Market Selection in Large Economies... Its just a matter of Luck! *

Filippo Massari

School of Banking and Finance, Australian School of Business

November 10, 2014

Preliminary and incomplete, please do not circulate

Abstract

This paper investigates, in a general equilibrium models with a continuum of traders, the hypothesis that markets favor traders with more accurate beliefs. Contrary to the known results for economies populated by finitely many traders, I find that risk attitudes have an effect on survival and that there are cases in which the market selects against traders with correct beliefs. Remarkably, even in these cases, asymptotic equilibrium prices reflect accurate beliefs. Thus, unlike the other known violations of the market selection hypothesis, my result corroborates Freedman's conjecture that the selection forces in the market support the adoption of rational expectation equilibria.

JEL Classification: D51, D01, G1

*I wish to thank Werner Ploberger, John Nachbar, Larry Blume, David Easley, Pietro Dindo, Alvaro Sandroni, Luciano Pomatto, Philip Dybvig, Aleks Yankelevich, Nick Papageorge and Stephanie Heger.

1 Introduction

According to the *market selection hypothesis*, henceforth MSH, the market selects for the traders with the most accurate beliefs. This hypothesis, first articulated by Alchian (1950) and Friedman (1953), is one of the key arguments for the adoption of rational expectations equilibria: as the consumption-share of traders with correct beliefs converges to one, financial markets can be understood, to a large extent, using models with a representative trader with correct beliefs.

The validity of the MSH has been studied extensively in various, general settings. Existing results (De Long et al (1990, 1991), Shleifer-Summers (1990) and Blume-Easley (1992)) have shown that the MSH can fail in partial equilibrium models or if the market contains some inefficiencies. Conversely, in general equilibrium models with finitely many traders, *small economies* henceforth, it has been shown that the MSH holds true (Sandroni (2000) and Blume-Easley (2006)), albeit mild assumptions on preferences (Kogan et al. (2006)) and on the aggregate endowment process (Kogan et al. (2011), Yan (2008)).

My paper enquires whether this fundamental result also applies to general equilibrium models that satisfy the regularity conditions of Sandroni (2000) and Blume-Easley (2006) but are populated by a continuum of traders, henceforth *large economies*. There are three main reasons to focus on large economies. First, a well-developed market typically contains a large number of traders with similar investment strategies and similar returns. Assuming a continuum of traders provide a good approximation of the case in which the number of trading period is not large enough to discriminate between similar investment strategies. Second, the *large economy* setting allows for the derivation of results that are probabilistically sharp (traders live forever), maintaining the aforementioned property that the most accurate trader cannot be uniquely identified after finitely many trading periods. Third, to justify the assumption that traders are price takers (i.e. the competitiveness of walrasian equilibrium Aumann (1965)). It is often argued that price taking behavior is not at odd with the small economy setting because of the equivalence between a small economy with n traders and a large economy with n groups of identical traders. Nevertheless, this level of homogeneity is hardly, if ever, met in financial markets. The large economy setting does not impose this restriction.

I find that *large economies* are qualitatively different from *small economies*. In small economies, the survival chances of a trader exclusively depend on his discount factor and on the accuracy of his beliefs. Whereas, in a large economy, the survival chances of a group of traders depends on the discount factor, on the accuracy of the beliefs of the most accurate trader in the group and, also, on the effect of risk attitudes on the aggregate investment strategies. Thus, only in large economies, risk attitudes can have an effect on survival (Section 4) and there are cases in which the market selects against traders with correct beliefs (Section 3), i.e. the MSH can fail. Risk attitudes have an effect on the survival chances of groups of traders because they affect the optimal aggregate saving rate. In the CRRA utility specification, the same parameter captures both traders attitudes toward risk and their attitudes toward intertemporal consumption. Risk attitudes have an effect on survival because, given two groups of traders with equivalent heterogeneous beliefs, the two groups have equivalent aggregate investment decisions but the group whose traders are less risk averse have stronger speculative incentives to invest, thus save more and come to dominate (see the example in Section 4). The same phenomenon is also present in small economies (see the example in Section 4.2), nevertheless in this setting, the best trader in the economy is selected faster and beliefs heterogeneity effectively disappear in finitely many periods eliminating the speculative incentives. Thus, in small economies, risk attitudes affect the asymptotic consumption-share distribution but their effect is not strong enough to drive the consumption-share of a trader all the way to 0 (1) and be captured by the standard, coarse notion, of trader survival.

The failure of the MSH I identify shows that there are market conditions under which to know the true distribution is not enough to survive. But who survive then? And what happens to equilibrium prices? My analysis shows that, if the MSH fails, it is impossible to know ex-ante which one of the remaining traders will survive. In other words, the only way for a trader to survive is to get lucky. This high degree of indeterminacy on traders' fate is compensated by a positive result about equilibrium price: irrespectively from who survives, asymptotic equilibrium prices reflect accurate beliefs. Thus, unlike the other known violations of the market selection hypothesis, my result corroborates Freedman's conjecture that the selection forces in the market support the adoption of rational expectation equilibria.

This seemingly counterintuitive result depends on the fact that a trader with incorrect beliefs can come to dominate only if his beliefs are, on the realized sequence, as good as the correct one. Specifically, the failure of the MSH I identify occurs if *i*) the true probability is such that the maximum likelihood parameter is a random variable with continuum support; *ii*) there is always a (positive mass) of traders whose believes are, by luck, arbitrarily close to the empirical maximum likelihood parameter and *iii*) traders with incorrect beliefs are investing aggressively enough for luck to pay off. The first condition is met, for example, if the true data generating process is a mixture of iid processes (aka, exchangeable process, see Appendix A). These processes are a natural generalization of iid processes and model the realistic situations in which the true probability is an iid process whose parameters are, ex-ante, unknown (See Appendix A for discussion). In Kreps (1988) words: *"...exchangeability is the same as "independent and identically distributed with a prior unknown distribution function"..."*. The last two conditions simply require that there is a trader that gets lucky and that the money in the market move fast enough for luck to pay off.

The paper proceeds as follows. Section 2 describes the setting and the assumptions. Section 3 studies economies in which all traders have identical utilities. This setting is already reach enough illustrate the role played by risk attitudes on survival (Section 3.1); to show a case in which the MSH fails (Section 3.2); to highlight that, if the MSH fails, to be lucky is the only way for a trader to survive (Section 3.3) and that asymptotic equilibrium prices reflect accurate beliefs even if the MSH fails (Section 3.4). Section 4 extends my finding to economies with different risk attitudes, characterize equilibrium prices and provide a general sufficient condition for a group of traders to vanish that covers both the small and the large economy setting. In Section 4.2, I discuss the relation between the small and the large setting.

2 The model

2.1 The probabilistic environment and beliefs accuracy

The model is an infinite horizon Arrow-Debreu exchange economy with complete markets with a unique perishable consumption good. Time is discrete and begins at date 0. At each date there is a finite set of states $\mathcal{S} \equiv \{1, ..., S\}$ with cardinality $|\mathcal{S}|$=S. The set of all infinite sequences of states is $\mathcal{S}^\infty$ with representative sequence of realizations $\sigma = (\sigma_1, ...)$. Let $\sigma^t = (\sigma_1, ..., \sigma_t)$ denote the partial history through date t of path σ, $\mathcal{S}^t$ be the set containing all of the different sequences of length t and Σ^t be the algebra that consists of all the finite unions of sequences of length t. Σ is the smallest σ-algebra on $\cup_{t=1}^{\infty}\Sigma^t$. The true probability measure on Σ is P. For any probability measure p on Σ, $p(\sigma^t)$ is the marginal probability of the partial history σ^t; that is, $p(\sigma^t) = p(\{\sigma_1 \times ... \times \sigma_t\} \times S \times S \times ...)$. In the next Sections I will introduce a series of economic variable that depends on σ^t. All of them are assumed to be Σ^t measurable. Each trader has a subjective probabilistic view p^i on Σ. Following the tradition in the market selection literature, I assume that p^i is *dogmatic*: each trader "agree to disagree" with the other traders in the market and trades for speculative reasons.

I rank beliefs' accuracy according to a standard statistical criterion: the likelihood ratio test.

Definition 1. *Given a true probability measure P, trader i's beliefs, p^i, are more accurate than trader j's if* $\frac{p^j(\sigma^t)}{p^i(\sigma^t)} \to^{P\text{-}a.s.} 0$.

This definition, unlike the one adopted by Sandroni (2000) and Blume-Easley (2006), does not use any approximation of the likelihoods of traders' beliefs. There are two reasons for this departure. First, as already observed in Blume-Easley (2006), Sandroni's definition (average accuracy), is to coarse to discriminate between different learning rates[1]. Second, Blume-Easley's definition cannot be applied because, as showed in Massari (2013), it can lead to incorrect results.

