\beta_0 &= \frac{1}{2}\left(1 + \beta_0\right)\left(1 + \frac{2^\lambda}{1 + 2^\lambda}\right) + \beta_1 \\
R\beta_1 &= \frac{1}{2}\left(1 + \beta_0\right)\left(\frac{1}{1 + 2^\lambda}\right)
\end{aligned}
$$

By solving the above system of equations, we obtain the price constant and the loadings of present and past dividends on prices:

$$\alpha = -\frac{\gamma(1 + \beta_0)^2\sigma^2}{R - 1} \tag{11}$$

$$\beta_0 = \frac{2R^2}{(R-1)\left(1 + 2R - \frac{2^\lambda}{1+2^\lambda}\right)} - 1 \tag{12}$$

$$\beta_1 = \frac{R\left(1 - \frac{2^\lambda}{1+2^\lambda}\right)}{(R-1)\left(1 + 2R - \frac{2^\lambda}{1+2^\lambda}\right)} \tag{13}$$

The toy model illustrates that, as the recency bias increases, prices become more responsive to the most recent dividend, $\frac{\partial \beta_0}{\partial \lambda} > 0$, and less responsive to past dividends, $\frac{\partial \beta_1}{\partial \lambda} < 0$. The intuition is straightforward: Under higher recency bias, both cohorts put more weight on the most recent dividend realization. Thus, prices become more responsive to recent dividends, and less responsive to past observations.

In addition, we can derive expressions for price volatility and auto-correlation:

$$
\begin{aligned}
Var(p_{t+1}) &= \frac{1 + 2R(1 + R + 2^{1+2\lambda}R + 2^{1+\lambda}(1+R))}{(R-1)^2(1 + 2(1+2^\lambda)R)^2}\sigma^2 \\
Corr(p_t, p_{t+j}) &= \begin{cases} \frac{R(1+R+2^{1+\lambda}R)}{(R-1)^2(1+2(1+2^\lambda)R)^2} & \text{for } j = 1 \\ 0 & \text{for } j > 1. \end{cases}
\end{aligned}
$$

where $Var(\cdot)$ and $Corr(\cdot)$ denote the unconditional variance and correlation respectively. It can be shown that the variance of prices is increasing in the recency bias while the price auto-correlation is decreasing in the recency bias. When recency bias is large, beliefs and prices are more dependent on the most recent dividend realization. Since both agents have observed the most recent dividend, recency bias reduces belief heterogeneity and induces correlated responses to shocks that generates price volatility and reduces price auto-correlation.

Dividends in this model predict prices, and both actual excess returns:

$$\frac{p_{t+1}+d_{t+1}}{p_t}-R=\frac{\alpha+(1+\beta_0)\,d_{t+1}+\beta_1 d_t}{\alpha+\beta_0 d_t+\beta_1 d_{t-1}}-R,$$

and (adding expectation operators E_t^t and E_{t-1}^t for generation t and $t-1$, respectively) expected excess returns. This equation is a first illustration how our model links demographics and market participation (which generations are trading in the market) are linked to return predictability.

4 Results

We now return to the general case, with $q>1$, and characterize the portfolio choice and resulting demand for the risky asset of the different cohorts under affine prices. We then use market clearing to verify the affine prices guess, and fully characterize demands and prices.

4.1 Characterization of Equilibrium Demands

For any $s,t\in\mathbb{Z}$, let $d_{s:t}=(d_s,...,d_t)$ denote the history of dividends from time s up to time t. For simplicity and WLOG, we assume that the initial wealth of all generations is zero, i.e., $W_n^n=0$ for all $n\in\mathbb{Z}$. At time $t\in\{n,...,n_q\}$, an agent of generation n determines her demand for the risky asset maximizing $E_t^n\left[-\exp\left(-\gamma x s_{t+1}\right)\right]$, as described in (3).

Proposition 4.1. *Suppose $p_t=\alpha+\sum_{k=0}^{K}\beta_k d_{t-k}$ with $\beta_0\neq-1$. Then, for any generation $n\in\mathbb{Z}$ trading in period $t\in\{n,...,n_q\}$, demands for the risky asset are given by*

$$x_t^n=\frac{E_t^n[s_{t+1}]}{\gamma V[s_{t+1}]}=\frac{E_t^n[s_{t+1}]}{\gamma(1+\beta_0)^2\sigma^2}. \tag{14}$$

Proof of Proposition 4.1. The result follows by Lemma B.1 in Appendix B. ☐

4.2 Characterization of Equilibrium Prices

To derive equilibrium prices, we note that equation (14) implies that demands at time t are affine in $d_{t-K:t}$. It is easy to see, then, that beliefs about future dividends are linear functions of the dividends observed by each generation participating in the market and thus prices depend on the history of dividends observed by the oldest generation in the market:

Proposition 4.2. *The price in any linear equilibrium is affine in the history of dividends observed by the oldest generation participating in the market, i.e., for any $t \in \mathbb{Z}$*

$$p_t = \alpha + \sum_{k=0}^{q-1} \beta_k d_{t-k}. \tag{15}$$

with

$$\alpha = -\frac{1}{\left(1 - \sum_{j=0}^{q-1} \frac{w_j}{R^{j+1}}\right)^2} \frac{\gamma\sigma^2}{R-1} \tag{16}$$

$$\beta_k = \frac{\sum_{j=0}^{q-1-k} \frac{w_{k+j}}{R^{j+1}}}{1 - \sum_{j=0}^{q-1} \frac{w_j}{R^{j+1}}} \qquad k \in \{0, ..., q-1\} \tag{17}$$

where $w_k \equiv \frac{1}{q} \sum_{age=0}^{q-1} w(k, \lambda, age)$.[9]

Proof of Proposition 4.2. See Appendix B. ☐

For each $k = \{0, 1, ..., q-1\}$, one can interpret w_k as the average weight placed on the dividend observed at time $t-k$ by all trading generations at time t.

The idea of the proof is as follows. By Lemma D.4, demands at time t are affine in dividends $d_{t-K:t}$. However, from these dividends, only $d_{t-q+1:t}$ may matter for forming beliefs; the dividends $d_{t-K:t-q}$ only enter through the definition of linear equilibrium. The proof shows that under market clearing, the coefficients accompanying the dividends $d_{t-K:t-q}$ are zero.

[9] It is understood that $w(k, \lambda, age) = 0$ for all $k > age$.

The Proposition also implies that we can apply the same restriction to demands and conclude that demands at time t only depend on $d_{t-q+1:t}$.

This result captures the belief channel described by Friedman and Schwartz: prices are a function of past dividends solely due to the fact that generations form their beliefs using past data. By studying a general equilibrium model, however, we provide a more nuanced view. Since observations of older generations affect current prices, they also affect the demand of younger generations, that did not necessarily experience those observations. As such, we provide a link between the factors influencing asset prices and demographic composition.

We also observe that $\frac{\partial \beta_k}{\partial R} < 0$ and $\frac{\partial \alpha}{\partial R} > 0$ for any λ. Thus, the theory predicts that, if the interest rate is higher, the equilibrium price of the risky asset is higher and less volatile, as the variance of prices is given by $\sigma_P^2 = \left(\sum_{k=0}^{q-1} \beta_k^2\right) \sigma^2$. Furthermore, higher risk aversion γ decreases the equilibrium price by lowering α.

The following proposition establishes that when agents form their beliefs by using non-decreasing weights (i.e., $\lambda \geq 0$) prices are more sensitive to more recent dividends.

Proposition 4.3. *For $\lambda > 0$, $0 < \beta_{q-1} < .... < \beta_1 < \beta_0$.*

Proof of Proposition 4.3. See Appendix B. □

This result reflects the fact that the dividends at time t are observed by all generations whereas past dividends are only observed by older generations.

Lemma 4.1. *β_0 is increasing in λ with $\lim\limits_{\lambda\to\infty} \beta_0 = \frac{(Rq)^{-1}}{1-(Rq)^{-1}}$ and $\lim\limits_{\lambda\to\infty} \beta_k > 0$.*

Proof of Lemma 4.1. See Appendix B. □

As $\lambda \to \infty$, it follows that w_k defined in Proposition 4.2 converges to $1_{\{k=0\}}$ for all $k = 0, 1, ..., K$. Therefore, $\beta_k \to 0$ for all $k > 0$ and $\beta_0 \to \frac{(rq)^{-1}}{1-(rq)^{-1}}$. In other words, under extreme recency bias (i.e., $\lambda \to \infty$), only the current dividend affects prices in equilibrium, and at its maximal value, while the weights of all past dividends vanish.

We now describe some implications of EBL for equilibrium prices and returns.

Predictability of Excess Returns. We note that the equilibrium excess return at time $t+j$ is given by:

$$\frac{p_{t+j+1}+d_{t+j+1}}{p_{t+j}} - R = \frac{\alpha + (1+\beta_0)d_{t+j+1} + \sum_{k=1}^{q-1}\beta_k d_{t+j+1-k}}{\alpha + \sum_{k=0}^{q-1}\beta_k d_{t+j-k}} - R.$$

Thus, at time t and for $j \leq q-1$, the dividends $d_{t+j-(q-1)}, ..., d_t$ can be used as factors for predicting the excess returns. For $j > q-1$, our model predicts that excess returns are independent from dividends at time t. It is worth noting that the predictability of excess returns is an equilibrium phenomenon that stems solely from our learning mechanism and not from, say, a build-in dependence in dividends. In fact, our model provides a link between age profile of agents participating in the stock markets and factor for predicting stock returns. This theory provides a nuance mechanism that connects past realizations to future returns through the former's impact on the level of disagreements across market participants.

Price Dynamics. Our results imply that the variance of prices is given by $\sigma_P^2 = \left(\sum_{k=0}^{q-1}\beta_k^2\right)\sigma^2$ and that the autocorrelation structure for prices is

$$Cov(p_{t+j}, p_t) = \begin{cases} \sigma^2\left(\sum_{k=0}^{q-1-j}\beta_k\beta_{k+j}\right) & \text{for any } j \leq q-1 \\ 0 & \text{otherwise.} \end{cases}$$

A direct implication is that, as $\lambda \to \infty$, the autocorrelation of prices vanishes. That is, as the recency bias becomes stronger, the prices tend to be uncorrelated.

4.3 Cross-Section of Asset Holdings

The next proposition establishes that younger generations react more optimistically (pessimistically) than older generations to positive (negative) changes in current dividends.

Proposition 4.4. *For any $t \in \mathbb{Z}$ and any generation n alive at t, there is a threshold $j_0 \leq t-n-1$ such that for dividends that date back up to j_0 periods, younger generations* *react stronger to changes than older generations, while for dividends that date back more than j_0 periods the opposite effect holds, i.e.,*

1. $\frac{\partial x_t^{n+1}}{\partial d_{t-j}} \geq \frac{\partial x_t^n}{\partial d_{t-j}}$ *for* $0 \leq j \leq j_0$ *and*

2. $\frac{\partial x_t^{n+1}}{\partial d_{t-j}} \leq \frac{\partial x_t^n}{\partial d_{t-j}}$ *for* $j_0 < j \leq t$.

Proof. See Appendix B. □

In our model, the younger generation puts more weight on current dividends when forming beliefs, so when d_t increases, the younger are "overly optimistic" relatively to the older generation. This reflects that an increase (decrease) in current dividends makes younger agents more optimistic (pessimistic) about the return of the risky asset than older agents because they put more weight on recent realizations. This term is only zero when both agents have the same belief formation (e.g. $w(0, \lambda, 1) = 1$). We use Lemma 2.1 to extend this intuition to the more recent dividends as opposed to just the current one.

Moreover, let $\xi(n, k, t) \equiv x_t^n - x_t^{n+k}$ be the discrepancy between positions of generation n and $n + k$. By Proposition 4.1, and some algebra it follows that:

$$\xi(n, k, t) = \frac{E_t^n[\theta] - E_t^{n+k}[\theta]}{\gamma(1 + \beta_0)\sigma^2} \tag{18}$$

for any $k = \{0, ..., t - n\}$. So the discrepancy between positions of different generations is entirely explained by the discrepancy in beliefs. Note that as indicated by the indicator function, $1_{\{j \leq t-n-k\}}$, the younger generation does not form beliefs about dividends prior to their birth. For instance, if for some $a > 0$, $d_{n:t} \approx d_{n+a:t+a}$, then $\xi(n+a, k, t+a) \approx \xi(n, k, t)$.[10]

In addition, the next result shows that during "expansions", understood as periods with increasing dividends, for any two generations born during the expansions, the young generation has a relatively higher demand for risky assets than the old one.

[10] This last claim follows since the inter-temporal change in discrepancies between sets of generations of the

Proposition 4.5. *Suppose $\lambda \geq 0$ and $t_0 \leq t_1$ are two points in time such that dividends are non-decreasing from t_0 up to t_1. Then for any two generations n and $n+k$ born between t_0 and t_1, the older generation has lower demand of the risky asset (x_t^n) than the younger generation (x_t^{n+k}) at any point $n \leq t \leq t_1$, i.e., $\xi(n,k,t) \leq 0$. On the other hand, if dividends are non-increasing then $\xi(n,k,t) \geq 0$.*

Proof. See Appendix B. □

4.4 Trade Volume

We now study how learning and disagreements affect the volume of trade observed in the market. We define the total volume of trade in the economy as

$$TR_t \equiv \left(\sum_{n=t-q}^{t} \frac{1}{q} \left(x_t^n - x_{t-1}^n \right)^2 \right)^{\frac{1}{2}} \tag{19}$$

with $x_{n-1}^n = 0$. That is, trade volume is the weighted sum (squared) of the change in positions of all agents in the economy. It can be characterized as follows:

Proposition 4.6. *The trade volume defined in (19) can be expressed as*

$$TR_t = \chi \left\{ \frac{1}{q} \sum_{n=t-q}^{t} \left[\left(\theta_t^n - \theta_{t-1}^n \right) - \frac{1}{q} \sum_{n=t-q}^{t} \left(\theta_t^n - \theta_{t-1}^n \right) \right]^2 \right\}^{\frac{1}{2}}, \tag{20}$$

where $\chi = \frac{1}{\gamma\sigma^2(1+\beta_0)}$.

same age, $\xi(n+a,k,t+a) - \xi(n,k,t)$ for $a > 0$, is given by

$$\begin{aligned}
&\frac{\sum_{j=0}^{t-n-k}\{w(j,\lambda,t-n) - w(j,\lambda,t-n-k)\}d_{t+a-j}}{\gamma(1+\beta_0)\sigma^2} + \frac{\sum_{j=t-n-k+1}^{t-n} w(j,\lambda,t-n)d_{t+a-j}}{\gamma(1+\beta_0)\sigma^2} \\
&- \frac{\sum_{j=0}^{t-n-k}\{w(j,\lambda,t-n) - w(j,\lambda,t-n-k)\}d_{t-j}}{\gamma(1+\beta_0)\sigma^2} - \frac{\sum_{j=t-n-k+1}^{t-n} w(j,\lambda,t-n)d_{t-j}}{\gamma(1+\beta_0)\sigma^2} \\
=&\frac{\sum_{j=0}^{t-n-k}\{w(j,\lambda,t-n) - w(j,\lambda,t-n-k)\}(d_{t+a-j} - d_{t-j})}{\gamma(1+\beta_0)\sigma^2} \\
&+ \frac{\sum_{j=t-n-k+1}^{t-n} w(j,\lambda,t-n)(d_{t+a-j} - d_{t-j})}{\gamma(1+\beta_0)\sigma^2}.
\end{aligned}$$

Proof. See Appendix C. □

The previous Proposition shows that the presence of learning and disagreements induces trade volume through changes in beliefs, that in our framework are driven by changes in the observed history of dividends. Note that the trade volume measure TR_t proxies for the volatility of changes in beliefs. We can see that when the change in each cohorts beliefs is different from the average change in beliefs, trade volume is increased.

Therefore, to understand the drivers of trade volume, we need to understand the changes in beliefs across cohorts to a given shock. In our framework, an increase (decrease) in dividends impacts the belief of both generations in the market, but the effect on beliefs is stronger for the younger generations. Therefore, an increase (decrease) in dividends should induce trade if it makes young agents more optimistic (pessimistic) than old agents. This mechanism is solely due to the presence of experience-based learners, since it is essential that each generation reacts differently to the same realization of dividends. We can see that if all agents adjust their beliefs equally, trade volume is zero.

Thought Experiment. Suppose $d_{t_0} = d_{t_0+1} = ... = d_{t-1} = \bar{d}$ for $t - t_0 > q$ and that $d_t \neq \bar{d}$. This thought experiment is supposed to capture this economy's reaction to a shock after a long period of stability. First, note that all generations alive at time $t-2$ and $t-1$ have only observed a constant stream of dividends: $\bar{d}$. Therefore, $\theta^n_{t-2} = \theta^n_{t-1} = \bar{d}$ for all $n = \{0, ..., q-1\}$. As a result, we know that trade volume in $t = 1$ should be zero: $TR_{t-1} = 0$. What happens when dividend $d_t \neq \bar{d}$ is observed? Note that now, for each generation $n = \{0, .., q-1\}$, beliefs are given by $\theta^n_t = w(0, \lambda, t-n)(d_t - \bar{d}) + \bar{d}$ which implies the following change in beliefs:

$$\theta^n_t - \theta^n_{t-1} = w(0, \lambda, t-n)(d_t - \bar{d}) \tag{21}$$

Trade volume in t is then given by:

$$TR_t = \left|d_t - \bar{d}\right| \chi \left[\frac{1}{q} \sum_{n=0}^{q-1} \left(w(0,\lambda,t-n)^2 - \left(\frac{1}{q} \sum_{n=0}^{q-1} w(0,\lambda,t-n) \right)^2 \right) \right]^{\frac{1}{2}} \tag{22}$$

First, note that trade volume increases proportionally to the change in dividends, independently of whether the latter is positive or negative, and to a function that reflects the dispersion of the weights agents assign to the most recent observation in their belief formation process. Second, the level of trade volume generated by a given change in dividends will depend on the level of recency bias of the economy. For example, as $\lambda \to \infty$, the dispersion in weights decreases as $w(0,\lambda,a) \to 1$ for all $a \in \{0, ..., q-1\}$. Thus, our results suggest that higher recency biases (reflected in higher λ), should generate lower trade volume responses for a given shock to dividends, and vice-versa.

