If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets

Lawrence Blume
David Easley

SFI WORKING PAPER: 2001-06-031

If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets

Lawrence Blume and David Easley
Department of Economics
Cornell University

March 2001
Today: June 5, 2001

First version, 1 June 1999. The authors thank YoungMiao Hong, Carsten Krabbe Nielsen, Werner Ploberger, Alvaro Sandroni, Bill Zame, and seminar audiences at Erasmus University, The Hebrew University of Jerusalem, the University of Iowa, the University of Rochester, Stanford University, the Santa Fe Institute, the Stockholm School of Economics, UCLA, the University of Wisconsin and Yale University for stimulating conversation. We thank The National Science Foundation for research support under grant SES 9808690 and the Santa Fe Institute. Blume gratefully acknowledges research support from the Cowles Foundation for Economic Research and the Department of Economics at Yale University.

Abstract: This paper provides an analysis of the asymptotic properties of consumption allocations in a stochastic general equilibrium model with heterogeneous consumers. In particular we investigate the *market selection hypothesis*, that markets favor traders with more accurate beliefs. We show that in any Pareto optimal allocation whether each consumer vanishes or survives is determined entirely by discount factors and beliefs. Since equilibrium allocations in economies with complete markets are Pareto optimal, our results characterize the limit behavior of these economies. We show that, all else equal, the market selects for consumers who use Bayesian learning with the truth in the support of their prior and selects among Bayesians according to the size of the their parameter space. Finally, we show that in economies with incomplete markets these conclusions may not hold. Payoff functions can matter for long run survival, and the market selection hypothesis fails.

Keywords: Market selection hypothesis, subjective beliefs, general equilibrium, incomplete markets, complete markets.

JEL Classification: D46, D51, D52, D81

Correspondent:
Professor Lawrence Blume
Department of Economics
Uris Hall
Cornell University
Ithaca NY 14850

1 Introduction

General equilibrium models of macroeconomic and financial phenomena commonly assume that traders maximize expected utility with *rational*, which is to say, correct, beliefs. The *expected utility hypothesis* places few restrictions on traders' behavior in the absence of *rational expectations*, and so much attention has been paid to the validity of assuming accurate beliefs. However, an adequate explanation of how traders come to correctly forecast endogenous equilibrium rates of return is lacking.

Two kinds of explanations have been offered. One posits that correct beliefs can be learned. That is, rational expectations are stable steady states of learning dynamics — Bayesian or otherwise. In our view learning cannot provide, a satisfactory foundation for rational expectations. In models where learning works, the learning rule is tightly coupled to the economy in question. Positive results are delicate. Robust results are mostly negative. See Blume, Bray, and Easley (1982), Blume and Easley (1998b) and Marimon (1997) for more on learning and its limits.

The other approach posits "natural selection" in market dynamics. The *market selection hypothesis*, that markets favor rational agents over irrational agents, has a long tradition in economic analysis. Alchian (1950) and Friedman (1953) believed that market selection pressure would eventually result in behavior consistent with maximization; those who behave irrationally will be driven out of the market by those who behave *as if* they are rational. Cootner (1964) and Fama (1965) argued that in financial markets, investors with incorrect beliefs will lose their money to those with more accurate assessments, and will eventually vanish from the market. Thus in the long run prices are determined by traders with rational expectations. This argument sounds plausible, but until recently there was no careful analysis of the market selection hypothesis; that wealth dynamics would select for expected utility maximizers, or, within the class of expected utility maximizers, select for those with rational expectations.

In two provocative papers, Delong, Shleifer, Summers and Waldman (1990, 1991) undertook a formal analysis of the wealth flows between rational and irrational traders. They argue that irrationally overconfident *noise traders* can come to dominate an asset market in which prices are set exogenously; a claim that contradicts Alchian's and Friedman's intuition. Blume and Easley (1992) address the same issue in a general equilibrium model. We showed that if savings rates are equal across traders, general equilibrium wealth dynamics need not lead to traders making portfolio choices as if they maximize expected utility using rational expectations. We did not study the emergence of fully intertemporal expected utility maximization, nor did we say much about the emergence of beliefs. Sandroni (2000) addressed the latter question. He built economies with intertemporal expected utility maximizers and studied the emergence of rational expectations. He showed in a Lucas trees economy that, controlling for discount factors, only traders with rational expectations, or those whose forecasts merge with rational expectations forecasts, survive. He also showed that if even if no such traders are present, no trader whose forecasts are persistently wrong survives in the presence of a learner.

In this paper we explore more completely when selection occurs, when it does not occur and why. To do this we build an intertemporal general equilibrium model with uncertainty. We show that in any Pareto optimal allocation of resources, whether an agent vanishes or survives is determined entirely by discount factors and beliefs. Attitudes toward risk are irrelevant to the long run fate of agents. In particular, controlling for discount factors, those consumers whose beliefs are closest to the truth end up with all the resources. By the first theorem of welfare economics, this implies that correct beliefs are selected for whenever markets are complete, or in any competitive equilibrium in which markets are dynamically complete. This conclusion is robust to the asset structure, of course, so long as markets are complete at the equilibrium prices. So for economies with complete markets the market selection hypothesis is correct. But the assumption that some trader has correct beliefs at the outset limits the interest of this result. Therefore we also examine how the market selects over learning rules.

In studying learning we first assume that all traders have the same discount factor. Our first learning result is that a Bayesian almost surely survives for almost all possible truths in the support of her prior. Furthermore, in the presence of such a Bayesian, any traders who survive are not too different from Bayesians. They use a forecasting rule that asymptotically looks like Bayes forecasts. Second we show that the market selects over Bayesians according to the size of the support of their prior beliefs. We consider Bayesian traders whose belief supports are open sets containing the true parameter value, and whose prior belief has a density with respect to Lebesgue measure. We show that survival prospects are indexed by the dimension of the support: Those traders with the lowest dimension supports survive, and all others vanish.[1] Having more prior information favors survival only if it affects the dimension of the support of prior beliefs. Finally, we show that discount factor differences dominate differences in belief supports. Among Bayesians, the survivors are always the most patient Bayesians.

Without complete markets, the market selection hypothesis may fail. Traders with incorrect beliefs may “drive out” those with more accurate beliefs. In economies with incomplete markets, a trader who is overly optimistic about the return on some asset in some state can choose to save enough to more than overcome the poor asset allocation decision that his incorrect expectations create. This result is particularly significant because when traders’ beliefs are heterogeneous, some market incompleteness is natural. With heterogeneous beliefs, market completeness implies that traders can bet on any differences in beliefs. This amounts to opening a new set of markets every time a new trader with sufficiently different beliefs enters the economy. In the context of the model, the relevant state space contains the union of the supports of each trader’s beliefs. Thus adding a new trader can require expanding the state space, and therefore adding new markets. We believe that in economic models that take agent heterogeneity seriously, market incompleteness is the natural assumption.

2 The Model

Our model and examples are concerned with infinite horizon exchange economies which allocate a single commodity. In this section we establish basic notation, list the key assumptions and characterize Pareto optimal allocations.

2.1 Notation and Basics

Formally, we assume that time is discrete and begins at date 0. The possible states at each date form a finite set $\{1, \ldots, S\}$. The set of all sequences of states is Σ with representative sequence $\sigma = (\sigma_0, \ldots)$, also called a *path.* $\sigma^t = (\sigma_0, \ldots, \sigma_t)$ denotes the partial history through date t of the path σ, and $1_t^s(\sigma)$ is the indicator function defined on paths which takes on the value 1 if $\sigma_t = s$ and 0 otherwise.

The set Σ together with its product sigma-field is the measurable space on which everything will be built. Let p denote the "true" probability measure on Σ. Expectation operators without subscripts intend the expectation to be taken with respect to the measure p. For any probability measure q on Σ, $q_t(\sigma)$ is the (marginal) probability of the partial history $(\sigma_0, \ldots, \sigma_t)$. That is, $q_t(\sigma) = q(\{\sigma_0 \times \cdots \times \sigma_t\} \times S \times S \times \cdots)$.

In the next few paragraphs we introduce a number of random variables of the form $x_t(\sigma)$. All such random variables are assumed to be date-t measurable; that is, their value depends only on the realization of states through date t. Formally, $\mathcal{F}_t$ is the σ-field of events measurable at date t, and each $x_t(\sigma)$ is assumed to be $\mathcal{F}_t$-measurable. For a given path σ, σ_t is the state at date t and $\sigma^t = (\sigma_0, \ldots, \sigma_t)$ is the partial history through date t of the evolution of states.

An economy contains I consumers, each with consumption set $\mathbf{R}_{++}$. A *consumption plan* $c : \Sigma \to \prod_{t=0}^{\infty} \mathbf{R}_{++}$ is a sequence of $\mathbf{R}_{++}$-valued functions $\{c_t(\sigma)\}_{t=0}^{\infty}$ in which each c_t is $\mathcal{F}_t$-measurable. Each consumer is endowed with a particular consumption plan, called the *endowment stream.* Trader i's endowment stream is denoted e^i.