[1]The averaging factor, $\frac{1}{t}$, in the measure Sandroni adopts renders the $\frac{k}{2}\log t$ dimensionality component of the BIC approximation mute (see Section 3).

2.2 The traders in the economy

The measure space of traders is $(A, \mathcal{A}, i)$ where A is the unit interval, $\mathcal{A}$ its Borel subsets, and i is the Lebesgue measure. The economy is characterized by the aggregate preferences, $\succ_{\gamma_j}$, and by the aggregate time 0 consumption $C_0^{\gamma_j}$ of N sets of traders A_{γ_j}, $j = 1, ..., N$. $\succ_{\gamma_j}$ and $C_0^{\gamma_j}$ are constructed, respectively, by aggregating the preferences and the initial consumptions of *groups* of individual traders, i, with believes p^i, utilities u^i and infinitesimal time 0 consumption c_0^i. With an abuse of notation, A_{γ_j} represents, at the same time, a set of traders, $A_{\gamma_j} = \{i \in A_{\gamma_j}\}$ and a set of probabilities, $A_{\gamma_j} = \{p^i : i \in A_{\gamma_j}\}$ i.e. the beliefs of the traders in A_{γ_j}. With this in mind, I can introduce the fundamental notion of cluster:

Definition 2. *A cluster,A_{γ_j}, is a measurable open subset of A such that:*

- *cluster A_{γ_j} has strictly positive time 0 consumption: $C_0^{\gamma_j} = \int_{A_{\gamma_j}} c_0^i di > 0$*
- *probabilities in A_{γ_j} belong to the same, regular, probabilistic model $\mathcal{M}_{A_{\gamma_j}}$*
- *traders in A_{γ_j} have identical CRRA utility function $u(c) = \frac{c^{1-\gamma_j}}{1-\gamma_j}$ and identical discount factor β_j.*

The definition of cluster groups traders to ensure tractable aggregate preferences for sets of traders with positive aggregate time 0 consumption.[2] Clearly, a small economy with n traders is formally equivalent to a large economy with n clusters of traders with identical beliefs. The notation adopted, (c_0^i, p^i instead of $c_0(i)$,$p(.|i)$), is intended to ease the comparison between the two settings (i.e. $\sum_{i \in A} c_0^i p^i(\sigma^t) = \int_A c_0^i p^i(\sigma^t) di$).

Regularity of $\mathcal{M}_{A_{\gamma_j}}$, formally defined in Appendix B, is a technical assumption to ensure that cluster's investment strategies can be asymptotically approximated. This assumption is not probabilistically restrictive as most of the commonly adopted parametric models are regular (all of the members of the iid exponential family and

[2]The assumptions in the next Section ensure that the second welfare theorem applies to my setting. Thus I am entitled to make direct assumptions on the initial consumption shares. This is done with the understanding that, the assumptions are made on the pareto weight distribution of the social planner problem in the background of the competitive equilibrium. For the same reason I will not discuss the endowment processes.

most of its non-iid members). Requiring that all of the beliefs in cluster A_{γ_j} belong to the same probabilistic model class is also not restrictive: if it were not the case, A_{γ_j} could be divided into two sub-clusters that satisfy the requirement.

Let $C^{\gamma_j}(\sigma^t) = \int_{A_{\gamma_j}} c^i(\sigma^t)di$ be cluster A_{γ_j}'s consumption at time/event σ^t. In the tradition of the selection literature, the asymptotic fate of a cluster is coarsely characterize by the distinction between those clusters who disappear and those who do not.

Definition 3. *A cluster A_{γ_j} vanishes on σ^t if its aggregate consumption-share converges to 0:* $\lim_{t\to\infty} C^{\gamma_j}(\sigma^t) = 0$. *A cluster A_{γ_j} survives on σ^t if :* $\limsup_{t\to\infty} C^{\gamma_j}(\sigma^t) > 0$. *A cluster A_{γ_j} dominates on σ^t if :* $\lim_{t\to\infty} C^{\gamma_j}(\sigma^t) = 1$.

2.3 The assumptions

Throughout the paper I refer to these assumptions.

A1: All traders have CRRA utility functions $u^i(c) = \frac{c^{1-\gamma_i}}{1-\gamma_i}$ with $\gamma_i \in (0, \infty)$.

A2: The aggregate endowment is bounded above and below.

A3: For all traders i, all dates t and all paths σ, $p^i(\sigma^t) > 0 \Leftrightarrow P(\sigma^t) > 0$.

A4: The competitive equilibrium exists.

A5: In each cluster A_{γ_j}, $c_0^i = c_0(i)$, is a continuous, strictly positive, bounded, integrable function of i.

A6: All traders have identical discount factor β.

Assumptions A1-A3 and A6 are standard in the selection literature. If the traders in the economy can be organized in finitely many clusters with identical beliefs, the economy is formally equivalent to a small economy and Assumptions A1-A3 and A6 are implied by Sandroni (2000)'s Blume-Easley (2006)'s one. As usual, a competitive equilibrium is a sequence of prices $\{q(\sigma^t)\}_{t=1}^{\infty}$ and, for each cluster A_{γ_j}, a sequence of consumption choices $\{C^{\gamma_j}(\sigma^t)\}_{t=0}^{\infty}$ that is affordable, preference maximal on the budget set and mutually feasible. Assumption A4 is made for simplicity. If the economy is equivalent to a small economy, A1-A3 are sufficient for Peleg-Yaari's (1970) existence theorem; while, in properly large economies, it can be shown that the existence of the competitive equilibrium follows from the other assumptions as an implication of Olsroy's (1984) existence theorem. The prove is omitted because notationally intensive and tangent to the main contribution of the paper. Assumption A5 is a smoothness assumption needed for Theorem 1, which can be made WLOG because the second welfare theorem applies to this setting.

3 Homogeneous risk attitudes

This Section covers the case of large economies in which all traders have the same CRRA utility function and identical inter-temporal discount factor (A6). Let start by introducing a simple economy to use as intuitive reference for the results that follows.

The economy is a discrete time Arrow-Debreu exchange economy with complete markets, constant aggregate endowment and two states $\mathcal{S} = \{W, R\}$. There is a unit mass of *unskilled* traders , cluster A_U, with iid beliefs whose union covers the simplex (A_U=$\{\cup_{i\in U} p^i(R)\}$=$\{p \in (0,1)\}$) and a unit mass of identical *skilled* traders, cluster A_B, whose beliefs, p^B, coincide which the probabilities obtained via Bayes' rule from a uniform prior over A_U. Each trader has the same CRRA utility function with parameter γ, identical inter-temporal discount factor β and aims to solve:

$$U^i(c) = E_{p^i} \sum_{t=0}^{\infty} \beta^t u^i(c_t(\sigma))$$

$$s.t. \quad \sum_{t=0} \sum_{\sigma^t \in S^t} q(\sigma^t) \left(c_t^i(\sigma) - e_t^i(\sigma)\right) \leq 0.$$

The true probability, P, is, for now, left unspecified.

In equilibrium, traders maximize their subjective expected discounted utility subject to the budget constraints and markets clear. Traders' first order conditions of the maximization problem imply that, in every path σ^t, $\left(c^i(\sigma^t)\right)^\gamma = (c_0^i)^\gamma \frac{\beta^t p^i(\sigma^t)}{q(\sigma^t)}$, which, rearranging and (Riemann) summing over traders of the same cluster, gives

$$\int_{A_B} c^i(\sigma^t) di = \beta^{t\frac{1}{\gamma}} \frac{\int_{A_B} p^B(\sigma^t)^{\frac{1}{\gamma}} c_0^i di}{q(\sigma^t)^{\frac{1}{\gamma}}} \quad and \quad \int_{A_U} c^i(\sigma^t) di = \beta^{t\frac{1}{\gamma}} \frac{\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di}{q(\sigma^t)^{\frac{1}{\gamma}}}.$$

Taking the ratio of the aggregate consumptions of the two clusters, prices simplify out and we obtain

$$\frac{\int_{A_B} c^i(\sigma^t)di}{\int_{A_U} c^i(\sigma^t)di} = \frac{\int_{A_B} p^B(\sigma^t)^{\frac{1}{\gamma}} c_0^i di}{\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di}. \tag{1}$$

The following Lemma uses standard arguments in the selection literature to shows that Equation 1 is the fundamental quantity to determine which cluster vanishes.