Volume of Trade in the Toy Model. We can compute the trade volume of the economy with $q = 2$. At time t, the trade volume as stated in Proposition 4.6 can be re-written as:

$$\begin{aligned} TR_t &= \frac{1}{\gamma\sigma^2 (1+\beta_0)} \left[\frac{1}{2} \left(\theta_t^t - \bar{\theta}_t\right)^2 + \frac{1}{2} \left(\theta_t^{t-1} - \bar{\theta}_t\right)^2 \right]^{\frac{1}{2}} \\ &= \frac{|\theta_t^t - \theta_t^{t-1}|}{2\gamma\sigma^2 (1+\beta_0)} = \frac{\left(1 - \frac{2^\lambda}{1+2^\lambda}\right) |d_t - d_{t-1}|}{2\gamma\sigma^2 (1+\beta_0)} \end{aligned}$$

where $\bar{\theta} = 0.5 \left(\theta_t^t + \theta_t^{t-1}\right)$. We can see that the trade volume is decreasing in the recency bias, λ, and is increasing in the change in dividends, $|d_t - d_{t-1}|$.

5 Demographics and Equilibrium Prices

In this extension, we study the effect of one-time unexpected demographic shock in our toy economy with $q = 2$. We model this as a shock to the mass of young people entering the economy at time $t = \tau$. At any time t, we denote the mass of young agents by y_t and the total mass of agents by $m_t = y_t + y_{t-1}$. We assume that $y_t = y$ and thus $m_t = 2y = m$ for all $t < \tau$ and $t > \tau + 1$. We consider two types of demographic shocks: a positive shock, $y_\tau > y$, and a negative shock, with $y_\tau < y$. The former can be interpreted as a baby-boom and the latter as a war that occur at $t = \tau$.

We know from our previous results that the for $t < \tau$ and $t > \tau + 1$ prices are given by $p_t = \alpha + \beta_0 d_t + \beta_1 d_{t-1}$ where $\{\alpha, \beta_0, \beta_1\}$ are given by equations 11-13 since for these time periods the economy is as the one described in Section 3. Thus, we are left to characterize demands and prices for τ and $\tau+1$, where the shock generation is young and old respectively. To do so, we make the following guesses:

$$
\begin{aligned}
p_\tau &= a^y + b_0^y d_\tau + b_1^y d_{\tau-1} \\
p_{\tau+1} &= a^o + b_0^o d_{\tau+1} + b_1^o d_\tau
\end{aligned}
$$

We solve the problem by backwards induction. Note that the form of agent's demands remains unchanged. Market clearing in $\tau + 1$, with mass y of young agents and y_τ of old agents, states:

$$
1 = y \frac{E_{\tau+1}^{\tau+1}\left[p_{\tau+2} + d_{\tau+2}\right] - r p_{\tau+1}}{\gamma\left(1+\beta_0\right)^2 \sigma^2} + y_\tau \frac{E_{\tau+1}^{\tau}\left[p_{\tau+2} + d_{\tau+2}\right] - r p_{\tau+1}}{\gamma\left(1+\beta_0\right)^2 \sigma^2}
$$

where $\omega \equiv \frac{2^\lambda}{1+2^\lambda}$. Our guess is verified, and we obtain the following coefficients for the price function in $\tau + 1$:

$$
\begin{aligned}
a^o &= \alpha \frac{1}{r}\left[1 + \frac{r-1}{m_\tau}\right] \\
b_0^o &= \beta_0 \left[1 + \frac{1}{r}\left(\frac{m_\tau - y_\tau}{m_\tau} + \frac{y_\tau}{m_\tau}\omega - \frac{y}{m}(1+\omega)\right)\right] + \frac{1}{r}\left(\frac{m_\tau - y_\tau}{m_\tau} + \frac{y_\tau}{m_\tau}\omega - \frac{y}{n}(1+\omega)\right) \\
b_1^o &= \beta_1 \frac{y_\tau}{m_\tau}\frac{m}{y}
\end{aligned}
$$

where $m_\tau = y + y_\tau$. First, note that for $y_\tau = y$, the coefficients are as those in the baseline model in equations 11-13. Second, and most importantly, note that the total mass of agents, m_τ, only affects the constant a^o, while the loadings b_0^o and b_1^o are only a function of the fraction

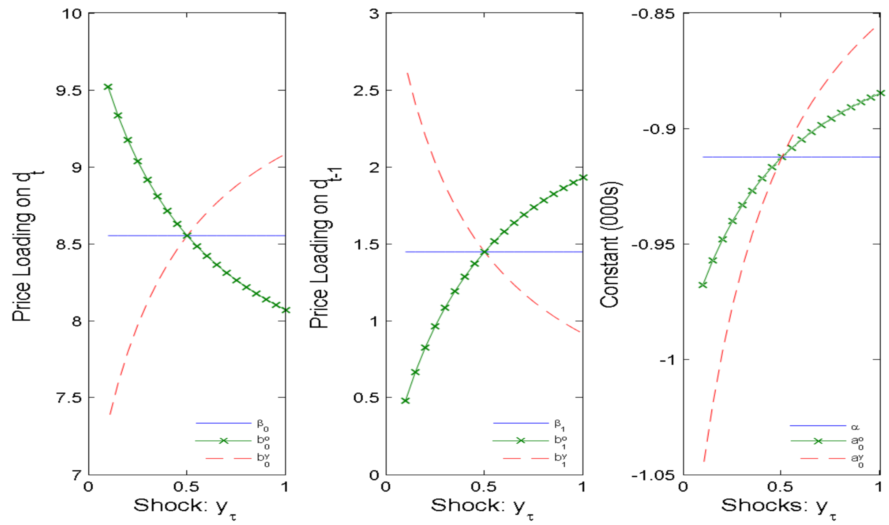


Figure 2: Demographic Shocks and Price Coefficients.

Note: This figure plots coefficients $\{\beta_0, b_0^y, b_0^o\}$, $\{\beta_1, b_1^y, b_1^o\}$, and $\{\alpha, a^y, a^o\}$, respectively, as a function of the demographic shock y_τ. The results are for $y = 0.5$, $\gamma = 1$, $\lambda = 3$, $\sigma = 1$, and $R = 1.1$.

of young agents in the economy. Finally, given this, the demands and market clearing at time τ imply:

$$1 = y_\tau \frac{E_\tau^\tau \left[p_{\tau+1} + d_{\tau+1}\right] - rp_\tau}{\gamma \left(1 + b_0^o\right)^2 \sigma^2} + y \frac{E_\tau^{\tau-1} \left[p_{\tau+1} + d_{\tau+1}\right] - rp_\tau}{\gamma \left(1 + b_0^o\right)^2 \sigma^2}$$

Our guess is verified, and we obtain the following coefficients for the price function at τ:

$$\begin{aligned}
a^y &= \frac{1}{r}\left[a^o - \frac{\gamma \left(1 + b_0^o\right)^2 \sigma^2}{m_\tau}\right] \\
b_0^y &= \frac{1}{r}\left(1 + b_0^o\right)\left(\frac{y_\tau}{m_\tau} + \frac{m_\tau - y_\tau}{m_\tau}\omega\right) + \frac{1}{r^2}\left(1 + \beta_0\right)\frac{y_\tau}{m_\tau}\left(1 - \omega\right) \\
b_1^y &= \frac{1}{r}\left(1 + b_0^o\right)\frac{m_\tau - y_\tau}{m_\tau}\left(1 - \omega\right)
\end{aligned}$$

Figure 2 shows how the the reliance of prices on dividend realizations changes as a function of the size and direction of the demographic shock. From the first two panels, we can see that a positive demographic shock generates a stronger response of prices to the contemporaneous dividend and a weaker response to past prices; and that this response increases in the size of

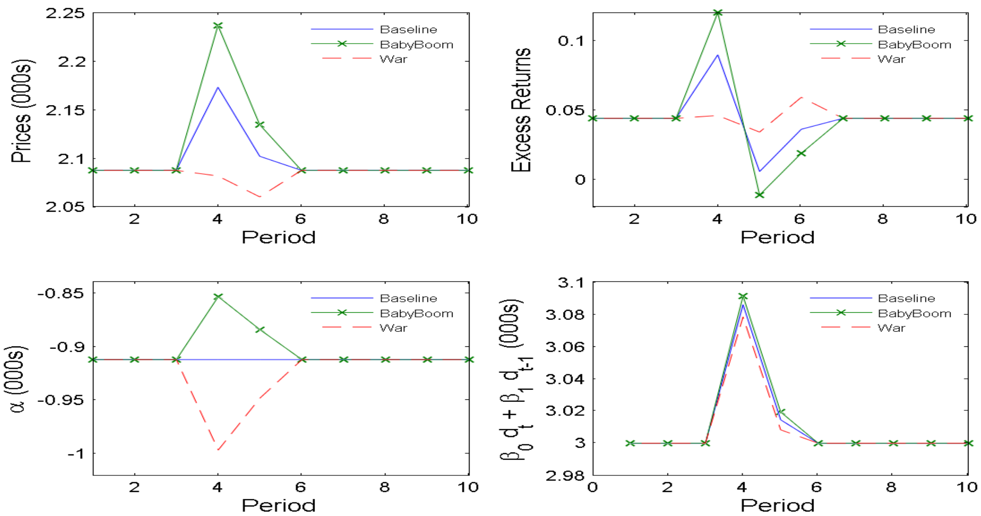


Figure 3: Demographics and Dividend Shock.

Note: This figure plots the response of prices and excess returns to a 1% one-period-only increase in dividends in $t = 4$. The results are for $\gamma = 1$, $\lambda = 3$, $\sigma = 1$, and $R = 1.1$. In the *Baby Boom* line there is a positive demographic shock in $t = 4$ as well, $y_4 = 0.75$, while the *War* line shows a negative demographic shock, $y_4 = 0.25$. In the *Baseline* case, there are no demographic shocks, $y = 0.5$ for all t.

the shock. This is because there is more young people in the market who pay no attention to past dividends. Consistent with this, when the τ-generation is old, prices depend less on contemporaneous dividends and more on past dividends than in the baseline case. In addition, the third panel shows that there is a level increase in prices that is captured in an increase in the price constant; this is because there is a higher overall demand for the risky asset since there are more people in the market. These predictions are reversed for a negative demographic shock.

Figure 3 shows how prices and excess returns respond to a positive dividend shock that is contemporaneous to a *Baby-Boom* or to a *War*. We can see that for the baby-boom case, prices over-react relative to the baseline case, and that this over-reaction is still present, but less severe, when the baby-boom generation is old. This is because when the baby-boomers are young, there is more people in the market that only pay attention to the present dividends, relative to the baseline case. When this generation ages, however, the over-reliance is softened because the old baby-boomers update their beliefs and reduce the weight they put on past dividends, and also a new generation of young agents that disregards past dividends will enter the market. Consistent with this, returns are positive in response to the shock, and negative following the shock. We can also see that the reaction of prices and returns is reversed when the positive shock occurs during a war. In this particular example, the fall in the overall demand for the risky asset (since less people are present in the market) is strong enough to generate a fall in prices in response to the chosen positive dividend shock.

6 Stylized Empirical Facts

In this section, we examine whether our model is consistent with aggregate facts about equity holdings and stock turnover. The model generates at least two predictions that are directly testable: first, the stronger response of the younger generations' risky asset demand to recent dividends; and second the relationship between differences in experience-based beliefs and trade volume.

To test these model implications, we combine historical data on stock-market performance, obtained from Robert Shiller's website, with data on stock holdings from the Survey of Consumer Finances (SCF) and stock turnover data from the Center for Research in Security Prices (CRSP).

Dividends in our model do not translate one-to-one to dividends in the real world. Rather, the role of dividends in this Lucas-tree economy capture news about firm performance. Real-world dividends, instead, may not reflect how well a firm is doing, for example, because a firm might decide to retain earnings rather than distributing them to shareholders, or because management might have incentives to smooth dividends.[11] We therefore turn to stock market returns rather than dividends as a measure of stock market performance.[12]

[11] We also note that dividends exhibit an increasing time trend, which would need to be corrected. Nevertheless our results remain very similar when we include dividends into the return calculation. The appendix contains figures based on returns including dividends.

[12] Another possible measure for (theoretical) dividends would be earnings but, similar to dividends, earnings do not necessarily reflect profitability, and they exhibit a similar time trend.

Correspondingly, we calculate the lifetime experiences of "dividends" of the different generations as the weighted average of the performance over their lifetimes, using linearly declining weights and $\lambda = 3$. As described above, stock market performance is measured by annual returns of the SP500 Index. All performance measures are de-trended using the consumer price index (CPI). We construct two measures of disagreement between cohorts based on their past experiences. First, we take the difference between the experienced performance by an older age group (60 years and older) and a younger age group (40 years and younger). For each age group, the experienced performance is calculated as the weighted average of each cohort within that age range.[13] Each cohort-year is weighted by the total number of people of that cohort in that year. Second, disagreement is approximated by the standard deviation of experienced performances across cohorts in a given year. Again, we use the total number of individuals in each cohort in a given year to weight the observations.

The SCF includes data on dollar stock holdings and liquid by individuals from 1960 to 2013 on a household level.[14] With these two variables, we code a measure for the extensive margin of stock holding, i.e., whether the household has invested a positive amount into stock, and a measure for the intensive margin, i.e., how much of the liquid assets are invested into stocks. For the intensive margin, we drop all households that have no money in stocks. Since the age of the head of the household is known, we can match our experienced performance measures to each household. We then aggregate the household into the aforementioned age groups by taking the unweighted average of the intensive and extensive margin over all households whose head falls into that particular age-group.[15]

Figure 4 depicts the relationship between the extensive and intensive margin of stock holdings and difference in experienced returns between the above-60 age-group and the below-40 age-group for weights of past returns. In graphs 4(a) and 4(c), experienced returns are

[13] We assume that every individual in the survey is born on January 1 and experiences those returns. We further assume that the SCF is also conducted on January 1. The results are unchanged if either, birthday or date of survey, or both are assumed to be December 31.

[14] Appendix E provides a more detailed description

[15] Note that the extent of the survey has changed over time. Hence, both, the intensive and extensive margin is weighted by sample weights included in the SCF.

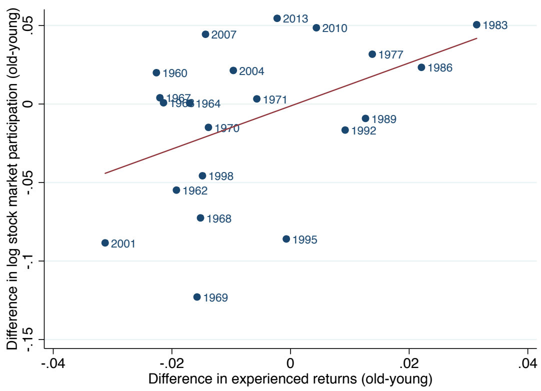
calculated from equation (5), with λ set to one, corresponding to linearly declining weights. For graphs 4(b) and 4(d), λ is set to three, which corresponds more closely to previous estimations. The results are in line with the predictions for either calibration. If the older age-group experienced higher stock returns, they are more likely to hold stock compared to the younger age-group as in graph 4(a)). Likewise, the above-60 age group invests a relatively higher share of the their liquid assets into stocks vis-à-vis the younger age-group if their experienced returns are higher than those of the younger age-group; see graphs 4(c) and 4(d)).[16]

Our model suggests that trade volume is high when disagreement among investors is high. To validate this postulate, we examine the co-movement of trade volume and the evolution of the standard deviation of experienced performance, the aforementioned measure of disagreement. We obtain monthly data on the number of traded shares, number of outstanding shares and stock price for every ordinary common share in CRSP from 1960 to 2007. For each month the traded value is calculate as the number of traded shares times the average share price.[17] To account for changes in capitalization, we scale the traded volume by the market capitalization of the firm. The market capitalization is calculated as the number of outstanding shares times the average stock price. We refer to the scaled traded volume as turnover ratio. To align the turnover ratio with the frequency of our disagreement variable, we calculate the annual turnover ratio as the average of the monthly turnover ratios. Since the trade volume is likely driven by a number of factors not related to disagreement, for instance technological progress, we de-trend the turnover ratio as follows. First, we take the log of the turnover ratio. Second, the logged turnover series are regressed on a linear time trend. The residuals are averaged for each year to obtain a measure of deviation of the trading activity from the trend.

The co-movement of trading volume and disagreement are generally in line with our pre-

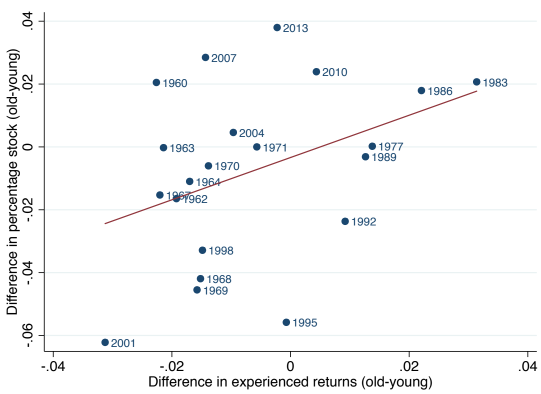
[16] The results are less clear-cut for the other performance measure, such as experienced dividends and earnings; see Appendix E.

[17] In particular, we obtain the items *prc* for the monthly stock price, "vol" for the number of traded shares in a month and *shrout* for the number of shares outstanding from CRSP.



(a) Stock Market Participation. ($\lambda = 1$)

(b) Stock Market Participation. ($\lambda = 3$)



(c) Fraction of Stock of Financial Assets. ($\lambda = 1$)

(d) Fraction of Stock of Financial Assets. ($\lambda = 3$)

Figure 4: Experienced Returns and Stock Holding

Difference in experienced returns is calculated as the experienced returns of the SP500 Index. More distant experienced receive a lower weight. The weights either decline linearly ($\lambda = 1$) or super-linear ($\lambda = 3$) as in equation (5). Stock Market Participation is measured as the fraction of households that either directly held stock or indirectly, e.g. via mutuals or retirement accounts. We classify households whose head is aged 60 or older as "old" and households whose head is younger than 40 as "young". Difference in stockholdings, the y-axis in graphs (a) and (b), is calculated as the difference between the logs of the fraction of stock-holding households of the old and young age group. Percentage stock, the y-axis in graphs (c) and (d), is the fraction of assets invested in stock. The red line depicts the linear trend.

diction. (See Figure 5.) If disagreement among investors, approximated by the standard deviation of experienced performance, is higher, the actual turnover ratio is higher than the trend turnover ratio.[18]

7 Extension: Model with Non-myopic Agents

In this section, we consider non-myopic agents who choose their portfolios by looking at their entire lifetime. We assume that agents consume only in their final period, i.e., they consume their final wealth. We first characterize the demands for risky assets. Even though demands for risky assets are linear (in accordance with Proposition 4.1), the functional form contains an additional term that accounts for the dynamic nature of the non-myopic problem. Due to the aforementioned linearity in risky demands, we are then able to show that the result in Proposition 4.2 continues to hold: Prices are affine functions of past dividends observed by the generations that are trading.