Consumer i has a utility function $U^i : c \mapsto \mathbf{R}$ which is the expected presented discounted value of some payoff stream with respect to some beliefs. Specifically, consumer i has beliefs about the evolution of states, which are represented by a probability distribution p^i on Σ. We call p^i a *forecast distribution*. She also has a payoff function $u^i : \mathbf{R}_{++} \to \mathbf{R}$ on consumptions and a discount factor β_i strictly between 0 and 1. The utility of a consumption plan is

$$U^i(c) = E_{p^i}\left\{\sum_{t=0}^{\infty} \beta_i^t u^i\big(c_t(\sigma)\big)\right\}.$$

This scheme is rather general in its treatment of beliefs. One obvious special case is that of iid forecasts. If trader i believes that all the σ_t are iid draws from a common distribution ρ, then p^i is the corresponding distribution on infinite sequences. Thus, for instance, $p^i_t(\sigma) = \prod_{\tau=0}^{t} \rho(\sigma_\tau)$. Markov models and other, more complicated stochastic processes can be accommodated as well.

This representation of beliefs nests Bayesian learning. Suppose that a measurable space Θ of parameters is given, and that each parameter represents a *model* of the stochastic process of states; that is, the conditional distributions $p(\cdot|\theta)$ give the probability distribution on Σ that would describe the stochastic process of states if the true parameter were θ. Assume that $p(\cdot|\cdot)$ is jointly measurable with respect to the product sigma-field associated with $\Theta \times \Sigma$. Suppose too that trader i has a *prior belief* given by a probability distribution μ^i on Θ. The conditional distributions on Σ given θ and the prior distribution μ^i together determine a joint distribution ν^i on $\Theta \times \Sigma$: We have for any measurable set $A \subset \Theta$ with indicator function $1_A(\theta)$ and measurable set $B \subset \Sigma$,

$$\nu^i(A \times B) = \int 1_A(\theta) p(B|\theta) d\mu^i(\theta)$$

The distribution ν^i contains everything of relevance for a description of Bayesian learning from the given models and prior belief. For instance, posterior beliefs from the first t observations of states assign probability $\nu^i(A \times \Sigma|\sigma_0, \ldots, \sigma_t)$ to the set $A \subset \Theta$. The forecast probability of a set $B' \subset S$ at date $t+1$ given observations $\sigma_0, \ldots, \sigma_t$ is $\nu^i(\Theta \times B' \times S \times S \times \cdots |\sigma_0, \ldots, \sigma_t)$. In fact, all of the forecast distributions can be constructed from the marginal distribution of ν^i on Σ. Take $p^i(B) = \nu^i(\Theta \times B)$ for any measurable $B \subset \Sigma$. Although every Bayesian learning process generates a p^i, one cannot in general go backwards. A given p^i is consistent with many different model sets and prior beliefs.

Seemingly more generally, we can view a learning rule as a sequence of $\mathcal{F}_t$-measurable functions which assigns to each history of states through t a probability distribution of states in period $t+1$. Beliefs at time 1 are simply a distribution on S. Beliefs at time 1 together with the learning rule determine through integration a probability distribution on $S \times S$ whose marginal distribution at time 1 is the time 1 beliefs. And in general, a given t-period marginal distribution and the learning rule determine through integration a probability distribution on partial histories of length $t+1$ whose marginal distribution on the first t periods is the given t-period marginal distribution. The Kolmogorov Extension Theorem (Halmos 1974, sec. 49) guarantees that there is a probability distribution p^i on paths whose finite-history marginal distributions agree with those we constructed. Notice, however, that the specification of a learning rule as a collection of conditional distributions is more detailed than the specification as an S-valued stochastic process, because from the process p^i a conditional distribution for a given partial history σ^t can be recovered if and only if $p^i(\sigma^t) > 0$. But the seeming loss of generality is inessential because Pareto optimality implies that consumption at any partial history should be 0 if that partial history is an impossible event.[2]

We will assume throughout the following properties of payoff functions:

A. 1. *The payoff functions u_i are C^1, strictly concave, strictly monotonic, and satisfy an Inada condition at 0.*

We assume that the aggregate endowment is uniformly bounded from above and away from 0:

A. 2. $\infty > F = \sup_{t,\sigma} \sum_i e_t^i(\sigma) \geq \inf_{t,\sigma} \sum_i e_t^i(\sigma) = f > 0.$

Finally, we assume that traders believe to be possible anything which is possible.

A. 3. *For all consumers i, all dates t and all paths σ, if $p_t(\sigma) > 0$ then $p_t^i(\sigma) > 0$.*

If $p_t^i(\sigma) = 0$ for some trader i and date t, it is not optimal to allocate any consumption to trader i after date $t-1$ on path σ. Traders like this who vanish after only a finite number of periods have no impact on long run outcomes, and so we are not interested in them.

2.2 Pareto Optimality

Standard arguments show that in this economy, Pareto optima can be characterized as maxima of weighted-average social welfare functions. Because of the Inada condition, each trader's allocation in any Pareto optimum is either p^i almost surely interior or it is 0. We are not interested in traders who do not play any role in the economy so we focus on Pareto optima in which each trader i's allocation is p^i-almost surely interior. If $c^* = (c^{1*}, \dots, c^{I*})$ is such a Pareto optimal allocation of resources, then there is a vector of welfare weights $(\lambda^1, \dots, \lambda^I) \gg \mathbf{0}$ such that c^* solves the problem

$$\begin{aligned} \max_{(c^1,\dots,c^I)} \quad & \sum_i \lambda^i U^i(c) \\ \text{such that} \quad & \sum_i c^i - e \leq \mathbf{0} \\ & \forall t, \sigma \; c_t^i(\sigma) \geq 0 \end{aligned} \tag{1}$$

where $e_t = \sum_i e_t^i$. The first order conditions for problem 1 are:

For all σ and t,

(i) there is a number $\eta_t(\sigma) > 0$ such that if $p_t^i(\sigma) > 0$, then

$$\lambda^i \beta_i^t u^{i\prime}\big(c_t^i(\sigma)\big) p_t^i(\sigma) - \eta_t(\sigma) = 0 \tag{2}$$

(ii) If $p_t^i(\sigma) = 0$, then $c_t^i(\sigma) = 0$.

These conditions will be used to characterize the long-run behavior of consumption plans for individuals with different preferences, discount factors and forecasts. Our method is to compare marginal utilities of different consumers which we derive from the first order conditions. This idea was first applied to the equilibrium conditions for a deterministic production economy by Blume and Easley (1998a), and to the equilibrium conditions for an asset model by Sandroni (2000). All of our results are based on the following simple idea:

Lemma 1. *On the event* $\{u^{i\prime}\big(c_t^i(\sigma)\big)/u^{j\prime}\big(c_t^j(\sigma)\big) \to \infty\}$, $c_t^i(\sigma) \downarrow 0$. *On the event* $\{c_t^i(\sigma) \downarrow 0\}$, *for some trader* j, $\limsup_t u^{i\prime}\big(c_t^i(\sigma)\big)/u^{j\prime}\big(c_t^j(\sigma)\big) = \infty$.

Proof. $u^{i\prime}\big(c_t^i(\sigma)\big)/u^{j\prime}\big(c_t^j(\sigma)\big) \to \infty$ iff either the numerator diverges to ∞ or the denominator converges to 0. The denominator, however, is bounded below by the marginal utility of the upper bound on the aggregate endowment, $u^{j\prime}(F)$. So the hypothesis of the lemma implies that $u^{i\prime} \uparrow \infty$, and so $c_t^i \downarrow 0$. Going the other way, in each period some consumer consumes at least f/I, and so some trader j must consume at least f/I infinitely often. Then $u^{j\prime}\big(c_t^j(\sigma)\big) \leq u^{j\prime}(f/I)$ infinitely often. If $\lim_t c_t^i = 0$, then $\limsup u^{i\prime}\big(c_t^i(\sigma)\big)/u^{j\prime}\big(c_t^j(\sigma)\big) = \infty$. □

Limsup is the best that can be done for the necessary condition because the surviving trader j may have fluctuating consumption.

