Markets Favor Bayesian Estimators over Maximum Likelihood Estimators.

Alvaro Sandroni*
Department of Economics, University of Rochester, Rochester NY 14627
(alsn@troi.cc.rochester.edu) and
Department of Managerial Economics and Decision Sciences,
Kellogg School of Management,
Northwestern University,
Evanston IL 60208 (sandroni@nwu.edu)

December 4, 2001

Abstract

In this paper, it is shown that, under certain conditions, agents who forecast according to a Maximum Likelihood model are driven out of the market by agents who forecast according to a Bayesian model. Hence, asset prices are eventually determined by Bayesian models.

*I gratefully acknowledge the financial support from the National Science Foundation. I also thank Larry Blume, David Easley, Bruce Hansen, Werner Ploberger and Shakeeb Khan for many useful comments. All errors are mine.

1. Introduction

Expectations play a central role in determining prices. However, the expectations of different agents do not influence prices equally. In fact, prices are eventually determined by the actions of those agents who accumulate wealth during the trading process. But which forecasting methods lead to greater wealth and, hence, become predominant in the market?

The two most widely spread statistical methods are the Bayesian model and the Maximum Likelihood model. In both cases, it is postulated that the data generating process belongs to a given parametric family. The Bayesian paradigm defines a mixture of the processes in the family, with weights given by a prior over the parameter space. The forecasts are determined by the conditional distributions of this mixture. The Maximum Likelihood model does not require a prior. The data is used to infer the value of the parameter by the Maximum Likelihood estimator. The forecasts are determined by the conditional distribution of the stochastic process associated with this parameter.

A natural question is whether Bayesian or Maximum Likelihood forecasts will become predominant in the market. To address this question, I assume that some agents' beliefs are given by the Bayesian paradigm and others by Maximum Likelihood. All agents maximize expected discounted utility according to their beliefs. I also assume that markets are dynamically complete and that agents have the same discount factor. These assumptions ensure the following result: assume that a path (i.e., an infinite string of data) is much less likely according to agent 1's beliefs than according to agent 2's beliefs. Then, agent 1 will be driven out of the market on this path, i.e. will eventually have no wealth (see Sandroni (2000)). This result shows that it suffices to analyze the likelihood ratios of a path to determine who will be driven out of the market on this path. It is not necessary to solve for equilibrium.

Dawid (1982) and Ploberger and Phillips (1998) argue that empirical models assigning a higher likelihood to the realized paths better describe the data.[1] Among the reasons offered is the Neyman-Pearson lemma which suggests the use of the likelihood ratio test. Therefore, assigning high likelihood to the realized paths is desirable from a predictive point of view and is a property that an empirical model must have in order to be influential in the market.

[1] An empirical model is a probability distribution on the space of paths. The realized paths are those in the support of the data generating process.

Ploberger and Phillips (1998) show that, for almost all parameters, Bayesian models do not assign arbitrarily lower likelihood to the realized paths than any other model. The prior, although somewhat arbitrary, must be defined in a parameter space with a dimension not higher than necessary. These results give a formal justification for the maxim of parsimony in the number of estimated parameters and for the claim that Bayesian models are difficult to beat from a predictive point of view. However, they also have an important implication in economics (that was simultaneously observed by Blume and Easley (2001)). They imply that an agent whose beliefs are given by a parsimonious Bayesian model survive in the market.

On the other hand, agents whose beliefs are given by a Maximum Likelihood model may be driven out of the market. Assume that there are two states a and b. The probability of a is θ in every period. It is known that θ is a rational number. The Maximum Likelihood forecast for a is the empirical frequency of a, whether or not θ is known to be rational. In this sense, the Maximum Likelihood forecasts ignores that θ is rational, but the Bayesian model may take this information into account. Let the Bayesian agents' priors put positive weight only on the rational parameters. I show that Maximum Likelihood agents will be driven out of the market by these Bayesian agents.

In the example above, Bayesian forecasts converge to θ faster than Maximum Likelihood forecasts. In other frameworks, Bayesian and Maximum Likelihood forecasts converge to the correct probabilities at the same rate. Then, the main differences between these forecasts occur in the initial periods. Hence, it is desirable to extend this theory of belief selection to the initial periods of the economy. I assume that there are K independent sources of uncertainty. Each of them is given by a coin k that, in each period, can either take value a_k or b_k. There are two types of coins, high and low. The probability of a_k is high if coin k is high and low if coin k is low. Before the economy starts each coin is flipped S times. This assumption makes the results more interesting because with a large data set Maximum Likelihood is expected to perform well (at least in a simple framework like this one in which Maximum Likelihood estimators are consistent and efficient).

An agent vanishes if, in period one, his wealth goes to zero as K increases. A Uniform Bayesian prior is such that, for every coin, both types are equally likely to be selected. I show that no matter which coins are selected, Maximum Likelihood agents vanish in the presence of an Uniform Bayesian agent.

Nothing ensures that a prior assumption like "all coin types are selected with equal probability" is correct (e.g. all coins could be low). However, even if the Uniform Bayesian prior is wrong and the coin types are selected with different probabilities, Maximum Likelihood agents vanish in the presence of these Bayesian agents. Moreover, Maximum Likelihood agents estimate the coin types quite precisely when there is a large data set available. Nevertheless, markets favor agents who, a priori, assume that both coin types are selected with equal odds.

In fact, this result is more general. Forecasts are said to be neutral if the forecast for a_k given an observed frequency of a_k, say f, is the same as the forecast for b_k when the observed frequency of b_k is f. Neutrality ensures that differences in forecasts comes from the data and not from prior beliefs. Several forecasting methods satisfy this condition. Any agent whose forecasts are neutral (e.g. Maximum Likelihood agents) vanish in the presence of Uniform Bayesian agents.

The probabilities of the states of nature are not known because coins are high or low with unknown odds. However, under neutrality, asset prices can be derived as a function of endowments and preferences. This follows because if K is large the relevant probabilities are given by the Uniform Bayesian model.

As the available data set increases, Bayesian forecasts depend more on the data and less on the prior. This makes it difficult to determine the properties of different priors. On the other hand, the Bayesian forecasts depend on the prior in the initial periods. So, priors do matter. If the prior of a Bayesian agent is not uniform then, for some coins, this non-Uniform Bayesian agent may vanish in the presence of a Maximum Likelihood agent. Hence, in this example, the Uniform Bayesian model is the only Bayesian model that dominates the Maximum Likelihood model.

This paper is organized as follows: The economy is defined in section 2. Section 3 contains a critical lemma relating likelihood ratios of agents' beliefs to survival in the market. Section 4 contains general results connecting standard properties of estimators with survival. In particular, it is shown that agents who forecasts according to inconsistent estimators are driven out of the market. In sections 5 it is shown that Bayesian agents survive, whereas in section 6, it is shown that Maximum Likelihood agents may be driven out of the market. In section 7, it is demonstrated that even when Maximum Likelihood is efficient, Maximum Likelihood agents may vanish from the economy. Conjectures for future work and the conclusion are in sections 8 and 9, respectively.