7.1 Characterization of risky demands for non-myopic agents.

For any $s, t \in \mathbb{Z}$, let $d_{s:t} = (d_s, ..., d_t)$ denote the history of dividends from time s up to time t. At time n, a n-generation agent solves the following problem:

$$\max_{\mathbf{x} \in \mathbb{R}^q} E_n^n \left[-\exp\left(-\gamma W_{n+q}^n(\mathbf{x})\right) \right] \tag{23}$$

$$\text{s.t. } W_{n+q}^n(\mathbf{x}) = \sum_{\tau=n}^{n_q} R^{n_q - \tau} x_\tau s_{\tau+1} \tag{24}$$

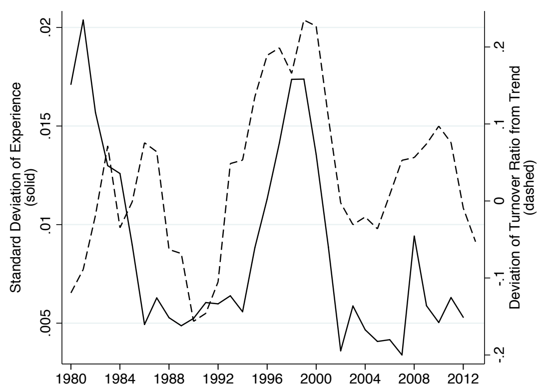
where $\mathbf{x} \in \mathbb{R}^q$ are the q trading decisions from n up to n_q. Note that, by moving from maximizing next period's wealth under the myopic formulation to maximizing final-period wealth, the non-myopic formulation introduces discounting with factor $R^{n_q - \tau}$.

We continue to assume that the initial wealth of all generations is zero, i.e., $W_n^n = 0, \forall n$.

[18] Again, the results are less clear for experienced dividends and earnings.

(a) Moving Average with 1 Lag

(b) Moving Average with 3 Lags



(c) Moving Average with 5 Lags

Figure 5: Turnover Ratio and Disagreement in Experienced Returns

The dashed line depicts the turnover ratio from its trend. The turnover ratio is calculated as the ratio of the value of traded stocks and the value of stocks outstanding. We de-trend the turnover ratio by first taking the log and then remove a linear trend. Finally, turnover ratio is smoothed by taking the moving average with 1, 3 and 5 lags.The solid line shows the standard deviation of experienced stock returns for a given year. For the the calculation of the standard deviation, we weight each age-group with the number of persons of that age.

We can cast this problem iteratively — by solving from n_q backwards — as

$$V^n_{n_q}(d_{n_q-K:n_q}) = \max_{x\in\mathbb{R}} E^n_{n_q}\left[-\exp\left(-\gamma s_{n+q}x\right)\right] \text{ and} \tag{25}$$

$$V^n_{\tau}(d_{\tau-K:\tau}) = \max_{x\in\mathbb{R}} E^n_{\tau}\left[V^n_{\tau+1}(d_{\tau+1-K:\tau+1})\exp\left(-\gamma s_{\tau+1}x\right)\right],\ \forall\ \tau\in\{n,...,n_q-1\} \tag{26}$$

Remark 7.1. *Notice that V^n_τ does not include the wealth at time τ, that is, from equation (**??**), the optimization problem can be cast as* $\max_{x\in\mathbb{R}}\exp\{-\gamma RW^n_{n_q}\}E^n_{n_q}\left[-\exp\left(-\gamma s_{n+q}x\right)\right]$. *However, our definition of $V^n_{n_q}$ omits the term* $\exp\{-\gamma RW^n_{n_q}\}$ *since it does not affect the maximization.*

This shows that, although the n-generation's problem at n_q is a static portfolio problem, for any other $\tau\in\{n,...,n_q-1\}$, it is not because $V^n_{\tau+1}$ is correlated with $s_{\tau+1}$ through dividends. That is, dividend realization $d_{\tau+1}$ impacts (i) the net payoff obtained from investing x_τ in the risky asset at time τ, and (ii) the continuation value $V^n_{\tau+1}(d_{\tau+1-K:\tau+1})$ by affecting the beliefs of the n-generation at $\tau+1$, and the resulting portfolio decision.

First, we characterize the portfolio choice and resulting demand for the risky asset of the different cohorts under affine prices. We begin by highlighting that the dynamic portfolio problem of agents in this economy cannot be expressed as a succession of static problems, as is standard in the literature (see Vives (2010)). This is because of learning and the fact that agents are sophisticated enough to understand how their beliefs evolve over their lifetime. These features introduce a correlation between future returns and continuation values that distorts the portfolio decisions. In what follows, however, we show that the agents dynamic portfolio problem can be expressed as an *adjusted static* problem where dividends follow a normal distribution with *adjusted* mean and variance. Intuitively, agents recognize that very high and very low realizations of future dividends will lead to more disagreement, which they will exploit in their future trades. As a result, extreme realizations are now associated with higher continuation values, leading to a downward adjustment of the variance.

Let $E_{N(\mu,\sigma^2)}[.]$ and $V_{N(\mu,\sigma^2)}[.]$ be the expectation and variance with respect to a Gaussian pdf with mean μ and σ^2.

Proposition 7.1. *Suppose* $p_t = \alpha + \sum_{k=0}^{K} \beta_k d_{t-k}$ *with* $\beta_0 \neq -1$. *Then, for any generation* n *in period* $n+j$ *for* $j \in \{0, ..., q-1\}$ *(the age of the generation), demands for the risky asset are given by:*

$$x_{n+j}^{n} = \frac{E_{N(m_j,\sigma_j^2)}[s_{n+j+1}]}{\gamma R^{q-1-j} V_{N(m_j,\sigma_j^2)}[s_{n+j+1}]} \tag{27}$$

where:

$$m_j \equiv \frac{\theta_{n+j}^{n} - \sigma^2 \left(b_j + \sum_{k=1}^{K} b_j(k) d_{n+j-k} \right)}{2c_j\sigma^2 + 1} \tag{28}$$

$$\sigma_j^2 \equiv \frac{\sigma^2}{2c_j\sigma^2 + 1} \tag{29}$$

for $\{\{b_j(k)\}_{k=1}^{q-1}, b_j, c_j\}$ *constants that change with the agent's age (j) (for exact expressions see the proof).*

Proof of Proposition 7.1. See Appendix D. □

The intuition of the proof is as follows. By solving the problem backwards we note that at time n_q the problem is in fact a static one (see equation (25)). In particular we show that $V_{n_q}^{n}$ is of the form exponential-quadratic in d_{n_q} (see Lemma B.1 in the Appendix). We then show that the exponential-quadratic term times the Gaussian distribution of dividends imply a new Gaussian distribution with an slanted mean and variance (see Lemma D.1 in the Appendix). Thus the problem at time $n_q - 1$ can be viewed as a static problem with a modified Gaussian distribution, and consequently (a) demands are of the form of 27 and $V_{n_q-1}^{n}$ is also of the exponential-quadratic form. The process thus continues until time n.

After straightforward algebra, we can cast equation (27), as [19]

$$
\begin{aligned}
x^n_{n+j} =& \frac{1}{R^{q-1-j}} \frac{E_{N(\theta^n_{n+j},\sigma^2)}[s_{n+j+1}]}{\gamma V_{N(\theta^n_{n+j},\sigma^2)}[s_{n+j+1}]} - \frac{(b_j + \sum_{k=1}^K b_j(k) d_{n+j-k})}{\gamma R^{q-1-j}(1+\beta_0)} \\
\equiv& \frac{1}{R^{q-1-j}} \tilde{x}^n_{n+j} + \Delta^n_{n+j}
\end{aligned}
\tag{30}
$$

The term $\tilde{x}^n_{n+j}$ coincides with the demand of a *static* portfolio problem for an agent with beliefs θ^n_{n+j}; see Proposition 4.1. We coin this term the *myopic component* of the demand for risky assets. The scaling by $1/R^{q-1-j}$ arise because agents discount the future by R. The second term $\Delta^n_{n+j} \equiv -\frac{(b_j+\sum_{k=1}^K b_j(k) d_{n-k})}{\gamma R^{q-1-j}(1+\beta_0)}$, is an adjustment which accounts for the dynamic nature of the problem, and thus, we call it the *dynamic component*. It arises because agents understand that they are learning about the risky asset, and thus understand that the value function is correlated with the one-period-ahead returns.

7.2 Characterization of equilibrium prices for non-myopic agents

The following proposition shows that in a linear equilibrium prices at any time t only depend on the dividends observed by the generations trading at time t. This result shows that the insights in Proposition 4.2 continue to hold in this setup with non-myopic agents.

Proposition 7.2. *For $R > 1$, the price in any linear equilibrium with $\beta_0 \neq -1$ is affine in the history of dividends observed by the oldest generation participating in the market. For any $t \in \mathbb{Z}, q \geq 1$,* [20]

$$
p_t = \alpha + \sum_{k=0}^{q-1} \beta_k d_{t-k}. \tag{31}
$$

Proof of Proposition 7.2. See Appendix D. □

The idea of the proof is as the one discussed for the myopic case.

[19] Note that $E_{N(b+a,s)}[s_{n+1}] = E_{N(a,s)}[s_{n+1}] + (1+\beta_0)b$.

[20] Heuristically, an equilibrium with $\beta_0 = -1$ is not well-defined since in this case the excess payoff, say, s_{t+q-1} is deterministic given the information at time $t+q-2$ and thus the agent will take infinite positions depending on $d_{t+q-1} + p_{t+q-1} - rp_{t+q-2}$.

7.3 The $q = 2$ Case.

We now specialize our results to the case with $q = 2$. By doing so, we are able to sharpen our previous results regarding the behavior of prices and risky demands in equilibrium.

The next lemma shows that $\{\alpha, \beta_0, \beta_1\}$ solve a complicated system of non-linear equations

Lemma 7.1. *For $R > 1$ in any linear equilibrium prices are given by:*

$$p_t = \alpha + \beta_0 d_t + \beta_1 d_{t-1} \qquad \forall t \in \mathbb{Z} \tag{32}$$

where the coefficients $\{\alpha, \beta_0, \beta_1\}$ are uniquely determined by a set of three non-linear equations specified in the appendix.

Proof of Lemma 7.1. See Appendix D. □

Although the equations in the lemma form a complicated system of non-linear equations, we are able to establish that prices react positively to dividends d_t and d_{t-1}. Formally,

Proposition 7.3. *For $\lambda > 0$, $\alpha \leq 0$ and $0 < \beta_1 < r\beta_0$.*

Proof of Proposition 7.3. See Appendix D. □

This proposition is analogous to Proposition 4.3 and establishes that when agents form their beliefs by using non-decreasing weights (i.e., $\lambda \geq 0$) $\beta_0 R$ is larger than β_1. This result reflects the fact that the dividends at time t are observed by both generations whereas d_{t-1} is only observed by the old generation; in fact it is not hard to see from the equations that in the case $w(1, \lambda, 0) = 0$ –agents do not put any weight on the previous dividend,– then $\beta_1 = 0$.

Figure 6 depicts the behavior of $\{\beta_0, \beta_1\}$ for different values of (λ, R). Note that the values of $\{\beta_0, \beta_1\}$ are independent of the process for dividends, σ^2, and of the coefficient of risk aversion, γ. Thus, the results shown in the figure do not depend on parameter values other than the ones used for comparative statics: (λ, R).

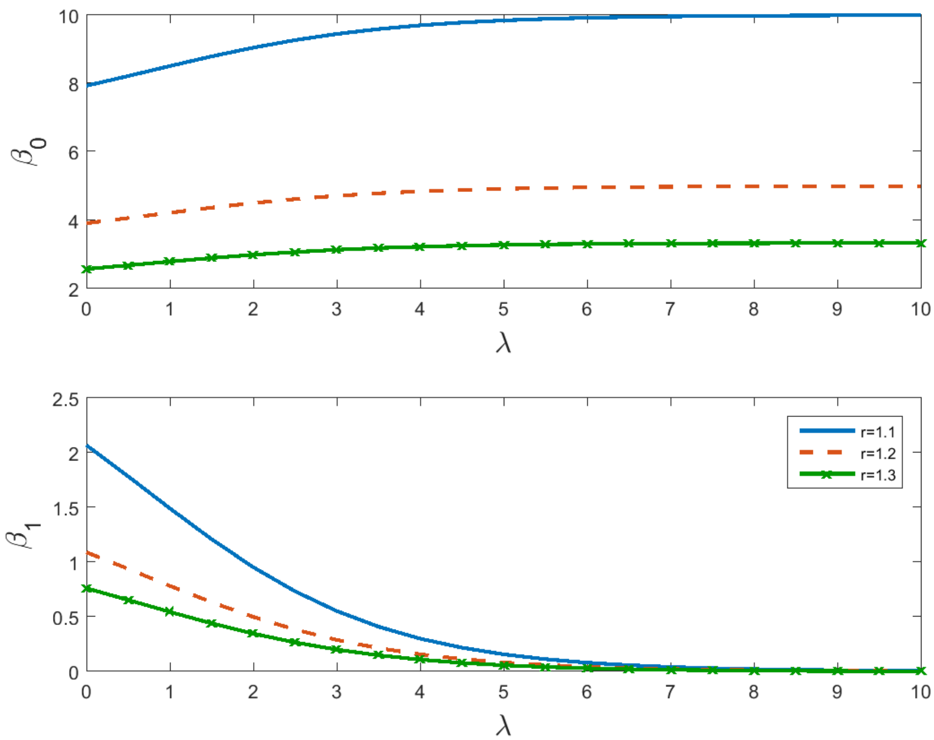


Figure 6: Comparative Statics: Sensitivity of Prices to Dividends for the q=2 Case.

The next proposition establishes that, as before, the demand of the young generation (decreases) increases, while the one of the old generation (increases) decreases, when current dividends (decrease) increase; and the opposite holds for the dividends last period.

Proposition 7.4. *For $\lambda > 0$: (1) $\frac{\partial x_t^t}{\partial d_t} > 0 > \frac{\partial x_t^{t-1}}{\partial d_t}$, and (2) $\frac{\partial x_t^t}{\partial d_{t-1}} < 0 < \frac{\partial x_t^{t-1}}{\partial d_{t-1}}$.*

Proof of Proposition 7.4. See Appendix D. $\square$

In our model, the young generation puts more weight on current dividends when forming beliefs, so when d_t increase, they young are "overly optimistic" relatively to the old generation. This effect contributes to the result (1) (and similar reasoning contributes to results (2)); however, this is not the only effect to consider. There additional effects due to the fact that the young are confronted with a different horizon investment.

In order to shed some light on the different effects, recall that in equation (30) we decomposed the demand for risky asset into two components: The myopic one and the dynamic one. For the particular case of $q = 2$ these decomposition yields $x_t^{t-1} = \tilde{x}_t^{t-1}$ and $x_t^t = \tilde{x}_t^t + \Delta_t^t$, where

$$\tilde{x}_t^{t-1} = \frac{\alpha\,(1-R)}{\gamma\,(1+\beta_0)^2\,\sigma^2} + \frac{(1+\beta_0)w(0,\lambda,1)+\beta_1-R\beta_0}{\gamma\,(1+\beta_0)^2\,\sigma^2}d_t + \frac{(1+\beta_0)(1-w(0,\lambda,1))-R\beta_1}{\gamma\,(1+\beta_0)^2\,\sigma^2}d_{t-1},$$

$$\tilde{x}_t^t = \frac{\alpha\,(1-R)}{\gamma R\,(1+\beta_0)^2\,\sigma^2} + \frac{1+\beta_0+\beta_1-R\beta_0}{\gamma R\,(1+\beta_0)^2\,\sigma^2}d_t + \frac{-R\beta_1}{\gamma R\,(1+\beta_0)^2\,\sigma^2}d_{t-1}$$

and

$$\Delta_t^t \equiv \frac{\alpha\,(1-R)+(\beta_1-R\beta_0)\,d_t - R\beta_1 d_{t-1}}{\gamma R\,(1+\beta_0)^2}\left(\frac{1}{s^2}-\frac{1}{\sigma^2}\right) + \frac{1}{\gamma R\,(1+\beta_0)}\left(\frac{m}{s^2}-\frac{d_t}{\sigma^2}\right)$$

where $s^2 = \sigma^2 \frac{(1+\beta_0)^2}{(1+\beta_0)^2+((1+\beta_0)w(0,\lambda,0)+\beta_1-R\beta_0)^2}$.

We focus first on understanding the changes in the myopic term. Let

$$\frac{\partial\left(\tilde{x}_t^t-\tilde{x}_t^{t-1}\right)}{\partial d_t} = \underbrace{\frac{(1+\beta_0)\,(1-w(0,\lambda,1))}{\gamma\,(1+\beta_0)^2\,\sigma^2}}_{\text{Beliefs Term}} + \underbrace{\frac{1+\beta_0+\beta_1-R\beta_0}{\gamma\,(1+\beta_0)^2\,\sigma^2}\left(-\frac{R-1}{R}\right)}_{\text{Discount Term}}$$

We refer to the first term as the *Beliefs Term*. This term is positive, and it reflects that an increase (decrease) in dividends makes young agents more optimistic (pessimistic) about the return of the risky asset than adult agents because the put more weight on recent realizations. This term is zero when both agents have the same belief formation (e.g. $w(0,\lambda,1) = 1$). The second term is the *Discount Term*, which is negative (see Lemma D.5 in Appendix). Even when agents share beliefs, young agents react less aggressively to a change in dividends (in their beliefs) because they discount the future more than old agents since $R > 1$.

