Market selection in large economies: A matter of luck

Filippo Massari

School of Banking and Finance, University of New South Wales

In a general equilibrium model with a continuum of traders and bounded aggregate endowment, I investigate the market selection hypothesis that markets favor traders with accurate beliefs. Contrary to known results for economies with (only) finitely many traders, I find that risk attitudes affect traders' survival and that markets can favor "lucky" traders with incorrect beliefs over "skilled" traders with accurate beliefs. My model allows for a clear distinction between luck and skills, and it shows that market selection forces induce efficient prices even when accurate traders do not survive in the long run.

Keywords. Market selection hypothesis, asset pricing, general equilibrium.

JEL classification. D50, D90, G12.

1. Introduction

According to the *market selection hypothesis* (MSH henceforth), the market rewards the traders with the most accurate beliefs. This hypothesis, first articulated by Alchian (1950) and Friedman (1953), is one of the key arguments that supports the efficiency of financial markets. Under the MSH, markets become efficient because, in the long run, it is the accurate traders who control most of the wealth and determine asset prices.

Since Milton Friedman's work, a number of papers have studied the connection between the MSH and efficiency. Sandroni (2000) and Blume and Easley (2006) show that in general equilibrium models with complete markets, bounded aggregate endowment and finitely many traders with time-separable preferences, the MSH holds and equilibrium prices eventually reflect the beliefs of the most accurate traders in the economy. However, there are several models in which the MSH fails and prices remain asymptotically inefficient. Negative results hold in partial equilibrium models with a continuum of traders (De Long et al. 1991). in temporal equilibrium models in which traders optimize on how to allocate consumption but do not optimize over their savings

(Blume and Easley 1992), in general equilibrium models in which the aggregate endowment either grows without bound or shrinks to zero (Yan 2008), and in general equilibrium models with nonseparable preferences (Borovička 2018, Dindo 2018).

While these findings suggest that the MSH and asymptotic efficiency are equivalent concepts, Kogan et al. (2006), Cvitanić and Malamud (2011), and Cvitanić et al. (2012) demonstrate the opposite. The MSH is not a sufficient condition for market efficiency. In economies with no intermediate consumption, they show that inaccurate traders who vanish can have an everlasting effect on equilibrium prices. In those cases in which inaccurate traders are overconfident on assets that pay on extremely unlikely events, markets can remain inefficient even if all inaccurate traders vanish.

In this paper, I show that the MSH is not necessary for market efficiency. I present a model in which accurate traders lose all their wealth, luck (which I define in Section 2.2) determines which trader survives, and yet markets become efficient in the sense that asymptotic prices of short-lived assets reflect correct beliefs.

My model generalizes Sandroni (2000) and Blume and Easley (2006) settings by allowing for a continuum of trader types (indexed by their beliefs). In what follows, I use *small economies* to refer to economies with finitely many trader types and *large economies* to indicate economies with a continuum of trader types. Allowing for a rich heterogeneity in trader beliefs alters some of the results of the standard model. While markets are asymptotically efficient in both settings, in large economies the MSH can fail because luck and risk attitudes can play a role in traders' survival.

I begin my paper by presenting an example in which the market selects against traders with correct beliefs, luck determines traders' survival, and asymptotic equilibrium prices reflect accurate beliefs. Two conditions are necessary for this result. First, luck can determine survival only if there is a sufficiently rich heterogeneity of beliefs. Imagine an environment with a continuum of traders who are incorrect in the sense that every trader concentrates his beliefs on a different set of paths that individually have a vanishing probability under the correct measure P, but collectively cover the whole set of paths. Then the group is sufficiently diverse so that for every path, there is a (lucky) trader who allocates consumption exactly along this path. Ultimately, it is always one trader from that group who accumulates consumption in the long run (a different trader along every path) and, thus, the group collectively accumulates all consumption in the long run along every path. Second, so as to generate this effect, one needs preferences that are sufficiently elastic. The reason is that each trader believes he is earning a high subjective return on his savings along the particular paths that he believes are likely. A constant relative risk aversion (CRRA) parameter smaller than 1 increases his chances of survival because it gives him enough incentive to save. Contrary to Sandroni (2000) and Blume and Easley (2006), the curvature of preferences matters because, although aggregate consumption is bounded, the consumption of an infinitesimal trader can become unbounded and so, individually, we are in the unbounded setting of Yan (2008).[1]

---

[1] I am thankful to an anonymous referee for providing this interpretation of my result.

In the rest of the paper, I develop the formal theory needed to discuss selection in large economies and to understand my leading example. In the tradition of the selection literature, I propose a survival index and use it to derive a sufficient condition for a positive mass of traders (henceforth a *cluster* of traders) to vanish that applies to both small and large economies. My survival index generalizes those found in the literature by including a new term that captures the effect of risk attitudes on the aggregate savings of those clusters where traders have heterogeneous beliefs. Everything else equal, a cluster where traders have a higher CRRA parameter vanishes against a cluster where traders have a lower CRRA parameter because it has a lower aggregate savings rate. In Section 6, I present my main result. Even in those cases in which the MSH fails, markets are asymptotically efficient: the prices of short-lived assets eventually reflect correct beliefs whenever there is a positive mass of traders with correct beliefs. There are three appendices. In Appendix A, I reconcile the apparent contrast between the selection results in small and large economies. Proofs are provided in Appendices B and C.

2. A precise definition of luck

2.1 *Probabilistic environment*

The model is an infinite horizon Arrow–Debreu exchange economy with complete markets for a unique perishable consumption good. Time is discrete and begins at date 0. At each date, the economy can be in S mutually exclusive states: $\mathcal{S} = \{1, \ldots, S\}$, with Cartesian product $\mathcal{S}^t = \times^t \mathcal{S}$. The set of all infinite sequences of states, paths, is $\mathcal{S}^\infty = \times^\infty \mathcal{S}$, with representative path $\sigma = (\sigma_1, \ldots)$. The equality $\sigma^t = (\sigma_1, \ldots, \sigma_t)$ denotes the partial history until period t, $\mathbf{C}(\sigma^t)$ is the cylinder set with base σ^t, $\mathbf{C}(\sigma^t) = \{\sigma \in \mathcal{S}^\infty | \sigma = (\sigma^t, \ldots)\}$, Σ^t is the σ-algebra generated by the cylinders, and Σ is the σ-algebra generated by their union. By construction, $\{\Sigma^t\}$ is a filtration. Next, I introduce a number of economic (random) variables with time index t. These variables are adapted to the filtration Σ^t.

The true probability measure on $(\mathcal{S}^\infty, \Sigma)$ is P, while each trader has a subjective, possibly incorrect, probabilistic view p^i on $(\mathcal{S}^\infty, \Sigma)$. For any probability measure p on $(\mathcal{S}^\infty, \Sigma)$, $p(\sigma_t | \sigma^{t-1})$ is the probability of a generic outcome in period t conditional on observing history σ^{t-1}, while $p(\sigma^t)$ is the probability of the cylinder with base σ^t, that is, $p(\sigma^t) = p(\mathbf{C}(\sigma^t)) = p(\{\sigma_1\} \times \cdots \times \{\sigma_t\} \times S \times S \times \cdots)$. With an abuse of notation, $p(\sigma^t)$ also indicates the likelihood of p on the first t realization of path σ. For example, if $\mathcal{S} = \{0, 1\}$ and p^i is independent and identically distributed (iid) Bernoulli (for all t, $p^i(\sigma_t = 1) = i$), then $p^i(\sigma^t) = \prod_{\tau=1}^t p^i(\sigma_\tau | \sigma^{\tau-1}) = i^{t_1}(1-i)^{t_0}$, where t_1 and t_0 denote the number of realizations of states 1 and 0 on the first t realization of path σ, respectively.

2.2 *Belief accuracy and luck*

Following an established tradition in the selection literature, I rank traders' accuracy according to the relative likelihood of their beliefs.[2,3]

DEFINITION 1. Trader i is *more accurate* than trader j if $\lim_{t\to\infty} p^i(\sigma^t)/p^j(\sigma^t) = \infty$ P-a.s. He is *as accurate as* trader j if $\lim_{t\to\infty} p^i(\sigma^t)/p^j(\sigma^t) \in (0, \infty)$ P-a.s. He is *less accurate* than trader j if $\lim_{t\to\infty} p^i(\sigma^t)/p^j(\sigma^t) = 0$ P-a.s.

Because no trader beliefs can be more accurate than the true probability, I say that a trader has skills if his beliefs are as accurate as the true probability, while he has no skills if his beliefs are less accurate.

DEFINITION 2. Trader i

- has *skills* if $\lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) > 0$ P-a.s.

- has *no skills* if $\lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) = 0$ P-a.s.

Definition 2 does not rule out the possibility of learning. However, a learning trader is skilled only if he is able to learn the truth quickly. To gain intuition, it is useful to recall the notions of *merging* and *weak merging* (Blackwell and Dubins 1962, Kalai and Lehrer 1994). Trader i's beliefs *(weakly) merge* with the truth if he eventually learns the true probability of (events in the near future) tail events, P-a.s. Because learning the probabilities of tail events is harder than learning the probabilities of events in the near future, merging implies weak merging but not vice versa (Kalai and Lehrer 1994). According to Definition 2, a trader whose beliefs merge with the truth has skills, while a trader whose beliefs do not merge with the truth has no skills, even if his beliefs weakly merge with the truth.