Using the first order conditions, we can express the marginal utility ratios in three different ways. Consider two consumers i and j, with forecasts p^i and p^j. For generic consumers i, j and k, define the following random variables

$$Z_t^k = -\sum_{s\in S} 1_t^s(\sigma)\log p^k(s|\mathcal{F}_{t-1}) \qquad Z_t = -\sum_{s\in S} 1_t^s(\sigma)\log p^k(s|\mathcal{F}_{t-1})$$

$$Y_t^k = Z_t^k - Z_t \qquad L_t^{ij} = \frac{p_t^j(\sigma)}{p_t^i(\sigma)},$$

where $l_t^s(\sigma) = 1$ if $\sigma_t = s$ and 0 otherwise. For any two traders i and j, and for any path σ and date t such that $p_t^i(\sigma), p_t^j(\sigma) \neq 0$,

$$\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \frac{\lambda_j}{\lambda_i}\left(\frac{\beta_j}{\beta_i}\right)^t L_t^{ij} \tag{3}$$

$$\log \frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \log\frac{\lambda_j}{\lambda_i} + t\log\frac{\beta_j}{\beta_i} + \sum_{\tau=0}^{t} Z_\tau^i - Z_\tau^j \tag{4}$$

$$= \log\frac{\lambda_j}{\lambda_i} + t\log\frac{\beta_j}{\beta_i} + \sum_{\tau=0}^{t} Y_\tau^i - Y_\tau^j \tag{5}$$

To see why 4 and 5 are true note that

$$p_t^i(\sigma) = \prod_{\tau=0}^{t} p^i(\sigma_t|\mathcal{F}_{t-1}) = \prod_{\tau=0}^{t}\prod_{s\in S} p^i(s|\mathcal{F}_{t-1})^{1_\tau^s(\sigma)}.$$

3 Belief Selection in Complete Markets

In this section we establish that belief selection is a consequence of Pareto optimality. The intuition is simple: In any optimal allocation of resources, consumers are allocated more in those states they believe to be most likely. Consequently, along those paths which nature identifies as most likely, consumers who believe these paths to be most likely consume the most. From these results on optimal paths, results on the behavior of competitive equilibrium prices and allocations in complete markets follow immediately from the First Fundamental Theorem of Welfare Economics.[3] In contrast, Sandroni's (2000) results come from a direct characterization of equilibrium paths in the markets he studies. Proofs for results in this section and in the remainder of the paper can be found in the Appendix.

Our results are concerned with the long-run behavior of individuals' consumptions along optimal paths. Throughout most of the paper we will make only the coarse distinction between those traders who disappear and those who do not.

Definition 1. *Trader i* vanishes *on path σ iff* $\lim_t c_t^i(\sigma) = 0$. *She* survives *on path σ iff* $\limsup_t c_t^i(\sigma) > 0$.

We want to be clear that, in our view, survival is not a normative concept. We do not favor the ant over the grasshopper. It is not "better" to have a higher discount factor. We do not label as irrational those who say "It's better to burn out than to fade away." We simply observe that they will not have much to do with long run asset prices.

Survival is actually a weak concept. Trader i could survive and consistently consume a large quantity of goods. But trader i surviving and $\liminf c_t^i = 0$ are not inconsistent. And in fact a survivor could consume a significant fraction of goods only a vanishingly small fraction of time. These three different survival experiences have different implications for the role of trader i in the determination of prices in the long run.

By examining equation (3) we can distinguish two distinct analytical problems. When discount factors are identical, a given trader i will vanish when there is another trader j for which the likelihood ratio L_t^{ij} of j's model to i's model diverges. We can analyze this question very precisely. When discount factors differ, we need to compare the likelihood ratio to the geometric series $(\beta_j/\beta_i)^t$. This is more difficult, and our results will be somewhat coarser. In this case our results about the effects of differences in beliefs are phrased in terms of the relative entropy of conditional beliefs with respect to the true beliefs. The *relative entropy* of probability distribution q on S with respect to probability distribution p is defined to be

$$I_p(q) = \sum_{s \in S} p(s) \log \frac{p(s)}{q(s)}$$

It is easy to see that $I_p(q) \geq 0$, is jointly convex in p and q and $I_p(q) = 0$ if and only if $q = p$. In this sense it serves as a measure of distance of probability distributions, although it is not a metric.

3.1 An Example — IID Beliefs

We first demonstrate our analysis for an economy in which the true distribution of states and the forecast distributions are all iid. The distribution of states is given by independent draws from a probability distribution ρ on $S = \{0, 1\}$, and forecasts p^i and p^j are the distributions on infinite sequences of draws induced by iid draws from distributions ρ^i and ρ^j on S, respectively.

This is the leading example for the complete markets results of this paper. Dividing equation (5) by t and taking limits shows that

$$\frac{1}{t}\log\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \frac{1}{t}\log\frac{\lambda_j}{\lambda_i} + \log\frac{\beta_j}{\beta_i} + \frac{1}{t}\sum_{\tau=0}^{t} Y_\tau^i - Y_\tau^j$$

The Y_t^k are iid random variables with a common mean $I_\rho(\rho^k)$, so

$$\frac{1}{t}\sum_{\tau=0}^{t} Y_\tau^k \to I_\rho(\rho^k) \quad p\text{-almost surely.}$$

Consequently,

$$\frac{1}{t}\log\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} \to \Big(\log\beta_j - I_\rho(\rho^j)\Big) - \Big(\log\beta_i - I_\rho(\rho^i)\Big) \tag{6}$$

p-almost surely. If the rhs is positive, the ratio of marginal utilities diverges, and so by Lemma 1, $\lim_t c_t^i(\sigma) \to 0$ almost surely. This says nothing about the consumption of trader j. She may or may not do well, but whatever her fate, i is almost sure to disappear. The expression $\kappa_i = \log\beta_i - I_\rho(\rho^i)$ is a *survival index* that measures the potential for trader i to survive. This analysis shows that a necessary condition for trader i's survival is that her index be maximal in the population.

When traders have identical discount factors, traders with maximal survival indices are those whose forecasts are closest in relative entropy to the truth. A trader with rational expectations survives, and any trader who does not have rational expectations vanishes. When discount factors differ, higher discount factors can offset bad forecasts. A trader with incorrect forecasts may care enough about the future in general that she puts more weight on future consumption even in states which she considers unlikely, than does a trader with correct forecasts who considers those same states likely but cares little about tomorrow.

Maximality of the survival index is not, however, sufficient for survival. The analysis of iid economies in which more than one trader has maximal index is delicate, and we will not pursue this here.[4] The analysis below generalizes these results to dependent processes. This generalization is necessary for an analysis of learning, and selection over learning rules.

Our analysis proceeds in three steps. First we consider the survival of traders with rational expectations in an economy in which all traders have identical discount factors. This analysis differs from the iid example in that we place no assumptions on the true distribution or on forecast distributions. Next we consider the survival of Bayesian and other learners in the identical discount factor case. Finally, we consider the effect of differing discount factors.

3.2 Rational Expectations — Identical β_i

When traders i and j have identical discount factors, equation (3) simplifies to

$$\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \frac{\lambda_j}{\lambda_i} L_t^{ij} \tag{3a}$$

Lemma 1 implies that trader i vanishes on the event $\{\liminf_t L_t^{ij} = +\infty\}$. All analysis of economies in which traders have identical discount factors hinges solely on the behavior of likelihood ratios.

In order to state and prove the results of this section we need to make use of the Lebesgue decomposition of a probability distribution p on Σ with respect to a distribution q. Briefly, there is a measurable set $U \subset \Sigma$, a function $l \in L_1(q)$, and measures $p_{\sim q}$ and p_q such that

1. $q(U) = 0$,
2. $p_{\sim q}(B) = p(B \cap U)$ and $p_q(B) = \int_{B \cap U^c} l\, dq$,
3. for all measurable $B \subset \Sigma$, $p(B) = p_{\sim q}(B) + p_q(B)$,
4. $p_t/q_t \to l$ in the $L_1(q)$-norm on U^c,
5. $p_t/q_t \to \infty$ p-a.s. on U.

We call U the *singular set* for p with respect to q. The identity $p = p_{\sim q} + p_q$ is the *Lebesgue decomposition* of p into a measure $p_{\sim q} \perp q$ (singular with respect to q) and a measure $p_q \prec q$ (absolutely continuous with respect to q).

Our first result gives a survival condition for trader i that is based on absolute continuity of the actual distribution p with respect to the forecast distribution p^i.

Theorem 1. *Assume A.1–3. Trader i survives p-almost surely on U^c, the complement of the singular set for p with respect to p^i.*

That traders with rational expectations survive is an immediate consequence of this theorem.

Definition 2. *Trader i has* rational expectations *if $p^i = p$.*

Corollary 1. *A trader with rational expectations p-almost surely survives.*

Absolute continuity of the actual distribution p with respect to the forecast distribution p^i is a very strong condition. It is more than the absolute continuity of finite horizon marginal distributions. For example, all finite dimensional distributions of the process describing iid coin flips with Heads probability 1/4 are mutually absolutely continuous with those from the process describing iid coin flips with Heads probability 1/3, but clearly the distributions on infinite paths are not absolutely continuous processes since the Heads fraction converges almost surely to 1/4 in one process and 1/3 in the other. The absolute continuity of the truth with respect to beliefs is the *merging* condition of (Blackwell and Dubins 1962), who showed that it has strong implications for the mutual agreement of conditional distributions over time. This kind of condition has been important in the literature on learning in games, and we will have more to say about it in the next section.[5]

Corollary 1 has a converse. In the presence of a trader with rational expectations, absolute continuity of the truth with respect to beliefs is necessary for survival. This fact is a consequence of the following Theorem:

Theorem 2. *Assume A.1–3. Suppose that traders i and j both survive on some set of paths V with $p^j(V) > 0$. Then $p^i(V) > 0$ and the restriction of p^j to V is absolutely continuous with respect to the restriction of p^i to V.*

The assumption that $p^j(V) > 0$ is hardly innocuous. Consider an iid economy in which all discount factors are identical. The trader with beliefs nearest the true distribution survives almost surely. If her beliefs are not accurate, then she assigns probability 0 to the event that she survives.