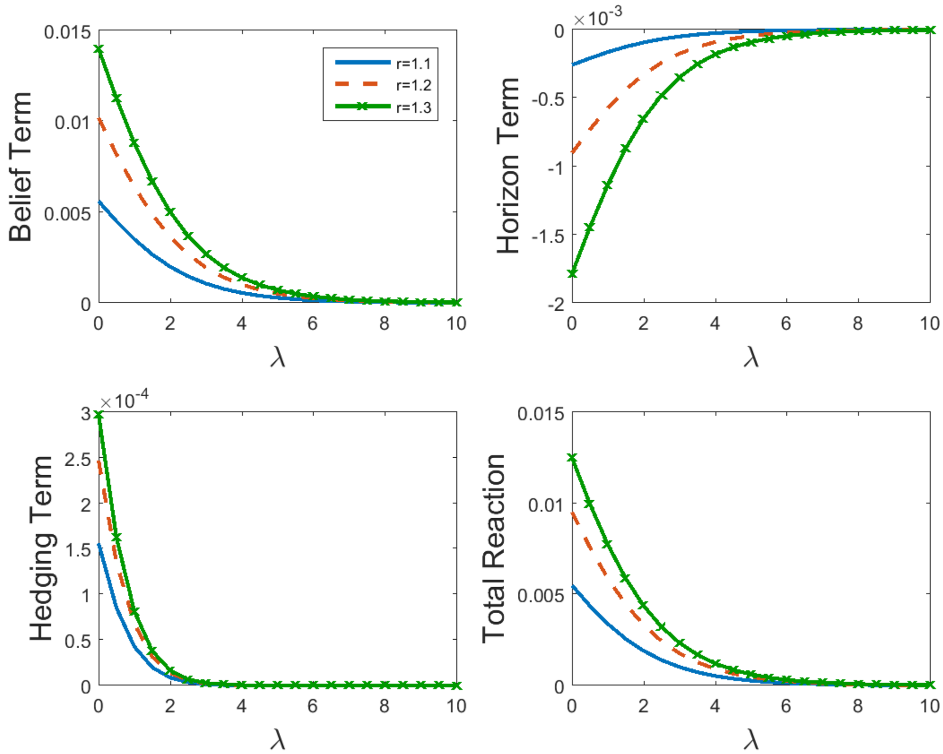


Figure 7: Comparative Statics: Sensitivity of Demands to Dividends for the q=2 Case.

Decomposition of $\frac{\partial\left(x_t^t - x_t^{t-1}\right)}{\partial d_t}$ into the Belief, Horizon, and Hedging Terms.

Regarding the hedging term, observe that

$$\frac{\partial \Delta_t^t}{\partial d_t} = \frac{\beta_1 - R\beta_0}{\gamma\left(1+\beta_0\right)^2 \sigma^2} \frac{1}{R}\left(\frac{\sigma^2}{s^2} - 1\right) - \frac{\left(1+\beta_0\right)}{\gamma\left(1+\beta_0\right)^2 \sigma^2}\left(\frac{l(1,1) l\,(0,1)}{\left(1+\beta_0\right)^2 R}\right)$$

Because the first term in the RHS is negative but the second term is positive (see the proof of Proposition 7.4), we can not pin down the sign of $\frac{\partial \Delta_t^t}{\partial d_t}$.

Therefore, even though the behavior $\frac{\partial(x_t^t - x_t^{t-1})}{\partial d_t}$ is affected by all these terms, we are able to show that the belief term dominates and thus the overall sign of $\frac{\partial(x_t^t - x_t^{t-1})}{\partial d_t}$ is positive.

In figure 7 we show the behavior of each of the terms for different values of (R, λ). Importantly, as λ increases, the "old" generation puts less weight to past dividends, and thus the discrepancy between the beliefs of the old and the young vanishes.

8 Conclusion

In this paper, we have proposed a simple OLG general equilibrium framework to study the effect of personal experiences of macroeconomic shocks on future economic outcomes such as the cross-section of asset holdings, asset prices, and market volatility. We have done so by incorporating the two main empirical features of experience effects, the over-weighing of lifetime experiences and recency bias, into the belief formation process of agents. We find that through our mechanism, macroeconomic shocks can have long-lasting effects on an economy, as suggested by Friedman and Schwartz (1963) and Blanchard (2012).

We highlight two channels through which shocks have long-lasting effects on economic outcomes. The first is the belief formation process, since all agents update their beliefs about the future after observing a given shock. The second is the cross-sectional heterogeneity in the population, since different experiences generate belief heterogeneity. Furthermore, we show that the demographic composition of an economy has important implications for the extent to which macroeconomic shocks can have long-lasting effects through the above described channels. Most importantly, we take our model predictions to the data, and find that they are consistent with empirical stylized facts on portfolio decisions and trade volume.

The results of this paper underline the importance of formally modeling the belief formation process of agents. This is not only relevant for improving our understanding of economic behavior, but also for effective policy making. We believe that the next step is two-fold. First, we need to continue improving our understanding of how agents form their beliefs about future economic outcomes. Second, it is important that these findings are formalized and incorporated to standard models to continue shaping our understanding of the way economies operate.

References

Adam, K., A. Marcet, and J. P. Nicolini (2012). Stock Market Volatility and Learning. *Journal of Finance*, forthcoming.

Agarwal, S., J. C. Driscoll, X. Gabaix, and D. Laibson (2013). Learning in the Credit Card Market. *working paper*.

Alesina, A. and N. Fuchs-Schundeln (2007). Good-bye Lenin (or Not?): The Effect of Communismon People s Preferences. *American Economic Review 97*, 1507 C1528.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015). X-CAPM: An extrapolative capital asset pricing model. *Journal of Financial Economics 115*(1), 1–24.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2016). Extrapolation and bubbles. Working Paper, National Bureau of Economic Research.

Barsky, R. and J. B. DeLong (1993). Why Does the Stock Market Fluctuate? *Quarterly Journal of Economics 108*(2), 291–311.

Blanchard, O. (2012). Sustaining a Global Recovery. *Finance and Development 46*(3), 9–12.

Brunnermeier, M. K. (2001). *Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding.* Oxford University Press on Demand.

Campbell, J. Y. (2008). Household Finance. *Journal of Finance 61*(4), 1553–1604.

Campbell, J. Y. and R. J. Shiller (1988). Stock Prices, Earnings, and Expected Dividends. *Journal of Finance 43*(3), 661–676.

Cecchetti, S. G., P. Lam, and N. C. Mark (2000). Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True? *American Economic Review 90*(4), 787–805.

Cogley, T. and T. J. Sargent (2008). The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression? *Journal of Monetary Economics 55*(3), 454–476.

Collin-Dufresne, P., M. Johannes, and L. A. Lochstoer (2014). Asset Pricing when'This Time is Different'. *Swiss Finance Institute Research Paper* (13-73), 14–8.

Delong, B. and L. Summers (2012). Fiscal Policy in a Depressed Economy. *Brookings Papers on Economic Activity 44*(1), 233–297.

Ehling, P., A. Graniero, and C. Heyerdahl-Larsen (2015). Asset Prices and Portfolio Choice with Learning from Experience. *working paper*, NYU.

Friedman, M. and A. J. Schwartz (1963). *A Monetary History of the United States.* Princeton, NJ: Princeton University Press.

Gabaix, X. (2011). Disasterization: A Simple Way to Fix the Asset Pricing Properties of Macroeconomic Models. *American Economic Review Papers and Proceedings 101*(3), 406–409.

Gabaix, X. (2012). Variable Rare Disasters: An Exactly Solved Framework For Ten Puzzles In Macro-Finance. *Quarterly Journal of Economics 127*, 645–700.

Giuliano, P. and A. Spilimbergo (2013). Growing Up in a Recession. *Review of Economic Studies 81*(2), 787–817.

Greenwood, R. and A. Shleifer (2014). Expectations of Returns and Expected Returns. *Review of Financial Studies*, 714–746.

Hertwig, R., G. Barron, E. U. Weber, and I. Erev (2004). Decisions from Experience and the Effect of Rare Events in Risky Choice. *Psychological Science 15*, 534 C539.

Jin, L. J. (2015). A speculative asset pricing model of financial instability. *Available at SSRN 2524762*.

Kaustia, M. and S. Knüpfer (2008). Do Investors Overweight Personal Experience? Evidence from IPO Subscriptions. *Journal of Finance 63*, 2679–2702.

LeRoy, S. F. (2005). Excess Volatility. *working paper*.

LeRoy, S. F. and R. D. Porter (1981). The present value relation: Tests based on implied variance bounds. *Econometrica 49*, 555–574.

Malmendier, U. and S. Nagel (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking? *Quarterly Journal of Economics 126*, 373–416.

Malmendier, U. and S. Nagel (2016). Learning from Inflation Experiences. *Quarterly Journal of Economics 131*, 53–87.

Malmendier, U. and L. S. Shen (2015). Experience Effects in Consumption. Working paper, UC-Berkeley.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates. *American Economic Journal: Applied Economics* (4), 1–29.

Schraeder, S. (2015). Information Processing and Non-Bayesian Learning in Financial Markets. *Review of Finance*, rfv002.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*, 421–436.

Simonsohn, U., N. Karlsson, G. Loewenstein, and D. Ariely (2008). The tree of experience in the forest of information: Overweighing experienced relative to observed information. *Games and Economic Behavior 62*(1), 263–286.

Timmermann, A. G. (1993). How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices. *Quarterly Journal of Economics 108*(4), 1135–1145.

Timmermann, A. G. (1996). Excess Volatility and Predictability of Stock Prices in Autoregressive Dividend Models with Learning. *The Review of Economic Studies 63*(4), 523–557.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 4157*, 1124–1131.

Vives, X. (2010). *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press.

Weber, E. U., U. Böckenholt, D. J. Hilton, and B. Wallace (1993). Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition 19*(5), 1151.

Woodford, M. (2013). Macroeconomic Analysis without the Rational Expectations Hypothesis. *Annual Review of Economics 5*(1), forthcoming.

Appendix A Proofs of Section 2

Proof of Lemma 2.1. Let $\Delta(k) \equiv w(k, \lambda, age) - w(k, \lambda, age')$ for all $k \in \{0, ..., age\}$. We need to show that $\exists k_0 \in \{0, ..., age'\}$ such that $\Delta(k) < 0$ for all $k \leq k_0$, and $\Delta(k) \geq 0$ for all $k > k_0$, with the last inequality holding strictly for some k.

For $k > age'$, $\Delta(k) > 0$ since $w(k, \lambda, age') \equiv 0$, and hence $\Delta(k) = w(k, \lambda, age) > 0$, for all $k \in \{age' + 1, ..., age\}$.

For $k \leq age'$, we note that

$$\Delta(k) > 0 \iff Q(k) := \frac{w(k, \lambda, age)}{w(k, \lambda, age')} > 1. \tag{33}$$

Hence, it remains to be shown that $\exists k_0 \in \{0, ..., age'\}$ such that $Q(k) < 1$ for all $k \leq k_0$, and $Q(k) \geq 1$ for all $k > k_0$. Since the normalizing constants used in the weights $w(k, \lambda, age)$ are independent of k (see the definition in (5)), we absorb them in a constant $c \in \mathbb{R}^+$ and rewrite

$$Q(k) = c \cdot \frac{(age + 1 - k)^\lambda}{(age' + 1 - k)^\lambda} = c \cdot \left[\frac{age + 1 - k}{age' + 1 - k}\right]^\lambda = c \cdot \alpha(k)^\lambda \;\; \forall k \in \{0, ..., age'\}. \tag{34}$$

The function $x \mapsto \alpha(x) = \frac{age+1-x}{age'+1-x}$ has derivative $\alpha'(x) = \frac{age-age'}{(age'+1-x)^2} > 0$ for $x \in [0, age' + 1)$, and hence $Q(\cdot)$ is strictly increasing over $\{0, ..., age'\}$. Thus, to complete the proof, we only have to show that $Q(k) < 1$ or, equivalently, $\Delta(k) < 0$ for some $k \in \{0, ..., age'\}$. We know that $\sum_{k=0}^{age} \Delta(k) = 0$ because $\sum_{k=0}^{age} w(k, \lambda, age) = \sum_{k=0}^{age'} w(k, \lambda, age') = 1$, and we also know that $\sum_{k=age'+1}^{age} \Delta(k) > 0$ since $\Delta(k) = w(k, \lambda, age) > 0$ for all $k \in \{age' + 1, ..., age\}$. Hence, it must be that $\Delta(k) < 0$ for some $k < age'$. □

Appendix B Proofs of Section 4

Proposition 4.1 directly follows from the following Lemma.

Lemma B.1. *Let $z \sim N(\mu, \sigma^2)$, then for any $a > 0$,*

$$x^* = \arg\max_x E[-\exp\{-axz\}] = \frac{\mu}{a\sigma^2}$$

and

$$\max_x E[-\exp -axz] = -\exp\left\{-\frac{1}{2}(\sigma a x^*)^2\right\} = -\exp\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right).$$

Proof of Lemma B.1. Since $z \sim N(\mu, \sigma^2)$, we can re-write the problem as follows:

$$\begin{aligned} x^* &= \arg\max_x -\exp\left(-axE[z] + \frac{1}{2}a^2x^2V[z]\right) \\ &= \arg\max_x \;\; ax\mu - \frac{1}{2}a^2x^2\sigma^2 \end{aligned}$$

From FOC, $x^* = \frac{\mu}{a\sigma^2}$. Plugging x^* in $-\exp\left(-ax^*\mu + \frac{1}{2}a^2(x^*)^2\sigma^2\right)$ the second result follows. $\square$

Proof of Proposition 4.2. We start from the proposed guess $p_t = \alpha + \beta_0 d_t + ... + \beta_{q-1} d_{t-q+1}$. From Lemma B.1, agents' demand for the risky asset is given by $x_t^n = \frac{E_t^n[s_{t+1}]}{\gamma V[s_{t+1}]}$. Plugging in our guess for prices, and for $\beta_0 \neq -1$, we obtain:

$$x_t^n = \frac{(1+\beta_0)\,\theta_t^n + \alpha + \beta_1 d_t + ... + \beta_q d_{t-q+1} - p_t R}{\gamma\,(1+\beta_0)^2\,\sigma^2} \tag{35}$$

By market clearing, $\frac{1}{q}\sum_{n=t-q+1}^{t} x_t^n = 1$, which implies that

$$\frac{(1+\beta_0)\,\frac{1}{q}\sum_{n=t-q+1}^{t}\theta_t^n}{\gamma\,(1+\beta_0)^2\,\sigma^2} + \frac{\alpha + \beta_1 d_t + ... + \beta_q d_{t-q+1} - p_t R}{\gamma\,(1+\beta_0)^2\,\sigma^2} = 1.$$

By straightforward algebra and the definition of θ_t^n, it follows that

$$(1+\beta_0)\,\frac{1}{q}\sum_{n=t-q+1}^{t}\left[\sum_{k=0}^{t-n} w\,(k,\lambda,t-n)\,d_{t-k}\right] + \left[\alpha - \gamma\,(1+\beta_0)^2\,\sigma^2\right] + \beta_1 d_t + ... + \beta_q d_{t-q+1} = p_t R.$$

Using the method of undetermined coefficients we find the expressions for α and the β's:

$$-\frac{\gamma\,(1+\beta_0)^2\,\sigma^2}{R-1} = \alpha \tag{36}$$

$$(1+\beta_0)\,\frac{1}{q}\sum_{n=t-q+1}^{t-k} w\,(k,\lambda,t-n) + \beta_{k+1} = \beta_k R \qquad \forall k \in \{0,1,...,q-1\} \tag{37}$$

$$0 = \beta_q R \tag{38}$$

As defined in Proposition 4.1, w_k is the average of the weights assigned to dividend d_{t-k} by each generation in the market, $w_k = \frac{1}{q}\sum_{n=t-q+1}^{t} w\,(k,\lambda,t-n)$. Given that a weight of zero is assigned to dividends that a generation did not observe, i.e., for $k > t-n$, we can rewrite $w_k = \frac{1}{q}\sum_{n=t-q+1}^{t-k} w\,(k,\lambda,t-n)$. Also using $\beta_q = 0$ we obtain:

$$(1+\beta_0)\,w_k + \beta_{k+1} = \beta_k R \qquad \forall k \in \{0,1,...,q-2\} \tag{39}$$

$$(1+\beta_0)\,w_{q-1} = \beta_{q-1} R \tag{40}$$

By solving this system of equations we obtain the expressions in the proposition. In particular, $(1+\beta_0)\,(w_{q-2} + w_{q-1}/R) = \beta_{q-2}R$ for $k = q-2$, $(1+\beta_0)\,(w_{q-3} + w_{q-2}/R + w_{q-1}/R^2) = \beta_{q-3}R$ for $k = q-3$, and so on, allow us to express (39) as

$$(1+\beta_0)\sum_{j=0}^{k-1} w_{q-(k-j)}/R^j = \beta_{q-k}R, \;\; for\ k = 2,...,q. \tag{41}$$

The last expression (41) implies, in particular, $\beta_0 = \frac{\sum_{j=0}^{q-1} w_j/R^j}{R-\sum_{j=0}^{q-1} w_j/R^j} = \frac{\sum_{j=0}^{q-1} w_j/R^{j+1}}{1-\sum_{j=0}^{q-1} w_j/R^{j+1}}$ (from plugging in $k = q$), which, plugged into (36) allows us to obtain the expression for α from (16) in Proposition 4.2. And expression (41) implies $\beta_k = \frac{\sum_{j=0}^{q-1-k} w_{k+j}/R^{j+1}}{1-\sum_{j=0}^{q-1} w_j/R^{j+1}}$ (from substituting k in (41) with $q-k$, and using the expression for β_0) as expressed in equation (17) of the Proposition. The latter also subsumes equation (40), solved for β_{q-1}, and the above formula for β_0 and hence holds for $k = 0, ...q-1$. $\square$

Proof of Proposition 4.3. For this proof, we will use equations (39) and (40). In addition, note that by construction, $w_k < w_{k-1}$ for $\lambda > 0$ since for all generations, $w(k, \lambda, age)$ is decreasing in k and more agents observe the realization of $d_{t-(k-1)}$ than d_{t-k}. Given this, it follows that since $\beta_0 > 0$ then $\beta_{q-1} > 0$ and

$$\beta_{q-1} = \frac{1}{R}[(1+\beta_0)\, w_{q-1}] < \frac{1}{R}[(1+\beta_0)\, w_{q-2} + \beta_{q-1}] = \beta_{q-2} \tag{42}$$

In addition, if $\beta_k < \beta_{k-1}$, then:

$$\beta_{k-1} = \frac{1}{R}[(1+\beta_0)\, w_{k-1} + \beta_k] < \frac{1}{R}[(1+\beta_0)\, w_{k-2} + \beta_{k-1}] = \beta_{k-2} \tag{43}$$

Thus, the proof that $\beta_k < \beta_{k-1}$ for all $k \in \{1, ..., q-1\}$ follows by induction. $\square$

Proof of Lemma 4.1. A. β_0 is increasing in λ. Let $G_q(\lambda) = \sum_{j=0}^{q-1} c^{j+1} w_j$ where $c = \frac{1}{R}$. Thus $\beta_0 = \frac{G_q(\lambda)}{1-G_q(\lambda)}$, and it suffices to show that $G_q'(\lambda) > 0$, $\forall q > 0$, $\forall \lambda > 0$. After some algebra, the terms in $G_q(\cdot)$ can be re-organized as follows:

$$G_q(\lambda) = \sum_{a=0}^{q-1} \frac{1}{q} \sum_{j=0}^{a} c^{j+1} w(j, \lambda, a) \tag{44}$$

Note that for any $a \in \{0, ..., q-1\}$: (i) $\sum_{j=0}^{a} w(j, \lambda, a) = 1$ and (ii) for any λ_1, λ_2 such that $\lambda_1 > \lambda_2 > 0$, $\sum_{j=k}^{a} w(j, \lambda_1, a) < \sum_{j=k}^{a} w(j, \lambda_2, a)$. Thus, the weight distribution given by λ_1 first-order stochastically dominates the weight distribution given by λ_2. Since $c > c^2 > c^3 > ... > c^{q-1}$ then stochastic dominance implies that for all $a \in \{0, ..., q-1\}$, $\sum_{j=0}^{a} c^{j+1} w(j, \lambda_1, a) > \sum_{j=0}^{a} c^{j+1} w(j, \lambda_2, a)$, and thus $G_q(\lambda_1) > G_q(\lambda_2)$.