Being skilled is a predetermined characteristic of a trader. It does not depend on empirical evidence because the likelihood ratio condition must hold on a set of sequences with true probability 1 to occur rather than on the realized path. Skilled traders are of interest because they are expected to be more accurate than others and thus survive (e.g., Sandroni 2000, Proposition 3). However, traders' performance depends on the likelihood their beliefs attach to the realized path, rather than an abstract notion of accuracy. The next definition refines the notion of skills by incorporating empirical evidence.

---

[2]Focusing on beliefs' likelihood is in the tradition of the selection literature; however, unlike Sandroni (2000) and Blume and Easley (2006), I cannot rely on approximate measures of it. My result captures $O(\log t)$ differences between traders' log likelihoods. Sandroni's definition (average accuracy) is too coarse to capture these differences because the averaging factor, $O(t^{-1})$, dominates this rate, while Blume–Easley's definition can lead to incorrect results when describing such small differences (Massari 2013, 2017).

[3]Definitions 1, 2, and 3, do not cover the case in which $\lim_{t\to\infty} p^i(\sigma^t)/p^j(\sigma^t)$ does not exist. This case is left unspecified because it does not play a role in my results and is potentially distracting.

DEFINITION 3. Trader i is

- *empirically accurate* on σ if $\lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) > 0$ on σ

- *empirically inaccurate* on σ if $\lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) = 0$ on σ.

Being empirically accurate depends on the path on which the condition is verified. Unlike skills, empirical accuracy is not a predetermined characteristic of a trader. The two definitions are similar but not equivalent. Although a trader with skills is empirically accurate on a set of sequences of true probability 1, there are many paths in which a trader with no skills is empirically accurate; for example, a measure 1 of sequences according to the unskilled trader beliefs.

I say that a trader is lucky if he has no skills and is empirically accurate, that is, if he is empirically accurate against the odds.

DEFINITION 4. Trader i is *lucky* on path σ if he has no skills and is empirically accurate on σ:

$$\lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) = 0 \quad P\text{-a.s.} \quad \text{and} \quad \lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) > 0 \quad \text{on } \sigma.$$

This definition of luck is very stringent: it requires an event of zero probability to occur. Furthermore, because the beliefs of a lucky trader are incorrect P-a.s., luck is not a sufficient condition for learning.

In small economies, the probability of observing at least one lucky trader is 0 because each unskilled trader has zero true probability to be lucky and the countable union of zero probability events has probability zero. On the contrary, there are large economies in which the probability of observing a lucky trader is 1. The set of sequences in which at least one unskilled trader is empirically accurate can be made large enough to cover a set of sequences that has true probability 1 because the uncountable union of zero probability events can have positive probability.

EXAMPLE 1. Suppose a fair coin is tossed t times and that there are 2^t traders. Each trader believes that the coin will deliver a distinct deterministic sequence of length t. Because the number of possible sequences (2^t) and the number of traders coincides, to every sequence, σ^t, there corresponds a trader who believes that σ^t will occur for sure. That is, for every $\hat{\sigma}^t$ there is a (lucky) trader, $\hat{i}(\hat{\sigma}^t)$, for whom the probability of obtaining a favorable realization is extremely low, $P\{\sigma^t \in S^t : p^{\hat{i}}(\sigma^t)/P(\sigma^t) > 0\} = 1/2^t$, but whose beliefs attach more likelihood to $\hat{\sigma}^t$ than the true probability does, $p^{\hat{i}}(\hat{\sigma}^t)/P(\hat{\sigma}^t) = 2^t$. With $t = \infty$, this belief structure illustrates a setting in which I have a lucky trader for every sequence.   ◇

While Example 1 illustrates a simple case of luck, my results do not apply to its setting. The belief structure of Example 1 is incompatible with the existence of the competitive equilibrium because it requires "too many" distinct beliefs and thus "too many"

distinct consumption allocations.[4] The beliefs structure I will use in the rest of the paper requires only one distinct allocation per frequency—a number that grows polynomially in t—rather than one distinct allocation per sequence—a number that grows exponentially in t.

3. The leading example

Consider a discrete time Arrow–Debreu exchange economy with two states $\mathcal{S} = \{W, R\}$, one perishable consumption good, dynamically complete markets, and no aggregate risk. There are two sets of traders with positive masses: A_U and A_S (clusters of traders, according to Definition 6 in Section 4). Individual traders, i, can have different beliefs, p^i, but share an identical CRRA utility function ($u^i(c) = (c^{1-\gamma} - 1)/(1 - \gamma)$) with parameter $\gamma < 1$ and discount factor β. Every individual trader in the economy solves

$$\max_{\{c_t^i(\sigma)\}_{t=0}^{\infty}} E_{p^i} \sum_{t=0}^{\infty} \beta^t u^i\big(c_t^i(\sigma)\big) \quad \text{s.t.} \quad \sum_{t=0}^{\infty}\sum_{\sigma} q_t(\sigma)\big(c_t^i(\sigma) - e_t^i(\sigma)\big) \leq 0,$$

where E_{p^i} is the expectation according to trader i's beliefs, and $c_t^i(\sigma)$, $e_t^i(\sigma)$, and $q_t(\sigma)$ are trader i's consumption, his endowment, and equilibrium prices (of a unit of consumption) in period t on the sequence of realizations σ, respectively.

For $j = U, S$, $C_t^j(\sigma) = \int_{A_j} c_t^i(\sigma)\, di$ is period t aggregate consumption on path σ of cluster A_j. In the tradition of the selection literature, the asymptotic fate of a cluster is coarsely characterized by the distinction between disappearance and nondisappearance.

DEFINITION 5. Cluster A_j *vanishes* on σ if $\lim_{t\to\infty} C_t^j(\sigma) = 0$; it *survives* on σ if $\limsup_{t\to\infty} C_t^j(\sigma) > 0$; it *dominates* on σ if the other cluster vanishes on σ.

The true probability of the states evolves according to the following (Pólya urn) process P_{Polya} (Pólya 1930, Mahmoud 2008). The process starts with an urn containing one white ball (W) and one red ball (R). At the beginning of each period, a ball is randomly selected 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.

Traders in A_S (skilled traders) are allowed to observe the composition of the urn before every draw. They have correct beliefs, for all $i \in A_S$ and $p^i = P_{\text{Polya}}$, and represent a group of traders with inside information.

Traders in A_U (unskilled traders) have heterogeneous iid beliefs about the probability of R. The union of unskilled trader beliefs covers the simplex so that, with an abuse of notation, $A_U = \{i \in \Theta_U = (0, 1)\}$, where i denotes both trader i and his iid beliefs: for all t and i, $p^i(R_t) = i$. The unskilled cluster collects the different opinions of those traders who, not having access to private information, never change their beliefs. Because the

---

[4]To ensure the existence of the competitive equilibrium, the cardinality of the consumption space must be at most countable (Ostroy 1984).

composition of the urn changes over time, P_{Polya} is not iid and all traders in A_U have incorrect beliefs.

Traders' first-order conditions of the maximization problem are sufficient for the Pareto optimum and, in every path σ^t, can be expressed as $(c_t^i(\sigma))^\gamma = (c_0^i)^\gamma \beta^t p^i(\sigma^t)/q_t(\sigma)$, where $p^i(\sigma^t)$ is the probability attached by trader i to path σ^t and c_0^i is his time 0 consumption. Rearranging and aggregating over traders of the same cluster yields

$$C_t^j(\sigma) = \int_{A_j} c_t^i(\sigma)\,di = \beta^{t\frac{1}{\gamma}} \frac{\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\,di}{q_t(\sigma)^{\frac{1}{\gamma}}}. \tag{1}$$

Exponentiating by γ and taking the ratio of clusters' aggregate consumption gives

$$\frac{C_t^S(\sigma)^\gamma}{C_t^U(\sigma)^\gamma} = \frac{\left(\int_{A_S} P_{\text{Polya}}(\sigma^t)^{\frac{1}{\gamma}} c_0^i\,di\right)^\gamma}{\left(\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\,di\right)^\gamma} = \frac{P_{\text{Polya}}(\sigma^t)\left(\int_{A_S} c_0^i\,di\right)^\gamma}{\left(\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\,di\right)^\gamma}. \tag{2}$$

The asymptotic value of (2) determines which cluster vanishes. Later in the paper I develop a method to determine its value. I will show (i) that de Finetti's theorem allows the Polya urn process to be turned into a probability that is parametrizable by $\Theta = [0,1]$,

$$\forall(t,\sigma), \quad P_{\text{Polya}}(\sigma^t) = \int_0^1 p^i(\sigma^t)\,di,$$

and (ii) that

$$\left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\,di\right)^\gamma = \left[\int_{A_j} p^i(\sigma^t)\,di\right] * \left[e^{-(\frac{\gamma-1}{2})\ln t + O(1)}\right].$$

These results (Corollaries 2 and 1, respectively) allows the rewriting of (2) as

$$\frac{C_t^S(\sigma)^\gamma}{C_t^U(\sigma)^\gamma} = \frac{\int_0^1 p^i(\sigma^t)\,di\left(\int_{A_S} c_0^i\,di\right)^\gamma}{\left[\int_{A_j} p^i(\sigma^t)\,di\right] * \left[e^{-(\frac{\gamma-1}{2})\ln t + O(1)}\right]}, \tag{3}$$

which shows that (2) converges to 0 (with probability arbitrarily close to 1). That is, the skilled cluster vanishes[5] and the MSH fails.

This example can be fairly surprising at first glance. All the skilled traders have correct beliefs, all the unskilled traders have incorrect beliefs, and yet skilled traders vanish. Next I give a preview of the results.

