The Implications of Heterogeneity and Inequality for Asset Pricing
Other titles in Foundations and Trends® in Finance

Initial Public Offerings: A Synthesis of the Literature and Directions for Future Research
Michelle Lowry, Roni Michaely and Ekaterina Volkova

Privatization, State Capitalism, and State Ownership of Business in the 21st Century
William L. Megginson
ISBN: 978-1-68083-338-6

Executive Compensation
Raghavendra Rau

Three Branches of Theories of Financial Crises
Itay Goldstein and Assaf Razin

The Economics and Finance of Hedge Funds: A Review of the Academic Literature
Vikas Agarwal, Kevin A. Mullally, and Narayan Y. Naik
ISBN: 978-1-68083-156-6
The Implications of Heterogeneity and Inequality for Asset Pricing

Stavros Panageas
UCLA Anderson School of Management and NBER
USA
stavros.panageas@anderson.ucla.edu
Editorial Scope

Topics

Foundations and Trends® in Finance publishes survey and tutorial articles in the following topics:

- Corporate Finance
  - Corporate Governance
  - Corporate Financing
  - Dividend Policy and Capital Structure
  - Corporate Control
  - Investment Policy
  - Agency Theory and Information

- Financial Markets
  - Market Microstructure
  - Portfolio Theory
  - Financial Intermediation
  - Investment Banking
  - Market Efficiency
  - Security Issuance
  - Anomalies and Behavioral Finance

- Asset Pricing
  - Asset-Pricing Theory
  - Asset-Pricing Models
  - Tax Effects
  - Liquidity
  - Equity Risk Premium
  - Pricing Models and Volatility
  - Fixed Income Securities

- Derivatives
  - Computational Finance
  - Futures Markets and Hedging
  - Financial Engineering
  - Interest Rate Derivatives
  - Credit Derivatives
  - Financial Econometrics
  - Estimating Volatilities and Correlations

Information for Librarians

Foundations and Trends® in Finance, 2020, Volume 12, 4 issues. ISSN paper version 1567-2395. ISSN online version 1567-2409. Also available as a combined paper and online subscription.
Contents

1 Introduction and Summary 2

2 Heterogeneous Preferences, Beliefs, and Limited Participation Models 11
   2.1 Heterogeneous Preferences 11
   2.2 Heterogeneous Beliefs 21
   2.3 Multiple Types of Agents 25
   2.4 Limited Participation 27
   2.5 Recursive Preferences and the Sharpe Ratio 31
   2.6 Empirical Implications 33

3 Heterogeneous Endowments 37
   3.1 Lack of Intra-Cohort Risk Sharing 37
   3.2 Lack of Inter-Cohort Risk Sharing 45
   3.3 Recursive Preferences 49
   3.4 Summary and Empirical Implications 51

4 Further Applications: Equilibrium Arbitrage, Heterogeneous Stochastic Discount Factors 61
   4.1 Equilibrium Arbitrage 61
   4.2 Asymmetric Information and Heterogeneous Stochastic Discount Factors 64
The Implications of Heterogeneity and Inequality for Asset Pricing

Stavros Panageas

UCLA Anderson School of Management and NBER, USA; stavros.panageas@anderson.ucla.edu

ABSTRACT

Does heterogeneity matter for asset pricing and in particular for risk premia? Starting with an irrelevance result, I classify the literature into two groups of papers according to how they link investor heterogeneity and risk premia. The first group contains models of investors who differ in terms of their preferences, beliefs, or access to markets. Despite their differences, these models have similar implications, and can be analyzed in a unified way. The second group of papers consists of models where investors experience uninsurable income shocks. The goal of this survey is to provide one unified framework to better understand this large literature, and especially to reconcile several of the seemingly inconsistent results found in some seminal papers.
Heterogeneity is all around us. Besides the obvious dimensions of heterogeneity (income and wealth), people of similar wealth and income have different savings rates and attitudes toward risk. While it is hard to argue with the existence of heterogeneity, this survey asks a question with a less obvious answer: Does heterogeneity matter for asset pricing? Specifically, does it matter for the market price of risk?