2. The Uncertainty

Let N be the set of natural numbers. Let N_+ be $N \cup \{0\}$. At period $t \in N$, a state of nature is publicly observed. The set of states of nature is finite and given by $T \equiv \{1, ..., L\}$. Let T^t, $t \in N \cup \{\infty\}$, be the $t-$Cartesian product of T. For every finite history $s_t \in T^t$, $t \in N$, a cylinder with base on s_t is the set $C(s_t) = \{s \in T^\infty \mid s = (s_t, ...)\}$ of all infinite histories whose t initial elements coincide with s_t. Let $\Im_t$ be the σ-algebra that consists of all finite unions of cylinders with base on T^t. The σ-algebras $\Im_t$ define a filtration $\Im_0 \subset ...\Im_t \subset ... \subset \Im$ where $\Im_0$ is the trivial σ-algebra, and $\Im$ is the σ-algebra generated by the algebra $\Im^0 \equiv \bigcup_{t \in N} \Im_t$, i.e. $\Im$ is the smallest $\sigma-$algebra which contains $\Im^0$. I also define $\Im_{-1} \equiv \Im_0$.

The data generating process belongs to a given family of stochastic processes $\mathcal{F} = \left\{P^\theta, \theta \in \Theta\right\}$, where Θ is a parameter space endowed with a σ-algebra $\mathcal{B}$. Let $\Delta(T)$ be the space of probability measures over T. Let $g : \bigcup_{t \in N} T^t \rightarrow \Delta(T)$ be a *forecasting function*. An *empirical model* is an arbitrary probability measure P on $(T^\infty, \Im)$ which may or may not belong to $\mathcal{F}$. There is a one to one correspondence between forecasting functions and empirical models. An empirical model can be thought of as an arbitrary sequence of forecasts, conditional on the available data.

A stochastic process P^B on $(T^\infty, \Im)$ is a *Bayesian model* if there exists a probability measure ν on $(\Theta, \mathcal{B})$ such that

$$P^B(D) = \int_\Theta P^\theta(D) d\nu \text{ for every } D \in \Im.$$

When $P^\theta\left(C(s_t)\right)$ is regarded as a function of θ, for a given finite history $s_t \in T^t$, it is a *likelihood function*. The *Maximum Likelihood estimator* $\theta_{s_t} \in \Theta$ maximizes the likelihood function, i.e.

$$P^{\theta_{s_t}}\left(C(s_t)\right) \geqslant P^\theta\left(C(s_t)\right) \text{ for all } \theta \in \Theta. \tag{2.1}$$

Given a finite history $s_t \in T^t$, let $g^{ML}(s_t) \in \Delta(T)$ be such that the probability of $k \in T$ is $P^{\theta_{s_t}}\left(C(s_t, k)\right)$ divided by $P^{\theta_{s_t}}\left(C(s_t)\right)$ (i.e., the conditional probability of $P^{\theta_{s_t}}$).[2] A *Maximum Likelihood model* (if well defined) is the probability measure P^{ML} associated with the forecasting function g^{ML}.

[2]If (2.1) has more than one solution then $g^{ML}(s_t)$ is any linear combination of these conditional probabilities.

2.1. The Economy

There are I long-lived agents. Agent i's belief, $i \in \{1, ..., I\}$, is given by a probability measure P^i on $(T^\infty, \Im)$. I assume that $P^i(B) > 0$ for any finite-time event $B \in \Im^0$. Bayesian agents' beliefs are given by Bayesian models. Maximum Likelihood agents' beliefs are given by Maximum Likelihood models.

There are M long-lived trees and a single consumption good c. Agents are born with shares of the trees and receive no other endowments. Let $k^i_{m,-1}$ be agent i's initial share of tree m. At period $t \in N_+$, dividends are given by $e_t = (e_{1,t}, ..., e_{M,t})$, share prices are given by $p_t = (p_{1,t}, ..., p_{M,t})$, agent i's consumption is given by c^i_t, agent i's share holdings are given by $k^i_t = (k^i_{1,t}, ..., k^i_{M,t})$, and agent i's wealth is given by $w^i_t \equiv (p_t + e_t)k^i_{t-1}$. The functions e_t, p_t, c^i_t, k^i_t, and w^i_t are $\Im_t$-measurable. I assume that dividends are uniformly bounded away from zero and infinity.

Let H_t be the $\Im_t-$measurable function defined by

$$H_t(s) = \begin{pmatrix} p_{t+1}(s(1)) + e_{t+1}(s(1)) \\ \vdots \\ p_{t+1}(s(L)) + e_{t+1}(s(L)) \end{pmatrix},$$

where $s = (s_t, ...)$, $s_t \in T^t$, $s(h) = (s_t, h, ...)$, $h \in T$.

I assume that markets are dynamically complete. That is, the number of assets is equal to the number of states of nature ($L = M$), and the rank of $H_t(s)$ is L. At period $t \in N_+$, agent i's expected discounted utility function is

$$E^{P^i}\left\{\sum_{r=0}^{\infty} \beta^r u^i(c^i_{t+r}) \mid \Im_t\right\},$$

where β is a common discount factor for all agents, E^{P^i} is the expectation operator associated with P^i, and u^i is a strictly increasing, strictly concave, continuously differentiable utility function such that $(u^i)'(c)$ goes to infinity as c goes to zero.

At period $t \in N_+$, agent i's observed and anticipated budget constraints are

$$c^i_{t+r} + p_{t+r}k^i_{t+r} \leq (p_{t+r} + e_{t+r})k^i_{t+r-1},\ w^i_{t+r} \geqslant 0,\ c^i_{t+r} \geqslant 0,\ r \in N_+.$$

Markets clear at period t if

$$\sum_{i=1}^{I} c^i_t = \sum_{m=1}^{M} e_{m,t} \text{ and } \sum_{i=1}^{I} k^i_{m,t} = 1\ m = 1, ..., M.$$

In equilibrium, agents maximize expected discounted utility subject to the budget constraints and markets clear in every period.

3. Basic Concepts

The accumulation of wealth is the main criteria for survival because only agents with positive wealth influence prices (see Blume and Easley (1992) and Sandroni (2000), section 7, for a formal analysis of this claim).

Definition 1. Agent i *is driven out of the market on a path* $s \in T^{\infty}$ if agent i's wealth, $w_t^i(s)$, converges to zero as t goes to infinity. Agent i *survives on a path* $s \in T^{\infty}$ if agent i is not driven out of the market on s.

Given any probability measure Q on $(T^{\infty}, \Im)$, let dQ_t be the $\Im_t-$measurable function defined by

$$dQ_t(s) \equiv Q(C(s_t)), \ t \in N, \text{ and } dQ_0 \equiv 1,$$

where $s = (s_t, ...) \in T^{\infty}$, $s_t \in T^t$. That is, $dQ_t(s)$ is the probability of the finite history s_t.

Lemma 1 below shows that the agents' survival on a path is determined by the ratio of the probabilities of finite histories.