$\square$

Proof of Proposition 4.4. From Proposition 4.1, for any k and t,

$$\frac{\partial x_t^k}{\partial d_t} = \frac{1}{V[s_{t+1}]} \left((1+\beta_0) \frac{\partial \theta_t^k}{\partial d_{t-j}} - R\beta_0 \right).$$

It follows that $\frac{\partial \theta_t^k}{\partial d_{t-j}} = w(j, \lambda, k)$. Hence, it suffices to show that $w(j, \lambda, t-n) < w(j, \lambda, t-$

$(n+1))$. First lets consider the case for $j=0$. For any age,

$$w(0,\lambda,age)=\frac{(age+1)^{\lambda}}{\sum_{k=0}^{age}(age+1-k)^{\lambda}}=\left(\sum_{k=0}^{age}\left(\frac{age+1-k}{age+1}\right)^{\lambda}\right)^{-1}=\left(1+\sum_{k=0}^{age-1}\left(\frac{age-k}{age+1}\right)^{\lambda}\right)^{-1},$$

and

$$\begin{aligned}w(0,\lambda,age+1)=&\frac{(age+2)^{\lambda}}{\sum_{k=0}^{age}(age+2-k)^{\lambda}}\\=&\left(\sum_{k=0}^{age+1}\left(\frac{age+2-k}{age+2}\right)^{\lambda}\right)^{-1}\\=&\left(1+\sum_{k=0}^{age}\left(\frac{age+1-k}{age+2}\right)^{\lambda}\right)^{-1},\end{aligned}$$

So to establish $w(0,\lambda,age+1)<w(0,\lambda,age)$ with $age=t-n-1$, it suffices to show that

$$\sum_{k=0}^{age-1}\left(\frac{age-k}{age+1}\right)^{\lambda}\quad<\quad\sum_{k=0}^{age}\left(\frac{age+1-k}{age+2}\right)^{\lambda}$$

We note that in the second expression, there are $age+1$ terms whereas in the first one there are age. We show that $\frac{age-l}{age+1}<\frac{age+1-l}{age+2}$ for any $l=0,...,age$. The inequality holds iff $(age-l)(age+2)<(age+1-l)(age+1)$ iff $age^2+2age-l(age+2)<age^2+(1-l)age+age+(1-l)$ iff $-2l<1-l$ for holds for any $l\geq 0$.

Therefore, $w(0,\lambda,t-n)<w(0,\lambda,t-(n+1))$. From Lemma 2.1, there exists a j_0 such that $w(j,\lambda,t-n)<w(j,\lambda,t-(n+1))$ for all $j\in\{0,...,j_0\}$ and $w(j,\lambda,t-n)\geq w(j,\lambda,t-(n+1))$ for the rest of the j's.

□

The proof of Proposition 4.5 relies on the following First order stochastic dominance result.

Lemma B.2. *Let $F(m,a)\equiv\sum_{j=0}^{m}w(j,\lambda,a)$. If $j\mapsto w(j,\lambda,a)-w(j,\lambda,a')$ for $a'<a$ is increasing in j, then $F(m,a)\leq F(m,a')$ for all $m\in\{0,...,a\}$.*

Proof. From Lemma 2.1, we know that there exists a unique j_0 where $w(j_0,\lambda,a')=w(j_0,\lambda,a)$. Thus, for $m\leq j_0$, the result is true because $w(j,\lambda,a')>w(j,\lambda,a)$ for all $j\in\{0,...m\}$. For $m>j_0$, the result follows from the fact that $w(j,\lambda,a')<w(j,\lambda,a)$ for all $j\in\{m,...a\}$ and $F(a,a)=F(a',a')=1$. □

Proof of Proposition 4.5. We first introduce some notation. For any $j=t-n-k,...,t-n$, let $w(j,\lambda,n-t-k)=0$; i.e., we define to be zero the weights of generation $n+k$ for time periods before they were born. Thus, $\sum_{j=0}^{t-n-k}w(j,\lambda,t-n-k)d_{t-j}=\sum_{j=0}^{t-n}w(j,\lambda,t-n)d_{t-j}$. In addition, we note that $(w(j,\lambda,t-n-k))_{j=0}^{t-n}$ and $(w(j,\lambda,t-n))_{j=0}^{t-n}$ are sequences of positive weights that add to one. Let for any $m = 0, ..., t-n$,

$$F(m, t-n-k) = \sum_{j=0}^{m} w(j, \lambda, t-n-k), \ \textit{and} \ F(m, t-n) = \sum_{j=0}^{m} w(j, \lambda, t-n).$$

It follows that the quantities above, as functions of m, are non-decreasing and $F(t-n, t-n-k) = F(t-n, t-n) = 1$ and $F(-1, t-n-k) = F(-1, t-n) = 0$. Moreover, $F(m+1, t-n-k) - F(m, t-n-k) = w(m+1, \lambda, t-n-k)$ and $F(m+1, t-n) - F(m, t-n) = w(m+1, \lambda, t-n)$.

By these observations and our previous derivations for $\xi(n, k, t)$, it follows that,

$$\begin{aligned}
&\xi(n,k,t) \\
=&\frac{\sum_{m=0}^{t-n}(F(m,t-n)-F(m-1,t-n))d_{t-m}-\sum_{m=0}^{t-n}(F(m,t-n-k)-F(m-1,t-n-k))d_{t-m}}{\gamma(1+\beta_0)\sigma^2} \\
=&\frac{F(0,t-n)d_t+(F(1,t-n)-F(0,t-n))d_{t-1}+...+(1-F(t-n-1,t-n))d_n}{\gamma(1+\beta_0)\sigma^2} \\
&-\frac{F(0,t-n-k)d_t+(F(1,t-n-k)-F(0,t-n-k))d_{t-1}+...+(1-F(t-n-1,t-n-k))d_n}{\gamma(1+\beta_0)\sigma^2} \\
=&\frac{(d_t-d_{t-1})F(0,t-n)+(d_{t-1}-d_{t-2})F(1,t-n)+...+(d_{n-1}-d_n)F(1,t-n)+d_n}{\gamma(1+\beta_0)\sigma^2} \\
&-\frac{(d_t-d_{t-1})F(0,t-n-k)+(d_{t-1}-d_{t-2})F(1,t-n-k)+...+(d_{n-1}-d_n)F(1,t-n-k)+d_n}{\gamma(1+\beta_0)\sigma^2} \\
=&\frac{\sum_{j=0}^{t-n+1}(d_{t-j}-d_{t-j-1})(F(j,t-n)-F(j,t-n-k))}{\gamma(1+\beta_0)\sigma^2}.
\end{aligned}$$

By assumption $d_{t-j} - d_{t-j-1} \geq 0$ for all $j = 0, ..., t-n+1$ in the case that dividends are non-decreasing. Thus, it suffices to show that $F(j, t-n) \leq F(j, t-n-k)$ for all $j = 0, ..., t-n+1$. To show this, we note that by Lemma 2.1 the hypothesis in Lemma B.2 holds. Thus, the result follows from applying the latter lemma with $a = t-n > t-n-k = a'$ and $j \in \{0, ..., t-n\}$.

Note that if the weights are non-increasing, then $d_{t-j} - d_{t-j-1} \leq 0$. Therefore, the sign of $\xi(n, k, t)$ changes accordingly. $\square$

Appendix C Proofs of Subsection 4.4

Proof of Proposition 4.6. By Proposition 4.1 and 4.2, it follows that

$$\begin{aligned}
x_t^n &= \frac{1}{\gamma\sigma^2\left(1+\beta_0\right)^2}\left(\alpha_0+(1+\beta_0)\theta_t^n+\sum_{k=1}^{q-1}\beta_k d_{t+1-k}-R\left(\alpha_0+\sum_{k=0}^{q-1}\beta_k d_{t-k}\right)\right) \\
&= \frac{1}{\gamma\sigma^2\left(1+\beta_0\right)^2}\left(\alpha_0(1-R)+(1+\beta_0)\theta_t^n-R\beta_0 d_t+\sum_{k=1}^{q-1}\beta_k(d_{t+1-k}-Rd_{t-k})\right)
\end{aligned}$$

and thus for $n \in [t-1, ..., t-q-1]$

$$x_t^n - x_{t-1}^n = \frac{(1+\beta_0)(\theta_t^n - \theta_{t-1}^n) + \mathcal{T}(d_{t:t-q})}{\gamma\sigma^2 (1+\beta_0)^2}$$

where $\mathcal{T}(d_{t:t-q}) \equiv \sum_{k=1}^{q-1} \beta_k (d_{t+1-k} - d_{t-k} - R(d_{t-k} - d_{t-1-k})) - R\beta_0(d_t - d_{t-1})$ is not cohort specific, that is, does not depend on n. Since $x_t^t - x_{t-1}^t = x_t^t$ and $x_t^{t-q} - x_{t-1}^{t-q} = -x_{t-1}^{t-q}$ we have instead:

$$x_t^t - x_t^{t-q} = \frac{(1+\beta_0)(\theta_t^t - \theta_{t-1}^{t-q}) + \mathcal{T}(d_{t:t-q})}{\gamma\sigma^2 (1+\beta_0)^2}$$

In addition, from market clearing and after some algebra, we have that:

$$\frac{1}{q}\left(\sum_{n=t-q}^{t} x_t^n - x_{t-1}^n\right) = \frac{1}{q}\sum_{n=t-q}^{t} \frac{(1+\beta_0)(\theta_t^n - \theta_{t-1}^n) + \mathcal{T}(d_{t:t-q})}{\gamma\sigma^2 (1+\beta_0)^2} = 0$$

$$\Rightarrow \quad \frac{1}{q}\sum_{n=t-q}^{t} \frac{(1+\beta_0)(\theta_t^n - \theta_{t-1}^n)}{\gamma\sigma^2 (1+\beta_0)^2} = -\frac{1}{q}\sum_{n=t-q}^{t} \frac{\mathcal{T}(d_{t:t-q})}{\gamma\sigma^2 (1+\beta_0)^2} = -\frac{\mathcal{T}(d_{t:t-q})}{\gamma\sigma^2 (1+\beta_0)^2}$$

where $\theta_{t-1}^t = 0$ and $\theta_t^{t-q} = 0$. As a result, we can express the change in individual demands as follows:

$$x_t^n - x_{t-1}^n = \chi\left[\left(\theta_t^n - \theta_{t-1}^n\right) - \frac{1}{q}\sum_{n=t-q}^{t}\left(\theta_t^n - \theta_{t-1}^n\right)\right]$$

where $\chi \equiv \frac{1}{\gamma\sigma^2(1+\beta_0)}$. From our TV formula, we focus on $\frac{TV^2}{\chi}$:

$$\frac{TV^2}{\chi} = \frac{1}{q}\sum_{n=t-q}^{t}\left[\left(\theta_t^n - \theta_{t-1}^n\right) - \frac{1}{q}\sum_{n=t-q}^{t}\left(\theta_t^n - \theta_{t-1}^n\right)\right]^2$$

$\square$

Appendix D Proofs for Section 7

To establish the results in this section we need the following lemmas (the proofs are relagated the end of this section).

Lemma D.1. *Suppose $z \sim N(\mu, \sigma^2)$, then for any $A, B \in \mathbb{R}$ and $C \geq 0$, $z \mapsto K^{-1}\exp\{-A - Bz - Cz^2\}\phi(z; \mu, \sigma^2)$ is Gaussian with mean $m \equiv -\Sigma^2 B + \Sigma^2\sigma^{-2}\mu$ and $\Sigma^2 \equiv \frac{\sigma^2}{2C\sigma^2+1}$, where*

$$K = E_{N(\mu,\sigma^2)}[\exp\{-A - Bz - Cz^2\}] = \frac{1}{\sqrt{2\sigma^2 C + 1}}\exp\{-(A + 0.5\sigma^{-2}\mu^2) + \frac{m^2}{2\Sigma^2}\}$$

Lemma D.2. *Demands for the risky asset in the last two period of an agent's life are given by: $x_t^{t-q} = 0$ and $x_t^{t-q+1} = \frac{E_t^{t-q+1}[s_{t+1}]}{\gamma\sigma_*^2}$, $\forall t \in \mathbb{Z}, q \geq 1$.*

Lemma D.3. *Let $z \sim N(\mu, \sigma^2)$. Let $A, B \in \mathbb{R}$ and $C \geq 0$, and $z \mapsto h(z) \equiv f + ez$ for any $e, f \in \mathbb{R}$. Then*

$$\max_x E[-\exp\{-A - Bz - Cz^2\}\exp\{-axh(z)\}] = -\frac{1}{\sqrt{2\sigma^2 C + 1}}\exp\left[-A - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} - \frac{m^2}{s^2}\right)\right]\exp\left[-\frac{1}{2}\frac{\tilde{\mu}(m,s^2)^2}{\tilde{\sigma}^2(m,s^2)}\right]$$

$$\arg\max_x E[-\exp\{-A - Bz - Cz^2\}\exp\{-axh(z)\}] = \frac{\tilde{\mu}(m,s^2)}{a\tilde{\sigma}^2(m,s^2)}$$

with $m = s^2\left[\sigma^{-2}\mu - B\right]$, $s^2 = \frac{\sigma^2}{2C\sigma^2+1}$, $\tilde{\mu}\left(m,s^2\right) = E_{N(m,s^2)}\left[h(z)\right]$, $\sigma^2\left(m,s^2\right) = V_{N(m,s^2)}\left[h(z)\right]$.

Let $\beta(k) = \beta_{k+1} - r\beta_k$ for $k \in \{0, ..., K-1\}$ and $\beta(K) = -r\beta_K$.

Lemma D.4. *Suppose $p_t = \alpha + \sum_{k=0}^{K}\beta_k d_{t-k}$ with $\beta_0 \neq -1$. Then the demand for risky assets of any cohort alive at time t is an affine function of past dividends, where the coefficients associated with a given dividend will depend on the agent's age, age. That is,*

$$x_t^{t-age} = \delta(age) + \sum_{k=0}^{K}\delta_k(age)d_{t-k}, \;\; for\ age \in \{0, ..., q\} \tag{45}$$

with

$$\delta(q) = \delta_k(q) = 0, \;\; \forall k \in \{0, ..., K\} \tag{46}$$

$$\delta(q-1) = \frac{\alpha(1-R)}{\gamma((1+\beta_0)\sigma)^2}, \;\; \delta_k(q-1) = \frac{(1+\beta_0)w(k,\lambda,q-1) + \beta(k)}{\gamma((1+\beta_0)\sigma)^2}, \;\; \forall k \in \{0, ..., q-1\} \tag{47}$$

$$\delta_k(q-1) = \frac{\beta(k)}{\gamma((1+\beta_0)\sigma)^2}, \;\; \forall k \in \{q, ..., K\}, \tag{48}$$

and for $age \in \{0, ..., q-2\}$,

$$\delta(age) = \frac{\alpha(1-R) - s_{age}^2(1+\beta_0)\delta_0(age+1)\delta(age+1)(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}, \tag{49}$$

$$\delta_k(age) = \frac{(1+\beta_0)s_{age}^2(\sigma^{-2}w(k,\lambda,age) - [(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\delta_{k+1}(age+1)\delta_0(age+1)]) + \beta(k)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2} \tag{50}$$

$$k \in \{0, ..., q-1\}, \tag{51}$$

$$\delta_k(age) = \frac{-(1+\beta_0)s_{age}^2[(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\delta_{k+1}(age+1)\delta_0(age+1)] + \beta(k)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}, \ k \in \{q, ..., K-1\} \tag{52}$$

$$\delta_K(age) = \frac{\beta(K)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}, \tag{53}$$

and $s_{q-1} = \sigma$ *and* $s_{age}^2 \equiv \frac{\sigma^2}{(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2(\delta_0(age+1))^2\sigma^2+1}$

The expressions for $b_j, b_j(k)$ and c_j for $j \in \{0, ..., q-1\}$ are:

$$\begin{aligned} b_j \equiv& (R^{q-1-j}\gamma)^2((1+\beta_0)\sigma_j)^2\delta(j)\delta_0(j) \\ b_j(k) \equiv& \delta_k(j)\delta_0(j)(R^{q-1-j}\gamma)^2((1+\beta_0)\sigma_j)^2 \end{aligned}$$

and, $c_{q-1} = 1$ and

$$c_{j-1} = 0.5(R^{q-1-(j+1)}\gamma)(1+\beta_0)\sigma_{j+1}\delta_0(j+1)$$

for $j \in \{0, ..., q-2\}$.

Proof of Proposition 7.1. By lemma B.1,

$$x_{t_q}^t = \frac{E_{N(m_{q-1},\sigma_{q-1}^2)}[s_{t+q}]}{\gamma V_{N(m_{q-1},\sigma_{q-1}^2)}[s_{t+q}]}$$

with $m_{q-1} = \theta_{t_q}^t$ and $\sigma_{q-1} = \sigma$, and

$$V_{t_q}^t = -\exp\{-0.5\left((1+\beta_0)\sigma\gamma x_{t_q}^t\right)^2\}.$$

By lemma D.4, $x_{t_q}^t$ is affine in $d_{t_q-K:t_q}$ and thus $V_{t_q}^t = -\exp\{-A - Bd_{t_q} - C(d_{t_q})^2\}$ where A, B and C depend on primitives and on $d_{t_q-K:t+q-2}$, in particular B is affine in $d_{t_q-K:t_q-1}$

and C is constant with respect to $d_{t_q-K:t_q}$:

$$\begin{aligned}
C \equiv& \frac{1}{2}\gamma^2((1+\beta_0)\sigma_{q-1})^2 \left(\delta_0(q-1)\right)^2 \\
B \equiv& \gamma^2((1+\beta_0)\sigma_{q-1})^2 \left(\delta(q-1) + \sum_{j=1}^{K} \delta_k(q-1) d_{t_q-j}\right) \delta_0(q-1) \\
A \equiv& \frac{1}{2}\gamma^2((1+\beta_0)\sigma_{q-1})^2 \left(\delta(q-1) + \sum_{j=1}^{K} \delta_k(q-1) d_{t_q-j}\right)^2 .
\end{aligned}$$

(see Lemma D.4 for the expressions for $\delta(q-1)$ and $(\delta_k(q-1))_{k=1}^K$).