Corollary 2. *If trader j has rational expectations and i almost surely survives, then p is absolutely continuous with respect to p^i.*

The Corollary does not require that i have rational expectations too. It does require that i's beliefs be not very different from the truth — in particular, i's forecasts of the future are asymptotically p-almost surely correct.

The proofs of the Theorems in this section all rest on the fact that, under the stated hypotheses, trader i survives on the set where the likelihood ratio of j's forecasts to i's forecasts remains bounded. This question is identical to the issue of efficiency in Dawid's (1984) development of *prequential forecasting systems.*

3.3 Bayesian Learning — Identical β_i

In this section we consider the survival possibilities for learners. To see that merging is of at least mild interest for learning, suppose trader i is a Bayesian learner. Models are parametrized by $\theta \in \Theta$, the parameter space. Suppose that $p^\theta = p$ for some $\theta \in \Theta$. If Θ is countable and trader j's prior belief on Θ has full support, then for all $\theta \in \Theta$, p^θ is absolutely continuous with respect to p^i. To see that merging is of at most mild interest, note that if Θ is an open subset of a Euclidean space, absolute continuity will fail. Consider the case of iid coin flips from a coin with unknown parameter $\theta \in (0, 1)$. If a decisionmaker holds a prior belief that is absolutely continuous with respect to Lebesgue measure, no p^θ is absolutely continuous with respect to her belief. She assigns probability 0 to the event that the frequency of Heads converges to θ, but distribution p^θ assigns this event probability 1. Nonetheless it is true that posterior beliefs are *consistent* in the sense that they converge to point mass at θ a.s.-p^θ. Absolute continuity of beliefs with the truth is thus a stronger statement than the claim that traders can learn the truth, at least in a Bayesian context.

Classes of models with richer parameterizations are larger, and therefore more likely to be representative of the world we are trying to model. The countable Θ case, with the possibility of hyper-learning agents, is not particularly interesting. We are led to study models with Θ a bounded, open subset of some Euclidean space. If the parameter is identified, there will be no beliefs which can be absolutely continuous with respect to p^θ for a large set of θ, say a set of positive Lebesgue measure. It is always possible that an economy will have a trader whose beliefs are tightly tuned to a finite or countable subset of parameters. If the "true θ" is in this set, she will survive and wipe out anyone else who is not like her. We dispense with this unlikely possibility, and look for the existence of learning rules that will guarantee almost sure survival for almost all $\theta \in \Theta$.

Our first result in this line is that any Bayesian will survive for almost all θ in the support of her prior.

Theorem 3. *Assume A.1–3. If trader i is a Bayesian with prior belief μ^i on Θ, then she survives for μ^i-almost all θ.*

At first reading the theorem may seem to contradict the results of the previous section. Consider a two person iid economy in which the states are flips of a coin. The true Heads probability is 1/2. One trader has rational expectations — she knows the coin is fair. The other does not. He is a Bayesian whose belief about the parameter can be represented by a uniform prior on $[0, 1]$. His forecast and the true state process are mutually singular. She, the rational trader, survives according to Theorem 1, while he, the learner, vanishes according to Theorem 2. But according to Theorem 3 he survives for almost all θ. The puzzle is resolved by noting that the exceptional set of the Theorem is the singleton $\{1/2\}$. With the forthcoming Theorem 5 one can see that if the first trader's beliefs are fixed, then for any θ except $\theta = 1/2$, the second trader survives and the first trader vanishes.

For Bayesians there is an analogue to Theorem 2 for prior beliefs. Suppose that the parameter can be identified from the data.

A. 4. *There is a partition of Σ into sets Σ_θ such that $p^\theta(\Sigma_\theta) = 1$.*

Theorem 4. *Assume A.1–4. Suppose that some trader j is a Bayesian with prior belief μ^j. Suppose that trader i is a Bayesian, and that trader i survives for μ^j almost all θ. Then trader j's prior belief is absolutely continuous with respect to i's prior belief μ^i.*

More generally, Theorem 2 applies as well. In the presence of a Bayesian, traders who almost surely survive make forecasts that are not too different from Bayesian, although they need not be Bayesians. In particular, their forecasts must merge with the Bayesian's forecasts in the sense of Kalai and Lehrer (1994). An example of such a forecasting rule is suggested by maximum likelihood estimation. Suppose that at each date there are two possible states, 1 and 2. States are distributed iid, and the probability of state 2 is θ. The trader forecasts using the maximum likelihood estimate of θ, $m_t(\sigma) = \sum_{t=0}^{T} (\sigma_t - 1)/(T + 1)$. The MLE converges to the Bayes estimate for the Dirichlet prior with mean 1/2 sufficiently quickly that it satisfies our survival criteria. But the MLE estimate is not Bayes. No Bayes forecast can predict 1 for sure after a finite string of all 1's *and* predict 2 for sure after a finite string of all 2's. We cannot use the MLE as an example however, because this boundary behavior violates Axiom A.3. In the following example we consider a trader investing according to a "trimmed" MLE. She survives in the presence of a Bayesian, and thus is nearly a Bayesian. But our trimmed MLE is not a Bayes forecast because, although it never assigns probabilities 0 or 1, it converges to the boundary faster than any Bayes forecast can on strings of identical observations.

Example: A Non-Bayesian Survivor.
Suppose that trader 1 is a Bayesian with a full support prior, and trader 2's forecasted probability of $s_{t+1} = 2$ given σ_t is $M_t(\sigma)$, which is defined as follows:

Choose $0 < \epsilon < 1$. Let $a_t = 1/(1 + \epsilon^{t^2})$ and $b_t = \epsilon^{t^2}$.

$$M_0(\sigma) = 1/2,$$

$$M_t(\sigma) = \begin{cases} a_t & \text{if } m_t(\sigma) > a_t, \\ m_t(\sigma) & \text{if } a_t \geq m_t(\sigma) \geq b_t, \\ b(t) & \text{if } b_t > m_t(\sigma). \end{cases}$$

This estimator does not take on the values 0 or 1, but in response to a string of all 2's it goes to 1 fast enough that it cannot be Bayes. To see this, suppose it was a Bayes posterior belief for prior μ. Since it can converge to any value in $[0, 1]$, $\operatorname{supp} \mu = [0, 1]$. Let x_t denote the initial segment of length t on the path $(2, 2, \dots)$. Choose $0 < \delta < 1$. Notice that the likelihood function is increasing in θ when the observation is x_t. It is easy to see that

$$\mu\{(\delta, 1]|x_t\} \leq \frac{\mu\{(\delta, 1]\}}{\mu(\delta, 1] + \mu\{[0, \delta]\}\delta^t}.$$

Consequently

$$1 - \mu\{\sigma_{t+1} = 2|x_t\} \geq \frac{\mu\{[0, \delta](1 - \delta)\delta^t}{\mu(\delta, 1] + \mu\{[0, \delta]\}\delta^t}$$

Since under the same conditions the forecasts from z_t converge to probability one on state 2 geometrically in *the square* of t, they cannot be Bayesian forecasts. On the other hand, they are identical with the MLE forecasts as soon as even one 1 is observed (or, in the case of all 1's, one 2), which is a probability one event for all $\theta \in (0, 1)$. Consequently, if θ is interior, trader 2 almost surely survives. □

So far we have shown that a Bayesian will survive almost surely with respect to her prior belief, and that Bayesians who almost surely survive on the same parameter set have mutually absolutely continuous beliefs. This would lead one to conjecture that Bayesians who consider larger model spaces are more likely to survive. This is true insofar as considering more possible models makes it more likely to consider the "true" model. But there is also a cost to a larger model space. We will see next that consideration of more parameters entails slower learning, and so Bayesians with large model spaces are at a disadvantage with respect to those with smaller spaces, when these smaller spaces also contain the true model.

