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

July 2002
Today: June 24, 2004

The authors thank YoungMiao Hong, Carsten Krabbe Nielsen, Werner Ploberger, Andrew Postlewaite, Alvaro Sandroni, Bill Zame, two anonymous referees for *Econometrica*, and seminar audiences at the University of Arizona, UCSB, Cornell University, Duke University, Erasmus University, The Hebrew University of Jerusalem, the University of Iowa, Johns Hopkins University, Notre Dame, 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 grants SES 9808690, SES0136307 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 Pareto optimal 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 their parameter space. Finally, we show that in economies with incomplete markets these conclusions may not hold. With incomplete markets payoff functions can matter for long run survival, and the market selection hypothesis fails.

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 explanations have been offered. One proposes 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 (1998) and Marimon (1997) for more on learning and its limits.

The other explanation posits "natural selection" in market dynamics. The *market selection hypothesis*, that markets favor rational traders over irrational traders, 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 has been 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 selection for rational expectations. He showed in a Lucas trees economy with some rational-expectations traders 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.

Sandroni's (2000) analysis stands in sharp contrast to that of DeLong, Shleifer, Summers and Waldman (1990,1991). Why is it that in one setting traders with rational expectations are selected for and in the other setting they are driven out of the market by those with irrational expectations? Answering this question, and in general understanding more completely when selection for rational traders occurs and when it does not, requires a more general analysis than that in the earlier papers. DeLong, Shleifer, Summers and Waldman do not undertake a full equilibrium analysis.[1] In their model attitudes toward risk and beliefs both matter in determining who will survive. Sandroni (2000) analyses a full general equilibrium model, with only Lucas trees for assets. In his world attitudes toward risk have no effect on survival.[2]

We show here that Pareto optimality is the key to understanding selection for or against traders with rational expectations. For any optimal allocation, the survival or disappearance of a trader is determined entirely by discount factors and beliefs. Attitudes toward risk are irrelevant to the long run fate of traders in optimal allocations. In particular, controlling for discount factors, only those traders with correct expectations survive. The first theorem of welfare economics implies that correct beliefs are selected for whenever markets are complete. This conclusion is robust to the asset structure, of course, so long as markets are complete at the equilibrium prices. Sandroni's (2000) analysis thus fits into this setting.[3] Evidently too, DeLong, Shleifer, Summers and Waldman's allocations are not Pareto optimal.

In economies with incomplete markets, opportunities for trade may be restricted; consequently, equilibrium allocations need not be Pareto optimal. We show that when markets are incomplete, the market selection hypothesis may fail. Discount factors, attitudes toward risk and beliefs all matter. Even when there is a common discount factor, traders with incorrect beliefs may drive out those with more accurate beliefs. We show that 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. We also show that a trader who is overly pessimistic 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. Whether optimism or pessimism leads to greater savings depends on the individual's utility function so there is no general result about selection for one type or the other.

If the market selection hypothesis addressed only selection for traders with correct beliefs, then it would be of little interest. We demonstrate its broader applicability by characterizing selection over learning rules in complete markets. In studying learning we first assume that all traders have the same discount factor. We show that a Bayesian almost surely survives for almost all possible truths in the support of her prior. Put somewhat differently, each Bayesian trader is sure that she will survive. Furthermore, in the presence of a Bayesian trader, any traders who survive are not too different from Bayesians. They use a forecasting rule that asymptotically looks like a Bayes forecasting rule. We also show that the market selects over Bayesians according to the size of the support of their prior beliefs. We consider Bayesians 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.[4] Thus, having more prior information favors the possibility of survival only if it affects the dimension of the support of prior beliefs. Our analysis of learning provides a rate of divergence (over time) of the Bayesian's marginal distributions on partial histories from the correct marginal distribution. We use this rate result to analyze the joint effect of differences in priors and discount factors. We show that among Bayesians whose prior belief supports contain the truth, the survivors are always among the most patient Bayesians, even if they all have large prior supports relative to less patient traders.

We conclude that for economies with complete or dynamically complete markets, and a common discount factor the market selection hypothesis hypothesis is correct. If any trader has correct beliefs that trader is selected for.[5] If there is a trader who is a Bayesian with the truth in the support of her prior, then all traders who survive will have asymptotically correct beliefs. So a more robust form of the market selection hypothesis than that posited by Alchian (1950), Friedman (1953) and Fama (1965) is true. They claim that markets select for traders with correct beliefs. But what happens when no trader has correct beliefs? We show that in the absence of traders with correct beliefs, the market selects for those traders whose beliefs are closest to correct in a sense that is formalized in section 3. In particular, markets favor learners over those traders who are persistently wrong. Among the class of learning traders, markets favor those traders who learn quickest. (These need not be the traders with the most correct initial beliefs!) We provide in section 3.1 for the iid economy and in section 3.4 for the general case a criterion for survival; a measure of economic "fitness". This criterion shows how discount factors and correctness of beliefs interact to determine survival. The aforementioned claims are applications of these criteria.

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 $U^i : c \mapsto [-\infty, \infty)$ which is the expected presented discounted value of some payoff stream with respect to some beliefs.[6] 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 $u^i : \mathbf{R}_+ \to [-\infty, \infty)$ 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, $p_t^i(\sigma) = \prod_{\tau=0}^{t} \rho(\sigma_\tau)$. Markov models and other, more complicated stochastic processes can be accommodated as well.

Individuals who learn from the past using a statistical learning rule also have beliefs which fit into our framework.[7] A statistical learning rule is a sequence of $\mathcal{F}_t$-measurable functions which assign to each history of states through t a probability distribution of states in period $t+1$. Time 1 beliefs are simply a distribution on S. Time 1 beliefs 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 belief. 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) implies 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 for our analysis because Pareto optimality implies that consumption at any partial history should be 0 if that partial history is an impossible (measure 0) event.[8]

We assume 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; that is, $u^{i\prime}(c) \to \infty$ as $c \downarrow 0$.*

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

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

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 $p^i_t(\sigma) > 0$.*

If $p^i_t(\sigma) = 0$ 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. The Inada condition implies that each trader's allocation in any Pareto optimum is either p^i almost surely positive 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 positive. If $c^* = (c^{1*}, \ldots, c^{I*})$ is such a Pareto optimal allocation of resources, then there is a vector of welfare weights $(\lambda_1, \ldots, \lambda_I) \gg \mathbf{0}$ such that c^* solves the problem

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

where $e_t = \sum_i e^i_t$. 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 $p^i_t(\sigma) > 0$, then

$$\lambda_i \beta_i^t u^{i\prime}\bigl(c^i_t(\sigma)\bigr)p^i_t(\sigma) - \eta_t(\sigma) = 0 \tag{2}$$

(ii) If $p^i_t(\sigma) = 0$, then $c^i_t(\sigma) = 0$.