---

[5]Because the aggregate endowment is bounded, $(C_t^S(\sigma)/C_t^U(\sigma))^\gamma \to 0 \Rightarrow C_t^S(\sigma) \to 0$.

3.1 *The role played by risk attitudes*

Risk attitudes affect cluster survival because of cluster A_U's belief heterogeneity. If all traders in A_U had identical and incorrect beliefs (as for small economies), their beliefs could be taken out of the integral in (2), the consumption ratio between the two clusters would be proportional to the likelihood ratio between cluster beliefs, and risk attitudes would not affect cluster survival. The skilled cluster would dominate because it is more accurate than the unskilled cluster.

However, because unskilled traders do not have identical beliefs, the right-hand side of (2) does not represent the ratio between two probabilities. de Finetti's theorem (Corollary 2) allows the exchange of the Polya urn process for a probability that is parametrizable by $\Theta = [0, 1]$, i.e., for all (t, σ), $P_{\text{Polya}}(\sigma^t) = \int_0^1 p^i(\sigma^t) \, di$, and allows us to rewrite (3) as

$$\frac{C_t^S(\sigma)^\gamma}{C_t^U(\sigma)^\gamma} = \frac{P_{\text{Polya}}(\sigma^t)\left(\int_{A_S} c_0^i \, di\right)^\gamma}{\left(\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i \, di\right)^\gamma} = \frac{P_{\text{Polya}}(\sigma^t)\beta^t\left(\int_{A_S} c_0^i \, di\right)^\gamma}{P_{\text{Polya}}(\sigma^t)\beta^t e^{-\frac{\gamma-1}{2}\ln t + O(1)}}. \tag{4}$$

The numerator is proportional to the product of the true probability of σ^t and the discount factor β^t, because $(\int_{A_S} c_0^i \, di)^\gamma$ is a finite positive constant. The denominator has an aggregate probability term, $P_{\text{Polya}}(\sigma^t)$, which coincides with the true probability of σ^t, and an aggregate discount factor term, $\beta^t e^{-\frac{\gamma-1}{2}\ln t}$, which also depends on cluster A_U's CRRA parameter. The comparison of cluster probability terms reveals that unskilled traders collectively behave as if they had correct beliefs, even if all unskilled trader has incorrect beliefs. Accordingly, the asymptotic fate of the two clusters is uniquely determined by their aggregate discount factor. With $\gamma < 1$, (4) implies that the unskilled cluster has a higher savings rate than the skilled cluster. The unskilled cluster dominates because its aggregate beliefs are identical to the skilled cluster's and it saves more.

3.2 *Who dominates?*

Equation (4) shows that both clusters have equivalent aggregate beliefs and that cluster selection solely depends on the effect of unskilled traders' risk attitudes on their aggregate discount factor. However, it is not informative enough to indicate how consumption shares are eventually distributed among unskilled traders. In Section 5.3, I demonstrate that within members of the unskilled cluster that dominates, the selection forces favor lucky traders. That is, those traders whose iid beliefs coincide with the empirical frequency of R. The intuition goes as follows.

de Finetti's (1937) theorem (Corollary 2) allows turning the Polya urn process into a probability that is parametrizable by $\Theta = [0, 1]$: for all (t, σ), $P_{\text{Polya}}(\sigma^t) = \int_0^1 p^i(\sigma^t) \, di$. Therefore, the Pólya urn process produces probabilities that are equivalent in distribution to the probabilities obtained using the following two-step procedure. In the first step, Nature randomizes according to a Uniform distribution on $(0, 1)$ to decide the

probability of red: $p(R)$. In the second step, Nature uses $p(R)$ to generate an iid sequence of length t. Skilled traders have skills because they know that Nature is choosing $p(R)$ at random according to a Uniform distribution, $\lim_{t\to\infty} P_{\text{Polya}}(\sigma^t)/P_{\text{Polya}}(\sigma^t) = 1$ P_{Polya}-a.s. Unskilled traders have no skills because each unskilled trader believes that there is a unique possible probability $p^i(R) = i$ and according to the randomization performed by Nature in the first step, each $i \in (0, 1)$ has 0 probability to be the realized value of $p(R)$ for all $i \in A_U$, $P_{\text{1st} step}(\{p(R) = i\}) = 0 \Rightarrow \lim_{t\to\infty} p^i(\sigma^t)/P_{\text{Polya}}(\sigma^t) = 0$ P_{Polya}-a.s. However, the union of unskilled trader beliefs covers the entire simplex. Thus, for every possible realization of $p(R)$, there is a (lucky) unskilled trader, $\hat{i}$, whose belief coincides with $p(R)$. Among unskilled traders, the market selects for $\hat{i}$ because, conditional on $p(R) = \hat{i}$, $\hat{i}$ is the only accurate trader in the economy.

Assume an exchangeable non-iid process (such as the Pólya urn described) plays a fundamental role in identifying luck at a theoretical level. If the true parameter were constant, it would be impossible to distinguish a trader who uses the correct parameter by chance from a trader who truly knows the true parameter. By contrast, in the Pólya urn described, there is no room for confusion. A trader with correct beliefs knows the true parameter in every period, while a trader is lucky if he incorrectly believes the true parameter to be constant and, by chance, his iid beliefs coincide with the realized frequency of red balls.

3.3 *Do markets become asymptotically efficient?*

Markets do become asymptotically efficient: the asymptotic equilibrium prices of the short-lived assets in a large homogeneous discount factor economy with a positive mass of skilled traders reflect correct beliefs even when the market selects against all skilled traders (Section 6). In the leading example, the intuition is as follows: By standard economic arguments, as the consumption share of lucky traders approaches 1, the equilibrium prices of the short-lived assets converge to their discounted beliefs. The result follows by noticing that as the number of trading periods increases, the number of balls in the urn also increases, making the composition of the urn more stable. Asymptotically, the effect of one extra ball per period becomes negligible and the Pólya urn process is indistinguishable from iid extractions from an urn whose composition coincides with the empirical frequency (i.e., the beliefs of the lucky trader).

This result holds in general. If there is a positive mass of traders with correct beliefs and the MSH holds, the market becomes efficient by standard arguments in market selection (Sandroni 2000). Otherwise, in Section 5.4, I show that violations of the MSH can occur only if the economy has a large number of traders, preferences are sufficiently elastic, and the data-generating process is such that the true maximum-likelihood parameter is a random variable with continuum support. That is, markets can select against accurate traders only in those cases in which Nature can be thought of as choosing its parameters at random. Markets are asymptotically efficient because in these cases, the next period beliefs of the lucky trader and the truth are eventually indistinguishable.[6]

---

[6]This result does not apply to the setting of Example 1, which violates Assumptions C3 and A4 in Sections 4.1 and 4.2, respectively. For instance, it is easy to verify that the lucky trader's beliefs never converge to the

4. The general model

4.1 *The traders in the economy*

The economy has N sets of traders with positive measure, clusters A_j, $j = 1, \ldots, N$. The (atomless) traders of each cluster are indexed by i, and are assumed to have identical utility function and discount factor. Each trader, i, has beliefs p^i, endowment processes $e_t^i(\sigma)$, and (infinitesimal) time 0 consumption c_0^i. The measure space of traders is $(A, \mathcal{A}, i)$, where $A := \times_{j=1}^N A_j$ is a subset of $\mathbb{R}^k$ for some $k < \infty$, $\mathcal{A}$ is its Borel subsets, and i is the Lebesgue measure.

DEFINITION 6. A *cluster*, A_j, is a positive measure of traders such that the following statements hold:

C1. Cluster A_j has strictly positive time 0 consumption: $C_0^j = \int_{A_j} c_0^i \, di > 0$.

C2. Traders in A_j have an identical discount factor $\beta_j \in (0, 1)$ and preferences.

C3. One of the following three conditions on trader beliefs holds:

   (i) All traders in A_j have identical beliefs.

  (ii) All traders in A_j are Bayesian with regular priors on the same support.

 (iii) All traders in A_j have multinomial beliefs with the same intertemporal structure (either iid or Markov with finitely many lags) and the union of their beliefs covers the whole simplex.

   Assumption C1 requires the initial (equilibrium) consumption of each cluster to be strictly positive.[7] Assumption C2 is necessary to obtain an analytical form for clusters' optimal consumption decisions as a function of its discount factor, preferences, and aggregate beliefs. Assumption C3 disciplines clusters' aggregate beliefs, and one of the following alternatives holds: (i) all traders have identical beliefs—a cluster of agents with identical beliefs can be treated as a representative trader with positive mass because its traders also have identical utilities and discount factors (C2); (ii) traders disagree on small sample probabilities but agree on tail event probabilities—the traders of a cluster of Bayesians with different (regular) priors on the same support disagree in small samples because of heterogeneity in the priors but agree in the long run because of learning; (iii) all traders disagree on both small sample and tail event probabilities—their beliefs

---

true probability: $\forall t, \|P(\sigma_t) - p^{\hat{i}}(\sigma_t)\|_\infty = |0.5 - 1| + |0.5 - 0| = 1 > 0$. The reason is technical: the space of all binary series with the sup norm has too many distinct elements; it is not separable. Therefore, the competitive equilibrium does not exist because there is no orthonormal basis for the space of consumption (Ostroy 1984). Back to my interpretation, my result does not hold with a belief structure like Example 1, because Nature cannot be thought of as randomizing among the set of distinct infinite paths since this space is not a Lebesgue space.