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

Yet, most leading asset-pricing models tend to be “representative” agent models, i.e., models where the distribution of wealth, income, consumption, and the associated dynamics of these quantities are all irrelevant for asset pricing.\footnote{Three leading such paradigms are Campbell and Cochrane (1999), Bansal and Yaron (2004), Barro (2006).} One possible justification for abstracting from all distributional considerations was given in a seminal paper by Grossman and Shiller (1982). That paper revisited and substantially
relaxed the assumptions of the Breeden (1979) aggregate consumption CAPM. Specifically, Grossman and Shiller (1982) showed that Breeden’s aggregate-consumption CAPM continues to determine risk premia, even if risk sharing is imperfect, and therefore each consumer experiences a different consumption growth.

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

$$E_t\{u'(c_{t+\delta}^i)(R_{t+\delta} - R_{t+\delta}^f)\} = 0, \quad (1.1)$$

where $u'(c_{t+\delta}^i)$ is the marginal utility of agent $i$. Assuming that the agent trades frequently, so that $\delta$ is small, and proceeding heuristically, a first-order expansion of $u'(c_{t+\delta}^i)$ around $c_t^i$ gives

$$-\frac{u'(c_t^i)}{u''(c_t^i)} E_t\{R_{t+\delta} - R_{t+\delta}^f\} \approx E_t\{(c_{t+\delta}^i - c_t^i)(R_{t+\delta} - R_{t+\delta}^f)\}. \quad (1.2)$$

Defining aggregate consumption as $C_t \equiv \int_i c_t^i di$ and integrating both sides of (1.2) across $i$ gives

$$E_t\{R_{t+\delta} - R_{t+\delta}^f\} \approx \left(-\int_i \frac{u'(c_t^i)}{u''(c_t^i)} di\right)^{-1} \times E_t\{(C_{t+\delta} - C_t)(R_{t+\delta} - R_{t+\delta}^f)\}$$

$$= \left(-\int_i \frac{u'(c_t^i)}{c_t^i u''(c_t^i)} C_t di\right)^{-1} \times E_t\left\{\left(\frac{C_{t+\delta} - C_t}{C_t}\right)(R_{t+\delta} - R_{t+\delta}^f)\right\}. \quad (1.3)$$

Equation (1.3) is remarkably similar to Breeden’s aggregate consumption CAPM. It states that the excess return on an asset is given by the product of two terms. The first term is a consumption-weighted “harmonic average” of the relative risk aversion coefficients $-\frac{c_t^i u''(c_t^i)}{u'(c_t^i)}$.

**Footnotes:**

2 A first-order Taylor expansion implies that $u'(c_{t+\delta}^i) \simeq u'(c_t^i) + u''(c_t^i)(c_{t+\delta}^i - c_t^i)$. Substituting this approximation into (1.1) and rearranging gives (1.2).

3 The harmonic average of $y_i$ with weights $x_i$ is defined as $(\int_i x_i y_i^{-1} di)^{-1}$. 
of the different investors. The second term is the expected product of aggregate consumption growth with the excess return.

One immediate implication of (1.3) is that even in the presence of heterogeneity, the conditional aggregate consumption CAPM continues to hold. Other than affecting the weights in the harmonic average of relative risk aversion, heterogeneity in individual consumption growth rates – whatever the reason for this heterogeneity – does not impact the risk premium. In particular, income risks that can cause idiosyncratic consumption fluctuations due to imperfect risk sharing are irrelevant for risk premia. To exaggerate for the sake of clarity, the risk premium in an economy with and without uninsurable idiosyncratic risks will be the same as long as all agents have the same risk aversion.

The Grossman and Shiller (1982) result relies on one approximation step (Equation (1.2)). This approximation step is innocuous if asset prices and consumption growth are both diffusions. Indeed, in the continuous-time limit that they consider, the approximation becomes exact, because of Ito’s Lemma.

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

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

The first strand of the first broad category (Subsection 2.1) contains models of risk aversion heterogeneity, typically in a framework where agents have expected utility preferences. These models are not departures from Grossman and Shiller (1982), since Equation (1.3) continues to hold. Indeed, most of the papers in this literature assume that agents trade continuously and therefore Equation (1.3) is exact, not approximate. The key feature of these models is the observation that if agents have different levels of risk aversion then the consumption distribution evolves dynamically, favoring the bold in good times and the meek in bad times.

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

The second strand of the first broad category (Subsection 2.2) comprises models of belief heterogeneity. These models assume that some investors may have different beliefs than others, possibly not resulting from superior information, but from different priors or plain irrationality. In such models, Equation (1.1) may not hold for some
investors, as they use a different expectation operator. In terms of asset-pricing implications, however, models of belief heterogeneity and preference heterogeneity are very closely related. In particular, the Sharpe ratio with heterogeneous beliefs is a consumption-weighted average of the Sharpe ratios that would obtain in homogeneous belief economies populated by only one of the constituent groups. Just as luck favors the bold in models of preference heterogeneity (in terms of increasing their wealth and consumption shares), luck favors the optimists in models of belief heterogeneity. Subsection 2.2 presents the formal connections between belief and preference heterogeneity.