Lemma 1. *Under A1-A6, if all traders have the same utility, cluster A_{γ_j} vanishes on σ if exists a cluster A_{γ_k} such that:* $\frac{\int_{A_{\gamma_j}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di}{\int_{A_{\gamma_k}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} \to 0$

Proof. By A2, $\int_{A_{\gamma_k}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di < \infty$. Thus, by Equation 1

$$\frac{\int_{A_{\gamma_j}} c^i(\sigma^t)di}{\int_{A_{\gamma_k}} c^i(\sigma^t)di} = \frac{\int_{A_{\gamma_j}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di}{\int_{A_{\gamma_k}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} \to 0 \Leftrightarrow \int_{A_{\gamma_j}} c^i(\sigma^t)di \to 0$$

□

Lemma 1 suggests that instead of focusing on aggregate beliefs accuracy, we should focus on risk adjusted aggregate beliefs. The main technical contribution I make is to provide an accurate approximation of risk adjusted aggregate beliefs and show that this distinction, which plays no role in small economies, is relevant in large economies. This approximation is done by generalizing a fundamental result about Bayesian accuracy: the BIC approximation (Schwarz (1978), Clarke-Barron (1990), Phillips-Ploberger(2003), Grünwald (2007)).

BIC approximation. *Let $\mathcal{M}$ be a regular parametric model and $p^B(\sigma^t)$ be the bayesian likelihood obtained from a prior distribution, g, that is continuous and strictly positive on a k-dimensional non-empty open strict subset, A_0, of the parameter space A, then,*

$$\forall \sigma^t \in \Sigma, \qquad p^B(\sigma^t) := \int_A p^i(\sigma^t) g^i di \approx e^{\ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k}{2}\ln t}.$$

Where $p^{\hat{i}(\sigma^t)}$ is the parameter choice in A_0 with the highest likelihood on σ^t.
Moreover, if $P \in A$, then $p^{\hat{i}(\sigma^t)} \to^{P\text{-a.s.}} P$.

The BIC approximation shows that the accuracy of the probabilities obtained via Bayes rule depends on the dimensionality of the prior support (k). It formalizes the intuition that there is an accuracy cost in using models with redundant parameters because some of the information of the sample is "wasted" to learn that their true value is 0.[3] It is important to stress the generality of the result. The first line, does not depends on the true probability P. It tells us that the Bayesian likelihood is, in every sequence, well approximated by the likelihood of the best model in the (interior of the) prior support corrected by a dimensionality term. The second line tells us that, if the true probability is in the prior support, then the maximum likelihood model in A converges P-a.s. to the true probability.

Lemma 2 obtains a similar approximation for risk adjusted aggregate probabilities. The approximation shows that risk attitudes do not interact with the dimensionality component of the BIC.

Lemma 2. *Under A1-A5, for any sequence $\sigma^t : p_{\gamma_j}^{\hat{i}(\sigma)} \in A_{\gamma_j}$, cluster A_{γ_j}'s risk adjusted aggregate beliefs satisfies:*

$$\forall \sigma^t \in \Sigma, \qquad \int_{A_{\gamma_j}} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di \approx e^{\frac{1}{\gamma_j} \ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k_j}{2} \ln t}$$

Where $p_{\gamma_j}^{\hat{i}(\sigma^t)}$ is the parameter choice in A_{γ_j} with the highest likelihood on σ^t and k_j is the dimensionality of A_{γ_j} .
Moreover, if $P \in A_{\gamma_j}$, then $p^{\hat{i}(\sigma^t)} \to^{P\text{-}a.s.} P$.

Proof. See Appendix □

[3] A classical example is the following.
Suppose the true probability is Bernoulli with parameter P. There are two Bayesians traders (B^1, B^2); B^1 has a smooth prior on the Bernoulli family (1 parameter: $k^1 = 1$) and B^2 has a smooth prior on the Markov (1) family (2 parameters: $k^2 = 2$). Since every iid model is also Markov 1, the next period forecasts of both traders converge to the true probability. Nevertheless, application of the BIC approximation reveals that the beliefs of B^1 are more accurate than the beliefs of B^2.

3.1 Risk attitudes and survival

In our reference economy, all *skilled* traders have identical beliefs, thus $p^B(\sigma^t)^{\frac{1}{\gamma}}$ can be taken out of the integral and Equation 1 becomes:

$$\frac{\int_{A_B} c^i(\sigma^t)di}{\int_{A_U} c^i(\sigma^t)di} = \frac{p^B(\sigma^t)^{\frac{1}{\gamma}} \int_{A_B} c_0^i di}{\int_{A_U} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di}. \tag{2}$$

Applying the BIC and Lemma 2 to Equation 2, [4]

$$\lim_{t\to\infty} \frac{\int_{A_B} c^i(\sigma^t)di}{\int_{A_U} c^i(\sigma^t)di} = \lim_{t\to\infty} \frac{p^B(\sigma^t)^{\frac{1}{\gamma}} \int_{A_B} c_0^i di}{\int_{A_U} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} = \lim_{t\to\infty} \frac{e^{\frac{1}{\gamma}\left(\ln p^{\hat{i}(\sigma^t)} - \frac{1}{2}\ln t\right)}}{e^{\frac{1}{\gamma}\ln p^{\hat{i}(\sigma^t)} - \frac{1}{2}\ln t}}$$

we see that, the risk adjusted aggregate likelihood of *skilled* traders has dimensionality term $\frac{1}{\gamma 2}\ln t$, while the risk adjusted aggregate likelihood of *unskilled*traders has dimensionality term $\frac{1}{2}\ln t$, thus risk attitudes have an effect on survival. More generally, the following Proposition holds:

Proposition 1. *Under A1-A6, if all traders have the same utility and the economy only contains a cluster of traders (A_U) with heterogeneous iid beliefs and a cluster of Bayesian traders, A_B with identical, continuous, strictly positive prior on A_U, then*

i) $\gamma \in (0,1) \Leftrightarrow$ *cluster A_B vanishes,* $\forall \sigma^t$

ii) $\gamma = 1 \Leftrightarrow$ *cluster A_B survives but does not dominate,* $\forall \sigma^t$

iii) $\gamma > 1 \Leftrightarrow$ *cluster A_B dominats,* $\forall \sigma^t$.

Proof. Application of Theorem 1 □

The economic intuition goes as follow. The aggregate beliefs of the two clusters are equivalent, thus the two clusters have equivalent aggregate investment strategies. Nevertheless, at an individual level, every *unskilled* trader has a dogmatic believe that the data generating process is iid with parameter p^i. Therefore, most of them believe that prices are incorrect and trade for speculative reasons. [5] In the CRRA utility specification, the γ parameter captures both traders' attitudes toward risk and their

[4] k=1 because for the Bernoulli model, we only need to estimate one parameter

[5] The speculative reasons exists also for *skilled* traders but are weaker because they share the same beliefs and this beliefs is close to equilibrium prices.

attitudes toward inter-temporal consumption. If $\gamma < (>) 1$ all of the *unskilled* traders optimally decide to postpone more (less) consumption than *skilled* traders to take advantage of the speculative opportunity. Thus, on aggregate, *unskilled* traders have an equivalent investment strategy but higher (lower) saving rate than *skilled* traders and dominate (vanish) in every path.

It is interesting to note that, if all traders have log utility ($\gamma = 1$), risk attitudes have no effect on aggregation. This is consistent with Rubinstein's (1974) finding that, if all traders have log utility, an heterogeneous beliefs economy can be equivalently represented as an economy with a representative trader whose beliefs are the wealth-weighted average of traders beliefs. In our case, $A_U = A_B$, thus the representative agents of the two clusters have equally accurate beliefs and both survives.

3.2 The role of the true probability

In Proposition 1, I make a comparison between a group of Bayesian's traders and a group of iid traders. The result holds in every path, thus it holds independent from the true probability distribution P. Depending on the assumptions we make on the true probability, Proposition 1 assumes different interpretations.

3.2.1 Do you learn fast enough?

If we assume that the true probability is iid, then there is one member of the *unskilled* cluster with correct beliefs and infinitesimal consumption. Proposition 1 can be interpret as describing the situation in which there is a "race" between the rate at which the Bayesian traders learn, which is independent from risk attitudes, and the rate at which the market moves consumption-shares to "reward" the members of the iid cluster whose beliefs are more accurate than the Bayesians at t, which it does depend on risk attitudes.

Proposition 1'. *Under the assumptions of Proposition 1, if the true probability is iid*

,

i) $\gamma \in (0,1) \Leftrightarrow$ *cluster* A_B *does not learn fast enough and vanish,* $\forall \sigma^t$

ii) $\gamma = 1 \Leftrightarrow$ *cluster* A_B *learns fast enough to survive, not enough to dominate,* $\forall \sigma^t$

iii) $\gamma > 1 \Leftrightarrow$ *cluster* A_B *learns fast enough to dominate,* $\forall \sigma^t$.

Note that, for $\gamma > 1$, the *skilled* cluster dominates even if the *unskilled* cluster contains a trader with the correct iid beliefs. Nevertheless, this result does not constitute an argument against the MSH because the consumption-share of this trader is 0 and his presence has no effect in the economy.