[7]Subsets of traders whose initial equilibrium consumption is zero can be ignored without loss of generality (WLOG) because they cannot affect any equilibrium quantity: since all traders' preferences satisfy the Inada condition at 0 (by A1), a cluster optimally chose to consume 0 at the first period only if the value of its endowment was 0.

can be aggregated because they at least agree on the intertemporal probabilistic structure. Assumption C3(iii) requires trader beliefs to be parametric and it ensures that each parameter represents a trader type. This structure naturally builds a bijection between the space of parameters and the space of traders, so that the topological properties of a cluster of traders that satisfies C3(iii), $(A_j, \mathcal{A}_j, i)$, and those of the space of the parameters characterizing its traders' beliefs, $(\Theta_j, \mathcal{B}_j, i)$, coincide: each trader is uniquely identified by his beliefs, which are uniquely identified by a vector of parameters. This observation allows using $\mathcal{A}_j$ and Θ_j interchangeably as the domain of integration.

In the rest of the paper, the Lebesgue dimensionality of the set of parameters Θ plays a fundamental role in my condition for a cluster to vanish. I use k^{BIC} to indicate the dimensionality of Θ when it refers to the prior support of a Bayesian cluster that satisfies C3(i) or (ii) and I use k^{MAR} to indicate the dimensionality of Θ when it refers to the heterogeneity of types among traders of a cluster that satisfy C3(iii). To become familiar with this construction, consider an economy with two states: $S = \{W, R\}$. A cluster A_j of traders with iid Bernoulli beliefs whose beliefs set covers the one-dimensional simplex ($\mathcal{A}_j := \bigcup_{i \in \Theta = (0,1)}$ trader i: $\forall t$, $p^i(\sigma_t = W) = i$) has heterogenous types ($k^{\mathrm{MAR}} = 1$) that do not learn ($k^{\mathrm{BIC}} = 0$), while a cluster A_B of Bayesian traders with Uniform priors on the one-dimensional simplex ($\Theta = (0,1)$) has homogenous types ($k^{\mathrm{MAR}} = 0$) that do learn ($k^{\mathrm{BIC}} = 1$).[8]

Finally, a special role in my condition for a cluster to vanish is played by the trader in the cluster whose beliefs have the highest likelihood on σ^t.

DEFINITION 7. Trader $\hat{i}_j(\sigma^t)$ is the *maximum-likelihood trader in A_j at σ^t* if

$$\hat{i}_j(\sigma^t) = \arg\sup_{i \in \Theta_j} p^i(\sigma^t).$$

4.2 *The assumptions*

Throughout the paper, I maintain the following assumptions:

A1. All traders have CRRA utilities: $\forall i \in \bigcup_{j=1}^N A_j$, $u^i(c) = (c^{1-\gamma^i} - 1)/(1 - \gamma^i)$ with $\gamma^i \in (0, \infty)$.

A2. The aggregate endowment is uniformly bounded from above and away from 0.

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

A4. For every cluster, j, c_0^i is a differentiable and strictly positive function of i for every i in the interior of $\mathcal{A}_j$.

Assumption A1 is made for tractability reasons as in Yan (2008). Assumptions A2 and A3 are standard in the selection literature. Assumption A4 is a smoothness assumption needed to ensure a good approximation of a cluster's beliefs around its maximum likelihood trader. If the traders in the economy can be organized into finitely many clusters

---

[8]In Lemma 3 we show that k^{MAR} is a measure of long-run type heterogeneity; as such, it is not affected by heterogeneity on the prior distributions: $k^{\mathrm{MAR}} = 0$ in all clusters satisfying C3(ii).

with identical beliefs, the economy is formally equivalent to a small economy that satis-fies the (more) general assumptions of Sandroni (2000) and Blume and Easley (2006).

A competitive equilibrium is a sequence of state prices $\{q_t(\sigma)\}_{t=0}^{\infty}$ and, for each cluster A_j, a sequence of consumption choices $\{C_t^j(\sigma)\}_{t=0}^{\infty}$ that is affordable, prefer-ence maximal on the budget set, and mutually feasible: $\forall \sigma, \forall t, \sum_{j=1}^{N} \int_{i \in A_j} e_t^i(\sigma)\, di = \sum_{j=1}^{N} \int_{i \in A_j} c_t^i(\sigma)\, di$. The existence of the competitive equilibrium follows from Ostroy's (1984) existence theorem (see Appendix C).

4.3 *The reference economy*

The economy is a discrete time Arrow–Debreu exchange economy with complete mar-kets, bounded aggregate endowment, S states, and N clusters A_j, $j = 1, \ldots, N$. Every individual trader in the economy solves

$$\max_{\{c_t^i(\sigma)\}_{t=0}^{\infty}} E_{p^i} \sum_{t=0}^{\infty} \beta_j^t u^i\big(c_t^i(\sigma)\big) \quad \text{s.t.} \quad \sum_{t=0}^{\infty}\sum_{\sigma} q_t(\sigma)\big(c_t^i(\sigma) - e_t^i(\sigma)\big)) \leq 0.$$

Traders' first-order conditions of the maximization problem are sufficient for the Pareto optimum and, in every path σ^t, can be expressed as $(c_t^i(\sigma))^{\gamma_j} = (c_0^i)^{\gamma_j} \beta^t p^i(\sigma^t)/q_t(\sigma)$. Rearranging and summing over traders of the same cluster yields

$$\int_{A_j} c_t^i(\sigma)\, di = \beta_j^{\frac{t}{\gamma_j}} \frac{\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di}{q_t(\sigma)^{\frac{1}{\gamma_j}}}.$$

Exponentiating both sides by the CRRA parameter and taking the ratio of clusters' risk-adjusted consumption, prices cancel:

$$\frac{C_t^j(\sigma)^{\gamma_j}}{C_t^k(\sigma)^{\gamma_k}} = \frac{\beta_j^t \left(\int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}}\, di\right)^{\gamma_j}}{\beta_k^t \left(\int_{A_k} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_k}}\, di\right)^{\gamma_k}}. \tag{5}$$

The following lemma uses standard arguments from the market selection literature to show that (5) is the fundamental quantity to determine which cluster vanishes.

LEMMA 1. *Under A1–A4, A_j vanishes on σ if A_k exists:*

$$\frac{\beta_j^t \left(\int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}}\, di\right)^{\gamma_j}}{\beta_k^t \left(\int_{A_k} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_k}}\, di\right)^{\gamma_k}} \to 0.$$

PROOF. By A2, $C_t^k(\sigma)^{\gamma_k} < \infty \ \forall k, \forall \sigma^t$. Thus by (5), $\forall \gamma_j, \gamma_k \in (0, \infty)$,

$$\frac{C_t^j(\sigma)^{\gamma_j}}{C_t^k(\sigma)^{\gamma_k}} = \frac{\beta_j^t \left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j}}{\beta_k^t \left(\int_{A_k} p^i(\sigma^t)^{\frac{1}{\gamma_k}} c_0^i \, di \right)^{\gamma_k}} \to 0 \quad \Leftrightarrow \quad C_t^j(\sigma) \to 0.$$

$\square$

Lemma 1 demonstrates that it is the ratio of risk-adjusted aggregate beliefs that determines cluster survival rather than the ratio of cluster aggregate beliefs.

4.4 *Technical background*

The distinction between risk-adjusted and not risk-adjusted aggregate beliefs is moot in small economies because for clusters with homogeneous beliefs, the common belief can be factored out of the integral. On the contrary, I find that this distinction does affect cluster survival in large economies. My results rely on an approximation of the asymptotic value of (5). This approximation can be seen as a generalization of a fundamental result about Bayesian accuracy; i.e., the Bayesian information criterion (BIC) approximation (Schwarz 1978, Clarke and Barron 1990, Ploberger and Phillips 2003, Grünwald 2007). To formally state these results, the following definition is necessary.

DEFINITION 8. Let $\mathcal{M}$ be a member of the exponential family parametrized by Θ.

- Let $\Theta_0 \subset \Theta$. I say that Θ_0 is *regular* if
  - the interior of Θ_0 is nonempty
  - the closure of Θ_0 is a compact subset of the interior of Θ.

- A sequence σ is Θ_0-*regular* if the maximum-likelihood parameter $\hat{i}(\sigma^t)$ exists, it is unique, and it belongs to the regular set Θ_0 for all large t.

- The set of all Θ_0-*regular* sequences is $\hat{S}$.

- A prior is Θ_0-*regular* if it is continuous and strictly positive in Θ_0.

BIC APPROXIMATION, GRÜNWALD (2007). Let $\mathcal{M}$ be a member of the exponential family parametrized by Θ, let Θ_0 be a regular subset of Θ, and let $p^B(\sigma^t)$ be the Bayesian likelihood obtained from a Θ_0-regular prior. Then[9]

$$\forall \sigma \in \hat{S}, \quad p^B(\sigma^t) = \int_\Theta p^i(\sigma^t) g^i \, di = e^{\ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k^{\mathrm{BIC}}}{2} \ln t + O(1)},$$

where k^{BIC} is the Lebesgue dimensionality of Θ.

---

[9]In standard Bayesian notation, $\int_\Theta p^i(\sigma^t) g^i \, di$ is expressed as $\int_\Theta p(\sigma^t | \theta) g(\theta) \, d\theta$.

The BIC approximation shows that the likelihood of the probabilities obtained via Bayes' rule depends on the dimensionality of the prior support (i.e., on the number of parameters that need to be learned). It formalizes the intuition that there is a likelihood cost in using models with redundant parameters because some of the information is "wasted" on learning that the true value of these parameters is zero.[10]

The following lemma obtains a similar approximation for risk-adjusted aggregate beliefs.