We examine the survival question more closely in a special class of economies in which the process of states $\{\sigma_t\}_{t=0}^{\infty}$ is iid. We have two comments on this assumption. First, it does not imply that endowments or any other observables we may wish to add to the model are iid. It simply means that these processes are built on an iid state space. This is not terribly restrictive. Any process which has only a finite number of possible conditional one-step-ahead distributions can be built on a finite iid state space. (If we did not require a finite state space, *any* process can be built on a state space of iid draws from the uniform distribution on $[0, 1]$.) On the other hand, the iid assumption does have implications for updating rules. Second, we make this assumption because we want to make use of some particular results about the behavior of Bayesian forecasts. These results have been shown to hold for a number of very general stochastic processes, but it is off the point of our paper to establish for exactly which discrete state processes they do hold.[6] Since processes will be iid, we let p^θ refer, as before, to the entire process, and $p(\cdot|\theta)$ refer to the distribution of a single draw. Without further apology we assume

A. 5. *The model set Θ is a bounded open set of a d-dimensional Euclidean space, and the processes p^θ are all iid. For each $\theta_0 \in \Theta$, suppose that for each $s \in S$, $p(s|\theta)$ is C^2 in θ in a neighborhood of θ_0. Suppose too that*

$$E_{\theta_0} \sup_{||\theta-\theta_0||<\delta} \left| \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p(s|\theta) \right| < \infty$$

and

$$E_{\theta_0}\left|\frac{\partial}{\partial\theta_i}\log p(s|\theta_0)\right|^2 < \infty$$

for some $\delta > 0$ *and all* i *and* j *from 1 to* d*.*

The conditions involving derivatives all have to do with how the parameters describe the models. The natural choice is that the parameters are the selection probabilities, that $p(s|\theta) = \theta_s$, and in this case the assumptions are satisfied.

Suppose that a decision maker has prior beliefs which have a density q with respect to Lebesgue measure on Θ, which is continuous and positive at θ_0. Let $I(\theta)$ denote the Fisher information matrix at θ. Let $\rho(\sigma^t) = \int_\Theta p^\theta(\sigma^t)q(\theta)\,d\theta$ denote the predicted distribution of the partial history σ^t. The following Theorem is due to Clarke and Barron (1990):

Clarke and Barron's Theorem. *For all* θ*,*

$$\log\frac{p^\theta(\sigma^t)}{\rho(\sigma^t)} - \left(\frac{d}{2}\log\frac{t}{2\pi} + \frac{1}{2}\log\det I(\theta) - \log q(\theta)\right) \xrightarrow{prob} \chi^2(d)$$

a χ^2 *random variable with* d *degrees of freedom.*

Notice that the conclusion of this theorem holds for all, and not just almost all θ. This result has some interesting consequences for the survival of learners. We already know from Theorem 2 that if trader j knows the truth or has a prior with countable support containing the truth, a Bayesian trader i with a prior absolutely continuous with respect to Lebesgue measure does not almost surely survive. In fact, her belief will be singular with respect to the truth, and so the likelihood ratios L_t^{ij} will almost surely diverge. Consequently, she will almost surely vanish. That is, $\lim_t c_t^i = 0$, p^θ almost surely. But suppose we have no informed or nearly informed trader. Suppose that trader j instead has a prior belief which is concentrated on a lower-dimensional subset of Θ. If she is correct, then in fact she has less to learn than does trader i. The following theorem shows that in this situation, dimension matters. Let Θ' denote an open manifold of dimension $d' < d$ contained in Θ.

Theorem 5. *Assume A.1–3 and A.5. Suppose that trader j has a prior belief which has positive density with respect to Lebesgue measure on Θ', and that trader i has a prior belief with a similar representation on Θ. Then for μ^i-almost all $\theta \in \Theta/\Theta'$, trader j vanishes p^θ-almost surely, while for μ^i-almost all $\theta \in \Theta'$, trader i vanishes in probability.*

Just as low dimensional learning, when right, is better than high dimensional learning, learning in a countable parameter space, when right, is better than learning in a continuum.

Corollary 3. *Suppose traders i and j are Bayesians. The support of trader j's prior belief is a countable set Θ, and trader i's prior belief has a positive density with respect to Lebesgue measure on some bounded open set $\Theta' \supset \Theta$ in some Euclidean space. Then trader i vanishes for all $\theta \in \Theta$.*

In conclusion, we have seen that traders with correct beliefs survive, and if such traders exist, other survivors are not to different from them in the sense that their opinions merge (in the Blackwell-Dubins sense). Absent traders who know the truth, Bayesians survive robustly. That is, for almost all parameters they believe to be possible, they will almost surely survive. Again, other survivors must merge with them. Finally, we can rank-order Bayesians by the dimension of their uncertainty — the dimension of the support of their beliefs. All Bayesians will learn, but when the truth falls into the lower-dimensional supports, those traders will learn faster than those with higher dimensional supports, and enough faster that they drive the high-dimensional traders out. Of course, when the truth does not lie in their supports, they in turn are almost surely driven out.

3.4 Heterogeneous Discount Factors

When discount factors differ, the effect of bad forecasts can be offset by a higher discount factor. One trader may make worse investments than another, but because he consumes less and saves more, he consumes more in the long run. When discount factors differ, our concern becomes the rate at which likelihood ratios diverge. Our main tool for comparing divergence rates with discount factors will be the *relative entropy*, or *Kullback-Leibler distance* between probability distributions. Unfortunately this measure is not well-behaved as probabilities approach 0 and 1, so we we will need an additional boundedness condition on probabilities.

A. 6. *There is a $\delta > 0$ such that for all paths σ, dates t, $s \in S$ and traders i, $p(s|\mathcal{F}_{t-1})(\sigma) > 0$ implies $p(s|\mathcal{F}_{t-1})(\sigma) > \delta$ and $p^i(s|\mathcal{F}_{t-1})(\sigma) > \delta$.*

The analysis of the iid economy in section 3.1 extends generally in the following manner.

Theorem 6. *Assume A.1–3 and A.6. On the event*

$$\left\{\log\frac{\beta_j}{\beta_i} > \limsup_t \frac{1}{t}\left(\sum_{\tau=0}^{t} \mathrm{E}(Y_\tau^j|\mathcal{F}_{\tau-1}) - \sum_{\tau=0}^{t} \mathrm{E}(Y_\tau^i|\mathcal{F}_{\tau-1})\right)\right\}$$

$\limsup_t c_t^i(\sigma) \to 0$ *p-a.s.*

The conditional expectation $E(Y_\tau^k|\mathcal{F}_{\tau-1})$ is the entropy of trader k's forecast about date τ given previous history with respect to the true conditional distribution of date τ's state given history. Consequently this Theorem is a direct generalization of the calculation in equation (6).

One implication of the Theorem is that if traders i and j have the same discount factors and trader i has uniformly less accurate forecasts than does j, then i vanishes. The proof technique of Theorem 6 provides some further characterizations of sample path behavior. If $\limsup$ is replaced with $\liminf$ in the hypothesis, then the same replacement can be made in the conclusion. That is, $\liminf c_t^i = 0$ on the event that the log of the discount factor ratio exceeds the liminf of the difference of the time average of relative entropies.

The following example demonstrates how the Theorem can be used in more general settings than the iid economy.

Example: Markov States.
Suppose that the true distribution of states and all forecasts are Markov. Suppose too that the true distribution of states is ergodic with unique invariant distribution ρ on S. Agent i's (j's) forecasts are represented by a transition matrix P^i (P^j) while the true transition matrix is P. For a given transition matrix Q, let $Q(s)$ denote the row of Q corresponding to state s. In other words, $Q(s)$ is the conditional distribution of tomorrow's state if today's state is s. Suppose that for traders i and j,

$$\mathrm{E}_\rho\, I_{P(\sigma_{t-1})}\big(P^i(\sigma_{t-1})\big) - \log \beta_i > \mathrm{E}_\rho\, I_{P(\sigma_{t-1})}\big(P^j(\sigma_{t-1})\big) - \log \beta_j$$

Since the Markov process of states is ergodic, $(1/t)\sum_{\tau=0}^{t} Z_\tau^k$ converges to the expectation of the relative entropy of k's conditional forecasts with respect to the true conditional distribution under the invariant distribution. Consequently, $\lim_t c_t^i = 0$ almost surely. If the Markov process is not ergodic, one carries out this exercise on each communication class. □

Theorem 6 lets us explore the tradeoff between learning speed and discount factors. We saw that when discount factors were identical, not all successful learners survived. The market favored faster learners. This is not true when discount factors differ. The following result does not require an iid state process or other process where learning rates can be measured.

Corollary 4. *Suppose Axioms 1–3 and 6. Let p^θ denote the true state process. Suppose too that traders i and j are Bayesians and for both of them, posterior distributions converge to point mass at θ p^θ-almost surely. If $\beta_j > \beta_i$, then trader i almost surely vanishes.*

The theorem says that when discount factors differ, different rates at which Bayesians learn are irrelevant to survival. Trader i could know the true distribution while trader j could be updating on a high-dimensional parameter space. In the iid case, the dimension of the parameter spaces is irrelevant here because the effects of differences in dimension are of order $\log t$, while the effect of discount factors is of order t. But this result does not rely on the $\log t$ rate. Since the relative entropies with respect to p of conditional forecasts almost surely converge, the time average of their difference is 0, and so the linear-in-time effect of discount rate differences determines survival.