Lemma 1. Fix an agent $i \in \{1, ..., I\}$ and a path $s \in T^{\infty}$. In every equilibrium, if there exists an agent $j \in \{1, ..., I\}$ such that

$$\frac{dP_t^i(s)}{dP_t^j(s)} \underset{t \to \infty}{\longrightarrow} 0. \tag{3.1}$$

then agent i is driven out of the market on s (by agent j). Moreover, if for all agents $j \in \{1, ..., I\}$ there exists $\varepsilon > 0$ such that

$$\frac{dP_t^i(s)}{dP_t^j(s)} > \varepsilon$$

then agent i survives on s.

Proof : See Sandroni (2000), lemma 2.

■

Lemma 1 depends on the assumption that agents have the same discount factor (more patient agents tend to accumulate more wealth, see Sandroni (2000) for results with heterogeneous discount factors) and that markets must be dynamically complete (see Blume and Easley (2001) and Becker (2001) for counterexamples when markets are incomplete). However, it is not necessary to assume that agents have the same preferences over risk.

The intuition in lemma 1 is that agents allocate relatively more wealth to events they believe more likely to occur than to events they believe less likely to occur. If the likelihood ratio of the finite histories converges to zero as in (3.1) then agent i will allocate much less wealth to this path than agent j. In equilibrium, agent j's wealth is bounded above and so, on this path, agent i must be driven out of the economy.

By Lemma 1, agents' beliefs determine who survives and who is driven out of the economy. This result greatly simplifies the analysis because it is not necessary to solve for the equilibrium. It suffices to analyze the likelihood ratios of the paths.

4. Empirical Models

Lemma 1 can be used as a bridge connecting results in statistical forecasting and economic theory. Consider the prequential approach of Dawid (1982) which, in his own words, is "founded on the premise that the purpose of statistical inference is to make sequential probability forecasts for future observations rather than to express information about parameters." Dawid's objective is normative: To determine good empirical models. The motivation in this paper is positive: To determine which empirical models survive in the economy. It turns out that there is a close connection between the two agendas.

Dawid (1982) argues that two probability measures Q and $\tilde{Q}$ on $(T^{\infty}, \Im)$ can be compared, for data $s \in T^{\infty}$, by the *prequential likelihood ratio*

$$\frac{d\tilde{Q}_t(s)}{dQ_t(s)}.$$

Under Q, the sequence of prequential likelihood ratios form a non-negative martigale and, hence, converge to a finite limit. That is,

$$\lim_{t\to\infty} \frac{d\tilde{Q}_t}{dQ_t} < \infty \qquad Q - a.s. \tag{4.1}$$

So, if the observed sequence of prequential likelihood ratios converge to infinite then Q can be rejected in favor of $\tilde{Q}$. Ploberger and Phillips (1998) offer other justifications for the use of the prequential likelihood ratio. Their arguments support the prequential likelihood ratios as a criteria of goodness for forecasting. Lemma 1 demonstrates that, in dynamically complete market economies, this is the correct criteria of fitness in the market. The examples below demonstrate how this connection can be explored.

Assume that there are two states of nature a and b. The data generating process P^{θ}, $\theta \in [0,1]$, is such that the probability of a is always θ. An empirical model P^{c} is *consistent* if for every $\theta \in [0,1]$,

$$\lim_{t \to \infty} g^{c}(s_t) = (\theta, 1-\theta),$$

for $P^{\theta}-$almost all paths $s = (s_t, ...)$. (Clearly, g^{c} is the forecasting function associated with P^{c}). An empirical model P^{nc} is *not consistent* if, for every $\theta \in [0,1]$, on P^{θ}-almost all paths $s = (s_t, ...)$, the limit of $g^{nc}(s_t)$ either does not exist or it is not $(\theta, 1-\theta)$. It is possible to show that, for all $\theta \in [0,1]$,

$$\lim_{t \to \infty} \frac{dP_t^{nc}}{dP_t^{c}} = 0 \qquad P^{\theta} - a.s.^{3} \tag{4.2}$$

Equation 4.2 connects the traditional concept of consistency with the prequential likelihood ratios. By lemma 1, an agent whose beliefs are not consistent will be driven out of the market by an agent with consistent beliefs. Hence, consistency is not only a good property from an intuitive view point, it is also a necessary property for an empirical model to be influential in the market.

The agenda proposed in this section can be summarized as follows: empirical models can be compared using prequential likelihood ratios. This approach has a counterpart with traditional concepts such as consistency and efficiency.[4] By lemma 1, agents' ultimate fate can be deduced from these ratios. This determines the beliefs of the agents who will eventually affect prices.[5]

[3]For a general result on these lines that do not require a fixed probability θ in every period, see Sandroni (2000) proposition 4.

[4]See Dawid (1982) for the connection between efficiency and prequential likelihood ratios.

[5]The results in the next section (5) show that Bayesian agents survive in the economy. These results were inspired by Ploberger and Phillips (1998) who had a normative motivation in mind. They were also simultaneously obtained by Blume and Easley (2001).

5. Bayesian Models

In this section, the parameter space Θ is assumed to be a subset of $\Re^l$. I assume that the Bayesian prior ν over Θ has a strictly positive density. That is,

$$\nu(D) = \int_D f(\theta)\partial\theta,\ D \subset \Theta,\ \ f(\theta) > 0 \text{ for all } \theta \in \Theta. \tag{5.1}$$

Recall that a Bayesian model P^B is defined as follows:

$$P^B(D) = \int_D P^\theta(D) f(\theta)\partial\theta \text{ for every } D \in \Im. \tag{5.2}$$

By equation 4.1,

$$\lim_{t\to\infty} \frac{d\tilde{Q}_t}{dP_t^B} < \infty \qquad P^B - a.s.$$

By equation 5.2, if a set has P^B-full measure, it must also have $P^\theta-$full measure for all $\theta \in \Theta$ except, possibly, for a Lesbegue-measure zero set of parameters. This demonstrates proposition 1.

Proposition 1. *Let $\tilde{Q}$ be any probability measure on $(T^\infty, \Im)$ and let P^B be a Bayesian model that satisfies (5.1). For almost all parameters $\theta \in \Theta$,*

$$\lim_{t\to\infty} \frac{d\tilde{Q}_t}{dP_t^B} < \infty \qquad P^\theta - a.s. \tag{5.3}$$

Corollary 1 below follows directly from proposition 1 and lemma 1.

Corollary 1. *In every equilibrium, a Bayesian agent (with a prior that satisfies (5.1)) survives on P^θ-almost all paths, for almost all parameters $\theta \in \Theta$.*

For almost all parameters, a Bayesian agent survives no matter what the beliefs of the other agents are. However, the proviso "almost" cannot be disposed. If $\tilde{Q} = P^{\bar{\theta}}$, $\bar{\theta} \in \Theta$, then the limit in (5.3) could be infinite, $P^{\bar{\theta}} - a.s.$[6]

[6]A partial converse of proposition 1 is as follows: Assume that an agent survives on a set of paths Ω_θ such that $P^\theta(\Omega_\theta) = 1$ for almost all parameters $\theta \in \Theta$. Then, this agent survives on the set $\Omega = \bigcup_{\theta\in\Theta} \Omega_\theta$. By equation 5.2, $P^B(\Omega) = 1$. So, this agent survives on a set of paths that a Bayesian assigns probability one. It follows that, on the paths $s = (s_t, ...) \in \Omega$, the forecasts $g(s_t)$ of this agent will eventually approach the forecasts $g^B(s_t)$ of the Bayesian agent (see Sandroni (2000), proposition 2 for a proof).