3.2.2 If you're so rich, why aren't you smart?

A proper violation of the MSH would require a case in which there is a cluster of traders with correct beliefs who vanishes against a cluster of traders with incorrect beliefs. This can be done by assuming that the true probability in our reference economy coincides with *skilled* traders' beliefs. In this case, the interpretation is that the "race" is between the rate at which the market reward the correct beliefs of *skilled* traders and the rate at which the lucky traders among the *unskilled* cluster accumulate wealth.

Proposition 1". *Under the assumptions of Proposition 1, if the true probability* $P = p^B$ *and* $\gamma \in (0,1)$*, skilled traders vanishes in every path: the MSH fails.*

To assume that the true probability coincides with the beliefs of a Bayesian learner may seem odd. Nevertheless this type of probabilities are well defined (Exchangeable processes) and, by De-Finetti's theorem, constitute a natural generalization of iid processes (see Appendix A).

For a concrete case in which $P = p^B$ consider the following **example:**

In our reference economy, let the true probability P evolves according to this (Polya urn) process: the process starts with an urn containing one White ball (W) and one Red ball (R). At each discrete time (trial), we randomly select a ball from the urn to determine the state of the economy. The selected ball is then returned to the urn along with one new ball of the same color. It can be easily verified that:

- the composition of the urn changes over time[6] according to the iterative formula: $P(R_{t+1}|\sigma^t) = \frac{1+\sum_{\tau=1}^{t} I_{\sigma_\tau=R}}{t+2}$
- no *unskilled* trader has correct beliefs (because P is not iid)
- *skilled* traders have correct beliefs: *skilled* traders are Bayesian with uniform prior on (0,1), a routine application of Bayes' rule verifies that their beliefs coincide in every path with the composition of the urn: $p^B(R_{t+1}|\sigma^t) = \frac{1+\sum_{\tau=1}^{t} I_{\sigma_\tau=R}}{t+2} = P(R_{t+1}|\sigma^t)$.

Nevertheless, this economy satisfies the assumptions of Proposition 1, thus, for $\gamma < 1$, *skilled* traders vanishes and the MSH fails.

The interested reader can find a direct proof of this result that only uses basic algebra in Appendix C.

3.3 If the MSH fails, the market selects for luck

Proposition 3.2.2 shows that there are cases in which a cluster whose members have correct beliefs is driven out of the market by a cluster whose members have incorrect beliefs. Here I discuss the consumption-share distribution between the traders in the cluster that dominates. It turns out that, among the group of *unskilled* traders, the market selects for those traders whose beliefs falls, by luck, into a shrinking interval around the empirical maximum likelihood. Luck is defined as follows.

Definition 4. *Trader i is lucky if these conditions hold:*
- the maximum likelihood parameters of the true probability are a random variables
- he believes them to be deterministic constants
- their realized value coincides with trader i's beliefs.

The definition of luck is stringent but unambiguous. It requires the true maximum likelihood parameter to be a RV, instead of a constant. The reason is that if the maximum likelihood parameters are random variables, it is impossible for a trader to

[6] For example, if the first ball extracted is Red the composition of the urn before the second extraction becomes 2 R and 1 W: $P(R|R) = \frac{2}{3}$. If the second ball is again Red, then $P(R|R,R) = \frac{3}{4}$. If then we have a White ball: $P(R|W,R,R) = \frac{3}{5}$; iteratively $P(R_{t+1}|\sigma^t) = \frac{1+\sum_{\tau=1}^{t} I_{\sigma_\tau=R}}{t+2}$.

know their value before making investment decisions. Thus ruling out the possible confusion between a trader that knows the true parameter because of his skills and a trader that, as a result of a random devise, adopts the true parameters by chance.

Proposition 2. *In a large economy that satisfies A1-A6 in which the MSH fails, to be lucky is the only way to survive.*

Proof. See Appendix □

In the reference economy of Section 3.2.2, the intuition goes as follows. Given an exchangeable process such as the Polya urn described the empirical maximum likelihood parameter is the realized frequency. If $\gamma < 1$ *skilled* traders vanish, thus a trader with correct beliefs vanish for sure. Among the *unskilled* traders, the market selects for traders with high empirical likelihood, which is to say for the traders whose beliefs are in a shrinking interval around the empirical frequency. Since the empirical frequency is a random variable (the composition of the urn changes stochastically), these traders cannot have any particular merit beside the fact the they made, by luck, the correct guess. Thus, to be lucky is the only way a trader can survive.

3.4 Asymptotic equilibrium prices reflect accurate beliefs

If the MSH holds, convergence to rational expectation equilibria follows from standard economic arguments, but what happens when the market does not select for the traders with correct beliefs? Here I show that, contrary to the other failure of the MSH identified in the literature, the convergence occurs even when the MSH fails: by selecting for lucky traders, the market brings equilibrium prices to reflect beliefs that are, ex post, as accurate as the beliefs of the most accurate trader in the economy.

Proposition 3. *In a large economy that satisfies A1-A6, asymptotic prices reflect the most accurate beliefs among traders*

Proof. If the MSH holds, the result follows from standard arguments. If it does not hold, by Proposition 2 , consumption-shares concentrates on traders whose beliefs are, in every σ^t, close to the belief with maximal empirical likelihood: $p^{\hat{i}(\sigma^t)}$. Thus, by standard arguments, prices reflects beliefs that becomes arbitrarily close to $p^{\hat{i}(\sigma^t)}$. $\square$

In our reference economy, the result can be interpreted as follows:
The Polya urn process can be equivalently thought of as representing the case in which Nature randomizes at time 0 to decide which iid model to use (i.e. the asymptotic frequency $\hat{p}$). Ex-ante, skilled traders' beliefs are correct because they know that Nature is choosing the parameter at random; while each unskilled traders incorrectly believe that there is a unique possible parameter. Ex-post, the market selects for the lucky iid traders whose beliefs are close to the realized parameter, i.e for the traders that have rational expectations, conditionally on the realized value of $\hat{p}$. Therefore, even if the MSH fails, equilibrium prices are in the long run correct and convergence to rational expectations occurs.

4 Heterogeneous utilities and discount factors

The main difficulty to extend the results of the previous sections to the case of different risk attitudes is that, at this level of generality, I need to be able to approximate not only aggregate risk adjusted beliefs but also asymptotic equilibrium prices. In this Section I precisely characterize equilibrium prices and provide a general sufficient condition for a cluster to vanish. My condition shows that the fate of a cluster depends on three components: clusters' inter-temporal discount factors, the accuracy of its most accurate member and the interaction between cluster's dimensionality and risk attitudes.

Let's start with an **Example:**
Consider an Arrow-Debreu exchange economy with two states $\mathcal{S} = \{W, R\}$. The economy contains a cluster of traders A_γ whose traders' iid beliefs cover the simplex: $A_\gamma = \{p \in (0,1)\}$, and a cluster of traders A_η whose traders' iid beliefs cover the same simplex: $A_\eta = A_\gamma$. all traders have identical discount factor β.
Rearranging the FOC and aggregating over clusters as in Section 3,

$$\frac{\int_{A_\eta} c^i(\sigma^t)di}{\int_{A_\gamma} c^i(\sigma^t)di} = \frac{\beta^t \int_{A_\eta} c_0^i p^i(\sigma^t)^{\frac{1}{\eta}} di}{\beta^t \int_{A_\gamma} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} q(\sigma^t)^{\frac{1}{\gamma}-\frac{1}{\eta}}. \tag{3}$$

Equation 3 generalizes Equation 1 to the case of different risk attitudes. It shows that if two clusters have different risk parameter, their asymptotic fate depends on the ratio of their inter-temporal discount factors, the ratio of their risk adjusted aggregate probability, and also on risk adjusted equilibrium prices, which, in this model are an endogenous quantity. Proposition 4, gives us a characterization of asymptotic equilibrium prices that is precise enough to serve to our purposes.

Proposition 4. *In an large economy that satisfy A1-A6, with n clusters, equilibrium prices satisfies:*

$$q(\sigma^t) \approx \max_{j \in n} e^{t \ln \beta_j + \ln \hat{p}^j(\sigma^t) - \frac{\gamma_j k_j}{2} \ln t}$$

Where, $\hat{p}_{\gamma_j}(\sigma^t)$, k_j *and* γ_j *are, respectively, the beliefs of the most accurate trader in cluster* A_{γ_j}, *the dimensionality of* A_{γ_j} *and the IES parameter of cluster* A_{γ_j}.