These conditions will be used to characterize the long-run behavior of consumption plans for individuals with different payoff functions, 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 (2002), and to the equilibrium conditions for an asset model by Sandroni (2000).[9]

All of our results are based on the following simple idea:

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

Proof. $u^{i\prime}\bigl(c^i_t(\sigma)\bigr)/u^{j\prime}\bigl(c^j_t(\sigma)\bigr) \to \infty$ 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 $c^i_t \downarrow 0$. 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 $u^{j\prime}\bigl(c^j_t(\sigma)\bigr) \leq u^{j\prime}(f/I)$ infinitely often. If $\lim_t c^i_t = 0$, then $\limsup u^{i\prime}\bigl(c^i_t(\sigma)\bigr)/u^{j\prime}\bigl(c^j_t(\sigma)\bigr) = \infty$. □

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. For generic consumers i, j and k, with forecasts p^i, p^j and p^k, define the following random variables

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

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

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

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

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

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

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 Theorem of Welfare Economics.[10] 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 Appendix B.

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 $S = \{1, 2\}$, 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.

Dividing equation (5) by t yields

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

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,

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

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.

This analysis also shows that the ratio of marginal utilities of traders i and j converges exponentially at rate equal to the difference of the traders survival indices. If we specify a form for the utility functions we can determine the rate of convergence of consumptions.

Example: Risk Aversion and the Rate of Convergence.
Suppose that traders i and j each have utility function $u^i(c) = \alpha^{-1}c^\alpha$ with $\alpha < 1$. In this case equation (6) simplifies to

$$\frac{1}{t} \log \frac{c_t^i(\sigma)^{\alpha-1}}{c_t^j(\sigma)^{\alpha-1}} \to \kappa_j - \kappa_i$$

So for large t the ratio of consumptions is almost surely

$$\frac{c_t^i(\sigma)}{c_t^j(\sigma)} \approx \exp\left(\frac{-t(\kappa_j - \kappa_i)}{1-\alpha}\right)$$

By this last statement we mean that for all $\epsilon > 0$ there is a T such that for all $t \geq T$,

$$\exp\left(\frac{-t(\kappa_j - \kappa_i + \epsilon)}{1-\alpha}\right) < \frac{c_t^i(\sigma)}{c_t^j(\sigma)} < \exp\left(\frac{-t(\kappa_j - \kappa_i - \epsilon)}{1-\alpha}\right)$$

If $\kappa_j - \kappa_i > 0$ then the consumption of i relative to j has an asymptotic rate of convergence to 0 that is exponential with coefficient $\frac{(\kappa_j - \kappa_i)}{1-\alpha}$. Note that if the traders preferences are nearly risk neutral, α near 1, then convergence is fast. More nearly risk neutral traders take more extreme asset positions so those with incorrect beliefs lose faster and for more nearly risk neutral preferences the exponential rate of convergence of marginal utilities drives faster convergence of consumptions. $\square$

In the iid economy with 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. If some trader has rational expectations, then 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.[11] The analysis below generalizes these results to dependent processes. This generalization is necessary for an analysis of learning, and selection over learning rules.

Our general 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 Discount Factors

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

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

Lemma 1 implies that trader i vanishes on the event $\{\liminf_t L_t^{ij} = +\infty\}$. So our analysis of economies in which traders have identical discount factors hinges solely on the asymptotic behavior of likelihood ratios.[12]

Our first result is that each trader is almost sure that she will survive. This follows immediately from the fact that from trader i's point of view the likelihood ratio is a non-negative martingale meeting the conditions of the martingale convergence theorem.

Theorem 1. *Assume A.1–3. Trader i survives p^i-almost surely.*

Whether a trader in fact survives depends on the relation between her beliefs, others beliefs and the truth. But if trader i has rational expectations then for any trader j the likelihood ratio is in fact a non-negative martingale meeting the conditions of the martingale convergence theorem. So it converges p-almost surely. That traders with rational expectations survive is thus an immediate consequence of Theorem 1.

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

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

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 p^i.*

The converse to Corollary 1 is:

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

The intuition behind Theorem 2 is that if j survives on a set of sample paths which he believes to be possible then i would be allocated zero consumption on this set of paths unless she too believed it to be possible. The assumption that j believes the set of sample paths to be possible, $p^j(V) > 0$, 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.

The necessary condition for i's survival in the presence of a trader with rational expectations is not that i have rational expectations too. But i will survive only if her beliefs are not very different from the truth — in particular, i's forecasts of the future must be asymptotically p-almost surely correct. We will have more to say about the restrictiveness of this conclusion in the next section.

3.3 Learning with Identical Discount Factors

In this section we consider the survival possibilities for learners. We first deal with the trivial case in which the truth is absolutely continuous with respect to some trader's beliefs. The following theorem, which is an immediate consequence of Theorems 1 and 2, shows that absolute continuity is sufficient for survival and that in the presence of a trader whose beliefs satisfy absolute continuity it is also necessary for survival. Unless otherwise stated the theorems in this section assume identical discount factors for all traders.

Theorem 3. *Assume A.1–3. Suppose that the truth p is absolutely continuous with respect to trader j's beliefs. Then:*

1. *Trader j survives p-almost surely.*

2. *If trader i survives p-almost surely then p is absolutely continuous with respect to p^i.*

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.[13]

To see that merging (absolute continuity of the truth with respect to beliefs) is of at least mild interest for learning, suppose trader i is a Bayesian learner. The Bayesian's models of the stochastic process on states are parameterized by $\theta \in \Theta$. Suppose that for some $\theta \in \Theta$ the model p^θ is correct, that is, $p^\theta = p$. 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.[14] 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 decision maker 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.

Our first result in this line is that any Bayesian will survive for almost all θ in the support of her prior. That is, any Bayesian is almost certain that she will survive. This is just a reinterpretation of Theorem 1 in the Bayesian context.

Theorem 4. *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 this theorem may seem to contradict Theorem 2. Consider a two person iid economy in which the states are flips of a coin. One knows the coin is fair. Her beliefs are iid with parameter 1/2. The other trader is not certain about the parameter. He is a Bayesian whose belief about the parameter can be represented by a uniform prior on $[0, 1]$. His forecast and trader 1's forecast are mutually singular. Now suppose the the true parameter value is indeed 1/2. The first trader has rational expectations. According to Theorem 1 she survives, while according to Theorem 2 trader 2 vanishes. But this is not inconsistent with Theorem 4. For almost all possible parameter values, all but value 1/2, trader 2 will almost surely learn the parameter value, and so he will survive while trader 1, who has beliefs which are initially incorrect and never improve, will vanish. The same conclusion applies if trader 1's beliefs have countable support. For parameter values in the support set trader 2 vanishes. This set is negligible with respect to his prior beliefs. Outside of that set trader 1 vanishes and trader 2 survives.

There is an analogue of Theorem 2 which provides a necessary condition for survival in the presence of a Bayesian. If there is a Bayesian in the economy, then 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).