The intuition in corollary 1 is as follows: any agent (Bayesian or not) survives on a set of paths that has full measure according to his own beliefs. This is intuitive because if this agent had correct beliefs then he would survive. This set of paths has full measure according to almost all probabilities in the family $\mathcal{F}$. So, for almost all parameters, a Bayesian agent survives.

Assume that a Bayesian agent (1) believes that $\bar{\theta}$ is likely to be the true parameter (and $\bar{\theta}$ is the true parameter). Agent 1 has the bulk of his prior around $\bar{\theta}$, which satisfies 5.1. Agent (2) does not share agent 1's belief. His prior is, say, the uniform prior. Agent 1 may eventually have more wealth than agent 2, but agent 2 will not be driven out of the economy by agent 1. On the other hand, if agent 1 knows that $\bar{\theta}$ is the true parameter then agent 1 has reduced, from the outset, the dimension of the parameter space (from 1 to 0). It follows that, $P^{\bar{\theta}} - a.s.$, agent 2 will be driven out of the economy by agent 1.[7] This is a more general phenomena. Consider example 2, below, in which the Bayesian models are defined on parameters spaces of different dimensions.

Example 1. There are two states of nature a and b, and two families of stochastic processes. In family $L = \left\{P^{\theta},\ \theta \in [0,1]\right\}$, P^{θ} is such that the probability of a next period is always θ. In family $H = \left\{P^{(\theta_1,\theta_2)},\ (\theta_1,\theta_2) \in [0,1]^2\right\}$, $P^{(\theta_1,\theta_2)}$ is the Markov process such that if a occurs then a has probability θ_1 and if b occurs then a has probability θ_2. Let P^{B_L} and P^{B_H} be Bayesian models associated with the families L (lower dimensional) and H (higher dimensional), respectively. Assume (5.1) in both models.

Assume that the true parameter belongs to the lower dimensional family L. A Bayesian agent has a model defined on L. Another Bayesian agent's model is defined on the higher dimension space H (a non-parsimonious model). The results below show that Bayesian agents whose model is not parsimonious will be driven out of the market.

Proposition 2. *Consider example* 2. *For almost all parameters* $\theta \in [0,1]$ *in family* L,

$$\frac{dP_t^{B_H}}{dP_t^{B_L}} \to 0 \qquad P^{\theta} - a.s.$$

Proof: See Appendix.

[7]See Sandroni (2000), page 1319, for a proof.

Corollary 2. *Consider example 2. In every equilibrium, for almost all parameters $\theta \in [0,1]$ in family L, in $P^{\theta}-$almost all paths, any Bayesian agent with a model on the (higher dimensional) space H will be driven out of the market by a Bayesian agent with a model on the (lower dimensional) space L.*

Proposition 2 shows that a Bayesian models with irrelevant parameters achieve lower prequential likelihood. This gives a formal justification for the maxim of parsimony in the number of estimated parameters. It follows that Bayesian agents whose empirical models are not parsimonious will be driven out of the market. This result is easy to generalize. An analogous argument shows that a Bayesian agent who believes that the data generating process is a Markov process of length $r+1$ will be driven out of the market by a Bayesian agent who believes (correctly) that the data generating process is a Markov process of length r.

The results in this section show that Bayesian models have good normative properties and that these models are influential in the market, but they do not imply that alternative models do not have similar properties. This motivates the comparison with the Maximum likelihood model conduced in the next section.

6. Maximum Likelihood Models

Assume that there are two states of nature a and b and the data generating process P^{θ}, $\theta \in [0,1]$, is such that the probability of a is always θ. Let P^{ML} be the Maximum Likelihood model. It is well known that the Maximum Likelihood forecast $g^{ML}(s^t)$ is equal to the frequencies of a's and b's in s_t. This follows because the likelihood function of θ is $\theta^{n_t}(1-\theta)^{t-n_t}$, where n_t is the number of a's in s_t. The Maximum Likelihood estimate of θ is

$$\frac{n_t}{t}. \tag{6.1}$$

The measure P^{ML} assigns zero probability to finite histories such as (a,b) or (b,a). So, I assume that before the economy started, $c>0$ states of nature a and $d>0$ states of nature b were observed. This avoids a violation of the assumption that no agent assigns zero probability to finite histories, which is relevant for the existence of an equilibrium. So, on a finite history $s_t \in T^t$, the Maximum Likelihood forecast for a next period is

$$\frac{n_t+c}{t+c+d}. \tag{6.2}$$

It is well known that this is also a Bayesian model, where the prior is determined by a Beta distribution $B(c,d)$ (see Lehman (1991), chapter 4.1). After conditioning on the history that precedes the economy, the posterior belief of a Bayesian model with a Beta distribution prior is given by another Beta distribution. So, in this example, the Maximum Likelihood model is a Bayesian model with a particular prior. The result in proposition 2 do not depend upon the prior (as long as (5.1) is satisfied). Therefore, a Maximum Likelihood agent also survives (almost surely, for almost all parameters).

Blume and Easley (2001) presented an example like this one. They modified the Maximum Likelihood model so that it does not forecast any finite-time event with probability zero. The difference is that instead of assuming that some data was available before the economy started, they "trimmed" the Maximum Likelihood forecast for a so that it is equal to the empirical frequency of a if this frequency is within bounds. Otherwise, the forecast is given by these bounds. This trimmed Maximum Likelihood model is similar (but not identical) to a Bayesian model and survives in the economy. Now assume that the probability of a is known to be a rational number.

Example 2. There are two possible states of nature a and b. The parameter space is $\bar{\Theta} \subset [0,1]$, the rational numbers between 0 and 1. The measures P^{θ}, $\theta \in \bar{\Theta}$, are such that the probability of a is always θ. Before the economy started, $c > 0$ states of nature a and $d > 0$ states of nature b were observed.

In example 2, the Maximum Likelihood forecasts are also given by (6.2) because $\frac{n_t+c}{t+c+d}$ is a rational number. The Bayesian models are given by

$$P^B \equiv \sum_{\theta \in \bar{\Theta}} y^{\theta} P^{\theta}, \ \sum_{\theta \in \bar{\Theta}} y^{\theta} = 1, \ y^{\theta} \geqslant 0.$$

I assume that the weights y^{θ} are strictly positive;

$$y^{\theta} > 0 \text{ for all } \theta \in \bar{\Theta}. \tag{6.3}$$

The Maximum Likelihood model can be thought of as a Bayesian model in an one dimensional parameter space, $[0,1]$, when the true parameter space has dimension zero (the set $\bar{\Theta}$ of rational numbers). The results below show that a Maximum Likelihood agent will be driven out of the market by a Bayesian agent.

Proposition 3. *Consider example* 2*. Let P^B be a Bayesian model that satisfies* (6.3)*. Then, for any $\theta \in \bar{\Theta}$,*

$$\frac{dP_t^{ML}}{dP_t^B} \to 0 \qquad P^\theta - a.s.$$

Proof: See Appendix.