Proof. See Appendix □

Application of Proposition 4 and Lemma 2 to Equation 3 gives us

$$\begin{aligned}
\frac{\int_{A_\eta} c^i(\sigma^t)di}{\int_{A_\gamma} c^i(\sigma^t)di} &= \frac{\int_{A_\eta} c_0^i p^i(\sigma^t)^{\frac{1}{\eta}} di}{\int_{A_\gamma} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} q(\sigma^t)^{\frac{1}{\gamma} - \frac{1}{\eta}} \\
&\approx \frac{e^{\frac{1}{\eta} \ln \hat{p}(\sigma^t) - \frac{1}{2} \ln t}}{e^{\frac{1}{\gamma} \ln \hat{p}(\sigma^t) - \frac{1}{2} \ln t}} e^{\left(\frac{1}{\gamma} - \frac{1}{\eta}\right)\left(\ln \hat{p}(\sigma^t) - \frac{\gamma}{2} \ln t\right)} \\
&\approx e^{-\frac{1}{2}(1 - \frac{\gamma}{\eta}) \ln t} \rightarrow^{\text{for } \gamma < \eta} 0
\end{aligned}$$

and, in line with the interpretation of the other results, the more risk averse cluster vanishes in every sequence. More generally:

Proposition 5. *In a large economy that satisfy A1-A6 with n clusters with identical aggregate beliefs, the least risk averse cluster dominates in every sequence.*

Proof. Application of Theorem 1 □

4.1 A sufficient condition for a cluster to vanish

The intuition of the example can be used to construct a general sufficient condition to vanish that applies to both small and large economies:

Theorem 1. *In a large economy that satisfy A1-A5, cluster j vanishes P-a.s. if there is a cluster k such that:*

$$(t\ln\beta_j - t\ln\beta_k) + \left(\ln\hat{p}_{\gamma_j}(\sigma^t) - \ln\hat{p}_{\gamma_k}(\sigma^t)\right) + \left(-\frac{\gamma_j k_j}{2}\ln t + \frac{\gamma_k k_k}{2}\ln t\right) \to^{P\text{-}a.s.} -\infty$$

Where, for $n = i, k$, $\hat{p}_{\gamma_n}(\sigma^t)$, k_n and γ_n are, respectively, the beliefs of the most accurate trader in cluster A_{γ_n}, the dimensionality of A_{γ_n} and the IES parameter of cluster A_{γ_n}.

Proof. By the FOC: $C^{\gamma_j} = \frac{\beta_j^t \int_{A_{\gamma_j}} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di}{q(\sigma^t)^{\frac{1}{\gamma_j}}}$. The result follows using Lemma 2 and Proposition 4 to approximate risk adjusted beliefs and equilibrium prices. □

Theorem 1 highlights that the survival of a cluster depends on three exogenous components. The first two components are standard in the selection literature as they do not depends on the dimensionality of cluster's beliefs. The last one is new and only appears in the large setting. The interpretation of these components is straightforward. The first component captures the role of inter-temporal discount factors, the second component captures the accuracy of the most accurate trader in the cluster and the last components capture the interaction between the heterogeneity of opinions and risk attitudes: if, within a cluster, traders invest aggressively, the wealth-shares concentrate fast around the most accurate trader in the cluster and the aggregate strategy of the cluster becomes accurate fast, thus the cluster is hard to beat.

Keeping the other two components equal, differences in the first components indicate that the least patient cluster vanishes P-a.s.; differences in the second components indicate that a cluster vanishes P-a.s. if its most accurate trader is less accurate than the most accurate trader of another cluster; and differences in the last component indicate that a cluster vanishes P-a.s. if there is another cluster which invest more aggressively (i.e. with lower risk adjusted dimensionality ratio).

The example of Section 3 illustrates the way the second and the third components can interact and compensate each other. The accuracy components are respectively: $\ln\hat{p}^B(\sigma^t) = \ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{1}{2}\ln t$ (because of the BIC approximation) and $\ln p^{\hat{i}(\sigma^t)}(\sigma^t)$ (the most accurate *unskilled* trader has iid beliefs whose probability coincides with the empirical frequency), while the risk/dimensionality components are respectively, 0 (because all *skilled* traders have the same beliefs), and $\frac{\gamma}{2}\ln t$ (By Lemma 2). Thus, as shown, Theorem 1 implies that for $\gamma < 1$ *skilled* traders vanish.

4.2 Small economies

A large economy with finitely many clusters of identical traders is formally equivalent to a small economy, thus the condition of Theorem 1 also applies to this setting. In this case, the beliefs of the most accurate trader in the cluster coincides with trader's beliefs and the risk/dimensionality component becomes mute (k=0). For this reason, risk attitudes do not play a significant role on aggregation and do not have an effect on survival. This result is consistent with Sandroni (2000) and Blume-Easley (2006) finding and is summarized in the following corollary:

Corollary 1. *In a small economy that satisfy A1,A2,A3 and A6 the market selects for the most accurate traders P-a.s..*

The qualitative difference between *large and small economies* can be puzzling. My assumptions are implied by Sandroni's (2000) and Blume-Easley's (2006), and the only difference between the two setting is on the cardinality of the set of traders. Their results apply to economies with an arbitrarily large number of traders and yet here I show that are not valid in large economies. It turns out that this discontinuity is only apparent, as it is generated by the dichotomic definition of survival, not by a qualitative difference between the two settings. If instead of focusing on 0 Vs positive asymptotic consumption we were focusing on the size of the asymptotic consumption-shares, we would have found no discontinuity between the two settings: in small economies, risk attitudes have an effect on asymptotic consumption-shares that has the same direction as the one found in large economies. The following "small economy adaptation" of Section 3's setting can serve as an illustrative **example**.[7]

[7]The example also shows that uncountably many traders are already enough for an economy to behave like a large economy, albeit complications due to the way set of beliefs A_{γ_j} and the prior support of the Bayesian traders are constructed.

The economy is a discrete time Arrow-Debreu exchange economy with complete markets, constant aggregate endowment and two states $\mathcal{S} = \{W, R\}$. Traders 1,...,n have heterogeneous iid beliefs p^i, cluster A_U, and traders n+1,...,2n are Bayesian traders, cluster A_B, whose beliefs p^b are derived from a uniform prior on $A_U = \{\cup_{i=1}^n p^i\}$: $p^b(\sigma^t) := \sum_{i=1}^n \frac{1}{n} p^i(\sigma^t)$. All trader have the same CRRA utility function with parameter γ and identical inter-temporal discount factor β. For the sake of the argument, assume that $\forall i, c_0^i = \frac{1}{2n}$ and that p^1 is the most accurate trader[8].

Rearranging the FOC,

$$\frac{\sum_{b\in A_B} c^b(\sigma^t)}{\sum_{i\in A_U} c^i(\sigma^t)} = \frac{\sum_{i=n+1}^{2n} \frac{1}{2n} \left(\sum_{i=1}^n \frac{1}{n} p^i(\sigma^t)\right)^{\frac{1}{\gamma}}}{\sum_{i=1}^n \frac{1}{2n} p^i(\sigma^t)^{\frac{1}{\gamma}}} = \frac{\frac{1}{2}\left(\frac{1}{n} + \frac{1}{n}\sum_{i=2}^n \frac{p^j(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma}}}{\frac{1}{2n} + \frac{1}{2n}\sum_{i=2}^n \left(\frac{p^j(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma}}} \to^{P\text{-a.s.}} \frac{n}{n^{\frac{1}{\gamma}}}.^{9}$$

Which shows that *i*) risk attitudes have an effect on the asymptotic consumption share of the Bayesian trader and *ii*) for $\gamma > 0$ and $n < \infty$, this effect is not strong enough to be detected by the definition of survival.

The reason is intuitive: the consumption-share of the trader with correct beliefs initially grow faster than the consumption-share of the Bayesian traders because the beliefs of the Bayesian traders are initially incorrect. Nevertheless, as the beliefs of the Bayesian concentrate around the true probability, this difference disappears. Risk attitudes have an effect on asymptotic consumption-shares because determine how fast consumption-shares move: the lower the gamma, the fastest consumption-shares move and the lower will be the asymptotic consumption-shares of the Bayesians. The cardinality of I has an effect on survival because it determine both the convergence rate of Bayesian posterior (for p^B) and the convergence rate of wealth-shares: if $|I| < |\mathbb{R}|$, both convergence rates are exponential, thus, in finitely many periods, the Bayesian learns the true probability and heterogeneity disappears from the market. While, if $|I| = |\mathbb{R}|$, both convergence rates are slower than exponential (they are respectively $O(\frac{1}{t^{\frac{k}{2}}})$ and $O(\frac{1}{t^{\frac{k}{\gamma 2}}})$); in this case the market tolerates a mild form of long run beliefs heterogeneity and the posterior never exactly concentrates on the true probability.

[8]According to the definition of cluster, A_U fails to be clusters: c_0 is not continuous. Traders are groped in this fashion to maintain the parallel with the approach followed in Section 3.