Theorem 5. *Suppose that trader j is a Bayesian with prior belief μ^j on Θ. If trader i survives for μ^j-almost all θ then p^j is absolutely continuous with respect to p^i.*

An example of a forecasting rule that is not Bayesian, but which generates forecasts that merge with Bayesian forecasts, is suggested by maximum likelihood estimation.[15] 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 the state space is $\{1, 2\}$. States are iid and the probability of state 2 is θ. There are two traders. 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 $(2, 2, \ldots)$. 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 x_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. $\square$

Since a Bayesian will survive almost surely with respect to her prior belief one might 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. Learning 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. More general results are possible — see the discussion below. We will make use of some particular results about the behavior of Bayesian forecasts. These results — well known in the statistics literature — 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. Recent versions of this result include Phillips and Ploberger (1996) and Ploberger and Phillips (1998).[16] The version we use is due to Clarke and Barron (1990). We assume that traders know the state process is iid. The model space Θ parameterizes the distribution from which the current state is drawn. We let p^θ refer, as before, to the entire process, and $p_\theta(\cdot \,|\, \cdot)$ refer to the distribution of a single draw. Let E_θ denote the expectation operator given θ. We assume

A. 4. *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.

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.

Although results like this have been known in statistics for decades, most economic theorists (including, initially, us) are surprised by them. One piece of intuition goes as follows. Consider Bayesian updating on a finite set of models of an iid process. The updating formula says that for any two models, the log of the posterior odds is the sum of the log of the prior odds and the sum of the log-likelihood ratios. The log-likelihood ratio for n observations is the sum of n independent log-likelihood ratios, one for each observation. How fast beliefs converge with one prior versus another is determined by the rate of divergence of the infinite series of log-likelihood ratios. This rate has nothing to do with the first term, which is the only place the prior appears, and everything to do with the likelihood functions. Clarke and Barron's Theorem and similar results explore this rate for likelihood functions on parameter spaces of different dimensions. We provide some additional intuition on why it is that dimension matters in Appendix A.

The expectation of the random variable $\log p^\theta(\sigma^t)/\rho(\sigma^t)$ under the distribution $p^\theta(\sigma^t)$ is the relative entropy of the Bayesian trader's marginal distribution on states from period 1 through t with respect to the true distribution. Although the terms of this series converge to 0 (assuming the truth is in the support of the trader's prior beliefs), the series diverges. Clarke and Barron's Theorem tells us the rate at which the series diverges.

This result has interesting consequences for the survival of Bayesians. Suppose that trader j's prior belief is concentrated on a lower-dimensional subset of Θ than is trader i's prior belief. If j 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 6. *Assume A.1–4. 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 all $\theta \in \Theta'$, trader i vanishes in probability.*

If two traders have prior beliefs with supports of different dimension, the trader with the higher dimensional support will vanish, even if the high-dimensional prior is concentrated around the truth and the low-dimensional prior is quite diffuse. If the supports of the two distributions are identical, then both traders will survive or fail together, and if they both survive, how well they do relative to each other will be determined by the shape of their prior densities.

An alternative approach to understanding Theorem 6 is to relate it to Theorem 2. Suppose that, in contrast to the conclusion of Theorem 6, traders i (with the higher dimensional parameter set) and j (with the lower dimensional set) both survive. Let $V^j \subset \Sigma$ denote the support of j's beliefs. Both traders survive on V^j, so conclude from Theorem 2 that $p^i(V^j) > 0$. But under our assumptions, or even much weaker assumptions, it is easy to see that $p^i(V^j) = 0$, which is a contradiction.[17]

Theorem 6 has stronger assumptions than does Theorem 2, but it has more powerful conclusions since it gives rate information. To show one use for this rate information we consider the effect of differing discount factors. Theorem 6 combined with Equation (3) demonstrates that the effects of belief differences work much slower than the effects of discount factors. The effect of discount factor differences on the log of the ratio of i's and j's marginal utilities is linear in t, while the effect of differences in the dimension of parameter spaces is $\log t$. This insight can be pushed further, to show that when discount factors are heterogeneous, discount rate effects dominate parameter space effects. We have not found a way to prove the following theorem with absolute continuity arguments.

Theorem 7. *Suppose the assumptions of Theorem 6 and, in addition, that* $\beta_i > \beta_j$. *Then for all* $\theta \in \Theta$, *trader* j p^θ*-almost surely vanishes.*

In the next section we consider the effect of differing discount factors and differing beliefs in the general case in which learning rules need not be Bayesian.

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 saves more, he consumes more in the long run. With differing discount factors the rate at which likelihood ratios diverge is crucial. 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.[18]

A. 5. *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* $p^i(s|\mathcal{F}_{t-1})(\sigma) > \delta$*.*

Define $I^k(\sigma^t) = I_{p(\cdot|\sigma^t)}\big(p^i(\cdot|\sigma^t)\big)$. That is, $I^k(\sigma^t)$ is the relative entropy of trader k's conditional beliefs about date $t+1$ given states through date t, relative to the true conditional probability. The analysis of section 3.1's iid economy generalizes in the following manner:

Theorem 8. *Assume A.1–3 and A.5. On the event*

$$\limsup_t \frac{\sum_{\tau=1}^t \log \beta_j - I^j(\sigma^{\tau-1})}{\sum_{\tau=1}^t \log \beta_i - I^i(\sigma^{\tau-1})} < 1$$

$c_t^i(\sigma^t) \to 0$ *p-almost surely.*

In the iid case the ratio on the left hand side is κ_j/κ_i. Since these are both negative numbers, it follows that if $\kappa_j > \kappa_i$, the ratio will be less than 1, and trader i will vanish.

One implication of Theorem 8 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 8 provides some further characterizations of sample path behavior. If $\limsup$ is replaced with $\liminf$ in the hypothesis, then on that event, $\liminf_t c_t^i(\sigma^t) = 0$.

Both the numerator and the denominator of the condition provided by Theorem 8 are diverging. The statement of the theorem makes clear the dependence of survival on rates: discount rates and rates of learning.

In economies where the traders are learning, the relative entropy terms are converging to 0. But the sum of the relative entropies may be diverging. This is the situation in the learning examples of the previous section. Traders learn, but the sum of relative entropies of the conditional distributions diverge at $\log t$ rates. The comparison of those rates determines the survivor.

It is easy to compute the survival ratio in Theorem 8 for iid economies and the following example demonstrates how it can be computed for Markov economies.

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,

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

Since the Markov process of states is ergodic, $\frac{1}{t}\big(\sum_{\tau=1}^{t} \log \beta_k - I^k(\sigma^{\tau-1})\big)$ converges to $\log \beta_k$ minus the expectation of the relative entropy of k's conditional forecasts with respect to the true conditional distribution under the invariant distribution. Consequently, the lim sup of the ratio in Theorem 8 is less than one and $\lim_t c_t^i = 0$ almost surely. If the Markov process is not ergodic, one carries out this exercise on each communication class. □

Theorem 8 lets us explore the trade-off between learning speed and discount factors. We saw in Theorem 6 that when discount factors were identical, not all successful learners survived. The market favored faster learners. Theorem 7 shows that in the iid smooth case 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 easily measured, but it does use the uniform bounds of Axiom 5.