Corollary 3. *Consider example* 2*. In every equilibrium, for any parameter $\theta \in \bar{\Theta}$, in $P^\theta-$almost all paths, a Maximum Likelihood agent will be driven out of the market by a Bayesian agent whose model satisfies* (6.3)*.*[8]

Corollary 3 shows Maximum Likelihood agents may be driven out of the market. This is in contrast with corollary 2 which shows that, for almost all parameters, Bayesian agents survive in the economy, regardless of the other agents' beliefs.

Under standard assumptions in statistical inference such as independent, identically distributed stochastic processes, Maximum likelihood and Bayesian forecasts eventually become identical. Consider example 2. The Maximum Likelihood and the Bayesian forecasts converge to the true parameter. A Bayesian agent makes the convergence faster, but convergence is inevitable even if there were only Maximum Likelihood agents in the economy. Hence, it is desirable to extend this theory of belief selection to the initial periods of the economy where the Maximum Likelihood and the Bayesian forecasts differ. In the next section, I show that as the number of states of nature increases, the wealth of Maximum Likelihood agents vanish in the first period.

7. Many States of Nature per Period

Let coin k be a random variable that in each period can take two values, a_k or b_k. Every period, K coins are flipped simultaneously. The realizations of a coin are independent from those of any other coin. There are two types of coins: high (h) and low (l). In every period, the probability of a_k is $h = 0.5 + x$ if coin k is high and $l = 0.5 - x$ if the coin k is low, where $x \in (0, 0.5)$.

[8] The same result obtains if, instead of assuming that data was available before the economy started, the Maximum Likelihood forecasts were "trimmed" as in Blume and Easley (2001).

Each coin is selected from a different urn. A draw from each urn is done just once. The outcome is a (non-observable) coin type. The proportion of high coins in urn k is $\theta_k \in [0,1]$, a realization of a random variable $\tilde{\theta}_k$. The distribution of $\tilde{\theta}_k$ is unknown and may depend on k, but it is known that $\tilde{\theta}_k$ is independent from $\tilde{\theta}_{k'}$, $k \neq k'$.[9] Each coin is flipped S times before the economy starts. Let $\omega_k \in \{a_k, b_k\}^S$ be the outcome of the S flips. Let η_k be the number of a_k's in ω_k.

The Bayesian and Maximum Likelihood models are defined as in section 2, but below I describe the initial forecasts in greater detail. Given an arbitrary prior, let p_k be the expected value of $\tilde{\theta}_k$. The Bayesian forecast for a_k in the first period of the economy, $g_k^B(a_k|\, \omega_k)$, is given by

$$\frac{p_k(0.5+x)^{\eta_k+1}\,(0.5-x)^{S-\eta_k}+(1-p_k)\,(0.5-x)^{\eta_k+1}\,(0.5+x)^{S-\eta_k}}{p_k(0.5+x)^{\eta_k}\,(0.5-x)^{S-\eta_k}+(1-p_k)\,(0.5-x)^{\eta_k}\,(0.5+x)^{S-\eta_k}}. \tag{7.1}$$

Let the *Uniform Bayesian model* be given by the uniform prior on $[0,1]^K$. The *Uniform Bayesian forecast* $g_k^{UB}(a_k|\, \omega_k)$ is defined replacing p_k with 0.5 in 7.1.

Given the available data, the likelihood of $\theta = (\theta_1, ..., \theta_K)$ is

$$\prod_{k=1}^{K}\left(\theta_k(0.5+x)^{\eta_k}\,(0.5-x)^{S-\eta_k}+(1-\theta_k)\,(0.5-x)^{\eta_k}\,(0.5+x)^{S-\eta_k}\right).$$

Thus, the Maximum Likelihood estimate of θ_k is 1 if the empirical frequency of a_k is greater than the frequency of b_k and 0 if, conversely, the frequency of a_k is smaller than of b_k. This estimate is arbitrary if these frequencies are identical. I will assume it to be 0.5. Hence, the Maximum Likelihood forecast for a_k in the first period of the economy, $g_k^{ML}(a_k|\, \omega_k)$, is given by

$$\begin{array}{ll} 0.5+x, & \text{if } \eta_k > \frac{S}{2}; \\ 0.5, & \text{if } \eta_k = \frac{S}{2}; \\ 0.5-x, & \text{if } \eta_k < \frac{S}{2}. \end{array} \tag{7.2}$$

[9] So, the realizations of a coin do not provide useful information about another coin. This captures the idea that different coins represent different variables like an election in one state and the weather in another.

Let $r_k \in \{a_k, b_k\}$ be the realization of coin k in the first period of the economy. Given data $\omega^K = (\omega_1, ..., \omega_K)$, forecasts for $r^K = (r_1, ..., r_K)$ are defined by

$$g(r^K | \ \omega^K) \equiv \prod_{k=1}^{K} g_k(r_k | \ \omega_k).^{10} \tag{7.3}$$

7.1. Comments on the Forecasts and the Concept of Neutrality

Assume that coin k is high. By the law of large numbers, the empirical frequency of a_k converges to $0.5 + x$ as the number of observations increase. So, after some finite, but random, number of observations the frequency of a_k remains above 0.5. The Maximum Likelihood forecasts for a_k will then be exactly correct $(0.5 + x)$.

Given finite number of observations S, the Maximum Likelihood forecasts for a_k in the first period of the economy will be correct for a fraction of the coins y_S that is arbitrarily close to one if S is sufficiently large. In these y_S coins, the Maximum Likelihood forecasts are strictly closer to the truth than the Uniform Bayesian forecasts because those are never exactly correct. In the remaining coins, the Uniform Bayesian forecasts are closer to the truth than the Maximum Likelihood forecasts because the Uniform Bayesian forecast for a_k is between the Maximum Likelihood forecast and the true probability of a_k. So, it is not intuitively clear which method is better from a predictive point of view.

Forecasts are *neutral* if

$$g_k(a_k | \ \omega_k) = g_k(b_k | \ \tilde{\omega}_k), \tag{7.4}$$

when ω_k has η_k a_k's and $S - \eta_k$ b_k's and $\tilde{\omega}_k$ has η_k b_k's and $S - \eta_k$ a_k's.

Neutrality is a property that the forecasts should satisfy if the agent does not have a reason to believe, a priori, that a_k is neither more nor less likely than b_k. Uniform Bayesian and Maximum Likelihood forecasts are neutral.

If $S = 1$ the forecasts are neutral as long as $g_k(a_k | \ a_k) = g_k(b_k | \ b_k)$. The Uniform Bayesian forecasts

$$g_k^{UB}(a_k | \ a_k) = g_k^{UB}(b_k | \ b_k) = (0.5 + x)^2 + (0.5 - x)^2 \, ,$$

[10] Clearly, $g_k(b_k | \ \omega_k) = 1 - g_k(a_k | \ \omega_k)$. Equation 7.3 is valid for both Maximum Likelihood and Bayesian forecasts. To make this notation consistent with the previous sections, let the forecast $g(\omega^K)$ be a measure on $\{a_k, b_k\}^K$ such that the probability of r^K is $g(r^K | \ \omega^K)$.

and the Maximum Likelihood forecasts

$$g_k^{ML}(a_k|\ a_k) = g_k^{ML}(b_k|\ b_k) = (0.5 + x)$$

are special cases.