[9]The convergence occurs because p^1 is assumed to be the most accurate trader.

5 Conclusions

This paper extends the project started by Sandroni (2000) and Blume-Easley (2006) on market selection in general equilibrium complete markets to the large economy setting. I show that large economies are qualitatively different from small economies in that risk attitudes do play a role on survival. It follows that contrary to the standard result, markets can fail do identify the traders with correct beliefs. This failure of the MSH is qualitatively different from all of the other cases found in the literature in that it does not invalidate Freedman's conjecture that the selection forces in the market support the adoption of rational expectation equilibria. It turns out that equilibrium prices can be asymptotically correct even if the market selects against traders with correct beliefs. My result shows that risk attitudes and aggregation affect investment decisions in a non trivial way even when traders optimize on saving decisions and allocations decisions at the same time. Moreover, my setting allows to discuss the role played by luck in financial markets and its relation with risk attitudes. In particular I find cases in which to be lucky is the only way to survive.

A Exchangeability and De Finetti's theorem

In this Section, I introduce the notion of exchangeable sequences and De Finetti's Theorem. The scope of this Section is to illustrate that it is not only possible, but also natural to think about situations in which P is exchangeable but not iid.

Informally, a sequence of random variables is exchangeable if the probability of the sequence does not depends on the order of the realizations:

Definition 5. *An infinite sequence of realization σ^∞ is exchangeable if, for every finite t, $P(\sigma_1, ..., \sigma_t) = P(\sigma_{\pi(1)}, ..., \sigma_{\pi(t)})$ for any permutation π of the indices*

It follows from the definition that, every sequence of iid random variables, conditional on some underlying distributional form, is exchangeable. De Finetti's Theorem ensures that the converse statement is also true, for infinite sequences, and that every infinite exchangeable sequence can be characterized as a mixture of iid sequences. For illustrative purposes, I make a small departure from the standard formulation of De-Finetti's theorem, which is normally stated with respect to exchangeable sequences, and I formulate it with respect to P: the distribution according to which the sequence is exchangeable.

De Finetti's Theorem. *A probability distribution P on Σ is exchangeable if and only if P is a mixture of iid distributions (Q): $P(A) = \int Q^{\infty}(A)\mu(dQ)$, for some probability distribution μ on the space of all probability distributions on $\boldsymbol{S}$.*

For an intuition of the relationship between exchangeable and iid processes consider these examples of Polya urn processes. Suppose we have an urn that contains N balls with a certain composition of Black balls and White balls. (i) Sampling from the urn with replacement is an iid, hence exchangeable process. (ii) Sampling from the urn, replacing each ball extracted with (n¿1) balls of the same color is exchangeable, not iid, because the probability of an outcome depends on the previous outcomes, and, by De Finetti's Theorem, there is a mixture of iid distribution that coincides with this model. (iii) Sampling from the urn without replacement is an exchangeable process that is not iid, but De Finetti's Theorem does not apply because the process cannot generate infinite sequences.

The importance of De-Finetti's theorem becomes evident in light of the following observation: Suppose we are Bayesian and we want to estimate the probability of Head on a possibly biased coin. The building block of our learning method is the Bayesian prior distribution, which is to say a probabilistic assessment on the possible values of the true probability of Head. Nevertheless, if we believe that the sequence of coin tosses is iid, we incur into a logical contradiction: on one hand we are assuming that there is a deterministic mathematical parameter describing the series of realizations, while on the other hand we are modeling this parameter as if it were a random variable. De-Finetti's Theorem provides an elegant solution to this conundrum introducing the notion of exchangeability. The Bayesian paradigm becomes free of logical contradictions if we assume that the sequence is exchangeable instead of iid: under this point view, the true parameters are indeed random variables.

Thus, unless we have infallible knowledge of the parameters of the data generating process, if we accept the Bayesian paradigm we are also assuming that the true probability is exchangeable.

B Appendix

In this appendix I make use of the notation O(.) and o(.) and $\approx$ with the following meanings. The big-O notation, $f(x) = O(g(x))$, means $\limsup_{x\to\infty} \frac{|f(x)|}{|g(x)|} < \infty$. The little-o notation, $f(x) = o(g(x))$, abbreviates $\lim_{x\to\infty} \frac{f(x)}{g(x)} \to 0$. The $\approx$ notation, $f(x) \approx g(x)$, is used, not conventionally, to abbreviates $\lim \frac{f(x)}{g(x)} \in (0, +\infty)$.

Definition 6. *A probabilistic model $\mathcal{M}$ is regular if it is parametric and, for every p^* in the interior of its parameter set A the following standard approximation holds:*

$$D(p^*||p) := E_{p^*} \ln \frac{p^*}{p} = \frac{1}{2}(p - p^*)^T I(p^*)(p - p^*) + o(||p - p^*||^2)$$

Where $I(p^*)$ is the Fisher information matrix evaluated at p^*.

This high order assumption is typically derived from more fundamental smoothness conditions on the behavior of the log likelihood (see Schwarz (1978), Clarke-Barron (1990), Phillips-Ploberger(2003), Grünwald (2007)). Its derivation is orthogonal to my proof, thus omitted. Also, there are some technical issues that arises when the realized sequence is such that the maximum likelihood parameter is on the boundary of A. This cases are rule out by assuming that all A_{γ_j} are open sets.

Proof of Lemma 2

Proof. First: $\int_{A_{\gamma_j}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di \approx e^{\frac{1}{\gamma}\ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k}{2}\ln t}$. It follows from Lemma 4 by substituting A_{γ_j} for A and exponentiating and ignoring constants.

Second: if $P \in A_{\gamma_j}$, then $p^{\hat{i}(\sigma_t)} \to^{P\text{-a.s.}} P(\sigma_t|\sigma^{t-1})$ by consistency of the maximum likelihood estimator. $\square$

The proof of Lemmas 3 and 4 almost coincide with Grünwald's (2007, pg. 248) proofs of the BIC. The only difference between my proof and Grünwald's is the $\frac{1}{\gamma}$ term and the fact that c_0^i need not to be a density. For, $\gamma = 1$ (log economies) the two proofs coincide. For simplicity, I assume that $\mathcal{M}$ is the iid Bernoulli model. The generalization to other families is straightforward.

Lemma 3. *Let $\mathcal{M}$ be a regular probabilistic model parametrized by A and c_0^i a function that satisfies A5, then:*

$$\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di = \ln \int_A e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di + \ln p^{\hat{i}(\sigma^t)}(\sigma^t)$$

Where $p^{\hat{i}(\sigma^t)}$ is the parameter choice with the highest likelihood on σ^t and $D(p^{\hat{i}(\sigma^t)}||p^i) := E_{p^{\hat{i}(\sigma^t)}} \ln \frac{p^{\hat{i}(\sigma^t)}}{p^i}$ is the one period Kullback-Leibler divergence between p^i and $p^{\hat{i}(\sigma^t)}$.

Proof.

$$\begin{aligned}
\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di &= \ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di + \ln p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}} - \ln p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}} \\
&= \ln \int_A \left. \frac{p^i(\sigma^t)}{p^{\hat{i}(\sigma^t)}(\sigma^t)} \right.^{\frac{1}{\gamma}} c_0^i di + \ln p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}} \\
&= \ln \int_A e^{\frac{1}{\gamma}\left(\ln p^i(\sigma^t) - \ln p^{\hat{i}(\sigma^t)}(\sigma^t)\right)} c_0^i di + \ln p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}} \\
&=^a \ln \int_A e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di + \frac{1}{\gamma} \ln p^{\hat{i}(\sigma^t)}(\sigma^t)
\end{aligned}$$

a: For the Bernoulli model, the result follows because:

$$\ln p^i(\sigma^t) - \ln p^{\hat{i}(\sigma^t)} = t \left(\frac{1}{t} \sum_{\tau=1}^{t} I_{\sigma_\tau = s} \ln \frac{p^i(s_\tau)}{p^{\hat{i}(\sigma^t)}(s_\tau)} \right) = -t E_{p^{\hat{i}(\sigma^t)}} \ln \frac{p^{\hat{i}(\sigma^t)}}{p^i} = t D(p^{\hat{i}(\sigma^t)}||p^i)$$

□

Lemma 4. *Let $\mathcal{M}$ be a regular probabilistic model parametrized by A and c_0^i a function that satisfies A5, then*

$$\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di = \frac{1}{\gamma} \ln p^{\hat{i}(\sigma^t)}(\sigma^t) + \ln \sqrt{\gamma} + \ln c_0^{\hat{i}} - \frac{1}{2} \ln \frac{t}{2\pi} - \ln \sqrt{\det I(p^{\hat{i}(\sigma^t)})} + o(1)$$

Where $I(p^{\hat{i}(\sigma^t)})$ is the Fisher information evaluated at $p^{\hat{i}(\sigma^t)}$.