At time $t+q-2$, by equation (26),

$$x^t_{t+q-2} = \arg\max_{x\in\mathbb{R}} E^t_{t+q-2}\left[V^t_{t_q}(d_{t_q-K:t_q}) \exp\left(-\gamma s_{t_q} x\right)\right]$$

where the expectation is taken with respect $N(\theta^t_{t+q-2}, \sigma^2)$. Hence, by lemma D.1, this problem can be cast as

$$x^t_{t+q-2} = \arg\max_{x\in\mathbb{R}} E_{N(m_{q-2},\sigma_{q-2})}\left[-\exp\left(-R\gamma s_{t_q} x\right)\right]$$

where $m_{q-2} = \sigma_{q-2}(\frac{\theta^t_{t+q-2}}{\sigma^2} - B)$ and $\sigma^2_{q-2} = \frac{\sigma^2}{2C\sigma^2+1}$. Hence, by lemma B.1

$$x^t_{t+q-2} = \frac{E_{N(m_{q-2},\sigma^2_{q-2})}[s_{t_q}]}{\gamma R V_{N(m_{q-2},\sigma^2_{q-2})}[s_{t_q}]},$$

Also, by lemma B.1, $V^t_{t+q-2} = -\exp\{-0.5\left(V_{N(m_{q-2},\sigma^2_{q-2})}[s_{t_q}] R\gamma x^t_{t+q-2}\right)^2\}$. By lemma D.4, x^t_{t+q-2} is affine and thus $V^t_{t+q-2} = -\exp\{-A - Bd_{t+q-2} - C(d_{t+q-2})^2\}$ where A, B and C depend on primitives and on $d_{t+q-2-K:t+q-3}$, in particular B is affine in $d_{t+q-2-K:t+q-3}$ and C is constant with respect to $d_{t_q-K:t_q}$:

$$\begin{aligned}
C \equiv& \frac{1}{2}(R\gamma)^2((1+\beta_0)\sigma_{q-2})^2 \left(\delta_0(q-2)\right)^2 \\
B \equiv& (R\gamma)^2((1+\beta_0)\sigma_{q-2})^2 \left(\delta(q-2) + \sum_{j=1}^{K} \delta_k(q-2) d_{t+q-2-j}\right) \delta_0(q-2) \\
A \equiv& \frac{1}{2}(R\gamma)^2((1+\beta_0)\sigma_{q-2})^2 \left(\delta(q-2) + \sum_{j=1}^{K} \delta_k(q-2) d_{t+q-2-j}\right)^2 .
\end{aligned}$$

(observe that the A and B and C are not the same as the previous ones; the expressions for $\delta(q-2)$ and $(\delta_k(q-2))_{k=1}^K$ can be found in the statement of lemma D.4).

The result for $j \in \{0, ..., q-3\}$ follows by iteration. $\square$

Proof of Proposition 7.2. Market Clearing and Lemma D.4 imply that, for all $k \in \{0, ..., K\}$,

$$\sum_{age=0}^{q-1} \delta_k(age) = 0 \tag{54}$$

and

$$\sum_{age=0}^{q-1} \delta(age) = q.$$

For $k = K$, it follows from equations 48 and 53

$$\sum_{age=0}^{q-1} \delta_K(age) = \beta(K) \left(\sum_{age=0}^{q-1} \frac{1}{R^{q-1-age}\gamma((1+\beta_0)s_{age})^2} + \frac{1}{\gamma((1+\beta_0)\sigma)^2} \right)$$

therefore $\beta(K) = 0$ which implies that $\beta_K = 0$ and $\beta(K-1) = -R\beta_{K-1}$ and $\delta_K(age) = 0$ for any age.

For $k = K-1$, by equations 48 and 52

$$\sum_{age=0}^{q-1} \delta_{K-1}(age) = \beta(K-1) \left(\sum_{age=0}^{q-2} \frac{1}{R^{q-1}\gamma((1+\beta_0)s_{age})^2} + \frac{1}{\gamma((1+\beta_0)\sigma)^2} \right)$$

and thus $\beta(K-1) = 0$ which implies that $\beta_{K-1} = 0$ and $\beta(K-2) = -R\beta_{K-2}$ and $\delta_{K-1}(age) = 0$ for any age.

By induction, for any $k \in \{q, ..., K-2\}$, taking $\beta_{k+1} = 0$, it follows by equations 48 and 52, that

$$\sum_{age=0}^{q-1} \delta_k(age) = \beta(k) \left(\sum_{age=0}^{q-2} \frac{1}{R^{q-1-age}\gamma((1+\beta_0)s_{age})^2} + \frac{1}{\gamma((1+\beta_0)\sigma)^2} \right)$$

and thus $\beta(k) = 0$ which implies $\beta_k = 0$ and $\beta(k-1) = -R\beta_{k-1}$ and $\delta_k(age) = 0$ for any $age \in \{q, ..., K\}$.

$\square$

Proof of Lemma 7.1. By Proposition 7.1, we have the following demands:

$$x_t^{t-2} = 0 \tag{55}$$

$$x_t^{t-1} = \frac{E_t^{t-1}\left[s_{t+1}\right]}{\gamma R\left(1+\beta_0\right)\sigma^2} = \frac{\alpha\left(1-R\right)+l\left(0,1\right)d_t+l\left(1,1\right)d_{t-1}}{\gamma\left(1+\beta_0\right)^2\sigma^2} \tag{56}$$

$$x_t^{t} = \frac{E_{\Phi(m,s^2)}\left[s_{t+1}\right]}{\gamma R\left(1+\beta_0\right)s^2} = \frac{\alpha\left(1-R\right)+\left(\beta_1-R\beta_0\right)d_t-R\beta_1 d_{t-1}+\left(1+\beta_0\right)m}{\gamma R\left(1+\beta_0\right)^2 s^2} \tag{57}$$

where $l(0,1) \equiv (1+\beta_0)w(0,\lambda,0)+\beta_1-R\beta_0$, $l(1,1) \equiv (1+\beta_0)w(1,\lambda,0)-R\beta_1$,

$$m = \frac{s^2}{\sigma^2}\left[d_t-\sigma^2 B_{t+1}\left(1\right)\right]$$

$$s^2 = \frac{\sigma^2}{2C\left(1\right)\sigma^2+1},$$

and

$$B_{t+1}\left(1\right) = \frac{\alpha\left(1-R\right)l\left(0,1\right)}{\left(1+\beta_0\right)^2\sigma^2}+\frac{l(1,1)l\left(0,1\right)}{\left(1+\beta_0\right)^2\sigma^2}d_t$$

$$C\left(1\right) = \frac{l\left(0,1\right)^2}{\left(1+\beta_0\right)^2\sigma^2}$$

Therefore:

$$m = \frac{s^2}{\sigma^2}\left[d_t-\frac{\alpha\left(1-R\right)l\left(0,1\right)}{\left(1+\beta_0\right)^2}-\frac{l(1,1)l\left(0,1\right)}{\left(1+\beta_0\right)^2}d_t\right] = \frac{s^2}{\sigma^2}\left[-\frac{\alpha\left(1-R\right)l\left(0,1\right)}{\left(1+\beta_0\right)^2}+\left(1-\frac{l(1,1)l\left(0,1\right)}{\left(1+\beta_0\right)^2}\right)d_t\right]$$

$$s^2 = \frac{\sigma^2}{2\frac{l(0,1)^2}{(1+\beta_0)^2\sigma^2}\sigma^2+1} = \frac{\left(1+\beta_0\right)^2}{l\left(0,1\right)^2+\left(1+\beta_0\right)^2}\sigma^2.$$

Plugging this in the expression for x_t^t, it follows that

$$\begin{aligned} x_t^t =& \frac{\alpha\left(1-R\right)+\left(\beta_1-R\beta_0\right)d_t-R\beta_1 d_{t-1}+\left(1+\beta_0\right)\frac{s^2}{\sigma^2}\left[-\frac{\alpha(1-R)l(0,1)}{(1+\beta_0)^2}+\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)d_t\right]}{\gamma R\left(1+\beta_0\right)^2 s^2} \\ =& \frac{\alpha\left(1-R\right)\left[1-\frac{s^2}{\sigma^2}\frac{l(0,1)}{(1+\beta_0)}\right]+\left[\beta_1-R\beta_0+\left(1+\beta_0\right)\frac{s^2}{\sigma^2}\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)\right]d_t-R\beta_1 d_{t-1}}{\gamma R\left(1+\beta_0\right)^2 s^2}. \end{aligned}$$

By Market clearing:

$$\begin{aligned}
1 =& \frac{1}{2}\left(\frac{\alpha(1-R)+l(0,1)\,d_t+l(1,1)\,d_{t-1}}{\gamma(1+\beta_0)^2\sigma^2}\right)\\
&+\frac{1}{2}\left(\frac{\alpha(1-R)\left[1-\frac{s^2}{\sigma^2}\frac{l(0,1)}{(1+\beta_0)}\right]+\left[\beta_1-R\beta_0+\frac{s^2}{\sigma^2}(1+\beta_0)\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)\right]d_t-R\beta_1 d_{t-1}}{\gamma R(1+\beta_0)^2 s^2}\right)\\
=& \frac{1}{2}\left(\frac{\alpha(1-R)+l(0,1)\,d_t+l(1,1)\,d_{t-1}}{\gamma(1+\beta_0)^2\sigma^2}\right)\\
&+\frac{1}{2}\left(\frac{\alpha(1-R)\frac{\sigma^2}{s^2}\left[1-\frac{s^2}{\sigma^2}\frac{l(0,1)}{(1+\beta_0)}\right]+\left[\frac{\sigma^2}{s^2}(\beta_1-R\beta_0)+(1+\beta_0)\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)\right]d_t-\frac{\sigma^2}{s^2}R\beta_1 d_{t-1}}{\gamma R(1+\beta_0)^2\sigma^2}\right),
\end{aligned}$$

which implies

$$\begin{aligned}
2\gamma(1+\beta_0)^2\sigma^2 =& \left(\alpha(1-R)+l(0,1)\,d_t+l(1,1)\,d_{t-1}\right)\\
&+\frac{1}{R}\left[\alpha(1-R)\frac{\sigma^2}{s^2}\left[1-\frac{s^2}{\sigma^2}\frac{l(0,1)}{(1+\beta_0)}\right]\right]\\
&+\frac{1}{R}\left[\left[\frac{\sigma^2}{s^2}(\beta_1-R\beta_0)+(1+\beta_0)\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)\right]d_t-\frac{\sigma^2}{s^2}R\beta_1 d_{t-1}\right]\\
=& \alpha(1-R)\frac{1}{R}\left[R+\frac{\sigma^2}{s^2}-\frac{l(0,1)}{(1+\beta_0)}\right]\\
&+\left[l(0,1)+\frac{1}{R}\frac{\sigma^2}{s^2}(\beta_1-R\beta_0)+\frac{1}{R}(1+\beta_0)\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)\right]d_t+\left[l(1,1)-\frac{\sigma^2}{s^2}\beta_1\right]d_{t-1}.
\end{aligned}$$

Therefore $\{\alpha,\beta_0,\beta_1\}$ solve the following system of equations:

$$0=\alpha(1-R)\left[R+\frac{\sigma^2}{s^2}-\frac{l(0,1)}{1+\beta_0}\right]-2R\gamma(1+\beta_0)^2\sigma^2 \tag{58}$$

$$0=l(0,1)+\frac{1}{R}\frac{\sigma^2}{s^2}(\beta_1-R\beta_0)+\frac{1}{R}(1+\beta_0)\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right) \tag{59}$$

$$0=l(1,1)-\frac{\sigma^2}{s^2}\beta_1 \tag{60}$$

where $l(0,1)\equiv[(1+\beta_0)w(0,\lambda,0)+\beta_1-R\beta_0]$ and $l(1,1)\equiv[(1+\beta_0)w(1,\lambda,0)-R\beta_1]$. $\square$

Proof of Proposition 7.3. Throughout the proof, let $w_0\equiv w(0,\lambda,0)$.

We know from Lemma 7.1 that $\{\alpha,\beta_0,\beta_1\}$ solve the system of equations given by (59) and (60) and 58.

STEP 1. By equation (58),

$$2R\gamma\left(1+\beta_0\right)^2\sigma^2=\alpha\left(1-R\right)\left[R+\frac{\sigma^2}{s^2}-\frac{\left[(1+\beta_0)w(0,\lambda,0)+\beta_1-R\beta_0\right]}{1+\beta_0}\right].$$

We note that $R>1\geq w(0,\lambda,0)$, thus, if $0<\beta_1<R\beta_0$ and $1+\beta_0>0$, then $\left[R+\frac{\sigma^2}{s^2}-\frac{[(1+\beta_0)w(0,\lambda,0)+\beta_1-R\beta_0]}{1+\beta_0}\right]>0$ and $\alpha\leq 0$.

STEP 2. We show that if $1+\beta_0>0$, then $0<\beta_1<R\beta_0$.

For $1+\beta_0>0$, equation (60) implies $\beta_1>0$ and $l\left(1,1\right)>0$. Now assume that $\beta_1-R\beta_0>0$, this implies that $l\left(0,1\right)>0$. For equation (59) to hold it must be that $1-\frac{1}{R}\frac{l(1,1)}{1+\beta_0}<0$.

$$\begin{aligned}1-\frac{1}{R}\frac{l(1,1)}{\left(1+\beta_0\right)^2}&=1-\frac{1}{R}\frac{\left(1+\beta_0\right)\left(1-w_0\right)-R\beta_1}{\left(1+\beta_0\right)} &(61)\\ &=1-\frac{1}{R}\left(1-w_0\right)+\frac{\beta_1}{1+\beta_0}>0 &(62)\end{aligned}$$

Since $R>1,\ w_0<1$, and $\beta_1>1$. Contradiction. Then, $1+\beta_0>0\Rightarrow\beta_1-R\beta_0<0$.

STEP 3. We now show that $1+\beta_0>0$. Let $\phi\equiv\frac{\sigma^2}{s^2}>1$. From equation (60):

$$\frac{\left(1+\beta_0\right)\left(1-w_0\right)}{\phi+R}=\beta_1.$$

We plug this into equation (59) and we obtain:

$$\begin{aligned}\phi\left(-\beta_0R+\frac{\left(1+\beta_0\right)\left(1-w_0\right)}{\phi+R}\right)+R\left[\frac{\left(1+\beta_0\right)\left(1-w_0\right)}{\phi+r}+\left(1+\beta_0\right)w_0-\beta_0R\right]+&\\ +\left[1+\beta_0-\frac{\phi(1-w_0)\left(1+\beta_0-\beta_0\phi R-\beta_0R^2+\left(1+\beta_0\right)\left(\phi+R-1\right)w_0\right)}{\left(\phi+R\right)^2}\right]&=0.\end{aligned}$$

Note that this is a linear equation on β_0, i.e.,

$$\begin{aligned}\beta_0\{\phi\left(\frac{1-w_0}{\phi+R}-R\right)+R\left[\frac{1-w_0}{\phi+R}+w_0-R\right]+1-\frac{\phi(1-w_0)\left(1-\phi R-R^2+\left(\phi+R-1\right)w_0\right)}{\left(\phi+R\right)^2}\}&\\ +\phi\left(\frac{1-w_0}{\phi+R}\right)+R\left[\frac{\left(1-w_0\right)}{\phi+R}+w_0\right]+\left[1-\frac{\phi(1-w_0)\left(1+\left(\phi+R-1\right)w_0\right)}{\left(\phi+R\right)^2}\right]&.\end{aligned}$$

Therefore,

$$\beta_0=-\frac{2-w_0(1-R)-\frac{\phi(1-w_0)(1+(\phi+R-1)w_0)}{(\phi+R)^2}}{2-w_0(1-R)-\frac{\phi(1-w_0)(1+(\phi+R-1)w_0)}{(\phi+R)^2}-(R\phi+R^2)\left[1-\frac{\phi(1-w_0)}{(\phi+R)^2}\right]}\equiv-\frac{A}{A-x}.$$

where $A\equiv 2-w_0(1-R)-\frac{\phi(1-w_0)(1+(\phi+R-1)w_0)}{(\phi+R)^2}$ and $x\equiv\left(R\phi+R^2\right)\left[1-\frac{\phi(1-w_0)}{(\phi+R)^2}\right]>0$.

Note that for $x = 0 \Rightarrow \beta_0 = -1$. Then, it suffices to show that $\frac{\partial \beta_0}{\partial x} = \frac{A}{(A-x)^2} \geq 0$, that is, $A \geq 0$. For $w_0 = 0.5$, which corresponds to $\lambda = 0$, A is positive, i.e., $A(0.5) > 0$. In addition, $\frac{\partial A}{\partial w_0} = \frac{(\phi+R-1)(R^2+\phi(R-2(1-w_0)))}{(\phi+R)^2} > 0$ for $w_0 \geq 0.5$. Therefore, $A > 0$ for $w_0 \geq 0.5$.

If we are interested in $\lambda < 0$ cases, since $A(0) > 0$, all we need to ensure that A is positive, and thus the result holds for $w_0 \in [0, 0.5)$, is that $R \geq 2(1-w_0)$.

□

In order to show Proposition 7.4, we need the following Lemmas (their proofs are relegated to the end of the section).

Lemma D.5. *For* $\lambda \geq 0, 1 + \beta_0 + \beta_1 - r\beta_0 > 0$.