LEMMA 2. *Under A1–A4, let $\mathcal{M}_j$ be the beliefs set of cluster j and let $\Theta_{j,0}$ be a regular subset of Θ_j. Then cluster j's risk-adjusted beliefs satisfy*

$$\forall \sigma \in \hat{S}_j, \quad \left(\int_{\Theta_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j} = e^{\ln \hat{p}^{i(\sigma^t)}(\sigma^t) - \frac{\gamma_j k_j^{\mathrm{MAR}}}{2} \ln t + O(1)},$$

where k_j^{MAR} is cluster j's type dimensionality and γ_j is its CRRA parameter.

See Appendix B for proofs not given in the main text.

When $\gamma = 1$(log), Lemma 2 coincides with the BIC approximation. However, for $\gamma \neq 1$ and $k^{\mathrm{MAR}} > 0$, Lemma 2 demonstrates that risk-adjusted aggregate probabilities are not mutually absolutely continuous with respect to their non-risk-adjusted counterparts. In particular, for $\gamma < \eta$, Lemma 2 implies that the ratio of the γ-risk-adjusted aggregate beliefs and the η-risk-adjusted aggregate beliefs diverge in every sequence. In economic terms, cluster γ has a higher savings rate than cluster η; thus it dominates.

Importantly, the BIC approximation and Lemma 2 do not depend on the true data-generating process. They hold, more generally, on every sequence in which the maximum-likelihood parameter (conditional on the model class $\mathcal{M}_j$) lies in a well behaved subset of the parameter space for all large t: $\hat{S}$. Under C3(iii), this set includes almost all paths. For example, for the multinomial (Markov) iid class with parameters covering the simplex, $\hat{S}$ contains the set of all sequences whose limsup and liminf of the (conditional) frequencies of events belong to the interior of the simplex. In particular, $P(\{\hat{S}_j\}) = 1$ for every measure P that does not eventually attach a (conditional) probability zero to one of the outcomes.

The approximation of Lemma 2 also holds when k_j^{BIC} and $k_j^{\mathrm{MAR}} = 0$, respectively, which shows that in small economies, risk-adjusted probabilities are mutually absolutely continuous with their non-risk-adjusted counterparts and risk attitudes have no effect on survival. Moreover, for this case, $\hat{S}_j$ is the set of all paths.

---

[10]A classical example is the following. Suppose the true probability is iid Bernoulli with parameter P. There are two Bayesian traders (B^1, B^2): B^1 has a smooth prior on the Bernoulli family (one parameter, $k_1^{\mathrm{BIC}} = 1$) and B^2 has a smooth prior on the Markov (1) family (two parameters: $k_2^{\mathrm{BIC}} = 2$). 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 B^1's beliefs are empirically more accurate than B^2's.

5. Main result

I am now ready to present a general condition for a cluster to vanish that depends only on exogenous quantities. In the tradition of the selection literature, I assign to every cluster a survival index. The asymptotic fate of each cluster can be determined by pairwise comparison of these indexes.

Definition 9. Cluster's A_j *survival index* on σ is

$$s_j = t \ln \beta_j + \left[\ln p_j^{\hat{i}(\sigma^t)}(\sigma^t) - k_j^{\mathrm{BIC}} \ln \sqrt{t} \right] - \gamma_j k_j^{\mathrm{MAR}} \ln \sqrt{t}.$$

The survival index has four terms. The first three terms are standard in the selection literature: $t \ln \beta_j$ is a cluster's discount factor; $[\ln p_j^{\hat{i}(\sigma^t)}(\sigma^t) - k_j^{\mathrm{BIC}} \ln \sqrt{t}]$ represents the empirical accuracy of the most accurate trader in the cluster. Specifically, $\ln p_j^{\hat{i}(\sigma^t)}(\sigma^t)$ is the most accurate trader likelihood and k_j^{BIC} is the BIC dimensionality term, which equals zero unless traders in A_j are Bayesian traders with k^{BIC}-dimensional prior support of positive Lebesgue measure. The last term, $\gamma_j k_j^{\mathrm{MAR}} \ln \sqrt{t}$, is new and appears only in the large economy setting. It captures the effect of risk attitudes and type dimensionality on each cluster's optimal saving decisions. This term is absent in small economies because k_j^{MAR} is zero in homogeneous belief clusters (and in clusters of Bayesians with regular priors on the same support). The survival indexes determine cluster survival as follows:

Proposition 1. *Under A1–A4, cluster A_j vanishes on $\hat{S}_j \cap \hat{S}_k$ if there is a cluster A_k such that $s_j - s_k \to -\infty$.*

The proof is by application of Lemma 1 using Lemma 2 and BIC to approximate the right-hand side (RHS) of (5).

Proposition 1 links cluster survival to the four components of its survival index. Keeping the other three components equal, differences in the first component indicate that the least patient cluster vanishes. Differences in the second component indicate that a cluster vanishes if its maximum-likelihood trader (parameter choice if it is a Bayesian cluster) is less accurate than the maximum-likelihood trader (parameter choice) of another cluster. Differences in the third component indicate that, given two Bayesian clusters whose support contains the true probability, the cluster that has to estimate more parameters vanishes (as per Theorem 6 in Blume and Easley 2006). In addition, differences in the last component indicate that the cluster with the lowest $\gamma_j k_j^{\mathrm{MAR}}$ term dominates because it saves more.

These four components have different intensities. The first two components diverge at rate t, while the last two diverge at rate $O(1) \ln \sqrt{t}$. Thus, differences in the first two components always dominate differences in the other components.[11] Therefore, if all traders have an identical discount factor, the leading term of the survival indexes is the

---

[11] Differences in the second two components would disappear if I were to use an average measure of accuracy as in Sandroni (2000) because they would be dominated by the $1/t$ term.

one that captures the empirical accuracy: the market selects for empirically accurate traders. For the cases in which there is more than one cluster with the most empirically accurate parameter choice/trader, my condition highlights that risk attitudes can affect survival not only via direct comparison of the last term of the survival indexes (Section 5.1), but also via the interaction between its third component, which captures the parameter dimensionality, and its last component, which captures the interaction between risk and agent type dimensionality (Section 5.2). This interaction can be responsible for failures of the MSH.

In the next sections, I discuss specific implications of Proposition 1. Because the first two components of the survival index are well understood, I focus on economies in which only the last two components differ, i.e., economies in which all clusters have a homogeneous discount factor (HDF) and the same maximum-likelihood trader/parameter.

DEFINITION 10. An economy is a *HDF* if $\forall i \in A$, $\beta_i = \beta \in (0,1)$.

5.1 *The role of risk attitudes*

To highlight the effect of risk attitudes on cluster survival, I start with the simple case in which clusters differ only in their risk attitudes.

PROPOSITION 2. *In an HDF economy that satisfies A1–A4 with N clusters, with identical belief sets Θ and $k^{\mathrm{MAR}} > 0$, the least risk-averse cluster dominates on $\hat{S}$.*[12]

PROOF. We have that $\gamma_j < \gamma_k \Rightarrow s_k - s_j \to -\infty \Rightarrow^{\mathrm{By\ Th.\ 1}} k$ vanishes.   □

EXAMPLE 2. Consider an Arrow–Debreu exchange economy with two states $\mathcal{S} = \{W, R\}$. The economy contains two clusters, A_γ and A_η, with an identical discount factor β but different risk attitudes $\gamma < \eta$. All traders have iid Bernoulli beliefs so that $\Theta_\eta = \Theta_\gamma = \{i \in (0,1)\}$ and $k_\eta^{\mathrm{MAR}} = k_\gamma^{\mathrm{MAR}} = 1$. It follows that $s_\eta - s_\gamma = \gamma \ln \sqrt{t} - \eta \ln \sqrt{t} \to -\infty$ and, by Proposition 1, the most risk-averse cluster, A_η, vanishes.   ◇

Because A_γ and A_η have an identical beliefs set, Example 2 highlights that risk attitudes affect cluster survival through their impact on cluster savings rate. In the CRRA utility specification, the CRRA parameter captures both trader attitudes toward risk and their attitudes toward intertemporal consumption. Everything else equal, a low CRRA parameter increases the survival chances of a cluster because it gives higher incentives to save to its empirically most accurate traders.

The following corollary decomposes the optimal consumption plans of clusters into their aggregate belief and aggregate discount factor (saving) components.

COROLLARY 1. *Under A1–A4, cluster A_j's risk-adjusted aggregate beliefs satisfy*

$$\beta_j^t \left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j} = \left[\int_{A_j} p^i(\sigma^t) \, di \right] * \left[\beta_j^t * e^{-(\frac{\gamma_j k^{\mathrm{MAR}} - k^{\mathrm{MAR}}}{2}) \ln t + O(1)} \right].$$

---

[12]Where $\hat{S} = \bigcap_{j \in N} \hat{S}_j = \hat{S}_j$ for $j = 1, \ldots, N$, because all clusters have the same belief support.

The first component represents cluster aggregate beliefs; the second expresses the cluster aggregate discount factor.