Proof. By Lemma 3

$$\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di = \ln \int_A e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di + \frac{1}{\gamma} \ln p^{\hat{i}(\sigma^t)}(\sigma^t)$$

For $0 < \alpha < \frac{1}{2}$ let $B_t = \{i : p^i \in [p^{\hat{i}(\sigma^t)} - t^{-\frac{1}{2}+\alpha}, p^{\hat{i}(\sigma^t)} + t^{-\frac{1}{2}+\alpha}]\}$.

By additivity of the integral:

$$\int_A e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di = \int_{A\setminus B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di + \int_{B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di$$

Since c_0^i is continuous on A and strictly positive in $int(A)$ there is a T, such that $\forall t > T\ c_0^i > 0, \forall i \in B_t$. In what follows I always assume $t > T$. The proof is done by performing a second order Taylor expansion of $D(p^{\hat{i}(\sigma^t)})||p^i)$ to bound the two integrals. For $t > T$ the error terms in the Taylor expansions are small for $i \in B$ and

$$D(p^{\hat{i}(\sigma^t)}||p^i) = \frac{1}{2}\left(p^{\hat{i}(\sigma^t)} - p^i\right)^2 I(p^{i^*}) \tag{4}$$

for some $i^* \in B_t$ such that p^{i^*} lies between p^i and $p^{\hat{i}}$.

† **First integral:** $\exists k, a < \infty : \mathcal{I}_1 = \int_{A\setminus B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di < ke^{-at^{2\alpha}} \to 0$

Remember that $D(p^{\hat{i}(\sigma^t)}||p^i)$ as a function of p^i is strictly convex, has a minimum at $p^i = p^{\hat{i}(\sigma^t)}$ and is increasing in $|p^i - p^{\hat{i}(\sigma^t)}|$, so that:

$$0 < \int_{A\setminus B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di < \int_{A\setminus B_t} e^{-\frac{t}{\gamma}\min_{i\in A\setminus B_t} D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di$$

By Equation 4

$$\min_{i\in A\setminus B_t} D(p^{\hat{i}(\sigma^t)}||p^i) \geq \frac{1}{2}t^{-1+2\alpha} \min_{i\in int(A)} I(p^i)$$

so that, since $I(p^i)$ is continuous and > 0 for all $i \in A$, and $\int_{A\setminus B_t} c_0^i di < \infty$,

$$0 < \int_{A\setminus B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di < \int_{A\setminus B_t} e^{-\frac{t}{\gamma}\left(\frac{1}{2}t^{-1+2\alpha}\min_{i\in int(A)} I(p^i)\right)} c_0^i di < ke^{-at^{2\alpha}}$$

For $a = \min_{i\in A\setminus B_t} \frac{I(p^i)}{2\gamma} > 0$ and $k = \int_{A\setminus B_t} c_0^i di < \int_A c_0^i di < \infty$.

‡ **Second integral:** $\mathcal{I}_2 = \int_{B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di \approx \frac{\sqrt{2\pi}c_0^{\hat{i}}}{\sqrt{t\frac{I(p^{\hat{i}})}{\gamma}}}$

Let $I_t^- := \inf_{i'\in B_t} I(p^{i'})$, $I_t^+ := \sup_{i'\in B_t} I(p^{i'})$, $c_t^- := \inf_{i'\in B_t} c_0^{i'}$, $c_t^+ := \sup_{i'\in B_t} c_0^{i'}$, by Equation 4

$$\mathcal{I}_2 = \int_{B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di = \int_{B_t} e^{-\frac{t}{\gamma}(p^{\hat{i}(\sigma^t)} - p^i)^2 I(i')} c_0^i di$$

Where i' depends on i. Using the definitions above, we get

$$c_t^- \int_{B_t} e^{-\frac{t}{\gamma}(p^{\hat{i}(\sigma^t)}-p^i)^2 I_t^+} di \leq \mathcal{I}_2 \leq c_t^+ \int_{B_t} e^{-\frac{t}{\gamma}(p^{\hat{i}(\sigma^t)}-p^i)^2 I_t^-} di.$$

We now perform the substitutions $z := (p^{\hat{i}(\sigma^t)} - p^i)\sqrt{t\frac{I_t^+}{\gamma}}$ on the left integral and $z := (p^{\hat{i}(\sigma^t)} - p^i)\sqrt{t\frac{I_t^-}{\gamma}}$ on the right integral, to get

$$\frac{c_t^-}{\sqrt{t\frac{I_t^+}{\gamma}}} \int_{|z|<t^\alpha\sqrt{I_t^-}} e^{-\frac{1}{2}z^2} dz \leq \mathcal{I}_2 \leq \frac{c_t^+}{\sqrt{t\frac{I_t^-}{\gamma}}} \int_{|z|<t^\alpha\sqrt{I_t^-}} e^{-\frac{1}{2}z^2} dz$$

We now recognize both integrals as standard Gaussian.

Since, as $t \to \infty$ $I_t^- \to I(p^{\hat{i}})$ and $I_t^+ \to I(p^{\hat{i}})$, the domain of integration tends to infinity for both integrals, so that they both converge to $\sqrt{2\pi}$. Since $c_t^+ \to c_0^{\hat{i}}$ and $c_t^- \to c_0^{\hat{i}}$ the constant in both integrals converges to $\frac{c_0^{\hat{i}}}{\sqrt{t\frac{I(p^{\hat{i}})}{\gamma}}}$ and we get $\mathcal{I}_2 \approx \frac{\sqrt{2\pi}c_0^{\hat{i}}}{\sqrt{t\frac{I(p^{\hat{i}})}{\gamma}}}$.

Putting † and ‡ together:

$$\begin{aligned}\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di &= \ln(\mathcal{I}_1 + \mathcal{I}_2) + \frac{1}{\gamma}\ln p^{\hat{i}(\sigma^t)}(\sigma^t) \\ &\to \frac{1}{\gamma}\ln p^{\hat{i}(\sigma^t)}(\sigma^t) + \ln\sqrt{\gamma} + \ln c_0^{\hat{i}} - \frac{1}{2}\ln\frac{t}{2\pi} - \ln\sqrt{\det I(p^{\hat{i}})} + o(1)\end{aligned}$$

Where the approximation holds uniformly for all $\sigma^t \in A_0$ because the bond on $\mathcal{I}_1$ does not depend on σ^t, whereas, because c_0^i and $I(p^i)$ are continuous functions of i over the compact set A convergence of $\mathcal{I}_2$ is also uniform. □

Proof of Proposition 2 In a large economy that satisfies A1-A6 in which the MSH fails, the market selects for luck.

Proof. Let A be the cluster that dominates and $p^{\hat{i}(\sigma^t)} \in A$ be the beliefs of the trader with the maximum likelihood parameter on σ^t. Let B_t be the following shrinking sub-cluster of A : $B_t = [i \in A : p^i \in p^{\hat{i}(\sigma^t)} - t^{-\frac{1}{2}+\alpha}, p^{\hat{i}(\sigma^t)} + t^{-\frac{1}{2}+\alpha}]$, for $0 < \alpha < \frac{1}{2}$. By the FOC and using † and ‡ in the proof of Lemma 4

$$\lim_{t\to\infty} \frac{\int_{i\in A\setminus B_t} c^i(\sigma^t)di}{\int_{i\in B_t} c^i(\sigma^t)di} = \lim_{t\to\infty} \frac{\int_{i\in A\setminus B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)}w(i)di}{\int_{i\in B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}||p^i)}w(i)di} \to 0$$

Thus, by Lemma 1 consumption-shares concentrate in the shrinking interval B_t around $p^{\hat{i}(\sigma^t)}$. Since $p^{\hat{i}(\sigma^t)}$ is a random variable, B_t is also a random variable and the market is selecting for the lucky traders whose beliefs are, by chance, in B_t. $\square$

Proof of Proposition 4 In an large economy that satisfy A1-A5, with n clusters, equilibrium prices satisfies: $q(\sigma^t) \approx \max_{j\in n} e^{t\ln\beta_j + \ln \hat{p}^j(\sigma^t) - \frac{\gamma_j k_j}{2}\ln t}$.