Corollary 3. *Suppose Axioms 1–3 and 5. 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.*

This result 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. Since the relative entropies with respect to p of conditional forecasts almost surely converge for both traders, 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 8 are false without the uniform bounds of Axiom 5. 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.5.
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 $\sum_{t=1}^\infty \big(I_{p_t}(p^i) - I_{p_t}(p^j)\big)$ diverges to $+\infty$. If the conclusions of Theorem 8 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 The Role of the Inada Condition

Throughout our analysis we have assumed that traders are strictly risk averse, and more importantly that their utility functions satisfy an Inada condition at zero consumption. Within this class of utility functions our selection results do not depend on agents' particular utility functions. But utility functions with finite first derivatives at the zero consumption boundary, in particular those representing risk neutrality, do not lie in this class of preference orders. Our results do not address the survival of these agents.

All our results are essentially of two kinds: Sufficient conditions for a trader to vanish and sufficient conditions for a trader to survive. In this section we will show that all results concerning sufficient conditions for vanishing carry over to the larger class of preferences with monotonic, concave and C^1 payoff functions that may or may not satisfy an Inada condition at the origin. We also show by means of a counterexample that the sufficient conditions for survival require the Inada condition.

One must be careful in interpreting the results we present here. With the Inada conditions, an economy in which Arrow securities are traded will always have a competitive equilibrium, and that equilibrium will be Pareto optimal. So we can identify many asset structures for which are results actually characterize competitive paths. This need not be true in the larger class of preferences. If two risk neutral traders attach different probabilities to two states then at any prices for Arrow securities paying off only in those two states at least one trader will want to take an infinite long or short position in the securities. In the larger class of preferences we cannot make claims about the properties of competitive paths.

The theorems giving sufficient conditions for a trader to vanish work by showing that likelihood ratios diverge and thus from the first order conditions for Pareto optimality we know that ratios of marginal utilities diverge. It follows from Lemma 1 that these traders vanish. Without the Inada condition the first order conditions must account for the boundary at 0. Nonetheless Lemma 1 can be extended to cover vanishing in this case.

First, we weaken Axiom 1.

A.1′. *The payoff functions u^i are C^1, concave, and strictly monotonic.*

The relevant modification of Lemma 1 is:

Lemma 1′. *Suppose Axioms* 1′, 2 *and* 3 *are satisfied, and consider any Pareto optimal allocation in which each trader has a strictly positive welfare weight. On the event* $\{\beta_j^t p_t^j(\sigma)/\beta_i^t p_t^i(\sigma) \to \infty\}$, $c_t^i(\sigma) \downarrow 0$.

Proof. The first order conditions for the Pareto problem (1) have to be modified to reflect the possibility that the inequality constraint on consumption is binding. The new first order conditions are

For all σ and t,

(*i*) there is a number $\eta_t(\sigma) > 0$ and numbers $\mu_t^i(\sigma) \geq 0$ for all i such that

$$\lambda_i \beta_i^t u^{i\prime}\left(c_t^i(\sigma)\right) p_t^i(\sigma) - \eta_t(\sigma) + \mu_t^i(\sigma) = 0 \tag{7}$$

(*ii*) $\mu_t^i(\sigma) c_t^i(\sigma) = 0$ and $c_t^i(\sigma) \geq 0$ for all i.

Suppose that $c_t^i(\sigma)$ does not converge to 0. Then there is subsequence of dates t' such that $c_{t'}^i(\sigma) > 0$ and $\mu_{t'}^i(\sigma) = 0$. Rewriting (7) using this fact and noting that $\mu_{t'}^j(\sigma) \geq 0$, we have:

$$\frac{\lambda^j}{\lambda^i}\frac{\beta_j^{t'}}{\beta_i^{t'}}\frac{p_{t'}^j(\sigma)}{p_{t'}^i(\sigma)} \leq \frac{u^{i\prime}(c_{t'}^i(\sigma))}{u^{j\prime}(c_{t'}^j(\sigma))}.$$

By hypothesis $\beta_j^{t'} p_{t'}^j(\sigma)/\beta_i^{t'} p_{t'}^i(\sigma) \to \infty$. Trader j's marginal utility is bounded below by $u^{j'}(F)$. So we have $u^{i'}(c_{t'}^i(\sigma)) \uparrow \infty$. This is possible only if $c_{t'}^i(\sigma) \downarrow 0$ which contradicts the hypothesis that $c_t^i(\sigma)$ does not converge to 0.

□

It is easy to see that in Theorems 2, 3.2, 5, and 6, Axiom 1 can be replaced by Axiom 1′. Thus if the sufficient conditions on discount factors and beliefs for vanishing are satisfied, these traders vanish in any Pareto optimal allocation even if they are risk neutral. The intuition for this series of results is simple. Suppose one trader (risk neutral or not) puts positive probability on some set of sample paths and another trader (risk neutral or not) puts zero probability on the same set of sample paths. In any Pareto optimal allocation the trader with zero probability on the set of sample paths should be allocated no consumption on this set of paths, that is he vanishes.

Our results giving sufficient conditions for survival require the Inada condition. Lemma 1 uses the Inada condition to show that in any Pareto optimal allocation a trader must be allocated some consumption on any sample path that has positive probability from his point of view. But a risk neutral trader may be allocated no consumption on paths that he believes to be possible as long as others believe them to be more likely. In the following example, trader 2 puts positive probability on a sample path in which he is allocated no consumption. This occurs because trader 1 puts higher probability on this path.

Example: Risk Neutrality.
The economy consists of two traders with a common discount factor and a

common linear utility function. The state space is $S = \{1, 2\}$. Trader 1 believes that state 1 always occurs; that is, $p^i_t(\sigma) = 1$ if and only if $\sigma = (1, 1, \dots)$, and 0 otherwise. Trader 2 believes that states are independent and the probability that he puts on state 1 at date t is $q_t = \exp -t/2$. Then $\prod_{t=1}^{\infty} q_t = 1/e$. So trader 2 believes that $\sigma = (1, 1, \dots)$ has positive probability, but unlike trader 1, at each date he places probability less than one on the next state being state 1.

Consider weights $\lambda_i = 1$ for all i. On $\sigma = (1, 1, \dots)$ trader 1 is allocated all of the consumption good. This occurs even though the likelihood ratios L^{21}_t are bounded. □

In summary, the Inada condition plays no role in our sufficient conditions for vanishing. But our sufficient conditions for survival depend on the Inada condition, which rules out risk neutrality. This concludes our analysis of complete markets. In the next section we take up the market selection hypothesis with incomplete markets.