The conclusions of Theorem 6 are false without the uniform bounds. The following example answers in the negative a question raised by Sandroni (2000) about the possibility of doing without uniform bounds across dates and states on ratios of forecasted and true state probabilities for results involving entropy calculations.

Example: The Need for A.6.
Consider a two person exchange economy with two traders, i and j. At each date there are two states, s_a and s_b. To save on notation, the economy begins at date 1. States are drawn independently over time, and at date t the probability of s_a is $1 - 1/t^2$ and the probability of s_b is $1/t^2$. Traders' utility functions satisfy A.1, and endowments are fixed at $e > 0$, and are independent of state. But traders have different forecasts. At date t trader i assigns probability $(\exp t^3 - 1)/(\exp t^3 - \exp -t)$ to state a, and trader j assigns probability $(\exp t^2 - \exp -t)/(\exp t^3 - \exp -t)$ to state a. Thus $Z_t^i - Z_t^j$ takes on the value $-t$ in state s_a and t^3 in state s_b. The entropy difference is $I_{p_t}(p^i) - I_{p_t}(p^j) = 1/t$, and so the series $\sum_{t=1}^{\infty} I_{p_t}(p^i) - I_{p_t}(p^j)$ diverges to $+\infty$. If the conclusions of Theorem 6 were true, trader i would disappear. Nonetheless the sum $\sum_{t=0}^{\infty} Z_t^i - Z_t^j$ diverges quickly to $-\infty$, implying that j disappears and, consequently, that i does not. That is, $c_t^j \to 0$ almost surely, and $c_t^i \to 2e$. To see why this is true, observe that $\{\sum_t Z_t^i - Z_t^j \to -\infty\}^c \subset \{s_t = s_b \quad i.o.\}$, and according to the Borel-Cantelli Lemma, this is a 0 probability event since $\sum_{t=1}^{\infty} 1/t^2$ converges. The magnitude of $\sum_{\tau=1}^{t} Z_\tau^i - Z_\tau^j$ grows at rate $O(t^2)$, and so trader i will survive no matter how small her discount factor and how large trader j's.

Intuitively, we should expect i to survive, as indeed she does. The probability of state s_a's occurrence is converging to 1, as is i's belief about s_a, while the probability j assigns to a is converging to 0. But trader i overshoots the mark, giving her the larger relative entropy with respect to the truth. This is possible even though trader i is forecasting more accurately than is j in any intuitive sense because of the asymmetry of the relative entropy function, which becomes extreme as the true distribution assigns negligible probability to some states. □

4 Belief Selection in Incomplete Markets

Results in the previous section showed that a form of belief selection is a consequence of Pareto optimality. The first welfare theorem can be used to draw implications for prices when markets are complete. When markets are incomplete, optimality no longer characterizes equilibrium allocations, and the belief selection properties of market equilibrium must be investigated directly. In this section we build two examples to show that the strong belief selection properties exhibited by complete markets fail in incomplete markets. They fail for two reasons, one trivial, one less so. The trivial reason, demonstrated in the second example, is that entropy does not match well with the asset structure — one distribution could be very far from the true distribution in ways that are totally irrelevant to the equilibrium investment problem, while another distribution could be quite near, but differ from the truth in ways which are critical. The less trivial reason for the failure of the market selection hypothesis has to do with savings behavior. Undue optimism or pessimism (depending upon the payoff function) can lead to excessive saving, so that the investor with the worst beliefs will come to dominate the market over time.

Example: Savings Effects.
The story of the first example is that two traders buy an asset from a third trader. The two traders hold different beliefs about the return of the asset. Trader 1 is correct, while trader 2 consistently overestimates the return.

At each date there are two states: $S = \{s_1, s_2\}$. The true evolution of states has state 1 surely happening every day. There is a single asset available at each date and state which pays off in consumption good in the next period an amount which depends upon next period's state. The asset available at date t pays off, at date $t+1$,

$$R_t(\sigma) = \begin{cases} \Big(1 + \big(\tfrac{1}{2}\big)^t\Big) & \text{if } \sigma_t = s_1, \\ \\ 2\Big(1 + \big(\tfrac{1}{2}\big)^t\Big) & \text{if } \sigma_t = s_2. \end{cases}$$

Traders 1 and 2 have CRRA utility with coefficient 1/2. Trader 3 has logarithmic utility. Traders 1 and 2 have common discount factor $(8)^{-\frac{1}{2}}$ and trader 3 has discount factor 1/2. Traders 1 and 3 believe correctly that state s_1 will always occur with probability 1, and trader 2 incorrectly believes that state s_2 will always occur with probability 1. All three traders have endowments which vary with time but not state. Traders 1 and 3 know the correct price sequence. Trader 2 believes that at each date, state s_2-prices will equal the (correct) state s_1 prices. Traders 1 and 2 have endowment stream $(1, 0, 0, \dots)$. Trader 3's endowment stream is $e_1^3 = 0$ and $e_t^3 = 5/4 + 3(1/2)^{t-1}$ for $t > 1$.

This model has an equilibrium in which the price of the asset (in terms of the consumption good) is, for every state,

$$q_t = \frac{1}{2}\left(1 + \left(\frac{1}{2}\right)^t\right).$$

In this equilibrium, at each date trader 3 supplies 1 unit of asset and traders 1 and 2 collectively demand 1 unit of asset. Trader 1's wealth at date t is $(1/2)^{t-1}$ and at each date he consumes 3/4 of this wealth. Trader 2's wealth is 1 at each date and at each date he consumes 1/2 of this wealth. Trader 3 consumes 3/4 at each date. So trader 1's wealth and consumption converges to 0 even though he has correct beliefs and trader 2 has incorrect beliefs. Although the details of the example are complicated, the intuition is simple. At each date, trader 1 believes that the rate of return on the asset is 2, while trader 2 believes it is 4. Trader 2's excessive optimism causes him to save more at each date, so in the end he drives out trader 1.

It is more enlightening to understand how this example was constructed than it is to go through the details of verification of the equilibrium claim. We constructed it as follows: Our idea was to fix some facts that would allow us to solve the traders' Euler equations, and then to derive parameter values that would generate those facts. Accordingly, we fixed the gross rates of return on the asset at 2 and 4 in states s_1 and s_2, respectively. We also assumed that traders 1 and 2's total asset demand would be 1. For an arbitrary gross return sequence the Euler equations pinned down prices. We then turned to the supply side and chose an endowment stream for trader 3 and a gross return sequence that would cause trader 3 to supply 1 unit of asset to the market at each date. □

In our example the over-optimistic trader drives out the trader on the same side of the market with rational expectations. If markets were complete, trader 1 would be able to bet with trader 2, and his more accurate forecasts would cause him to systematically benefit at trader 2's expense.

It is tempting to conclude that overly optimistic beliefs play a systematic role. In fact there is no general theorem here. Our method also allows us to construct examples for other risk aversion parameters. When the risk aversion coefficient is negative, optimism causes under-saving rather than over-saving. To drive out the rational trader in this case, the other trader would have to be overly-pessimistic.

Finally, this example illustrates the difference between "fitness" and happiness. Both traders have identical payoff functions and discount factors, and so there is some sense to asking who envies whom. Clearly trader 2 prefers the present discounted value of trader 1's realized utility stream to his own. Trader 2 is accumulating wealth share because he is consuming less than trader 1. He prospers through excessive saving. We see in this example a clear disconnect between utility maximization and survival.

The failure of the market selection hypothesis in this example is due to inefficient intertemporal allocation. Blume and Easley (1992) forced agents to have identical savings behavior and studied the effects of selection on portfolio choices. The next example shows how, even when investors have identical savings behavior, portfolio effects can cause incorrect beliefs to survive and even prosper.

Example: Portfolio Choice Effects.
Consider an economy with two assets and 3 states. Asset 1 pays off 1 unit of the consumption good in state 1 and 0 in the other two states. Asset 2 pays off 0 in state 1, but 1 unit in each of states 2 and 3. There are three traders with logarithmic payoff functions and common discount factor β. The state probabilities and beliefs are described in the following table:

	states		
	s_1	s_2	s_3
truth	$1/2$	$1/2-\epsilon$	ϵ
trader 1	$1/2$	$1/2-\epsilon$	ϵ
trader 2	$1/2$	ϵ	$1/2-\epsilon$
trader 3	$1/2$	$1/2-\epsilon$	ϵ

The parameter $\epsilon > 0$ is small. Traders 1 and 3 have rational expectations, while trader 2 does not. As before, trader 3's role is to sell assets to traders 1 and 2. Traders 1 and 2 have a state-independent endowment: $e_1^i = 1/2$ and $e_t^i \equiv 0$ for $t > 1$. Trader 3's endowment is also state independent: $e_1^3 = 0$ and $e_t^3 = 1$ for $t > 0$.