The third strand of the first broad category (Subsection 2.4) comprises models where Equation (1.1) holds for some investors but not for others. For instance, if investor $i$ is not even participating in the market for the risky asset, then Equation (1.1) does not apply to her. The implication is that when aggregating across all agents to get from Equations (1.2) to (1.3), one should only aggregate across the subset of agents that participate in the risky market. As Subsection 2.4 shows, models of this type closely resemble models of heterogeneous preferences, where non-participants are viewed as investors with infinite risk aversion for the purpose of pricing the risky assets.

Overall, the three strands of the literature mentioned so far belong to the same broad category, since they share more similarities than differences. Indeed, a novel aspect of this survey is to show that models of heterogenous preferences, beliefs, or participation opportunities can be analyzed with similar techniques as part of one unified framework. One could go as far as argue that these models do not invalidate the core of the Grossman and Shiller (1982) result: while Equation (1.1) may fail for some of the investors, ultimately a conditional version of the aggregate consumption CAPM continues to hold, with the consumption weights acting as “conditioning variables,” to put it in the jargon used by finance econometricians.

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

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

It would be natural to conjecture that the discrepancy between the two papers lies in the usage of discrete- versus continuous-time methods. If true, this would be a source of concern, since it would indicate that if one were to shrink the assumed time interval in Constantinides and Duffie (1996) to zero, the results of the paper could be jeopardized.

Subsection 3.1, which discusses the first strand of the second broad category of the literature, reconciles the results of Constantinides and Duffie (1996) and Grossman and Shiller (1982). Using a minor modification of Constantinides and Duffie (1996) that allows consideration of the continuous-time limit, this subsection shows that the key insight of Constantinides and Duffie (1996) is invariant to the assumed decision interval. The results of Grossman and Shiller (1982) do not apply because the continuous-time process is not a diffusion, but a process with discontinuous sample paths. Because of this, Equation (1.2) does not hold, even in continuous time. To the best of my knowledge, this is the first paper to develop a continuous-time version of Constantinides and Duffie (1996) and provide a reconciliation between Grossman and Shiller (1982) and Constantinides and Duffie (1996).

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

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

The reason for the similarity between the two types of models is
quite intuitive. The source of risk premia in models like Constantinides
and Duffie (1996) is an agent’s fear that if there is a large redistribution
among the existing cohorts of agents, she might end up being among the
losers rather than the winners. Because of risk aversion, she overweights
the possibility of being among the losers and demands a risk premium
for assets with bad payoffs when redistribution is high. In models of
imperfect inter-cohort risk sharing investors fear that an incoming cohort
of agents might introduce the next big company that will displace the
companies owned by current cohorts. Therefore any assets that are
prone to this displacement command a risk premium.

Subsection 3.3 discusses the third strand of the second broad cat-
egory. Models in this strand make endowment heterogeneity matter
by using recursive preferences rather than expected utility. In a semi-
nal paper, Bansal and Yaron (2004) highlighted that if agents are not
neutral to the timing of the resolution of uncertainty, then expected
returns reflect compensation not only for short run risk, but also for
risks that are associated with consumption growth in the “long run.”
Mathematically, this means that Equation (1.1) does not hold and the Grossman and Shiller (1982) argument fails at its origin.

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

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

The last section of the survey (Section 4) addresses two further observations relating to Equation (1.1). Subsection 4.1 discusses models where, for each risky asset, Equation (1.1) applies only to a subset of agents. Subsection 4.2 discusses the validity of (1.1) in the presence of asymmetric information.

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

Subsection 4.2 discusses the possibility that investors may have superior information compared to the econometrician. By itself, this is not a problem for the Grossman and Shiller (1982) argument, because
the Euler Equation (1.1) “conditions down” from the perspective of the econometrician. If, however, short-selling constraints prevent Equation (1.1) from holding for every investor and every asset, then agents may (endogenously) choose not to participate in certain asset classes. This means that the model features effectively heterogeneous stochastic discount factors. Models of this sort have important implications for portfolio biases, and performance evaluation.

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

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

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

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


References


References