Lemma D.6. *Given our linear guess for prices (8), when* $q = 2$, *at time* t*:*

$$x_t^{t-1} = \frac{E_t^{t-1}\left[s_{t+1}\right]}{\gamma R\left(1+\beta_0\right)\sigma^2} = \frac{\alpha\left(1-R\right)}{\gamma\left(1+\beta_0\right)^2\sigma^2} + \frac{l\left(0,1\right)}{\gamma\left(1+\beta_0\right)^2\sigma^2}d_t + \frac{l\left(1,1\right)}{\gamma\left(1+\beta_0\right)^2\sigma^2}d_{t-1} \tag{63}$$

$$x_t^{t} = \frac{E_{\Phi(m,s^2)}\left[s_{t+1}\right]}{R(1+\beta_0)s^2} = \delta(0) + \delta_0(0)d_t + \delta_1(0)d_{t-1} \tag{64}$$

with $l(0,1) \equiv [(1+\beta_0)w(0,\lambda,0) + \beta_1 - R\beta_0]$ *and* $l(1,1) \equiv [(1+\beta_0)w(1,\lambda,0) - R\beta_1]$, *and* $\delta(0) = \frac{\alpha(1-R)\left[1-\frac{s^2}{\sigma^2}\frac{l(0,1)}{(1+\beta_0)}\right]}{\gamma R(1+\beta_0)^2 s^2}$, $\delta_0(0) = \frac{\beta_1 - R\beta_0 + (1+\beta_0)\frac{s^2}{\sigma^2}\left(1-\frac{l(1,1)l(0,1)}{(1+\beta_0)^2}\right)}{\gamma R(1+\beta_0)^2 s^2}$, *and* $\delta_1(0) = -\frac{R\beta_1}{R\gamma(1+\beta_0)s^2}$.

Proof of Proposition 7.4. By lemma D.6 and Market Clearing, it follows that

$$\delta_0(0) + \frac{l\left(0,1\right)}{\gamma\left(1+\beta_0\right)^2\sigma^2} = 0,$$

and

$$\delta_1(0) + \frac{l\left(1,1\right)}{\gamma\left(1+\beta_0\right)^2\sigma^2} = 0.$$

And $\frac{\partial x_t^t}{\partial d_t} = \delta_0(0) = -\frac{\partial x_t^{t-1}}{\partial d_t}$, $\frac{\partial x_t^t}{\partial d_{t-1}} = \delta_1(0)$, and $\frac{\partial x_t^{t-1}}{\partial d_t} = \frac{l(0,1)}{\gamma(1+\beta_0)^2\sigma^2}$ and $\frac{\partial x_t^{t-1}}{\partial d_{t-1}} = \frac{l(1,1)}{\gamma(1+\beta_0)^2\sigma^2} = -\frac{\partial x_t^t}{\partial d_{t-1}}$.

Therefore, it suffices to show that $l(0,1) < 0$ and $\delta_1(0) < 0$.

By proposition 7.3, $\beta_1 > 0$ and $\beta_0 > 0$ and thus $\delta_1(0) = -\frac{R\beta_1}{R\gamma(1+\beta_0)s^2} < 0$. So it only remains to show that $l(0,1) < 0$.

We now show that $l(0,1) < 0$. From the equilibrium condition (59) we have:

$$0 = \left[R - \frac{l(1,1)}{(1+\beta_0)}\right] l\left(0,1\right) + \frac{l\left(0,1\right)^2}{\left(1+\beta_0\right)^2}\left(\beta_1 - R\beta_0\right) + \left[1+\beta_0+\beta_1-R\beta_0\right]$$

From Lemma D.5, $1 + \beta_0 + \beta_1 - R\beta_0 > 0$. Let $x = \frac{l(0,1)}{1+\beta_0}$, then

$$0 = \left[R\left(1+\beta_0\right) - l(1,1)\right]x + x^2\left(\beta_1 - R\beta_0\right) + \left[1+\beta_0+\beta_1-R\beta_0\right]$$

$$F(x) \equiv ax^2 + bx + c$$

with $a = \beta_1 - R\beta_0 < 0$ (by Proposition 7.3), $b = R(1+\beta_0) - l(1,1) = R(1+\beta_0) - (1+\beta_0)w(1,\lambda,0) + R\beta_1 > 0$ (by Proposition 7.3) and $c = 1+\beta_0+\beta_1 - R\beta_0 > 0$ (by Lemma D.5). Thus: F is convex and $F(0) = c > 0$. From FOC $a2x^* + b = 0 \Rightarrow x^* = -\frac{b}{2a} > 0$. Let's focus on x_2. Therefore, $F(x)$ has two roots x_1, x_2 with $x_1 < 0 < x^* < x_2$, where $x^* = \arg\max_{x\in\mathbb{R}} F(x)$.

We now show that $x_2 = \frac{l(0,1)}{1+\beta_0}$ cannot be a solution. Suppose not, that is assume that our solution is the positive root $\frac{l(0,1)}{1+\beta_0} = x_2$, then:

$$-\frac{b}{2a} < \frac{l(0,1)}{1+\beta_0} \tag{65}$$

$$\frac{R(1+\beta_0) - l(1,1)}{2[-(\beta_1 - R\beta_0)]} < \frac{l(0,1)}{1+\beta_0} \tag{66}$$

$$\frac{R(1+\beta_0) - l(1,1)}{2} < l(0,1)\frac{R\beta_0 - \beta_1}{1+\beta_0} \tag{67}$$

Let $Z \equiv -\frac{\beta_1 - R\beta_0}{1+\beta_0}$

$$\begin{aligned}
R(1+\beta_0) - (1+\beta_0)(1-w_0) + R\beta_1 &< 2l(0,1)Z \\
R(1+\beta_0) - (1+\beta_0)(1-w_0) + R\beta_1 &< 2Z[(1+\beta_0)w_0 + \beta_1 - R\beta_0] \\
R - 1 + w_0 + R\frac{\beta_1}{(1+\beta_0)} &< 2Z\left[w_0 + \frac{\beta_1 - R\beta_0}{(1+\beta_0)}\right] \\
Z(w_0 - Z) &> 0.5w_0 + \frac{1}{2}\left[R - 1 + R\frac{\beta_1}{1+\beta_0}\right] \\
\frac{w_0}{4} &> 0.5w_0 + \frac{1}{2}\left[R - 1 + R\frac{\beta_1}{1+\beta_0}\right].
\end{aligned}$$

Observe that $\frac{1}{2}\left[R - 1 + R\frac{\beta_1}{1+\beta_0}\right] > 0$ and thus a contradiction follows. The solution must be the negative root.

□

D.1 Proof of Lemmas D.1, D.2, D.3, D.4 and D.5 and D.6

Proof of Lemma D.1. Let $\varphi(z) \equiv K \exp\{-(A + Bz + Cz^2)\}\phi(z; \mu, \sigma^2)$. By definition of K, $\int \varphi(z)dz = 1$ and $\varphi \geq 0$, so it is a pdf. Moreover,

$$\begin{aligned}
\varphi(z) =& \frac{K^{-1}}{\sqrt{2\pi}\sigma} \exp\{-A - Bz - Cz^2 - 0.5\sigma^{-2}(z-\mu)^2\} \\
=& \frac{1}{K\sqrt{2\pi}\sigma} \exp\{-z^2(C + 0.5\sigma^{-2}) - 2z(0.5B - 0.5\sigma^{-2}\mu) - (A + 0.5\sigma^{-2}\mu^2)\} \\
=& \frac{1}{K\sqrt{2\pi}\sigma} \exp\{-(A + 0.5\sigma^{-2}\mu^2)\} \exp\{-0.5(2C + \sigma^{-2}) \left(z^2 - 2z\frac{(-B + \sigma^{-2}\mu)}{(2C + \sigma^{-2})}\right)\}.
\end{aligned}$$

Let $\Sigma^2 \equiv (2c+\sigma^{-2})^{-1}$, $m \equiv \Sigma^2(\sigma^{-2}\mu - b)$, and $K = \frac{1}{\sqrt{2\sigma^2 C+1}} \exp\{-(A+0.5\sigma^{-2}\mu^2) + \frac{m^2}{2\Sigma^2}\}$:

$$\begin{aligned}
\varphi(z) =& \frac{1}{K\sqrt{2\pi}\sigma} \exp\{-(a + 0.5\frac{\mu^2}{\sigma^2}) + \frac{m^2}{2\Sigma^2}\} \exp\{-\frac{z^2 - 2zm + m^2}{2\Sigma^2}\} \\
=& \frac{1}{K\sqrt{2\pi}\sigma} \exp\{-(a + 0.5\sigma^{-2}\mu^2) + \frac{m^2}{2\Sigma^2}\} \exp\{-\frac{(z-m)^2}{2\Sigma^2}\} = \quad \frac{1}{\sqrt{2\pi}\Sigma} \exp\{-\frac{(z-m)^2}{2\Sigma^2}\} \\
=& \frac{1}{\sqrt{2\pi\Sigma^2}} \exp\{-\frac{(z-m)^2}{2\Sigma^2}\}
\end{aligned}$$

□

Proof of Lemma D.2. At time $t + q$, an agent born in t is in the last period of his life, consuming all of its wealth. Therefore, he will sell all of its claims to the assets it holds and consume. The gain from saving is zero, and therefore the holding of financial assets is also zero by the end of this period: $x^t_{t+q} = 0, a^t_{t+q} = 0$. Given this, we can compute the portfolio choice of an agent with age $q-1$, who does want to save for next period when all wealth will be consumed. The agent's problem is a standard static portfolio problem, with initial wealth $W^t_{t_q}$:

$$\max_x E^t_{t_q} \left[-\exp\left(-\gamma\left(W^t_{t_q} + xs_{t+q}\right)\right)\right] = \max_x E^t_{t_q} \left[-\exp\left(-\gamma x s_{t+q}\right)\right] \tag{68}$$

At time t_q, the only random variable is d_{t+q}, which is normally distributed, and thus $s_{t+q} \sim N\left(E^t_{t_q}[s_{t+q}]; (1+\beta_0)\sigma^2\right)$. Given this, the agent's problem becomes:

$$V^{t-q}_{t-1} \equiv \max_x \left[-\exp\left(-\gamma x E^{t-q}_{t-1}[s_t] + \frac{1}{2}\gamma^2 x^2 (1+\beta_0)\sigma^2\right)\right] \tag{69}$$

$$\max_x \quad xE^{t-q}_{t-1}[s_t] - \frac{1}{2}\gamma x^2 (1+\beta_0)^2 \sigma^2 \tag{70}$$

And therefore, by FOC:

$$x^t_{t_q} = \frac{E^t_{t_q}[s_{t+q}]}{\gamma\sigma^2_*} \tag{71}$$

□

Proof of Lemma D.3. Note that $E[-\exp\{-A-Bz-Cz^2\}\exp\{-axh(z)\}]$ can be written as:

$$\int \exp\{-axh(z)\} - \exp\{-A-Bz-Cz^2\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{z-\mu}{\sigma^2}\right\}dz$$

By Lemma D.1, we know that his can be re-written as:

$$\frac{1}{\sqrt{2\sigma^2 C+1}}\exp\left\{-A-0.5\left(\frac{\mu^2}{\sigma^2}-\frac{m^2}{s^2}\right)\right\}\int -\exp\{-axh(z)\}\Phi\left(m,s^2\right)dz$$

with $m=-s^2B+s\sigma^{-2}\mu$ and $s^2=\frac{\sigma^2}{2C\sigma^2+1}$. Therefore, the maximization problem becomes:

$$\max_x E_{N(m,s^2)}[-\exp\{-axh(z)\}]$$

with $E_{N(m,s^2)}[\cdot]$ being the expectations operator over $z\sim N\left(m,s^2\right)$. Since $h(z)$ is linear, we know that $h(z)\sim N\left(\tilde{\mu}\left(m,s^2\right),\tilde{\sigma}\left(m,s^2\right)^2\right)$, with $\tilde{\mu}\left(m,s^2\right)=E_{N(m,s^2)}[h(z)]$, $\tilde{\sigma}\left(m,s^2\right)^2=V_{N(m,s^2)}[h(z)]$, by Lemma B.1, we know that

$$\begin{aligned}
\arg\max_x E[-\exp\{-A-Bz-Cz^2\}\exp\{-axh(z)\}] =&\frac{\tilde{\mu}\left(m,s^2\right)}{a\tilde{\sigma}\left(m,s^2\right)^2}\\
\max_x E[-\exp\{-A-Bz-Cz^2\}\exp\{-axh(z)\}] =&-\frac{1}{\sqrt{2\sigma^2C+1}}\exp\left[-A-0.5\left(\frac{\mu^2}{\sigma^2}-\frac{m^2}{s^2}\right)\right]\\
&\times\exp\left[-0.5\frac{\tilde{\mu}\left(m,s^2\right)^2}{\tilde{\sigma}\left(m,s^2\right)^2}\right]
\end{aligned}$$

□

Let $t\mapsto\rho(t)\equiv\gamma t^2$ and let

$$\begin{aligned}
\Lambda(d_{t-K},...,d_t)\equiv&\alpha(1-R)+\sum_{k=1}^{K}\beta_k d_{t+1-k}-R\sum_{k=0}^{K}\beta_k d_{t-k}\\
=&\alpha(1-R)+\sum_{j=0}^{K-1}\beta_{j+1}d_{t-j}-R\sum_{k=0}^{K}\beta_k d_{t-k}=\alpha(1-R)+\sum_{k=0}^{K}\beta(k)d_{t-k}
\end{aligned}$$

with $\beta(k)=\beta_{k+1}-R\beta_k$ for $k\in\{0,...,K-1\}$ and $\beta(K)=-R\beta_K$. We use Λ_τ to denote $\Lambda(d_{\tau-K},...,d_\tau)$.

Proof of Lemma D.4. We divide the proof into several steps.

STEP 1. It is straightforward that demand for risky assets can only be positive for a generation that is alive. From Lemma D.2, we know that $x_t^{t-q}=0$ and that $x_t^{t-q+1}=$

$\frac{E_t^{t-q+1}[s_{t+1}]}{\gamma((1+\beta_0)\sigma)^2}$. Therefore,

$$\delta(q) = \delta_k(q) = 0, \quad \forall k \in \{0, ..., K\} \tag{72}$$

$$\delta(q-1) = \frac{\alpha(1-R)}{\gamma((1+\beta_0)\sigma)^2}, \quad \delta_k(q-1) = \frac{(1+\beta_0)w(k,\lambda,q-1)+\beta(k)}{\gamma((1+\beta_0)\sigma)^2}, \quad \forall k \in \{0, ..., q-1\} \tag{73}$$

$$\delta_k(q-1) = \frac{\beta(k)}{\gamma((1+\beta_0)\sigma)^2}, \quad \forall k \in \{q, ..., K\}. \tag{74}$$

We also know from Lemma B.1 that

$$V^{q-1}(d_{t-K}, ..., d_t) = -\exp\left(-\frac{1}{2}\left(d_t\delta_0(q-1) + \delta(q-1) + \sum_{j=1}^{K}\delta_k(q-1)d_{t-j}\right)^2 \gamma^2((1+\beta_0)s_{q-1})^2\right)$$

where $s_{q-1} = \sigma^2$. Henceforth, we denote $V^{q-1}(d_{t-K}, ..., d_t)$ by V_t^{t-q+1}. In particular, $V_{t+1}^{t+1-q+1} = V_{t+1}^{t-q+2} = V^{q-1}(d_{t+1-K}, ..., d_{t+1})$.

STEP 2. We now derive the risky demand and continuation value for generation aged $q-2$. The problem of generation aged $q-2$ at time t is given by,

$$\max_x E_t^{t-q+2}\left[V_{t+1}^{t-q+2}\exp\left(-\gamma Rxs_{t+1}\right)\right]. \tag{75}$$

By the calculations in step 1, and using Λ_t as defined in (72), this problem becomes:

$$V^{q-2}(d_{t-K}, ..., d_t) \tag{76}$$

$$= \max_x E_t^{t-q+2}\left[-\exp\left(-\frac{1}{2}\left(x_t^{q-1}\right)^2\gamma^2((1+\beta_0)s_{q-1})^2 - \gamma Rx((1+\beta_0)d_{t+1}+\Lambda_t)\right)\right]. \tag{77}$$

with $x_t^{q-1} = d_{t+1}\delta_0(q-1) + \delta(q-1) + \sum_{j=1}^{K}\delta_k(q-1)d_{t+1-j}$.

Observe that

$$\begin{aligned}
&-\frac{1}{2}\left(d_{t+1}\delta_0(q-1) + \delta(q-1) + \sum_{j=1}^{K}\delta_k(q-1)d_{t+1-j}\right)^2\gamma^2((1+\beta_0)s_{q-1})^2 \\
&= -\frac{1}{2}\gamma^2((1+\beta_0)s_{q-1})^2\left(\delta(q-1) + \sum_{j=1}^{K}\delta_k(q-1)d_{t+1-j}\right)^2 \\
&\quad -\gamma^2((1+\beta_0)s_{q-1})^2\left(\delta(q-1) + \sum_{j=1}^{K}\delta_k(q-1)d_{t+1-j}\right)\delta_0(q-1)d_{t+1} \\
&\quad -\frac{1}{2}\gamma^2((1+\beta_0)s_{q-1})^2\left(\delta_0(q-1)\right)^2 d_{t+1}^2,
\end{aligned}$$

and that future dividends are the only random variable, with $d_{t+1} \sim N\left(\theta_t^{t-q+2}, \sigma^2\right)$. Therefore, by Lemma D.3, and with:

$$
\begin{aligned}
A =& \frac{1}{2}\gamma^2((1+\beta_0)s_{q-1})^2 \left(\delta(q-1) + \sum_{j=1}^{K} \delta_k(q-1) d_{t+1-j}\right)^2 \\
B =& \gamma^2((1+\beta_0)s_{q-1})^2 \left(\delta(q-1) + \sum_{j=1}^{K} \delta_k(q-1) d_{t+1-j}\right) \delta_0(q-1) \\
C =& \frac{1}{2}\gamma^2((1+\beta_0)s_{q-1})^2 \left(\delta_0(q-1)\right)^2
\end{aligned}
$$

we obtain:

$$
x_t^{t-(q-2)} = \frac{(1+\beta_0)s_{q-2}^2(\sigma^{-2}\theta_t^{t-(q-2)} - B) + \Lambda_t}{R\gamma((1+\beta_0)s_{q-2})^2}
$$

with $s_{q-2}^2 \equiv \frac{\sigma^2}{\gamma^2((1+\beta_0)s_{q-1})^2(\delta_0(q-1))^2\sigma^2+1}$. Therefore,

$$
\begin{aligned}
\delta(q-2) =& \frac{\alpha(1-R) - s_{q-2}^2(1+\beta_0)\delta_0(q-1)\delta(q-1)\gamma^2((1+\beta_0)s_{q-1})^2}{R\gamma((1+\beta_0)s_{q-2})^2} \\
\delta_k(q-2) =& \frac{(1+\beta_0)s_{q-2}^2(\sigma^{-2}w(k,\lambda,q-2) - [\gamma^2((1+\beta_0)s_{q-1})^2\delta_{k+1}(q-1)\delta_0(q-1)]) + \beta(k)}{R\gamma((1+\beta_0)s_{q-2})^2}, \ k \in \{0,...,q-1\} \\
\delta_k(q-2) =& \frac{-(1+\beta_0)s_{q-2}^2[\gamma^2((1+\beta_0)s_{q-1})^2\delta_{k+1}(q-1)\delta_0(q-1)] + \beta(k)}{R\gamma((1+\beta_0)s_{q-2})^2}, \ k \in \{q,...,K-1\} \\
\delta_K(q-2) =& \frac{\beta(K)}{R\gamma((1+\beta_0)s_{q-2})^2}.
\end{aligned}
$$

By lemma D.1, $d_{t+1} \sim N(m_t, s_{q-2}^2)$ with $m_t \equiv -s_{q-2}^2 B + s_{q-2}^2\sigma^{-2}\theta_t^{t-q+2}$. Thus, invoking lemma B.1 for this distribution for dividends and $a = R\gamma(1+\beta_0)$ implies that

$$
\begin{aligned}
V^{q-2}(d_{t-K},...,d_t) \asymp& -\exp\left(-\frac{1}{2}\left(x_t^{t-(q-2)}\right)^2 (R\gamma)^2((1+\beta_0)s_{q-2})^2\right) \\
=& -\exp\left(-\frac{1}{2}\left(d_t\delta_0(q-2) + \delta(q-2) + \sum_{j=1}^{K}\delta_k(q-2)d_{t-j}\right)^2 (R\gamma)^2((1+\beta_0)s_{q-2})^2\right)
\end{aligned}
$$

(the symbol $\asymp$ means that equality holds up to a positive constant).