There is an equilibrium such that for all t and σ, $q_t^1(\sigma) = q_t^2(\sigma) = \beta/2$. In equilibrium trader 3 supplies 1 unit of each asset. Each trader attributes to asset 2 the same distribution of returns, and so both traders hold identical amounts of both assets. Consequently the distribution of wealth between traders 1 and 2 remains unchanged.

To push this point farther, consider the following configuration of beliefs where $\delta > 0$ is small:

	states		
	s_1	s_2	s_3
truth	$1/2$	$1/2-\epsilon$	ϵ
trader 1	$(1-\delta)/2$	$(1/2-\epsilon)(1+\delta)$	$\epsilon(1+\delta)$
trader 2	$1/2$	ϵ	$1/2-\epsilon$

Because the traders have logarithmic utility, their portfolios maximize their expected growth rates of wealth. Trader 1 has slightly incorrect beliefs while trader 2 has grossly incorrect beliefs. But trader 2's beliefs lead him to make the same decisions that a trader knowing the truth would make, while trader 1 will do something else. Consequently trader 1 will vanish and trader 2 will dominate the market. □

In this example, relative entropy is simply the wrong measure. What we care about in the nearness of the chosen portfolio to the log-optimal portfolio at each date. When markets are complete this is measured by relative entropy of forecasts with respect to the truth. When markets are incomplete these measures do not coincide.

5 Conclusion

According to conventional wisdom, general equilibirum theory is devoid of positive implications. Our analysis shows that this is not correct. We do not claim that the Debreu-Mantel-Mas Colell-Sonnenschein Theorem is wrong. Rather our argument is that in economies with dynamically complete markets, there is structure to long run outcomes, and this structure does not require homogeniety of agents' preferences and beliefs. Instead, in some sense we get more structure the richer is the set of agent types in the economy; as then it is more likely that someone will for whatever reason have beliefs that are closer to rational expectations.

One could criticize this claim for its focus on the long run. But how long is the long run? The appropriate time scale is governed by the frequency of transactions. In some markets, such as housing and labor markets, individuals transact infrequently and our results do not have anything useful to say. In other markets, such as financial markets, transaction frequencies are high and it is plausible that the long run happens very quickly.

Here our focus is on which traders survive, but our results also have implications for long run asset pricing. In dynamically complete markets economies in which all traders have a common discount factor, if one trader has correct beliefs, then in the long run all assets will be priced efficiently. In fact, they will be priced using this trader's preferences and his correct beliefs. This claim stands in direct contradiction to much of the rapidly growing behavioral economics and behavioral finance literature. In that literature, traders are often assumed to behave irrationally in that they maximize expected utility with incorrect beliefs which are updated according to various psychologically motivated rules. Of course such irrational behavior exists, but economists used to believe that it did not matter for asset market aggregates such as prices. Our analysis shows that this old idea is correct in some settings and not correct in others. What matters is the completeness of markets. If markets are dynamically complete then irrationality cannot survive. If markets are incomplete then irrationality may be able to survive. But in any case there are selection pressures that cannot be ignored. Not all behaviors can survive in all market settings.

Appendix

Proof of Theorem 1. Choose an arbitrary trader j. From (3),

$$\begin{aligned}\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} &= \frac{\lambda_j}{\lambda_i}\frac{p_t^j(\sigma)}{p_t^i(\sigma)} \\ &= \frac{\lambda_j}{\lambda_i}\frac{p_t^j(\sigma)}{p_t(\sigma)}\frac{p_t(\sigma)}{p_t^i(\sigma)} \qquad (*)\end{aligned}$$

The first ratio on the rhs of $(*)$ is a fixed number. The limit of the second ratio is finite p-a.s. since the p_t^j are probability measures. The third ratio converges to a finite limit on U^c. So on U^c Lemma 1's necessary condition for vanishing fails to hold. □

Proof of Theorem 2. From equation (3),

$$\frac{\lambda_i}{\lambda_j}\frac{u^{i\prime}(c_t^i)}{u^{j\prime}(c_t^j)} = \frac{p_t^j}{p_t^i} = L_t^{ij} \qquad (7)$$

Suppose there is a measurable subset A of V such that $p^j(A) > 0$ and $p^i(A) = 0$. Then there is a measurable set $B \subset A$ such that $p^j(B) = p^j(A) > 0$ and $p^j_t(\sigma)/p^i_t(\sigma) \to \infty$ for all $\sigma \in B$. Consequently equation (7) implies that $c^i_t(\sigma) \to 0$ on B, which contradicts the hypothesis. □

Proof of Theorem 3. It is sufficient to show that the necessary condition for vanishing is p^θ-almost never met for ν-almost all θ. Consider trader j. The likelihood ratios L^{ij}_t are a non-negative martingale with mean 1 under p^i, and so they converge p^i-almost surely. Since p^i is a mixture of the p^θ, $p^\theta(\limsup_i L^{ij}_t = \infty) = 0$ for all but a set of zero measure with respect to trader i's prior belief. Extending this to all $j \neq i$, the necessary condition for vanishing fails as required. □

Proof of Theorem 4. This is an immediate consequence of Theorem 2. Observe that if the conclusion of Theorem 4 were false, then Axiom 4 implies that p^j is not absolutely continuous with respect to p^i, contradicting Theorem 2. □

Proof of Theorem 5. Consider the form of the first order conditions given in equation (5). Since discount factors are identical, this becomes

$$\log \frac{u^{i\prime}\big(c^i_t(\sigma)\big)}{u^{j\prime}\big(c^j_t(\sigma)\big)} = \log \frac{p^\theta_t(\sigma)}{p^i_t(\sigma)} - \log \frac{p^\theta_t(\sigma)}{p^j_t(\sigma)} \tag{$*$}$$

Choose $\delta \in (0, 1/2)$. From Clarke and Barron's Theorem we can assert that for almost all $\theta \in \Theta'$, for all $\epsilon > 0$ there is a T such tha for all $t > T$,

$$p^\theta \left\{ \log \frac{p^\theta_t(\sigma)}{p^i_t(\sigma)} < \frac{(d-\delta)}{2} \log t \right\} < \epsilon$$

and

$$p^\theta \left\{ \log \frac{p^\theta_t(\sigma)}{p^j_t(\sigma)} > \frac{(d'+\delta)}{2} \log t \right\} < \epsilon$$

Consequently for all $B > 0$ and $\epsilon > 0$ there is a T such that for $t \geq T$, $p^\theta\{u^{i'}/u^{j'} < B\} < \epsilon$. In other words, $u^{i'}/u^{j'} \uparrow \infty$ in probability, and so c_t^i converges to 0 in probability.

For almost all $\theta \in \Theta/\Theta'$, trader i will survive almost surely according to Theorem 3. To see that trader j vanishes, expand the logarithms in equation $(*)$ and divide by t. The right hand side becomes

$$\frac{1}{t}\sum_{\tau=0}^{t}\Big(\log p(\sigma_\tau|\theta) - \log p_\tau^i(\sigma_\tau|\sigma^{\tau-1})\Big) - \frac{1}{t}\sum_{\tau=0}^{t}\Big(\log p(\sigma_\tau|\theta) - \log p_\tau^j(\sigma_\tau|\sigma^{\tau-1})\Big)$$

Applying the SLLN, the first term in each parenthetical expression converges to $\sum_s p(s|\theta)\log p(s|\theta)$, the entropy of $p(\cdot|\theta)$.

Because Bayes learning is consistent for trader i, $\lim_t p^i(s|\sigma^t)$ converges p^θ-almost surely to $p(s|\theta)$. Therefore the time average of this term also converges to the entropy of $p(\cdot|\theta)$, and so the expression for trader i converges almost surely to 0. Bayes learning is inconsistent for trader j since the truth is outside the support of her prior belief. The standard convergence argument for the consistence of Bayes estimates from an iid sample shows in this case that the support of posterior beliefs converges upon those which minimize the relative entropy $\sum_s p(s|\theta)\log p(s|\theta)/p(s|\theta')$ over $\theta' \in \Theta'$. Suppose wlog that $\theta \notin \operatorname{cl}\Theta'$. Then this minimum is some $K > 0$. Consequently trader j's term converges to K. Thus $(1/t)\log u^{i'}(c_t^i)/u^{j'}(c_t^j)$ converges almost surely to $-K < 0$, and so $\log u^{j'}(c_t^j)/u^{i'}(c_t^i)$ converges almost surely to ∞. We see from Lemma 1 that trader j disappears almost surely. □

Proof of Corollary 3. Trader j's beliefs are absolutely continuous with respect to the truth. Axiom 4 implies that p^i and p^θ are singular, and so the claim follows from Theorem 2. □

Proof of Theorem 6. The following Lemma describes the behavior of the random variables Y_t^k. These results are necessary for the proof of Theorem 6 and are also of independent interest as a step towards proving other limit results.