The effect of risk attitudes on aggregate savings can be better understood by focusing on the recursive version of this competitive equilibrium. Because of the heterogeneity of trader opinions, in every period most traders subjectively believe that assets are mispriced and trade for speculative reasons. If traders have log utility ($\gamma = 1$), prices do not affect optimal investment choices and aggregation does not affect the cluster aggregate savings rate. The cluster's optimal choices can be equivalently modeled as those of a representative trader with positive mass and discount factor β, and whose beliefs coincide with the consumption share-weighted average of trader beliefs within the cluster (Rubinstein 1974). However, if $\gamma < (>)1$, the substitution effect is stronger (weaker) than the income effect and each member of the cluster optimally chooses to invest more (less) aggressively than if he/she had log utility. Because investing is the only way to save in this economy, this translates into a representative agent with a higher savings rate. Contrary to Sandroni (2000) and Blume and Easley (2006), the curvature of preferences matters because the consumption of an infinitesimal trader can become unbounded, even if the aggregate consumption is bounded. My results are qualitatively similar to Yan's (2008) because, at an infinitesimal level, the consumption shares dynamic of a large economy resembles that of a growing economy.

5.2 *The role of heterogeneity of opinions*

In this section, I analyze survival in economies that contain some clusters with heterogeneous beliefs (C3(iii)) and some clusters with Bayesian traders (C3(ii)). I show that cluster beliefs' dimensionality and risk attitudes, $\gamma_j k_j^{\mathrm{MAR}}$, have an effect on cluster survival that is of the same order as that of belief dimensionality for a Bayesian cluster, k_B^{BIC}. Therefore, they can offset each other. Let us start with the case in which all traders in the Bayesian cluster have identical priors (C3(i)).

PROPOSITION 3. *Under A1–A4, if the economy is HDF and contains only two clusters—A_U, whose traders have heterogeneous beliefs, and A_B, whose traders are Bayesian with identical, regular prior distribution on Θ_U—then*

(i) $\gamma_U \in (0, 1) \Leftrightarrow$ *cluster A_B vanishes $\forall \sigma \in \hat{S}$*

(ii) $\gamma_U = 1 \Leftrightarrow$ *cluster A_B survives but does not dominate $\forall \sigma \in \hat{S}$*

(iii) $\gamma_U \in (1, \infty) \Leftrightarrow$ *cluster A_B dominates $\forall \sigma \in \hat{S}$.*

The proof is achieved via application of Proposition 1 by noticing that $s_U - s_B = (k_B^{\mathrm{BIC}} - \gamma_U k_U^{\mathrm{MAR}}) \ln \sqrt{t}$.

EXAMPLE 3. Consider an Arrow–Debreu exchange economy with two states $\mathcal{S} = \{W, R\}$. There are two clusters, A_U and A_B, with identical risk attitudes γ and discount factor β. Traders in A_U have heterogeneous Bernoulli beliefs p^i such that $\Theta_U = \{i \in (0, 1)\}$

(i.e., $k_U^{\text{MAR}} = 1$ and $k_U^{\text{BIC}} = 0$), while traders in A_B have identical beliefs p^B that are obtained via Bayes' rule from a regular prior distribution on $\Theta_B = (0, 1)$ (i.e., $k_U^{\text{MAR}} = 0$ and $k_U^{\text{BIC}} = 1$). The result follows as an application of Proposition 1:

$$\forall \sigma \in \hat{S}, \quad s_B - s_U = \frac{\gamma k_U^{\text{MAR}}}{2} \ln t - \frac{k_B^{\text{BIC}}}{2} \ln t \to \begin{cases} 0 & \text{if } \gamma \in (0, 1) \\ r \in (0, \infty) & \text{if } \gamma = 1 \\ \infty & \text{if } \gamma \in (1, \infty). \end{cases}$$

The CRRA parameter does not affect the long-run aggregate savings of the Bayesian cluster because eventually all Bayesian traders agree with the probability implicit in equilibrium prices and use the market exclusively to smooth consumption rather than speculate. Conversely, traders in A_U never learn the truth and have speculative incentives to trade. Let $B_t(\hat{i}) \in A_U$ be a shrinking measure of traders with parameters around the lucky trader. For $\gamma < 1$, traders in $B_t(\hat{i})$ save more on the realized path than log traders would (as per Yan's (2008) result for growing economies), while traders outside $B_t(\hat{i})$ save less than what log traders would save, because they are investing more consumption on paths that do not realize. Proposition 1 ensures that at the aggregate level, the former effect dominates the latter: the aggregate savings rate of cluster A_U is higher than that of cluster A_B.

The next lemma allows us to generalize Proposition 3 to clusters in which traders have different regular priors over the same support (C3(ii)). Lemma 3 shows that the dimensionality term of a cluster of Bayesian traders with heterogeneous, regular priors on the same support equals 0 ($k_{\text{Bayesian}}^{\text{MAR}} = 0$): Bayesian traders with the same prior support and regular prior have zero type dimensionality because their beliefs eventually coincide.   ◇

Lemma 3. *A cluster of Bayesian traders with heterogeneous, regular priors on a common parameter support, Θ, can be treated as a cluster of Bayesian traders with an identical, regular prior on Θ.*

5.3 *The role of the true probability*

In Proposition 3, I characterize the long-run consumption share distribution between a cluster of Bayesian traders and a cluster of traders with heterogeneous beliefs. This result holds in every path in $\hat{S}$. Therefore, it does not depend on the true data-generating process and allows (almost) complete freedom in choosing the true distribution. Assuming that the true probability coincides with the beliefs of the Bayesian cluster and that $\gamma < 1$, we have an economy in which all traders in A_B have correct beliefs and yet cluster A_B vanishes: the MSH fails and luck determines which trader survives.

Proposition 4. *Under the assumption of Proposition 3, if we further assume that $P = p^B$ and $\gamma_U < 1$, then, with a probability arbitrarily close to 1, the MSH fails and luck determines which trader survives.*

But what does it mean that the true probability coincides with the probability obtained via Bayes' rule? The following corollary to de Finetti's (1937) theorem reminds the reader that the Pólya urn process in the leading example satisfies this requirement: de Finetti's theorem allows turning the Polya urn process into a probability that is parametrizable by $\Theta = [0, 1]$: for all (t, σ), $P_{\text{Polya}}(\sigma^t) = \int_0^1 p^i(\sigma^t)\,di$, which coincides with the probability of a Bayesian with a Uniform prior over the simplex.

COROLLARY 2. *The probability attached by the Pólya urn process in the leading example, P_{Polya}, coincides in every path with the probability obtained by Bayes' rule from a Uniform prior on the unit simplex of the Bernoulli iid family, p^B:*

$$\forall \sigma, \forall t, \quad P_{\text{Polya}}(\sigma^t) = \int_0^1 p^i(\sigma^t)\,di = p^B(\sigma^t).$$

The proof is established by a standard application of de Finetti's theorem (e.g., Mahmoud 2008, p. 30).

Similar examples can be constructed as long as the true data-generating process is exchangeable but not iid; that is, if the true data-generating process is not iid, but the probability of finite sequences does not depend on the order of the realizations (e.g., draws from a deck of cards without replacement).

DEFINITION 11. An infinite sequence of realizations σ^∞ is *exchangeable* if, for every finite t, $P(\sigma_1, \ldots, \sigma_t) = P(\sigma_{\pi(1)}, \ldots, \sigma_{\pi(t)})$ for any permutation π of the indexes.

From Definition 11, it follows that every sequence of iid random variables, conditional on some underlying distributional form, is exchangeable. de Finetti's (1937) theorem gives us a partially converse statement: every infinite exchangeable sequence can be characterized as a mixture of iid sequences; that is, every sequence of exchangeable random variables has a representation of the form $p(\sigma^t) = \int_\Theta p^i(\sigma^t)g^i\,di$, where the p^i are iid probability measures and g^i is the weight assigned to each model. Interpreting g^i as a prior distribution in the Bayesian sense, this representation implies that to every Bayesian model (with obvious generalization to a non-iid setting), there is a corresponding exchangeable (conditionally exchangeable) model and vice versa.

In the words of Kreps (1988), "…exchangeability is the same as 'independent and identically distributed with a prior *unknown* distribution function'…"

In the leading example, skilled traders have rational expectations because they know the *"unknown"* distribution function.

These observations can be used to construct other examples in which the MSH fails and luck is the sole determinant of trader survival.

EXAMPLE 4. Consider an Arrow–Debreu exchange economy with two states $\mathcal{S} = \{W, R\}$. The true probability P evolves according to the same Pólya urn process I used in my leading example. There are two clusters, A_S and A_U, with an identical discount factor β. Traders in A_S (skilled traders) are Bayesian with Uniform priors on $\Theta_S = (0, 1)$ ($k_S^{\text{BIC}} = 1$), so that by Corollary 2, $P = p^B$. Traders in A_U (unskilled traders), have heterogeneous

Markov 1 beliefs p^i such that $\Theta_U = \{i \in (0, 1)^2\}$. Note that $k^{\mathrm{MAR}} = 2$ (the Markov 1 model has two parameters to be estimated: $p(W|R)$ and $p(W|W)$) and that the Bernoulli model is nested in the Markov 1 model: $\Theta_S \subset \Theta_U$. It is easy to verify that $s_S - s_U = (-k_S^{\mathrm{BIC}} + \gamma_U k_U^{\mathrm{MAR}}) \ln \sqrt{t} = (-1 + 2\gamma_U) \ln \sqrt{t}$; thus, if $\gamma_U < 0.5$, skilled traders vanish, the MSH fails, and a lucky trader dominates.   ◇

In Example 4, a value of γ smaller than in the leading example is needed to determine a failure of the MSH. This reflects the intuition that a qualitatively equal amount of aggregate consumption needs to be shared between a qualitatively larger set of traders in a Markov 1 cluster ($k_{M1}^{\mathrm{MAR}} = 2$) rather than an iid cluster ($k_{\mathrm{IID}}^{\mathrm{MAR}} = 1$). Accordingly, the lucky trader in the Markov cluster must be given more incentives to save than the lucky trader in the iid cluster because he gets a smaller infinitesimal share of the cluster's initial consumption.