STEP 3. We now consider the problem for agents of age $age \leq q-3$. Suppose the problem at age $age+1$ is solved, that is, suppose

$$\begin{aligned}
V_{t+1}^{t-age-1} =& V^{age+1}(d_{t+1-K},...,d_{t+1}) \\
\asymp & -\exp\left\{-\frac{1}{2}\left(d_{t+1}\delta_0(age+1)+\delta(age+1)+\sum_{j=1}^{K}\delta_j(age+1)d_{t+1-j}\right)^2 (R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\right\}.
\end{aligned}$$

The maximization problem is given by:

$$V^{age}(d_{t-K},...,d_t) \equiv \max_x E_t^{t-age}\left[V_{t+1}^{t-age-1}\exp\left(-\gamma R^{q-1-age}x((1+\beta_0)d_{t+1}+\Lambda_t)\right)\right]. \quad (78)$$

By similar calculations to step 2 and Lemma D.3,

$$x_t^{t-age} = \frac{(1+\beta_0)s_{age}^2(\sigma^{-2}\theta_t^{t-age}-B)+\Lambda_t}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}$$

with $s_{age}^2 \equiv \frac{\sigma^2}{(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2(\delta_0(age+1))^2\sigma^2+1}$, and

$$B \equiv (R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\left(\delta(age+1)+\sum_{j=1}^{K}\delta_j(age+1)d_{t+1-j}\right)\delta_0(age+1).$$

Therefore

$$\begin{aligned}
\delta(age) =& \frac{\alpha(1-R)-s_{age}^2(1+\beta_0)\delta_0(age+1)\delta(age+1)(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}, \\
\delta_k(age) =& \frac{(1+\beta_0)s_{age}^2(\sigma^{-2}w(k,\lambda,age)-[(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\delta_{k+1}(age+1)\delta_0(age+1)])+\beta(k)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2} \\
& k\in\{0,...,q-1\}, \\
\delta_k(age) =& \frac{-(1+\beta_0)s_{age}^2[(R^{q-1-(age+1)}\gamma)^2((1+\beta_0)s_{age+1})^2\delta_{k+1}(age+1)\delta_0(age+1)]+\beta(k)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2},\ k\in\{q,...,K-1\} \\
\delta_K(age) =& \frac{\beta(K)}{R^{q-1-(age)}\gamma((1+\beta_0)s_{age})^2}.
\end{aligned}$$

By lemma D.1, $d_{t+1}\sim N(m_t,s_{age}^2)$ with $m_t\equiv -s_{age}^2B+s_{age}^2\sigma^{-2}\theta_t^{t-q+2}$. Thus, invoking lemma B.1 for this distribution for dividends and $a=R^{q-1-age}\gamma(1+\beta_0)$ implies that

$$\begin{aligned}
V^{age}(d_{t-K},...,d_t) \asymp & -\exp\left(-\frac{1}{2}\left(x_t^{t-(age)}\right)^2(R^{q-1-(age)}\gamma)^2((1+\beta_0)s_{age})^2\right) \\
=& -\exp\left(-\frac{1}{2}\left(d_t\delta_0(age)+\delta(age)+\sum_{j=1}^{K}\delta_k(age)d_{t-j}\right)^2(R^{q-1-(age)}\gamma)^2((1+\beta_0)s_{age})^2\right).
\end{aligned}$$

$\square$

Proof of Lemma D.5. Assume it is not: $1+\beta_0+\beta_1-R\beta_0 \leq 0$. This implies that $l(0,1) = (1+\beta_0)w_0+\beta_1-R\beta_0 \leq 0$. From condition (59) we have:

$$0 = \left[R - \frac{l(1,1)}{(1+\beta_0)}\right] l\left(0,1\right) + \frac{l\left(0,1\right)^2}{\left(1+\beta_0\right)^2}\left(\beta_1 - R\beta_0\right) + \left[1+\beta_0+\beta_1-R\beta_0\right]$$

Then, since $\beta_1 - R\beta_0 \leq 0$ by proposition 7.3, for the previous equation to hold it must be that $\left[R - \frac{l(1,1)}{(1+\beta_0)}\right] \leq 0$.

$$\left[R - \frac{\left(1+\beta_0\right)\left(1-w_0\right) - R\beta_1}{\left(1+\beta_0\right)}\right] = \left[R + \frac{R\beta_1}{1+\beta_0} - \left(1-w_0\right)\right] > 0$$

Thus, $[1+\beta_0+\beta_1-R\beta_0] > 0$. $\square$

Proof of Lemma D.6. From Lemma B.1, we know that $x_t^{t-1} = \frac{E_t^{t-1}[s_{t+1}]}{\gamma(1+\beta_0)\sigma^2}$. Therefore, given our guess for prices and Lemma 7.2, we have:

$$x_t^{t-1} = \frac{E_t^{t-1}[d_{t+1}+p_{t+1}-p_tR]}{\gamma(1+\beta_0)\sigma^2} \tag{79}$$

$$= \frac{(1+\beta_0)\theta_t^{t-1} + \alpha(1-R) + (\beta_1 - R\beta_0)d_t - R\beta_1 d_{t-1}}{\gamma(1+\beta_0)\sigma^2} \tag{80}$$

since $\theta_t^{t-1} = w_0 d_t + (1-w_0)d_{t-1}$, we obtain equation (63), where $l(0,1) = (1+\beta_0)w_0+\beta_1-R\beta_0$ and $l(1,1) = (1+\beta_0)(1-w_0) - R\beta_1$. We also know from Lemma D.2 that

$$\begin{aligned} V_t^{t-1} &= -\exp\left(-\frac{1}{2}\frac{E_t^{t-q+1}\left[s_{t+1}\right]^2}{\gamma(1+\beta_0)\sigma^2}\right) \\ &= -\exp\left(-\frac{1}{2}\frac{\left(\alpha(1-R) + l(1,1)d_{t-1} + l(0,1)d_t\right)^2}{\gamma(1+\beta_0)\sigma^2}\right) \\ &= -\exp\left(-\frac{1}{2}\frac{\left(L_t(1,1) + l(0,1)d_t\right)^2}{\gamma(1+\beta_0)\sigma^2}\right) \end{aligned}$$

where $L_t(1,1) \equiv \alpha(1-R) + l(1,1)d_{t-1}$. Thus, we can write the value function of the generation who is investing for the last time on the market as follows:

$$V_t^{t-1} = -\exp(-A_t - B_t d_t - C d_t^2) \tag{81}$$

where $A_t \equiv \frac{L_t(1,1)^2}{2\gamma(1+\beta_0)^2\sigma^2}$, $B_t \equiv \frac{L_t(1,1)l(0,1)}{\gamma(1+\beta_0)^2\sigma^2}$, $C \equiv \frac{l(0,1)^2}{2\gamma(1+\beta_0)^2\sigma^2}$. Using this results to obtain V_{t+1}^t, the problem of the young generation at time t is given by:

$$\max_x \quad E_t^t\left[V_{t+1}^t \exp\left(-\gamma R x s_{t+1}\right)\right] \tag{82}$$

From Lemma D.3:

$$x_t^t = \frac{\tilde{\mu}\left(m, s^2\right)}{\gamma R \tilde{\sigma}\left(m, s^2\right)^2}$$

Where,

$$\begin{aligned}\tilde{\mu}\left(m, s^2\right) &= E_{\Phi(m,s^2)}\left[h\left(z\right)\right] = \alpha(1-R) + (\beta_1 - R\beta_0)d_t - R\beta_1 d_{t-1} + (1+\beta_0)m \\ \tilde{\sigma}\left(m, s^2\right)^2 &= V_{\Phi(m,s^2)}\left[h\left(d_{t+1}\right)\right] = (1+\beta_0)^2 s^2\end{aligned}$$

with $m = \frac{\theta_t^t - \sigma^2 B_{t+1}}{2C\sigma^2+1}$, $s^2 = \frac{\sigma^2}{2C\sigma^2+1}$. Incorporating the fact that $B_{t+1} = \frac{(\alpha(R-1)+l(1,1)d_t)l(0,1)}{(1+\beta_0)^2\sigma^2}$ and $\theta_t^t = d_t$ we obtain equation (64) and the respective δs.

□

Appendix E Data Appendix

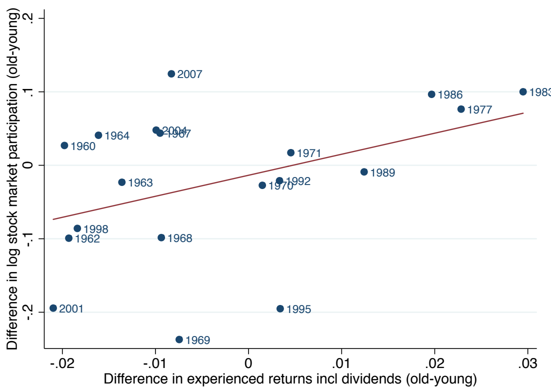
Our source of household-level microdata is the Survey of Consumer Finances (SCF), which provides repeated cross-section observations on asset holdings and various household background characteristics. Our sample has two parts. The first one is the standard SCF from 1983 to 2013, obtained from the Board of Governors of the Federal Reserve System and available every three years. The second source is the precursor of the ?modern? SCF, obtained from the Inter-university Consortium for Political and Social Research at the University of Michigan. The precursor surveys start in 1947, partly annually, but with some gaps. The data before 1960 contains information in stock holdings in some years, but age is measured in 5 or 10-year brackets, which would make our measurement of experienced returns imprecise, particularly for younger individuals. For this reason, we start in 1960 and use all survey waves that offer stock-market participation information, i.e., the 1960, 1962, 1963, 1964, 1967, 1968, 1969, 1970, 1971, and 1977 surveys.

The first measure is a binary variable for stock-market participation, available in each survey wave from 1960-2013. It indicates whether a household holds more than zero dollars worth of stocks. We define stock holdings as the sum of directly held stocks (including stock held through investment clubs) and the equity portion of mutual fund holdings. In our main tests, we include stocks held in retirement accounts (e.g., IRA, Keogh, and 401(k) plans). For 1983 and 1986, we need to impute the stock component of retirement assets from the type of the account or the institution at which they are held and allocation information from 1989. From 1989 to 2004, the SCF offers only coarse information on retirement assets (e.g., mostly stocks, mostly interest bearing, or split), and we follow a refined version of the Federal Reserve Board?s conventions in assigning portfolio shares. The Appendix provides the details. Online Appendix F reports robustness checks that exclude retirement account holdings from the analysis.

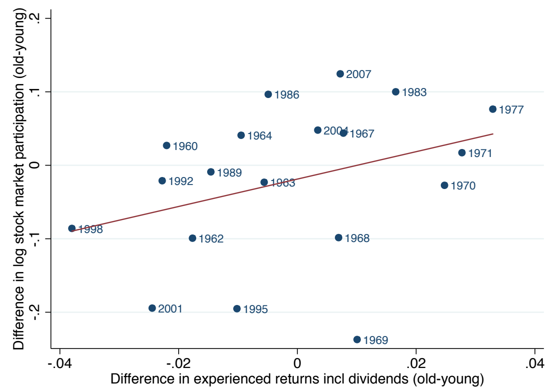
Our second measure of stock market participation is the fraction of liquid assets invested in stocks. The share of directly held stocks plus the equity share of mutual funds can be calculated in all surveys from 1960-2013 other than 1971. Liquid assets are defined as stock holdings plus bonds plus cash and short-term instruments (checking and savings accounts, money market mutual funds, certificates of deposit).

The 1983-2013 waves of the SCF oversample high-income households. The oversampling provides a substantial number of observations on households with significant stock holdings, which is helpful for our analysis of asset allocation, but could also induce selection bias. In our main tests, we weight the data using SCF sample weights, which undo the overweighting of high-income households and which also adjust for non-response bias. The weighted estimates are representative of the U.S. population.

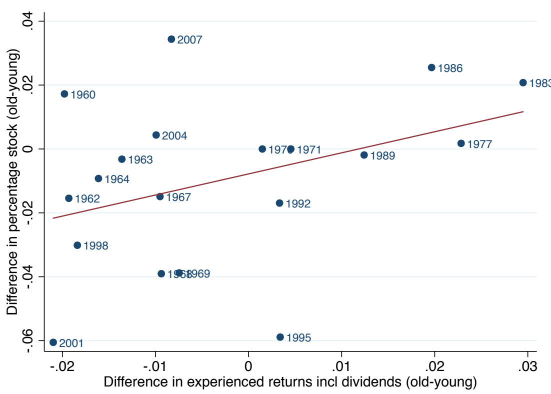
Appendix F Appendix Figures



(a) Stock Market Participation. ($\lambda = 1$)



(b) Stock Market Participation. ($\lambda = 3$)



(c) Fraction of Stock of Financial Assets. ($\lambda = 1$)

(d) Fraction of Stock of Financial Assets. ($\lambda = 3$)

Appendix Figure A.1 : Experienced Returns (including dividends) and Stock Holding

Difference in experienced returns is calculated as the experienced returns of the SP500 Index plus the dividend paid in that year. More distant experienced receive a lower weight. The weights either decline linearly ($\lambda = 1$) or super-linear ($\lambda = 3$) as in equation (5). Stock Market Participation is measured as the fraction of households that either directly held stock or indirectly, e.g. via mutuals or retirement accounts. We classify households whose head is aged 60 or older as "old" and households whose head is younger than 40 as "young". Difference in stockholdings, the y-axis in figures and , is calculated as the difference between the logs of the fraction of stock-holding households of the old and young age group. Percentage stock, the y-axis in figures and , is the fraction of assets invested in stock. The red line depicts the linear trend.

Online Appendix

In this Appendix, we explore the implications of having two-period live OLG where the mass of young agents born every period grows at rate g. To do so, we need to define an initial date for the economy, $t = 0$. Let y_t denote the mass of young agents born at time t, then $y_{t+1} = (1+g)\, y_t = y_0(1+g)^t$ and let $n_t = y_t + y_{t-1} = (2+g)\, y_{t-1}$ denote the total mass of people at any point in time $t > 0$. It is easy to check that $n_t = (1+g)\, n_{t-1}$; that is, total population grows at rate g.

The model is as the one presented in Section 3 in the paper. The main difference is that now population is growing over time. As a result, we make a different guess for the price function:

$$p_t = \alpha_0 \left(1+g\right)^{-t} + \beta_0 d_t + \beta_1 d_{t-1}$$

We verify this guess with our market clearing condition, where the demand of the young and the old need to add up to total supply of the asset, one:

$$\begin{aligned}
1 &= y_t \frac{E_t^y \left[p_{t+1} + d_{t+1}\right] - r p_t}{\gamma V \left[p_{t+1} + d_{t+1}\right]} + y_{t-1} \frac{E_t^o \left[p_{t+1} + d_{t+1}\right] - r p_t}{\gamma V \left[p_{t+1} + d_{t+1}\right]} \\
1 &= \frac{y_0 \left(1+g\right)^{t-1}}{\gamma \left(1+\beta_0\right)^2 \sigma^2} \left[\left(1+\beta_0\right) \left[\left(1+g\right) E_t^y \left[d_{t+1}\right] + E_t^o \left[d_{t+1}\right]\right] + \left(2+g\right) \left[\alpha_0 \left(1+g\right)^{-(t+1)} + \beta_1 d_t - r p_t\right]\right] \\
r p_t &= \left(1+\beta_0\right) \left\{ \frac{\left(1+g\right)}{\left(2+g\right)} d_t + \frac{1}{\left(2+g\right)} E_t^o \left[\left(1-\omega\right) d_{t-1} + \omega d_t\right] \right\} + \alpha_0 \left(1+g\right)^{-(t+1)} + \beta_1 d_t - \frac{\gamma V \left[p_{t+1} + d_{t+1}\right]}{y_0 \left(2+g\right) \left(1+g\right)^{t-1}}
\end{aligned}$$

We plug in $p_t = \alpha_0 \left(1+g\right)^{-t} + \beta_0 d_t + \beta_1 d_{t-1}$ and we use the method of undetermined coefficients to obtain:

$$\begin{aligned}
\alpha_0 &= -\frac{\gamma \left(1+\beta_0\right)^2 \sigma^2}{r - \frac{1}{1+g}} \frac{\left(1+g\right)}{y_0 \left(2+g\right)} \\
r\beta_0 &= \left(1+\beta_0\right) \left(\frac{1+g}{2+g} + \frac{1}{2+g} \omega \right) + \beta_1 \\
r\beta_1 &= \left(1+\beta_0\right) \frac{1-\omega}{2+g}
\end{aligned}$$

Let $\alpha_t \equiv \alpha_0 (1+g)^{-t}$ and $\gamma \equiv \frac{y_t}{n_t}$ denote the fraction of young agents, which is easy to verity is constant over time. Then, we see that the total mass of agents in the market is reflected only in the price constant, while the fraction of young people in the market determines the dividend loadings β_0 and β_1:

$$\begin{aligned}
\alpha_t &= -\frac{\gamma \left(1+\beta_0\right)^2 \sigma^2}{r - \frac{1}{1+g}} \frac{1+g}{n_t} \\
r\beta_0 &= \left(1+\beta_0\right) \left(\gamma + (1-\gamma)\omega\right) + \beta_1 \\
r\beta_1 &= \left(1+\beta_0\right) (1-\gamma)(1-\omega)
\end{aligned}$$