5 Belief Selection in Incomplete Markets

Results in the previous sections showed that a form of belief selection is a consequence of Pareto optimality. The first welfare theorem can be used to apply these results to economies with complete markets. 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 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 $t+1$,

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

where $k = (8^{1/2} - 1)^2$. 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 states s_1 and s_2 always occur with equal probability. Traders 1 and 2 have endowment stream $(1, 0, 0, \ldots)$ which does not vary with states. Trader 3's endowment stream is $e^3_1 = 0$ and $e^3_t(\sigma) = 3/4 + k^{N_t(\sigma)} R_{t-1}(\sigma) - k^{N_t(\sigma)}(1/2)(1 + (1/2)^t)$ for $t > 1$, where $N_t(\sigma)$ is the number of occurrences of state 2 through time t on σ.

Markets are incomplete because there are two states at each date, only one asset, and because traders would trade across states if they could. If we added an independent asset we would have dynamically complete markets, but no equilibrium would exist. Traders 1 and 3 are correctly certain that only state one occurs so if consumption in state two has a positive price, and if they can construct an asset that pays off only in sate two, then they would want to take an infinite short position in that asset. But in equilibrium, consumption in state two cannot have a zero price as then trader 2 would demand an infinite amount of it. So unless we (arbitrarily) constrain short sales we cannot have an equilibrium with two independent assets. If we were to constrain short sales then the asset structure would again not be equivalent to complete markets.

This model has an equilibrium of plans, prices and price expectations (Radner 1972) in which the price of the asset (in terms of the consumption good) is, for every state,

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

On the actual path of states $(s_1, s_1, \ldots)$ in this equilibrium, trader 3 supplies 1 unit of asset and traders 1 and 2 collectively demand 1 unit of asset at each date. 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 2 or $2k > 2$ with equal probability. 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 $2k$ in states s_1 and s_2, respectively. We also assumed that traders 1 and 2's total asset demand would be 1 on the actual path. 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 on the actual path. The details of the construction of this example are in Appendix C.

Because we have a stochastic infinite horizon economy in which traders with heterogeneous beliefs trade in incomplete markets the example is actually more complex than the intuition above might suggest. In an equilibrium we require that for any sequence of past states, each trader correctly forecast prices for all future states. This requires construction of endowments and demands for all partial histories, including partial histories to which a trader assigns probability 0. Traders 1 and 3 place zero probability on any history with an occurrence of state 2 so their beliefs and thus their demands could be defined arbitrarily. We have chosen to require a form of sequential rationality in which each trader continues to forecast prices correctly and applies his original i.i.d. beliefs to the future even after seeing an impossible (from his point of view) partial history of states. □

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 allow him to systematically benefit at trader 2's expense.

It is tempting to conclude that overly optimistic beliefs have a distinguished role in this analysis. In fact there is no general result. 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. Traders 1 and 2 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. This example demonstrates 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 is choosing the portfolio with the highest expected growth rated at each date. When markets are complete, the expected growth rate of the optimal portfolio increases as the relative entropy of beliefs with respect to the truth decreases. When markets are incomplete this need not be the case.

6 Conclusion

This paper has examined the long-run survival of traders with different beliefs and discount factors in complete and incomplete markets. Our complete market analysis proceeded by examining Pareto-optimal paths. In short, we have shown that the long-run fate of market participants is determined by discount factors and beliefs alone, and not payoff functions.[19] Trader j's survival is determined by the growth rate of the discounted value of the inverse of the likelihood ratio of trader j's marginal beliefs through period t to the true marginal distribution of the first t states. As unwieldy as it may seem, this object can be manipulated to evaluate the survival properties of Bayesian learning versus other rules, the impact of prior beliefs on the survival of Bayesian learners, and the tradeoff between discount factors and learning rates. In particular, the *market selection property*, that ceteris paribus, traders with the best beliefs survive, is valid for all complete markets, regardless of the asset structure. The results for incomplete markets are strikingly different. In simple examples we show that payoff functions matter for survival, and that even controlling for payoff functions, less accurate beliefs may survive while more accurate beliefs disappear.

Our results explain how various authors have obtained differing answers to the market selection question. Sandroni (2000) obtains a positive answer. Because the Lucas trees economy he considers has dynamically complete markets, his equilibrium allocations are Pareto optimal. Blume and Easley (1992) obtain a negative answer because with exogeneously fixed savings rates the allocations we analyzed are not Pareto optimal. DeLong, Shleifer, Summers, and Waldman (1990) does not address the question. They show that irrational traders can earn higher expected returns than rational traders, but as they note in their 1991 paper, this does not mean that irrational traders survive. DeLong, Shleifer, Summers, and Waldman (1991) claim a negative answer, but in fact they fail to address the question in an equilibrium model. They study an economy in which prices are kept fixed even though the wealths of rational and irrational traders are changing. The traders do not interact with each other — there are no markets in which an equilibrium is established. That the resulting allocations are not Pareto optimal is no surprise.

Our results have strong implications for long-run asset pricing. If discount factors are uncorrelated with beliefs, then in the long run of a complete markets economy, assets will be priced according to the beliefs of the trader with the most accurate beliefs. This claim stands in contradiction to those who have argued that the market is a super-aggregator of beliefs, and that market prices can be more accurate than the beliefs of any one trader. It also 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. In every market there are selection pressures that cannot be ignored in determining long-run asset prices. In complete markets, these pressures are strong enough that, in the long run, noise traders have no impact on asset prices. When markets are incomplete, however, it is conceivable that noise traders persist, and that assets may be mis-priced even in the long run. However even here market selection works, and some behaviors will be advantaged over others, with consequent implications for long-run prices. Further investigation of this topic is a promising subject for future research.

Our analysis is long-run, and so its relevance depends upon how long the long run actually is. We have two answers to this question. The first comes directly from the model. In the case of discount rate differences, convergence is driven by a geometric rate: $(\beta_j/\beta_i)^t$. When discount rates are identical and beliefs are asymptotically different, again the driving rate is geometric. When traders are learning at different rates, the rate at which any one trader can disappear is slow because the *differences* in learning rates can be small. This is the conclusion of Clarke's and Barron's Theorem, and its application to Theorem 6. But the rate of asset price convergence is determined not by the rate of divergence of their beliefs but by the rate at which their beliefs converge to the truth, and this could be faster. The second answer has to do with what time actually measures. If each period is a year, the long run is really long. If each date is a second, the long run may be quite short. We cannot know how long the long run is until we know the appropriate time scale of the model. We believe the relevant time scale is the rate at which transactions take place. Each date is an opportunity for portfolio change. In real, measured time, how fast can portfolio change take place? In residential housing markets, portfolio change is very slow. In securities markets, change is quite quick. Our point is that the answer to this question is empirical, and the answer varies enormously with the markets in question.

Since the long run fate of traders and long run prices can differ so much in between complete and incomplete markets, it is natural to ask which assumption is more nearly accurate. If markets are incomplete, some traders are restricted from making trades they believe to be mutually beneficial. One might expect markets to evolve to allow these trades under some conditions. See, for example, Allen and Gale (1994) and Pesendorfer (1995). Whether or not markets will end up being complete is an open question.