Proof. By Theorem 1 $\max_{j\in n} e^{t\ln\beta_j + \ln \hat{p}^j(\sigma^t) - \frac{\gamma_j k_j}{2}\ln t} \approx \max_{j\in n} \left\{ \beta_j^t \left(\int_{A_j} c_0^i p^i(\sigma)^{\frac{1}{\gamma_j}} di \right)^{\gamma_j} \right\}$.
The proof is done by contradiction: we have to consider two possible cases:
i) $\exists j, \exists \sigma^t$: $\frac{\beta^t \int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di}{q(\sigma^t)} \to \infty$. By the FOC $C_j(\sigma^t) = \frac{\beta^{\frac{t}{\gamma_j}} \int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di}{q(\sigma^t)^{\frac{1}{\gamma_j}}} \to \infty$ violating the bounded aggregate endowment assumption (A2).
ii) $\exists \sigma^t : \forall j = 1,...,n;$ $\frac{\beta^t \int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di}{q(\sigma^t)} \to 0$. By the FOC, $\forall j = 1,...,n;$ $C_j(\sigma^t) = \frac{\beta^{\frac{t}{\gamma_j}} \int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di}{q(\sigma^t)^{\frac{1}{\gamma_j}}} \to 0$ violating the positive aggregate endowment assumption (A2).
$\square$

C Appendix

I have to show that Equation 2 converges to 0. For the sake of the argument, let assume that all traders have the same initial equilibrium consumption. All traders in A_B are Bayesians with uniform prior on the Bernoulli (0,1), thus $p^B(\sigma^t) = \int_0^1 p(\sigma^t|\theta)d\theta$ renaming, WLOG, θ as i, and adopting the notation of the paper, Equation 2 becomes:

$$\frac{\int_{A_B} c^i(\sigma^t)di}{\int_{A_U} c^i(\sigma^t)di} = \frac{p^B(\sigma^t)^{\frac{1}{\gamma}} \int_{A_B} c_0^i di}{\int_{A_U} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di} = \frac{\frac{1}{2}\left(\int_0^1 p^i(\sigma^t)di\right)^{\frac{1}{\gamma}}}{\frac{1}{2}\int_0^1 p^i(\sigma^t)^{\frac{1}{\gamma}} di} \to 0. \tag{5}$$

An algebraic approximation of these quantities is given in Lemma 5. The result follows by substituting the approximations in Eq. 5.

Lemma 5. *Let σ^t be an arbitrary sequence of realizations with frequency of $R = p$, then, for $\gamma \in (0,\infty)$, $\int_0^1 p^i(\sigma^t)^{\frac{1}{\gamma}} di \approx \left(p^{pt}(1-p)^{(1-p)t}\right)^{\frac{1}{\gamma}} \frac{1}{\sqrt{t}}$.*

Proof. pt is by assumption, the number of R realizations on σ^t, thus

$$\int_0^1 p^i(\sigma^t)^{\frac{1}{\gamma}} di = \int_0^1 i^{\frac{\sum_{\tau=1}^t I_{\sigma_\tau = R}}{\gamma}} (1-i)^{\frac{t - \sum_{\tau=1}^t I_{\sigma_\tau = R}}{\gamma}} di = \int_0^1 i^{\frac{pt}{\gamma}}(1-i)^{\frac{(1-p)t}{\gamma}})di.$$

Partial integration gives

$$\int_0^1 i^{\frac{pt}{\gamma}}(1-i)^{\frac{(1-p)t}{\gamma}})di = \left[\frac{1}{\frac{pt}{\gamma}+1} i^{\frac{pt}{\gamma}+1}(1-i)^{\frac{(1-p)t}{\gamma}}\right]_0^1 + \frac{\frac{(1-p)t}{\gamma}}{\frac{pt}{\gamma}+1}\int_0^1 i^{\frac{pt}{\gamma}+1}(1-i)^{\frac{(1-p)t}{\gamma}-1}di,$$

where the middle term is equal to 0. Repeating this step $\frac{(1-p)t}{\gamma}$ times [10] leads to

$$\begin{aligned}
\int_0^1 i^{\frac{pt}{\gamma}}(1-i)^{\frac{(1-p)t}{\gamma}})di &= \frac{1*2*...*\frac{(1-p)t}{\gamma}}{(\frac{pt}{\gamma}+1)*...*(\frac{pt}{\gamma}+\frac{(1-p)t}{\gamma})}\int_0^1 i^{\frac{t}{\gamma}}di \\
&= \frac{(\frac{(1-p)t}{\gamma})! * (\frac{pt}{\gamma})!}{(\frac{t}{\gamma})!} * \frac{1}{\frac{t}{\gamma}+1} \\
&\approx^a \frac{e^{-\frac{pt}{\gamma}}\left(\frac{pt}{\gamma}\right)^{\left(\frac{pt}{\gamma}\right)+\frac{1}{2}} e^{-\left(\frac{(1-p)t}{\gamma}\right)}\left(\frac{(1-p)t}{\gamma}\right)^{\left(\frac{(1-p)t}{\gamma}\right)+\frac{1}{2}}}{e^{-\frac{t}{\gamma}}\left(\frac{t}{\gamma}\right)^{\frac{t}{\gamma}+\frac{1}{2}} * (\frac{t}{\gamma}+1)} \\
&= \frac{\left(\frac{t}{\gamma}\right)^{\frac{t}{\gamma}+1} p^{\frac{pt}{\gamma}+\frac{1}{2}}(1-p)^{\frac{(1-p)t}{\gamma}+\frac{1}{2}}}{\left(\frac{t}{\gamma}\right)^{\frac{t}{\gamma}+\frac{1}{2}} * (\frac{t}{\gamma}+1)} \\
&\approx \frac{\left(\frac{t}{\gamma}\right)^{\frac{1}{2}} p^{\frac{pt}{\gamma}}(1-p)^{\frac{(1-p)t}{\gamma}}}{(\frac{t}{\gamma}+1)} \\
&\approx \left(p^{pt}(1-p)^{(1-p)t}\right)^{\frac{1}{\gamma}}\frac{1}{\sqrt{t}}
\end{aligned}$$

a) By Sterling's approximation, for every integer x, $x! \in \left[\sqrt{2\pi}x^{x+\frac{1}{2}}e^{-x}, x^{x+\frac{1}{2}}e^{-x+1}\right]$).
Finite constants are disregarded, WLOG, every time I use the notation $\approx$.

$\square$

References

Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *The Journal of Political Economy*, **58** (3), 211–221.

Aumann, R. J. (1964). Markets with a continuum of traders. *Econometrica: Journal of the Econometric Society*, pp. 39–50.

[10] Assuming for simplicity and WLOG, since the difference in te approximation would be a finite quantity, that $\frac{(1-p)t}{\gamma}$ and $\frac{t}{\gamma}$ are integer.

Blume, L. and Easley, D. (1992). Evolution and market behavior. *Journal of Economic theory*, **58** (1), 9–40.

— and — (2006). If you're so smart, why aren't you rich? belief selection in complete and incomplete markets. *Econometrica*, **74** (4), 929–966.

Clarke, B. S. and Barron, A. R. (1990). Information-theoretic asymptotics of bayes methods. *Information Theory, IEEE Transactions on*, **36** (3), 453–471.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Cvitanić, J., Jouini, E., Malamud, S. and Napp, C. (2012). Financial markets equilibrium with heterogeneous agents. *Review of Finance*, **16** (1), 285–321.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, **98** (4), pp. 703–738.

—, —, — and — (1991). The survival of noise traders in financial markets. *The Journal of Business*, **64** (1), pp. 1–19.

Friedman, M. (1953). *Essays in positive economics*, vol. 231. University of Chicago Press.

Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.

Kogan, L., Ross, S., Wang, J. and Westerfield, M. M. (2009). *Market selection*. Tech. rep., National Bureau of Economic Research.

—, Ross, S. A., Wang, J. and Westerfield, M. M. (2006). The price impact and survival of irrational traders. *The Journal of Finance*, **61** (1), 195–229.

Kreps, D. (1988). *Notes on the Theory of Choice*. Westview press.

Ljungqvist, L. and Sargent, T. J. (2004). *Recursive macroeconomic theory*. MIT press.

Massari, F. (2013). Comment on if you're so smart, why aren't you rich? belief selection in complete and incomplete markets. *Econometrica*, **81** (2), 849–851.

Ostroy, J. M. (1984). On the existence of walrasian equilibrium in large-square economies. *Journal of Mathematical Economics*, **13** (2), 143–163.

Peleg, B. and Yaari, M. E. (1970). Markets with countably many commodities. *International Economic Review*, **11** (3), 369–377.

Ploberger, W. and Phillips, P. C. (2003). Empirical limits for time series econometric models. *Econometrica*, **71** (2), 627–673.

Rubinstein, M. (1974). An aggregation theorem for securities markets. *Journal of Financial Economics*, **1** (3), 225–244.

Sandroni, A. (2000). Do markets favor agents able to make accurate predictions? *Econometrica*, **68** (6), 1303–1341.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6** (2), 461–464.

Shleifer, A. and Summers, L. H. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, **4** (2), 19–33.

Yan, H. (2008). Natural selection in financial markets: Does it work? *Management Science*, **54** (11), 1935–1950.