Forecast are *different from Uniform Bayesian forecasts* if for some $\delta > 0$ there exists $\omega_k \in \{a_k, b_k\}^S$, $k = 1, ..., K$, such that

$$\left| g_k(a_k|\ \omega_k) - g_k^{UB}(a_k|\ \omega_k) \right| > \delta. \tag{7.5}$$

Maximum Likelihood forecasts are neutral and different from Uniform Bayesian forecasts. A superscript $*$ on g (i.e. $g_k^*(a_k|\ \omega_k)$) indicates that equations 7.4 and 7.5 are satisfied. So, $*$ indicates neutral and different from Uniform Bayesian forecasts.

7.2. The economy with many states per period

The economy remains defined as in section 2. Markets are dynamically complete. So, the number of trees is equal the number of states of nature ($L = 2^K$). The number of coins, K, is finite, but large. This assumption avoids unnecessary technical difficulties, but it implies that the parameter K must vary. Hence, the economic variables are indexed by K.

Let $\Lambda_k = \{a_k, b_k\}^S \text{x} \{a_k, b_k\}$ be the set of coin k realizations in the S stages before the economy and in the first period of the economy. Let $\sigma \in \prod_{k=1}^{\infty} \Lambda_k$ be an infinite sequence of coins realizations. Let $\sigma^K \in \prod_{k=1}^{K} \Lambda_k$ be the $K-$prefix of σ.

Definition 2. Agent i *vanishes* in period 1 *on* σ *if* $w_1^{i,K}(\sigma^K) \underset{K\to\infty}{\longrightarrow} 0$, $\sigma = (\sigma^K, ...)$.

That is, agent i vanishes if, as the number of coins increase, the wealth of agent i goes to zero in the first period of the economy. An agent who vanishes has arbitrarily small influence on prices (in period 1) if K is sufficiently large.

7.3. Main Results

Let $\lambda \in \{h, l\}^\infty$, $\lambda = (\lambda_1, ...)$, be an infinite sequence of coins types, where $\lambda_k \in \{h, l\}$ is the type of coin k. Given $\lambda \in \{h, l\}^\infty$, let P^λ be a measure on $\prod_{k=1}^{\infty} \Lambda_k$ such that, in each stage, coin k takes value a_k with probability λ_k.

Proposition 4. *For any (finite) length of data S, for every infinite sequence of coins $\lambda \in \{h,l\}^{\infty}$, for P^{λ}-almost every infinite sequence of coins realizations $\sigma \in \prod_{k=1}^{\infty} \Lambda_k$, $\sigma = (\sigma^K, ...)$, $\sigma^K = (\omega^K, r^K)$,*

$$\frac{g^{*}(r^K | \, \omega^K)}{g^{UB}(r^K | \, \omega^K)} \underset{K\to\infty}{\longrightarrow} 0.$$

In particular,

$$\frac{g^{ML}(r^K | \, \omega^K)}{g^{UB}(r^K | \, \omega^K)} \underset{K\to\infty}{\longrightarrow} 0.$$

Proof: See Appendix.

Corollary 4. *Assume a Uniform Bayesian agent in the economy. In every equilibrium, for any (finite) length of data S and for every infinite sequence of coins $\lambda \in \{h,l\}^{\infty}$, a Maximum Likelihood agent vanishes at period 1 on P^{λ}-almost every infinite sequence of coins realizations.*

Proof: The proof of this corollary is completely analogous to the proof of lemma 1 and, therefore, omitted.

In this example, the probabilities of the states of nature are not known because the coin selection process is unknown. However, if the number of coins is large, asset prices can be determined as a function of endowments and preferences. The relevant probabilities are given by the Uniform Bayesian model because agents whose forecasts are neutral and different from Uniform Bayesian forecasts vanish.

The popularity of the Maximum Likelihood model is partially explained by its good asymptotic properties (e.g. consistency and efficiency). In the simple framework of this section, the asymptotic properties of the Maximum Likelihood forecasts are as good as possible. After some random, but finite, number of observations the Maximum Likelihood forecasts are correct. However, proposition 4 shows that excellent asymptotic properties do not ensure relevance in the market. For any fixed (no matter how large) number of observations, Maximum Likelihood agents vanish in the presence of Uniform Bayesian agents.

Proposition 4 holds for every sequence of coins. This is significant because prior assumptions may be incorrect. A Uniform Bayesian agent assumes, a priori, that both types of coins are equally likely to be selected, but nothing ensures that this is true (e.g. only low coins may be selected). Maximum Likelihood agents estimate, from the data, which type of coin is more likely to have been selected and forecast accordingly (hence the intuitive appeal of Maximum Likelihood is greater when there is a large data set available). However, even if the Uniform Bayesian prior assumption is wrong and the coins are *not* selected with equal chance, Maximum Likelihood agents vanish in period one.

With an increasingly large number of observations, the prior is swamped by the data. This creates a well known difficulty for the development of a theory capable of differentiating the properties of priors. On the other hand, in the initial periods, the Bayesian forecasts depend on the prior. So, priors do matter. Let a non-Uniform prior be such that $p_k = p \neq 0.5$. The non-Uniform Bayesian forecast for a_k in the first period of the economy, $g_k^{NUB}(a_k|\ \omega_k)$, are given by 7.1 with $p_k = p \neq 0.5$. The results below show that, for some sequence of coins, a non-Uniform Bayesian agent vanishes, in period 1, in the presence of a Maximum Likelihood agent. In this sense, the Uniform Bayesian model is the only Bayesian model that dominates the Maximum Likelihood model.

Proposition 5. *Assume $S = 1$. For every $p \neq 0.5$ there exists $\bar{x} \in (0, 0.5)$ and an infinite sequence of coins $\bar{\lambda} \in \{h, l\}^\infty$, such that for $P^{\bar{\lambda}}$-almost every infinite sequence of coins realizations $\sigma \in \prod_{k=1}^{\infty} \Lambda_k$, $\sigma = (\sigma^K, ...)$, $\sigma^K = (\omega^K, r^K)$,*

$$\frac{g^{NUB}(r^K|\ \omega^K)}{g^{ML}(r^K|\ \omega^K)} \underset{K\to\infty}{\longrightarrow} 0.$$

Proof: *See Appendix.*

Corollary 5. *Assume $S = 1$. Assume a Maximum Likelihood and a non-Uniform Bayesian agent in the economy. For some parameter $\bar{x} \in (0, 0.5)$ and infinite sequence of coins $\bar{\lambda} \in \{h, l\}^\infty$, the non-Uniform Bayesian agent vanishes at period 1 on $P^{\bar{\lambda}}-$almost every infinite sequence of coins realizations.*

Proposition 5 holds for any finite S, but $S = 1$ is an interesting special case. With just one flip of coin k, the Maximum Likelihood forecast for a_k is the highest conceivable $(0.5 + x)$ if a_k is observed and the lowest $(0.5 - x)$ if b_k is observed. An appealing property of Bayesian forecasts is that they do not go to extremes based on a single observation. The downside of Bayesian forecasts is that they are based on prior assumptions that may be wrong. This drawback is greater when the prior is not uniform. Assume that the prior belief is that high coins are more likely than low coins, but high coins were selected sufficiently less often than low coins. This non-Uniform Bayesian agent may vanish in the presence of a Maximum Likelihood agent.