5.4 *Necessary conditions for a violation of the MSH*

I have presented two examples in which the MSH fails. These examples have three elements in common: a large number of traders, preferences that are sufficiently elastic, and a data-generating process such that the true maximum-likelihood parameter is a random variable with continuum support. All these requirements are necessary for a violation of the MSH.

PROPOSITION 5. *In an HDF economy that satisfies A1–A4, if a skilled cluster A_S vanishes P-a.s., then the following statements hold:*

(a) *At least one cluster, A_j, has heterogeneous traders.*

(b) *Cluster A_j's trader preferences are sufficiently elastic: $\gamma_j \leq 1$.*

(c) *The true maximum-likelihood parameter is a random variable with continuum support; that is, $k_P^{\mathrm{BIC}} > 0$.*

A large number of traders is necessary because luck can occur only if there is enough heterogeneity in trader beliefs; otherwise we are in the small economy setting of Sandroni (2000) and Blume and Easley (2006) in which skilled clusters survive P-a.s. Preferences that are sufficiently elastic are necessary to give the lucky traders in A_j enough incentive to save (as per Proposition 3). Condition (c) is necessary for A_j's survival index to be higher than that of the skilled cluster's, A_S. If the maximum-likelihood parameters of the true data-generating process were constants (or random variables with finite support), then A_S would have maximal survival index because both its BIC and MAR components equal 0.

6. MARKETS ARE ASYMPTOTICALLY EFFICIENT

In this section, I prove that the equilibrium prices of short-period assets eventually reflect correct beliefs whenever there is a skilled cluster.

PROPOSITION 6. *In an HDF economy that satisfies A1–A4, if there is a cluster of traders with correct beliefs, asymptotic prices are efficient: the prices of short-lived assets converge to the discounted, risk-adjusted beliefs of a trader with correct beliefs,*

$$\forall \sigma \in \hat{S}, \quad \left\| q(\sigma_t | \sigma^{t-1}) - \frac{u^{\hat{i}(\sigma^{t-1})} (c_{t-1}^{\hat{i}(\sigma^t)}(\sigma))'}{u^{\hat{i}(\sigma^{t-1})} (c_t^{\hat{i}(\sigma^t)}(\sigma))'} \beta P(\sigma_t | \sigma^{t-1}) \right\|_{\infty} \to 0,$$

where $q(\sigma_t | \sigma^{t-1}) = q_t(\sigma)/q_{t-1}(\sigma)$ is the price to move a unit of consumption from date/event σ^{t-1} to date event σ_t and $\| \cdot \|_{\infty}$ is the sup norm.

For the usual case in which the MSH holds, and the skilled cluster dominates, the result follows from standard economic arguments (Sandroni 2000). More interesting is the observation that markets become efficient even if the MSH fails and the skilled cluster vanishes. The result is implied by four intuitive claims: first, a cluster that vanishes does not affect next-period equilibrium prices, as per Sandroni (2000); second, among traders of the dominating cluster, consumption shares concentrate around the lucky trader (Proposition 5.3); third, the beliefs of non-lucky traders do not affect next-period equilibrium prices; fourth, the beliefs of the lucky trader are eventually accurate, because the leading component of the survival index is empirical accuracy and lucky traders are competing against a skilled cluster. Moreover, under the smoothness ensured by C3 and A4, the empirically accurate beliefs must weakly merge with the true probability.[13]

7. CONCLUSIONS

This paper extends the work started by Sandroni (2000) and Blume and Easley (2006) on market selection to the large economy setting. This generalization alters some of the basic implications of their model: in large economies, risk attitudes do affect trader survival and the MSH can fail. I provide a formal definition of luck and show that risk attitudes determine whether the market rewards for skills or luck. When the market selects for luck over skills, I have a violation of the MSH that is qualitatively different from cases found in previous literature. Although markets select against traders with correct beliefs, equilibrium prices of short-lived assets are asymptotically accurate.

APPENDIX A: RECONCILING SMALL AND LARGE ECONOMIES

A large economy in which all clusters have traders with identical beliefs is formally equivalent to a small economy. In this case, the risk/dimensionality component in the

---

[13]Proposition 6, together with Example 1, could foster the incorrect conjecture that my result implies that the market can achieve perfect foresight on iid coin tosses. This conjecture is incorrect and not consistent with my result for at least two reasons: first, the competitive equilibrium does not exist with a belief structure like that in Example 1 (Ostroy 1984); it violates my definition of cluster C3. Second, the approximation of the integral of Lemma 2 requires enough smoothness ((C3) and (A4)) around the maximum-likelihood trader in each cluster (Schwarz 1978). This assumption is violated when trader beliefs are Dirac deltas on single sequences.

survival indexes of every cluster is moot ($k^{\text{MAR}} = 0$) and, consistent with Sandroni (2000) and Blume and Easley (2006), I find that risk attitudes do not play a role in survival.

PROPOSITION 7. *In a small HDF economy that satisfies A1–A4, for all $\sigma \in \hat{S}$, irrespective of risk attitudes, the market selects for the most accurate trader.*

PROOF. Apply Proposition 1: in a small economy, for all $j = 1, \ldots, N$, $k_j^{\text{MAR}} = 0$.   □

The different implications of risk attitudes on survival for large and small economies can be puzzling. Although Proposition 7 applies to economies with an arbitrarily large number of traders, Proposition 3 implies that it is not valid in large economies. This apparent contradiction disappears if, instead of focusing on vanishing versus surviving (i.e., on the dichotomous distinction between zero versus nonzero asymptotic consumption shares), we look at the size of the asymptotic consumption shares.

Propositions 8 and 9 show that the results of Propositions 3 and 4 hold, approximately, for some small economies with a large number of traders. As intuition suggests, the discrepancy between the small and large setting is narrower when the small economy has a large number of traders.

PROPOSITION 8. *For all $\gamma_U \in (0, \infty)$ and for all $\epsilon > 0$, there exists a $\hat{n}(g, C_0)$ such that in every small HDF economy with $2n > 2\hat{n}$ traders that satisfies A1–A4 with a group of traders, A_U, with heterogeneous beliefs $A_U = \{p^1, \ldots, p^n\}$ and common CRRA utility with parameter γ_U and n Bayesian traders, A_B, with prior $G := [g^1, \ldots, g^n]$ on A_U, $c_0^i = O(n^{-1})$ for all i and $g^j = O(n^{-1})$ for all j,[14] such that the following inequalities hold $\forall \sigma \in \hat{S}^*$:*

   (i) $\gamma_U \in (0, 1) \Leftrightarrow \lim_{t \to \infty} \frac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} < \epsilon$,

   (ii) $\gamma_U = 1 \Leftrightarrow \lim_{t \to \infty} \frac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} \in (\epsilon, 1 - \epsilon)$,

   (iii) $\gamma_U \in (1, \infty) \Leftrightarrow \lim_{t \to \infty} \frac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} > 1 - \epsilon$,

where $\hat{S}^*$ is the set of sequences in which the Bayesian posterior eventually concentrates on a model in its support: $\hat{S}^* := \{\sigma : \exists i \in A_U : i \neq j \in A_U \Rightarrow \lim_{t \to \infty} p^i(\sigma^t)/p^j(\sigma^t) = 0\}$.

EXAMPLE 5. Small economy analog of Example 3. Consider a small economy with two states $S = \{W, R\}$, no aggregate risk, and $2n$ traders with time zero consumption $(2n)^{-1}$. Traders $1, \ldots, n$, group A_U, have a CRRA parameter γ_U and heterogeneous iid Bernoulli beliefs: $\{p^1(w), \ldots, p^n(w)\} = \{1/n, \ldots, (n-1)/n, 1\}$. Traders $n + 1, \ldots, 2n$ are Bayesian traders, group A_B, with Uniform prior G on $A_U = \{\bigcup_{i=1}^n p^i\}$ (i.e., for all (t, σ), $p^B(\sigma^t) = \sum_{i=1}^n p^i(\sigma^t)/n$) and CRRA parameter γ_B.   ◇

---

[14]We have that $c_0^i = O(n^{-1})$ for all i and $g^j = O(n^{-1})$ for all j are the finite analogs of the smoothness requirements on clusters' time zero consumption and on the Bayesian prior, respectively.

Rearranging the first order conditions (FOC) as for (5) and working through the notation

$$\frac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} = \frac{\left(\sum_{i\in A_B} c_t^i(\sigma)\right)^{\gamma_B}}{\left(\sum_{i\in A_U} c_t^i(\sigma)\right)^{\gamma_U}}$$

$$= \frac{\left(\sum_{i=n}^{2n} \frac{1}{2n} p^B(\sigma^t)^{\frac{1}{\gamma_B}}\right)^{\gamma_B}}{\left(\sum_{i=1}^{n} \frac{1}{2n} p^i(\sigma^t)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}}$$

$$= \frac{\left(\frac{1}{2}\right)^{\gamma_B} p^B(\sigma^t)}{\left(\sum_{i=1}^{n} \frac{1}{2n} p^i(\sigma^t)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}} = \frac{\left(\frac{1}{2}\right)^{\gamma_B}\left(\frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n} \frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)}{\left(\frac{1}{2n} + \frac{1}{2n}\sum_{i=2}^{n}\left(\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}},$$

which implies

$$\frac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} = \frac{\left(\frac{1}{2}\right)^{\gamma_B}\left(\frac{1}{n} + \frac{1}{n}\sum_{i=2}^{n} \frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)}{\left(\frac{1}{2n} + \frac{1}{2n}\sum_{i=2}^{n}\left(\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}} \to^{t\to\infty} (2^{\gamma_U - \gamma_B})n^{\gamma_U - 1} \quad \forall \sigma \in \hat{S}^*.[15]$$

Because $(2^{\gamma_U - \gamma_B}) \in (0, \infty)$, $\forall \epsilon > 0$, $\exists \bar{n} : n > \bar{n}$ implies the condition of Proposition 8.