Loosely speaking, markets are incomplete when there are bets traders would like to make that the market does not allow them to make. Whether such bets exist depends upon how traders construe their world. If we take belief heterogeneity seriously, we are required to recognize that there is a subjective element to the state space; two traders could have completely different world views, and those views could be characterized by two quite distinct state spaces. Admitting this subjectivity raises issues for how markets could actually achieve completeness. One possibility, raised by Ross (1976), is that in securities markets, for example, the only relevant bets are on asset movements, and so options are a kind of universal asset which completes markets regardless of how people construe the underlying states of the world. Another possibility is that this subjectivity makes empty the entire content of complete markets analysis. And this, obviously, is a very interesting subject for future research.

Appendix A

The proof of Theorem 6 follows easily from Clarke and Barron's Theorem, and so understanding Theorem 6 requires intuition about the Clarke and Barron's result (and the many theorems on the same theme which appear in the statistics and econometrics literature). And although the proof of theorems like Clarke and Barron's are complicated, the intuition is straightforward. What follows is not a proof, but a quick computation which explains intuitively why the theorem should be true.

Suppose there are $d+1$ states, s_0 through s_d. The likelihood function for n observations is $p^\theta(\sigma^t) = \prod_{j=0}^{d} \theta_j^{k_j(\sigma^t)}$, where θ_j is the probability of state s_j and $k_j(\sigma^t)$ is the number of s_j's in t observations. Since $\theta_k = 1 - \sum_{j=0}^{d-1} \theta_j$

there are only d independent parameters, and Θ is thus d-dimensional. Suppose that the true θ is the vector θ^0. Then

$$\begin{aligned}
\rho(\sigma^t) &= \int_0^1 \cdots \int_0^1 p^\theta(\sigma^t) q(\theta)\, d\theta \\
&= \int_0^1 \cdots \int_0^1 q(\theta) \exp\Big\{ \sum_j k_j(\sigma^t) \log \theta_j \Big\}\, d\theta \\
&\approx \int_0^1 \cdots \int_0^1 q(\theta) \exp\Big\{ t \sum_j \theta_j^0 \log \theta_j \Big\}\, d\theta \\
&= \exp\big\{ tH(\theta_0) \big\} \int_0^1 \cdots \int_0^1 q(\theta) \exp\Big\{ t\big(H(\theta) - H(\theta^0)\big) \Big\}\, d\theta
\end{aligned}$$

where $H(\theta) = \sum_{j=0}^k \theta_j^0 \log(\theta_j)$ and the approximation comes from the SLLN.

Suppose this approximation were exact. The function $H(\theta)$ is maximized at $\theta = \theta^0$, and so $H(\theta) = H(\theta^0) + \frac{1}{2}(\theta - \theta^0)^{\mathrm{T}} H''(\theta^0)(\theta - \theta^0) + \cdots$. Thus the integral is approximately

$$\exp\big\{ tH(\theta^0) \big\} \int_0^1 \cdots \int_0^1 q(\theta) \exp\Big\{ \frac{t}{2}(\theta - \theta^0)^{\mathrm{T}} H''(\theta^0)(\theta - \theta^0) \Big\}\, d\theta$$

and the symmetric matrix $H''(\theta^0)$ is negative definite. The Fisher information matrix is $I(\theta) = -H''(\theta)$.

Now expand the density: $q(\theta) = q(\theta^0) + q'(\theta^0)(\theta - \theta^0) + \cdot$. Substitute this and calculate to see that the integral is approximately

$$\begin{aligned}
&q(\theta^0) \exp\big\{ tH(\theta^0) \big\} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \exp\Big\{ \frac{t}{2} y^{\mathrm{T}} H''(\theta^0) y \Big\}\, dy + \\
&\qquad \exp\big\{ tH(\theta^0) \big\} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} q'(\theta^0) y \exp\Big\{ \frac{t}{2} y^{\mathrm{T}} H''(\theta^0) y \Big\}\, dy
\end{aligned}$$

after a change of variables (and observing that there is no cost to running the domain of integration over the whole real line taking $q(\theta)$ to be 0 outside the unit interval).

The second term, when multiplied by an appropriate scale factor, is the expectation of a linear combination of mean 0 normally distributed random variables with covariance matrix $-\big(tH''(\theta^0)\big)^{-1}$, and so its value is 0. Similarly, the first term after rescaling is the integral of the multivariate normal density over its entire range. Thus the first integral evaluates to

$$\left(\frac{2\pi}{t}\right)^{d/2}\left(\frac{1}{\det -H''(\theta^0)}\right)^{1/2}$$

Now $p^{\theta^0}(\sigma^t)$ is approximately $e^{tH(\theta^0)}$, and so

$$\frac{p^{\theta^0}(\sigma^t)}{\rho(\sigma^t)} \approx \left(\frac{t}{2\pi}\right)^{d/2}\sqrt{\det I(\theta^0)}\,\frac{1}{q(\theta^0)}.$$

Thus the value of the ratio in Clarke and Barron's Theorem is approximately

$$\log\frac{p^{\theta^0}(\sigma^t)}{\rho(\sigma^t)} \approx \frac{d}{2}\log\frac{t}{2\pi} + \frac{1}{2}\log\det I(\theta^0) - \log q(\theta^0)$$

The chi-squared term in Clark and Barron's Theorem is due to the remainders from the two SLLN approximations.[20]

Appendix B

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

$$\frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \frac{\lambda_j}{\lambda_i}\frac{p_t^j(\sigma)}{p_t^i(\sigma)} \tag{8}$$

The first ratio above is a fixed number. The second ratio is a non-negative martingale with mean 1 under p^i, and so converges p^i-almost surely. Thus, p^i-almost surely, Lemma 1's necessary condition for vanishing fails to hold. □

Proof of Theorem 2. ¿From equation (3),

$$\frac{\lambda_i}{\lambda_j}\frac{u^{i\prime}(c_t^i(\sigma))}{u^{j\prime}(c_t^j(\sigma))} = \frac{p_t^j(\sigma)}{p_t^i(\sigma)} = L_t^{ij} \tag{9}$$

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 $p_t^j(\sigma)/p_t^i(\sigma) \to \infty$ for all $\sigma \in B$. Consequently equation (9) implies that $c_t^i(\sigma) \to 0$ on B, which contradicts the hypothesis. □

Proof of Theorem 3. By Theorem 1, j survives p^j-almost surely. Since p is absolutely continuous with respect to p^j this implies that j survives p-almost surely. Theorem 2 implies that p is absolutely continuous with respect to p^i. □

Proof of Theorem 4. It is sufficient to show that the necessary condition for vanishing is p^θ-almost never met for ν-almost all θ. Consider an arbitrary trader j. The likelihood ratios L_t^{ij} 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, $p^\theta(\limsup_i L_t^{ij} = \infty) = 0$ 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 5. This is an immediate consequence of Theorem 2. Observe that if the conclusion of Theorem 5 were false, then p^j is not absolutely continuous with respect to p^i, contradicting Theorem 2. □