Remark 1. Assume that there is a Maximum likelihood and a Uniform Bayesian agent in the economy. Corollary 4 shows that prices will be determined by the Uniform Bayesian model in the first period. The same result is also true for every other period. Moreover, for every $\varepsilon > 0$, there is $\bar{K}(\varepsilon)$, (independent of the period t) such that if $K \geqslant \bar{K}(\varepsilon)$ then the wealth of the Maximum likelihood agent is smaller than ε in period t with probability $1-\varepsilon$ (a proof is available from the author upon request).[11]

Remark 2. There is nothing special about having the coins in parallel as opposed to a time series. Consider the following alternative framework. Assume that in each period there are two possible states of nature, a and b. The probability of a is determined by coin k from the period $(k-1)S$ until the period $kS-1$. The arguments in propositions 4 and 5 demonstrate that for every infinite sequence of coins $\lambda \in \{h,l\}^{\infty}$, a Maximum Likelihood agent is driven out of the market by a Uniform Bayesian agent on P^{λ}-almost every infinite sequence of coins realizations. A non-Uniform Bayesian agent is driven out of market by a Maximum Likelihood agent, for some $\bar{x} \in (0, 0.5)$ and infinite sequence of coins $\bar{\lambda} \in \{h,l\}^{\infty}$, on $P^{\bar{\lambda}}$-almost every infinite sequence of coins realizations.[12]

8. Extensions

In section 5, it was demonstrated that in finite dimensional parameter spaces, Bayesian agents survive in the economy. It would be interesting to know if this result generalizes to an infinite dimensional parameter space.

[11] This is why the assumption of an homogeneous discount factor was kept. Propositions 4 (and 5 as well) hold without it, but then the number of coins required to reduce the wealth of the Maximum Likelihood agent (to ε, with probability $1-\varepsilon$) would depend on the period t.

[12] This alternative modelling choice would keep the framework in this section closer to those in the other sections and, hence, would require less notation. I apologize to the reader, but I could not resist showing that the advantage of the Uniform Bayesian model over Maximum Likelihood is not related to forecasts produced when just a few flips of a coin had been observed.

In section 4, it was demonstrated that agents whose beliefs are not consistent will be driven out of the market by agents with consistent beliefs. Analogously, I conjecture that agents whose beliefs are given by inefficient estimators will be driven out of the market by agents whose beliefs are given by efficient estimators.

The example in Blume and Easley (2001) and the one at the start of section 6 show that Maximum Likelihood agents may survive. In example 2, however, Maximum Likelihood agents are driven out of the market. This difference has a counterpart in the statistical literature. For an i.i.d. data generating process, Maximum Likelihood are efficient in regular finite-dimensional parametric models. Bayesian estimators are also efficient under these regularity conditions (see Lehman (1991) for a detailed analysis). In some non-regular cases, Bayes' estimators are efficient, but Maximum Likelihood estimators are inefficient.[13] I conjecture that, under similar regularity conditions, Maximum Likelihood agents survive, but in some non-regular cases (like in example 2) Maximum Likelihood agents will be driven out of the market by Bayesian agents.

The assumption of just two types of coins simplified the analysis in section 7 enormously. However, it would be interesting to know if the results generalize to a continuum of coin types. I believe that other generalizations, e.g. to finitely many possible outcomes per coin, are straightforward.

9. Conclusion

An empirical model must assign high probability to the realized paths to be influential in the market. This property is also desirable from a predictive viewpoint.

Agents whose beliefs are given by a parsimonious Bayesian model survive in the market. Agents whose beliefs are given by either a Bayesian model with an unnecessarily large parameter space or by a Maximum Likelihood model may be driven out of the market.

In a simple framework where Maximum Likelihood is efficient, Maximum Likelihood agents may not significantly influence prices because if there are several independent sources of uncertainty, Maximum Likelihood agents vanish in the presence of Uniform Bayesian agents. However, Maximum Likelihood agents may not vanish in the presence of Bayesian agents with non-Uniform priors.

[13]For an example, see Hirano and Porter (2001). Unfortunately, they consider distributions that, to embed in a dynamically complete market economy, would require the technical difficulty of uncountable many states of nature in each period.

10. Appendix

Let Q and $\tilde{Q}$ be two arbitrary probability measures on $(T^{\infty}, \Im)$. The measure Q is absolutely continuous with respect to $\tilde{Q}$ (denoted $Q << \tilde{Q}$) if for every set $A \in \Im$, $\tilde{Q}(A) = 0 \Longrightarrow Q(A) = 0$. The measures Q and $\tilde{Q}$ are singular (denoted $Q \perp \tilde{Q}$) if there exists a set $\Omega \in \Im$ such that $\tilde{Q}(\Omega) = 0$ and $Q(\Omega) = 1$.

Proposition 6. *Q is absolutely continuous with respect to $\tilde{Q}$ if and only if*

$$\infty > \lim_{t\to\infty} \frac{d\tilde{Q}_t}{dQ_t} > 0,\ Q - a.s;$$

Q and $\tilde{Q}$ are singular if and only if

$$\lim_{t\to\infty} \frac{d\tilde{Q}_t}{dQ_t} = 0,\ Q - a.s.$$

Proof: See Shiryaev (1991), chapter $III - 9$, theorems 2 and 3.

■

Proof of proposition 2: Let Ω be the set of paths in which the frequency of a in the periods that follows a realization of a is the same as the frequency of a in the periods that follows a realization of b. By the lay of large numbers, if $\theta_1 = \theta_2 = \theta$ then $P^{\theta}(\Omega) = P^{(\theta_1,\theta_2)}(\Omega) = 1$. However, if $\theta_1 \neq \theta_2$ then $P^{(\theta_1,\theta_2)}(\Omega) = 0$.

The set of parameters $\{(\theta_1, \theta_2) \in [0,1]^2 \mid \theta_1 \neq \theta_2\}$ has full Lesbegue measure on $[0,1]^2$. So, $P^{B_L}(\Omega) = 1$ and $P^{B_H}(\Omega) = 1$. By proposition 6,

$$\lim_{t\to\infty} \frac{dP_t^{B_H}}{dP_t^{B_L}} = 0,\ P^{B_L} - a.s. \Rightarrow \text{for almost } \theta \in [0,1],\ \frac{dP_t^{B_H}}{dP_t^{B_L}} = 0\ P^{\theta} - a.s.$$

■

Proof of proposition 3: Given $\theta \in [0,1]$, let Ω_θ be the set of paths such that the frequency of a is θ. By the law of large numbers, $P^{\theta}(\Omega_\theta) = 1$ and $P^{\tilde{\theta}}(\Omega_\theta) = 0$ if $\tilde{\theta} \neq \theta$. So, $P^{ML}(\Omega_\theta) = 0$. Hence, for every $\theta \in [0,1]$, $P^{\theta} \perp P^{ML}$. It is immediate that for every $\theta \in \bar{\Theta}$, $P^{\theta} << P^{B}$. The conclusion follows from proposition 6.