Lemma 2. *Suppose Axioms 1–3 and 6.*

1. *On the set* $\{\sigma : \sum_t \mathrm{E}(Y_t^i|\mathcal{F}_{t-1}) < \infty\}$, $\sum_t Y_t^i$ *converges to a finite limit p-a.s.*

2. *On the set* $\{\sigma : \sum_t \mathrm{E}(Y_t^i|\mathcal{F}_{t-1}) = +\infty\}$,

$$\lim_{t\to\infty} \frac{1}{t}\left(\sum_{\tau=1}^{t} Y_\tau^i - \mathrm{E}(Y_\tau^i|\mathcal{F}_{\tau-1})\right) = 0 \qquad p\text{-}a.s.$$

Proof of Lemma 2. Observe that $\mathrm{E}(Y_\tau^k|\mathcal{F}_{\tau-1})(\sigma)$ is the relative entropy of $p^k(\cdot|\sigma^t)$ with respect to $p(\cdot|\sigma^t)$.

Fix a support $S' \subset S$, a (true) probability distribution ρ on S and a belief q on S which are both elements of $B_\epsilon(S')$, the set of all probability distributions q on S which assigns mass at least $\epsilon > 0$ to every state in the support of p. Then $\mathrm{E}(\log\rho - \log q) = I_\rho(q)$ and $\mathrm{Var}(\log\rho - \log q)$ are both bounded, non-negative, C^2 functions of q which take the value of 0 if and only if $q = \rho$. Furthermore, both are strictly convex at $q = \rho$, and so there are bounds $k_\rho, K_\rho > 0$ such that $k_\rho \mathrm{E}(\log\rho - \log q) \le \mathrm{Var}(\log\rho - \log q) \le K_\rho \mathrm{E}(\log\rho - \log q)$. These bounds are uniform on $B_\epsilon(S')$, and since S is finite, they can be chosen uniformly over all $S' \subset S$. According to axiom 6, each $p_t(s|\sigma^{t-1})$ and $p_t^i(s|\sigma^{t-1})$ are in one of the sets $B_\epsilon(S')$.

To prove the first claim, note that the bounds imply $\mathrm{Var}(Y_t^i|\mathcal{F}_{t-1}) \le K\,\mathrm{E}(Y_t^i|\mathcal{F}_{t-1})$. Summability of the conditional means implies the summability of the conditional variances. Kolmogorov's inequality implies that the sum $\sum_{t=1}^\infty Y_t^i - \mathrm{E}(Y_t^i|\mathcal{F}_{t-1})$ is almost surely finite, from which the result follows.

To prove claim 2, note that a consequence of the bounds is that if $\sum_t \mathrm{E}(Y_t^i|\mathcal{F}_{t-1}) = +\infty$, then $\sum_t \mathrm{Var}(Y_t^i|\mathcal{F}_{t-1}) = +\infty$. It follows from a Theorem of Neveu (1972, p. 150) that

$$\lim_{t\to\infty} \frac{\sum_{\tau=1}^t Y_\tau^i - \mathrm{E}(Y^i|\mathcal{F}_{\tau-1})}{\sum_{\tau=1}^t \mathrm{Var}(Y_\tau^i|\mathcal{F}_{\tau-1})} = 0 \qquad p\text{-a.s.}$$

Replacing the variances in the denominator by their upper bound and multiplying both sides by that bound gives the result. □

Now we prove the Theorem. In either case of the Lemma,

$$\frac{1}{t}\left(\sum_{\tau=0}^{t} Y_\tau^i - \mathrm{E}(Y_\tau^i|\mathcal{F}_{t-1})\right) = 0 \tag{$*$}$$

Dividing equation (5) by t gives

$$\frac{1}{t}\log\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \frac{1}{t}\log\frac{\lambda_j}{\lambda_i} + \log\frac{\beta_j}{\beta_i} + \frac{1}{t}\left(\sum_{\tau=0}^{t} Y_\tau^i - Y_\tau^j\right) \tag{$**$}$$

According to equation ($*$), we can replace each Y_τ^i in ($**$) with its conditional expected value, and the result follows. □

Proof of Corollary 4. The consistency of Bayesian updating implies that for p^θ-almost all σ, $p^k(s|\sigma^t) - p^\theta(s|\sigma^t) \to 0$ for all s and $k = i, j$. For both traders, then, the relative entropy of tomorrow's forecast relative to the true distribution of states tomorrow, $\mathrm{E}(Y_t^k|\mathcal{F}_{t-1})(\sigma)$, converges to 0 almost surely. Time averages converge to 0, and so

$$\frac{1}{t}\log\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} \to \log\frac{\beta_j}{\beta_i} > 0$$

p^θ-almost surely. From Lemma 1 conclude that trader i vanishes. □

Notes

[1]This does not contradict the previous statement about survival of Bayesians as the lower dimensional set has prior measure zero for the Bayesian with higher dimensional support.

[2] One implication of this construction is that *every* forecasting rule is Bayesian with respect to some prior belief. Given the sequence of prediction functions, construct the stochastic process of states as described. Then an individual who was sure this process was the true process — a special case of a point mass prior — would have these forecasting rules as conditional distribution of tomorrow's state given information available through today. The Bayesian assumption by itself is quite weak. Its power comes from ancillary assumptions about the set of processes which could be generating the data, which are embedded in the prior beliefs on processes, which this construction would not satisfy.

[3]We characterize Pareto optimal allocations in which each trader is allocated at some time, on some path, some of the good. So we only characterize competitive equilibrium allocations in which each trader's endowment has positive value. A trader whose endowment has zero value clearly has no effect on the economy and we ignore such traders.

[4]For a complete discussion of the iid economy, see (Blume and Easley 2000).

[5]Kalai and Lehrer (1994) is an excellent discussion of the implications of merging.

[6]See Dawid (1984) and Ploberger and Phillips (1998). The key requirements are suitable differentiability of the model and asymptotic normality of the maximum likelihood estimator of the parameters, which are certainly quite broad.

References

Alchian, A. (1950): "Uncertainty, Evolution and Economic Theory," *Journal of Political Economy*, 58, 211–221.

Blackwell, D., and L. Dubins (1962): "Merging of Opinions with Increasing Information," *Ann. Math. Statist.*, pp. 882–86.

Blume, L., M. Bray, and D. Easley (1982): "Introduction to the Stability of Rational Expectations," *Journal of Economic Theory*, 26(2), 313–17.

Blume, L., and D. Easley (1992): "Evolution and Market Behavior," *Journal of Economic Theory*, 58(1), 9–40.

——— (1998a): "Optimality and Natural Selection in Markets," forthcoming, *Journal of Economic Theory.*

——— (1998b): "Rational Expectations and Rational Learning," in *Organizations with Incomplete Information: Essays in Economic Analysis*, ed. by M. Majumdar, pp. 61–109. Cambridge University Press, Cambridge, UK.

——— (2000): "Optimal Paths in Markets with Long-Run Consumer Heterogeneity," unpublished, Cornell University.

Clarke, B. S., and A. R. Barron (1990): "Information Theoretic Asymptotics of Bayes Mathods," *IEEE Transactions on Information Theory*, 36(3), 453–71.

Cootner, P. (1964): *The Random Character of Stock Market Prices.* MIT Press, Cambridge, MA.

Dawid, A. P. (1984): "Present Position and Potential Development: Some Personal Views: Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society. Series A*, 147(2), 278–92.

DeLong, J. B., A. Shleifer, L. Summers, and R. Waldman (1990): "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98(4), 703–38.

——— (1991): "The Survival of Noise Traders in Financial Markets," *Journal of Business*, 64(1), 1–19.

Fama, E. (1965): "The Behavior of Stock Market Prices," *Journal of Business*, 38(1), 34–105.

Friedman, M. (1953): *Essays in Positive Economics*. University of Chicago Press, Chicago.

Halmos, P. R. (1974): *Measure Theory*, Graduate Texts in Mathematics, 18. Springer-Verlag.

Kalai, E., and E. Lehrer (1994): "Weak and strong merging of opinions," *J. Math. Econom.*, 23(1), 73–86.

Marimon, R. (1997): "Learning from Learning in Economics," in *Advances in Economics and Econometrics: Theory and Applications, Volume I, Seventh World Conggress*, ed. by D. Kreps, and K. Wallis, pp. 61–109. Cambridge University Press, Cambridge, UK.

Neveu, J. (1972): *Martingales à Temps Discret*. Masson, Paris.

Ploberger, W., and P. C. B. Phillips (1998): "Empirical Limits for Time Series Econometrics Models," unpublished, Yale University, http://butterfly.econ.yale.edu/P/cd/d12a/d1220.pdf.

Sandroni, A. (2000): "Do Markets Favor Agents Able to Make Accurate Predictions," *Econometrica*, 68(6), 1303–42.