Proof of Theorem 6. Consider the form of the first order conditions given in equation (5). Since discount factors are identical, this becomes

$$\log\frac{u^{i\prime}\left(c_t^i(\sigma)\right)}{u^{j\prime}\left(c_t^j(\sigma)\right)} = \log\frac{\lambda_j}{\lambda_i} + \log\frac{p_t^\theta(\sigma)}{p_t^i(\sigma)} - \log\frac{p_t^\theta(\sigma)}{p_t^j(\sigma)} \tag{$*$}$$

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

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

and

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

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 4. To see that trader j vanishes, expand the logarithms in equation $(*)$ and divide by t. The right hand side becomes

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

The first term converges to 0. 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 consistency 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 $\theta \notin \mathrm{cl}\, \Theta'$. Then this minimum is some $K > 0$. Consequently trader j's term converges to K. Thus $(1/t)\log u^{i\prime}(c_t^i)/u^{j\prime}(c_t^j)$ converges almost surely to $-K<0$, and so $\log u^{j\prime}(c_t^j)/u^{i\prime}(c_t^i)$ converges almost surely to ∞. We see from Lemma 1 that trader j disappears almost surely. □

Proof of Theorem 7. When discount rates differ, equation $(*)$ must be modified as follows:

$$\log\frac{u^{j\prime}\big(c_t^j(\sigma)\big)}{u^{i\prime}\big(c_t^i(\sigma)\big)}=\log\frac{\lambda_i}{\lambda_j}+t\log\frac{\beta_i}{\beta_j}+\log\frac{p_t^\theta(\sigma)}{p_t^j(\sigma)}-\log\frac{p_t^\theta(\sigma)}{p_t^i(\sigma)}\qquad(**)$$

Suppose that $\beta_i>\beta_j$. We see from the proof of Theorem 6 that in probability the last two terms diverge to $-\infty$ at a rate no faster than $\log t$, while the discount factor term diverges to $+\infty$ at rate t. Consequently for all $B>0$ and $\epsilon>0$ there is a T such that for $t\geq T$, $p^\theta\{u^{j\prime}/u^{i\prime}<B\}<\epsilon$. In other words, $u^{j\prime}/u^{i\prime}\uparrow\infty$ in probability, and so c_t^j converges to 0 in probability. □

Proof of Theorem 8. Beginning with equation (5) we have

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

where $B=\max_{k,t,\sigma}Y_t^k(\sigma)$. Assumption A.5 guarantees that $B<\infty$. Let $W^k(\sigma^t)=\log\beta_k-Y_\tau^k-B$. The expected value under p of $W^k(\sigma^t)$ given σ^{t-1} is $\log\beta_k-I^k(\sigma^{t-1})$. A.5 and the definitions together guarantee that there is a number $C>-\infty$ such that $C<W^k(\sigma^t)<0$. A SLLN due to Freedman (1973) implies that $\sum_{\tau=1}^t W_\tau^k(\sigma^\tau)/\sum_{\tau=1}^t E\{W_\tau^k(\sigma^\tau)|\sigma^{\tau-1}\}\to 1$ almost surely, and so there are random variables $A_t^k(\sigma^t)$ converging to 1 almost surely such that

$$\log \frac{u^{i\prime}\big(c_t^i(\sigma)\big)}{u^{j\prime}\big(c_t^j(\sigma)\big)} = \log \frac{\lambda_j}{\lambda_i} + A_t^j \sum_{\tau=0}^{t} \big(\log \beta_j - I^j(\sigma^{\tau-1})\big) - A_t^i \sum_{\tau=0}^{t} \big(\log \beta_i - I^i(\sigma^{\tau-1})\big) - B(A_t^j - A_t^i) \quad (10)$$

The last term converges to 0 almost surely. The stochastic terms can be rewritten as

$$\left(A_t^j \frac{\sum_{\tau=0}^{t}\big(\log \beta_j - I^j(\sigma^{\tau-1})\big)}{\sum_{\tau=0}^{t}\big(\log \beta_i - I^i(\sigma^{\tau-1})\big)} - A_t^i \right) \sum_{\tau=0}^{t} \big(\log \beta_i - I^i(\sigma^{\tau-1})\big)$$

On the event that the lim sup of the ratio is less than 1, the right hand side almost surely diverges to$+\infty$ and from Lemma 1 it follows that $c_t^i \to 0$. □

Proof of Corollary 3. 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. □

Appendix C

We will adopt a different notation just for this section. The expression $c(\sigma^t)$ refers to consumption after the partial history σ^t has been realized. The expression $c(\sigma^t, s)$ refers to consumption after the $t+1$-length partial history consisting of initial segment σ^t followed by state s. The same notation

$w(\sigma^t)$, $a(\sigma^t)$, $e(\sigma^t)$ and $q(\sigma^t)$ will be used for wealths, asset demands, trader 3's endowment and asset prices. It will be clear from the context whose consumption any expression refers to.

We first need to solve each trader's decision problem given the conjectured price process. We assume that even after partial histories which are impossible given trader 1's beliefs, he continues to believe that only state 1 can occur in the future. For each partial history the Euler equation for trader 1 is

$$c(\sigma^t)^{-1/2} = 2\beta c(\sigma^t, 1)^{-1/2}$$

The 2 on the right hand side in in this equation is the state 1 rate of return on the asset at the proposed prices.

With $\beta = 8^{-1/2}$, the Euler equation implies that $c(\sigma^t, 1) = c(\sigma^t)/2$. Let $a_t(\sigma^t)$ denote the trader's asset demand and $w(\sigma^t)$ denote his wealth. Suppose that $c(\sigma^t) = 3w(\sigma^t)/4$. We will check that this is a solution to trader 1's optimization problem. At each partial history, the budget constraint requires that the value of current consumption plus current asset transactions must equal current wealth. So $a(\sigma^t) = w(\sigma^t))/\big(2(1 + (1/2)^t)\big)$. Then a computation shows that trader 1's wealth at date $t + 1$ will be $w(\sigma^t)/2$ if state 1 occurs and $kw(\sigma^t)/2$ if state 2 occurs. So consumption at date $t + 1$ in the only state that matters to trader 1, state 1, will be $c(\sigma^t, 1) = (3/4)(1/2)w(\sigma^t) = c(\sigma^t)/2$. So the conjectured consumption plan satisfies the Euler equation. Since it is feasible and satisfies the Euler equation, it is optimal. The evolution of trader 1's wealth produced by this plan is

$$w(\sigma^t) = (1/2)^{t-1} k^{N_t(\sigma)}$$

where $N_t(\sigma)$ is the number of occurrences of state 2 on path σ by time t. Using this it is immediate that trader 1's budget constraint is satisfied at each date. Given the proposed prices trader 1's decision problem clearly has a solution and so the conjectured solution is in fact a solution to his problem.