■

Proof of proposition 4: By equation 7.5 and Lemma 2 in Lehrer and Smorodinsky (1996),

$$g^{UB}(a_k|\ \omega_k)\log\left(\frac{g^*(a_k|\ \omega_k)}{g^{UB}(a_k|\ \omega_k)}\right)+g^{UB}(b_k|\ \omega_k)\log\left(\frac{g^*(b_k|\ \omega_k)}{g^{UB}(b_k|\ \omega_k)}\right)\leqslant\bar{\delta}<0. \quad (10.1)$$

Let P^h and P^l be the measures in Λ_k such that a_k always occurs with probability h and l, respectively. Let $\bar{P}\equiv 0.5P^h+0.5P^l$.

By equation 10.1 and the law of iterated expectations,

$$E^{\bar{P}}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}\leqslant\bar{\delta}<0, \quad (10.2)$$

where $E^{\bar{P}}$ is the expectations operator associated with $\bar{P}$. By equation 7.4,

$$E^{P^h}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}=E^{P^l}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}. \quad (10.3)$$

Equations 10.2 and 10.3 imply that

$$E^{P^h}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}=E^{P^l}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}\leqslant\bar{\delta}<0. \quad (10.4)$$

Hence, for every infinite sequence of coins $\lambda\in\{h,l\}^\infty$,

$$\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}E^{P^{\lambda_k}}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}\leqslant\bar{\delta}<0. \quad (10.5)$$

By the law of large numbers, P^λ-almost surely,

$$\lim_{K\to\infty}\frac{1}{K}\sum_{k=1}^{K}\left(\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)-E^{P^{\lambda_k}}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}\right)=0. \quad (10.6)$$

Equations 10.5 and 10.6 imply that

$$\sum_{k=1}^{\infty}\left\{\log\left(\frac{g^*(r_k|\ \omega_k)}{g^{UB}(r_k|\ \omega_k)}\right)\right\}=-\infty\Longrightarrow\frac{g^*(r^K|\ \omega^K)}{g^{UB}(r^K|\ \omega^K)}\underset{K\to\infty}{\longrightarrow}0.$$

■

Remark 3. The key result in proposition 4 is equation 10.4. Once it is demonstrated that the expected value of the log of the forecast ratio is strictly negative, it is a direct consequence of the law of large numbers that the sum of the log of these ratios is minus infinite and, hence, their product is zero. I believe that the intuition behind 10.4 is related to the concavity of the log function. The expected value of the log of Uniform Bayesian forecasts are greater than (say) the expected value of the log of Maximum Likelihood forecasts partially because Uniform Bayesian forecasts are "centered" by the prior, whereas Maximum Likelihood forecasts tend to be more "extreme."

Proof of proposition 5: Assume that $p = 0.5 + y$, where $y \in (0, 0.5)$. Let $\bar{\lambda}$ be the infinite sequence of low coins (the proof is identical if $p = 0.5 - y$ with an infinite sequence of high coins). By 7.1 and 7.2, if coin k is low then values and probabilities of $\log\left(\frac{g^{NUB}(r_k|\ \omega_k)}{g^{ML}(r_k|\ \omega_k)}\right)$ are given by:

$$\log\left(\frac{g^{NUB}(a_k|\ a_k)}{g^{ML}(a_k|\ a_k)}\right) = \log\left(\frac{(0.5+x)^2(0.5+y)+(0.5-x)^2(0.5-y)}{((0.5+x)(0.5+y)+(0.5-x)(0.5-y))(0.5+y)}\right), \quad w.p.\ \ (0.5-x)^2\,;$$

$$\log\left(\frac{g^{NUB}(a_k|\ b_k)}{g^{ML}(a_k|\ b_k)}\right) = \log\left(\frac{(0.5+x)(0.5-x)/(0.5-y)}{((0.5+x)(0.5+y)+(0.5-x)(0.5-y))}\right), \quad w.p.\ \ (0.5-x)\,(0.5+x)\,;$$

$$\log\left(\frac{g^{NUB}(b_k|\ a_k)}{g^{ML}(b_k|\ a_k)}\right) = \log\left(\frac{(0.5+x)(0.5-x)/(0.5-y)}{((0.5-x)(0.5+y)+(0.5+x)(0.5-y))}\right), \quad w.p.\ \ (0.5-x)\,(0.5+x)\,;$$

$$\log\left(\frac{g^{NUB}(b_k|\ b_k)}{g^{ML}(b_k|\ b_k)}\right) = \log\left(\frac{(0.5-x)^2(0.5+y)+(0.5+x)^2(0.5-y)}{((0.5+x)(0.5+y)+(0.5-x)(0.5-y))(0.5+y)}\right), \quad w.p.\ \ (0.5+x)^2\,.$$

The limit of the expected value of $E^{P^l}\left\{\log\left(\frac{g^{NUB}(r_k|\ \omega_k)}{g^{ML}(r_k|\ \omega_k)}\right)\right\}$ as x goes to 0.5 is

$$\log\left(\frac{0.5-y}{0.5+y}\right) < 0.$$

Choose $\bar{x}$ sufficiently close 0.5 such that this expected value is strictly negative. The proof now can be concluded exactly as the proof of proposition 4 was completed after equation 10.4. Hence, I omit this part of the proof.

■

References

[1] Becker, P. (2001) "Are Inefficient Entrepreneurs Driven out of the Market?" mimeo, Cornell University.

[2] Blume, E. and D. Easley (1992) "Evolution and Market Behavior," *Journal of Economic Theory*, 58, 9-40.

[3] Blume, E. and Easley, D. (2001) "If You're So Smart, Why Aren't You Rich? Belief Selection in Complete and Incomplete Markets" mimeo, Cornell University.

[4] Dawid, A. (1984) "Present Position and Potential Developments: Some Personal Views. Statistical Theory. The Prequential Approach." *Journal of the Royal Statistical Society, Series A,* 147, Part 2, 278-292.

[5] Lehrer, E. and R. Smorodinsky (1996): "Compatible Measures and Merging," *Mathematics of Operations Research*, 21-3, 697-706.

[6] Lehmann, E. (1991): *Theory of Point Estimation.* Wadsworth & Books/Cole.

[7] Hirano, K. and J. Porter (2001) "Asymptotic Efficiency in Parametric Structural Models with Parameter-Dependent Support," mimeo, University of California Los-Angeles.

[8] Ploberger, W. and P. Phillips (1998), "Empirical Limits for Time Series Econometric Models," mimeo, University of Rochester.

[9] Sandroni, A. (2000) "Do Markets Favor Agents Able to Make Accurate Predictions?" *Econometrica*, 68-6, 1303-1341.

[10] Shiryaev, A. (1991) "*Probability,*" second edition, Springer -Verlag.