Trader 2 believes that both states are possible, so his Euler equation is

$$c(\sigma^t)^{-1/2} = \beta\Big(\frac{1}{2}2c(\sigma^t, 1)^{-1/2} + \frac{1}{2}2kc(\sigma^t, 2)^{-1/2}\Big)$$

Now we suppose that $c(\sigma^t) = w_t(\sigma^t)/2$. Then $a(\sigma^t) = w_t(\sigma^t)/(1 + (1/2)^t)$, and a computation shows that trader 2's wealth at date $t+1$ will be $w(\sigma^t, 1) = w(\sigma^t)$ if state 1 occurs and $w(\sigma^t, 2) = kw(\sigma^t)$ if state 2 occurs. So the Euler equation will be satisfied if and only if

$$\left(\frac{1}{2}w(\sigma^t)\right)^{-1/2} = \beta\left(\left(\frac{1}{2}w(\sigma^t)\right)^{-1/2} + k^{1/2}\left(\frac{1}{2}w(\sigma^t)\right)^{-1/2}\right)$$

For the given values of β and k, this equation is satisfied. So the conjectured consumption plan satisfies the Euler equation and as before it is a solution to the decision problem. The evolution of trader 2's wealth produced by this plan is

$$w(\sigma^t) = k^{N_t(\sigma)}$$

The total demand for the asset from traders 1 and 2 is $k^{N_t(\sigma)}$. We will construct trader 3's endowment stream so that supplying this amount of the asset is optimal for him. Trader 3's Euler equation is

$$c(\sigma^t)^{-1} = 2\beta c(\sigma^t, 1)^{-1}$$

Since for trader 3, $\beta = 1/2$, this implies that $c(\sigma^t)$ is constant. Suppose that $c(\sigma^t) = 3/4$. Using the asset demand from traders 1 and 2 we have that trader 3's endowment, $e(\sigma^t)$, must solve for each $t > 1$ and σ^t,

$$e(\sigma^t) = 3/4 + k^{N_{t-1}(\sigma)}R_{t-1}(\sigma^t) - k^{N_t(\sigma)}q(\sigma^t)$$

and $e_1 = 0$. The right hand side of this equation is strictly positive, so we will take it to be trader 3's endowment, guaranteeing that for him it is optimal to provide 1 unit of asset in every partial history.

The prices $q(\sigma^t)$ clear the asset market at each date since at these prices trader 3 supplies the amount of asset demanded by traders 1 and 2.

Notes

[1]DeLong, Shleifer, Summers, and Waldman (1991) is a partial equilibrium model and makes many approximations that, we think, limit the value of the analysis. DeLong, Shleifer, Summers, and Waldman (1990) only analyzes expected returns which is not sufficient to determine who will survive.

[2]More precisely, attitudes toward risk within the class of strictly concave utility functions satisfying an Inada condition have no effect on survival. See section 4 for more on this point.

[3]We do not address existence of competitive equilibrium in our setting. The Pareto optimal allocations we consider do exist so our results provide a characterization of any complete or dynamically complete markets equilibria that exist. Existence of complete market equilibria is well understood, but the existence of equilibria when the market structure is endogenous can be difficult to establish. Sandroni (2000) does not address existence of the equilibria he considers.

[4]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.

[5]When multiple traders have correct beliefs then having correct beliefs is necessary for survival, but it is not sufficient.

[6]We allow utility functions to take on the value $-\infty$ at 0 in order to accommodate CRRA utility with risk-aversion coefficient at least 1.

[7]We consider statistical learning rules which modify beliefs and individuals who act optimally given their beliefs. Individuals who learn by by directly modifying their actions according to some adaptive process, such as reinforcement learners, do not fit into our framework.

[8]This construction includes forecasting rules generated by Bayesian learning.

[9]All of these analyses work off of the Euler equations which characterize optimality for the owner of a firm in Blume and Easley (2002), an investor in an asset economy for Sandroni (2000) and for Pareto optimal allocations here. The technique in Blume and Easley (2002) is similar to that used here, but the economic question is very different. In Blume and Easley (2002) the economy is deterministic and the selection is over firms rather than traders. Both Blume and Easley (2002) and Sandroni (2000) apply this technique to Euler equations for individuals in an equilibrium setting and so are concerned with equilibrium prices. The approach here is more direct.

[10]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.

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

[12]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.*

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

[14]The effect of this assumption is similar to Kalai and Lehrer (1993)'s Grain of Truth assumption for learning to play a Nash equilibrium.

[15]More precisely this rule is not Bayesian given the maintained hypothesis that the true process is iid.

[16]See also Dawid (1984) and Rissanen (1986). 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.

[17]See Sandroni (2001).

[18]This condition requires that traders agree about zero conditional probability events in every partial history. This would be satisfied for example if the world is iid with positive probability on some set of states S', traders are Bayesians who know that the world is iid, and all of their models put positive probability on S'. Alternatively, the world could be T-step Markov (conditional probabilities depend on the last T states) and traders could be Bayesians who consider a class of Markov models each of which puts positive conditional probability on the states that can actually occur. What it rules out is true conditional probabilities that converge to zero or forecasts that converge to zero for states that have positive probability. We show in an example in the text that this condition is necessary for our analysis.

[19]This conclusion is modified in an obvious way if the Inada conditions are not satisfied.

[20]Another way to understand this theorem is to see that it follows from the SLLN approximations and Laplace's method of integrating functions of the form $\int g(x)\exp\{tf(x)\}\,dx$ where x has an interior maximum at $x = c$. Laplace was the first to show that the integral just computed has an asymptotic expansion whose highest order term in t is

$$\sqrt{\frac{2\pi}{-tf''(c)}}\, g(c)\exp\{tf(c)\}$$

References

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

ALLEN, F., AND D. GALE (1994): *Financial Innovation and Risk Sharing.* MIT Press, Cambridge, MA.

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.

——— (1998): "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.

——— (2002): "Optimality and Natural Selection in Markets," *Journal of Economic Theory*, 107(1), 95–130.

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.

Freedman, D. A. (1973): "Another Note on the Borel-Cantelli Lemma and the Strong Law, with the Poisson Approximation as a By-product," *Annals of Probability*, 1(6), 910–925.

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 (1993): "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61(5), 1019–1045.

——— (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.

Pesendorfer, W. (1995): "Financial Innovation in a General Equilibrium Model," *Journal of Economic Theory*, 65(1), 79–116.

Phillips, P. C. B., and W. Ploberger (1996): "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica*, 64(2), 381–412.

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.

Radner, R. (1972): "Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets," *Econometrica*, 40(2), 289–303.

Rissanen, J. (1986): "Stochastic Complexity and Modelling," *Annals of Statistics*, 14(3), 1080–1100.

Ross, S. A. (1976): "Options and Efficiency," *Quarterly Journal of Economics*, 90(1), 75–89.

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

——— (2001): "Markets Favor Bayesian Estimators over Maximum Likelihood Estimators," unpublished, the University of Rochester.