PROPOSITION 9. *Under the conditions of Proposition 8, if I further assume $P = p^B$, $\max_i g^i < \epsilon$, and $\gamma_U < 1$, then*

(i) $\forall i \in A_U, P\{\sigma : \lim_{t\to\infty} \frac{p^i(\sigma^t)}{P(\sigma^t)} > 0\} < \epsilon;$

(ii) $\lim_{t\to\infty} \frac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} < \epsilon$ *P-a.s.;*

(iii) $\exists \hat{i}(\sigma) \in A_U : \lim_{t\to\infty} \frac{c_t^{\hat{i}(\sigma)}(\sigma)}{C_B(\sigma^t) + C_U(\sigma^t)} > 1 - \epsilon$ *P-a.s.*

The intuition mimics that presented in Section 3.2. The data-generating process can be understood as describing this two-step procedure. In the first step, Nature randomizes according to g to decide the probability of red: $P(R)$. In the second step, Nature uses $P(R)$ to generate an iid sequence of length t. While traders in A_B know that Nature is randomizing over A_U according to g, each trader in A_U is dogmatically sure that

---

[15]The convergence occurs by definition of $\hat{S}^*$.

his model is the correct one, an event whose true probability is smaller than ϵ. Because Nature's choice is restricted to models in A_U, exactly one trader in A_U is empirically accurate, $\hat{i}$. For large n, this trader is "almost lucky" (his ex ante probability of being empirically accurate is at most ϵ) and "almost dominates" (his asymptotic consumption share is above $1 - \epsilon$).

Examples 3 and 5 highlight the role played by risk attitudes and belief set dimensionality in small and large economies.

Risk attitudes affect the asymptotic consumption shares distribution through their effect on the concentration rate of consumption shares: lower values of γ determine a faster consumption shares concentration rate and thus lower asymptotic consumption shares for the Bayesian cluster. The dimensionality of A_U affects cluster survival through its effect on the concentration rate of both the Bayesian posterior and the consumption shares as follows. If $|A| < |\mathbb{N}|$, both convergence rates are exponential; the Bayesian measure and the aggregate risk-adjusted measure are mutually absolutely continuous and the Bayesian survives without dominating. If $|A| = |\mathbb{R}|$, both convergence rates are slower than exponential (they are, respectively, $O(t^{-\frac{k^{\mathrm{BIC}}}{2}})$ and $O(t^{-\frac{k^{\mathrm{MAR}}}{2\gamma}})$), the two measures are not mutually absolutely continuous, and clusters' survival depends on γ, k^{BIC} and k^{MAC}.

Appendix B: Proofs

I make use of the notations $o(\cdot)$, $O(\cdot)$, and $\sim$ with the meanings

$$f(x) = o\big(g(x)\big) \quad \text{abbreviates} \quad \lim_{x \to \infty} \frac{f(x)}{g(x)} \to 0$$

$$f(x) = O\big(g(x)\big) \quad \text{abbreviates} \quad \limsup \frac{f(x)}{g(x)} < +\infty$$

$$f(x) \sim g(x) \quad \text{abbreviates} \quad \lim \frac{f(x)}{g(x)} = 1.$$

Proof of Lemma 2. We need to show that

$$\left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j} = e^{\ln \hat{p}^{i(\sigma^t)}(\sigma^t) - \frac{\gamma_j k_j^{\mathrm{MAR}}}{2} \ln t + O(1)}.$$

The result follows from Lemma 5 by substituting A_j for A, multiplying by γ_j, exponentiating, and ignoring the constant terms. □

The proof of Lemmas 4 and 5 follows the steps of Grünwald's (2007, p. 248) proof of the BIC (if $\gamma = 1$ and c_0 is a density, the two proofs coincide).

Lemma 4. *Let $\mathcal{M}$ be a member of the exponential family parametrized by A and let c_0^i be a function that satisfies A4. Then*

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

where $D(p^{\hat{i}(\sigma^t)} \| p^i) = E_{p^{\hat{i}(\sigma^t)}} \ln p^{\hat{i}(\sigma^t)}/p^i$ is the Kullback–Leibler divergence from p^i to $p^{\hat{i}(\sigma^t)}$.

PROOF.

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

$$= \ln \int_A \left(\frac{p^i(\sigma^t)}{p^{\hat{i}(\sigma^t)}(\sigma^t)} \right)^{\frac{1}{\gamma}} c_0^i \, di + \ln p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}}$$

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

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

For example, if $p^i(\sigma_t = 1) = i$ is iid Bernoulli, the result follows because

$$\ln p^i(\sigma^t) - \ln p^{\hat{i}(\sigma^t)}(\sigma^t) = t \left(\frac{1}{t} \sum_{\tau=0}^{t} \sum_{s=0,1} I_{\sigma_\tau = s} \ln \frac{p^i(s)}{p^{\hat{i}(\sigma^t)}(s)} \right)$$

$$= -t E_{p^{\hat{i}(\sigma^t)}} \ln \frac{p^{\hat{i}(\sigma^t)}(s)}{p^i(s)} = -t D(p^{\hat{i}(\sigma^t)} \| p^i). \qquad \square$$

LEMMA 5. *Let $\mathcal{M}$ be a member of the exponential family parametrized by A and let c_0^i be a function that satisfies A4. Then, $\forall \sigma \in \hat{S}$,*

$$\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i \, di = \frac{1}{\gamma} \ln p^{\hat{i}(\sigma^t)}(\sigma^t) + \ln \sqrt{\gamma} + \ln c_0^{\hat{i}}$$

$$- \frac{k_A}{2} \ln \frac{t}{2\pi} - \ln \sqrt{\det I(p^{\hat{i}(\sigma^t)})} + o(1),$$

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

PROOF. By Lemma 4,

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

WLOG, let us focus on the case in which $\mathcal{M}$ is the iid Bernoulli family, so that $p^i = i$ and $k_A = 1$.[16] Let $B_t = \{i \in [\hat{i}(\sigma^t) - t^{-\frac{1}{2}+\alpha}, \hat{i}(\sigma^t) + t^{-\frac{1}{2}+\alpha}]\}$ with $0 < \alpha < \frac{1}{2}$. To gain intuition, take α very small, so that B_t is a neighborhood of the maximum likelihood that shrinks to 0 at a rate slightly slower than $(\sqrt{t})^{-1}$. Because $\sigma \in \hat{S}$, B_t concentrates around $\hat{i}$, and because c_0^i is continuous and strictly positive in A, there is a $T: \forall t > T$, $B_t \subset A_0$ where A_0 is a compact subset of A in which $c_0^i > 0$. I always assume $t > T$.

---

[16]The generalization to the multinomial case and non-iid intertemporal structures is straightforward.

By additivity of the integral,

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

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. The $\mathcal{M}$ is a member of the exponential family (Bernoulli in my case); thus, by the results in Chapter 19 of Grünwald (2007), $D(p^i\|P)$ can be well approximated in B as

$$D(p^{\hat{i}(\sigma^t)}\|p^i) = \frac{1}{2}(\hat{i}(\sigma^t)-i)^2 I(p^{i*}) \tag{6}$$

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

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

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

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

By (6) and the definition of B_t,

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

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

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

for $a = (2\gamma)^{-1}\min_{i\in\text{int}(A)} I(p^i) > 0$ and $k = \int_{A\backslash B_t} c_0^i\,di < \int_A c_0^i\,di < \infty.$

(‡) **Second integral:** $\mathcal{I}_2 = \int_{B_t} e^{-\frac{t}{\gamma}D(p^{\hat{i}(\sigma^t)}\|p^i)}c_0^i\,di \sim \sqrt{2\pi}c_0^{\hat{i}}(\sqrt{tI(p^{\hat{i}})/\gamma})^{-1}.$

Letting $I_t^- = \inf_{i'\in B_t} I(p^{i'})$, $I_t^+ = \sup_{i'\in B_t} I(p^{i'})$, $c_t^- = \inf_{i'\in B_t} c_0^{i'}$, and $c_t^+ = \sup_{i'\in B_t} c_0^{i'}$, by (6),

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

where i' depends on i. Using the definitions above, I get

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

I now perform the substitutions $z = (\hat{i}(\sigma^t) - i)\sqrt{tI_t^+/\gamma}$ on the left integral and $z = (\hat{i}(\sigma^t) - i)\sqrt{tI_t^-/\gamma}$ on the right integral to get

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

I now recognize both integrals as standard Gaussian. Because, 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 $c_0^{\hat{i}}(\sqrt{tI(p^{\hat{i}})/\gamma})^{-1}$ and I get $\mathcal{I}_2 \sim \sqrt{2\pi}c_0^{\hat{i}}(\sqrt{tI(p^{\hat{i}})/\gamma})^{-1}$.

Putting † and ‡ together yields

$$\ln \int_A p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\, di = \ln(\mathcal{I}_1 + \mathcal{I}_2) + \frac{1}{\gamma}\ln p^{\hat{i}(\sigma^t)}(\sigma^t)$$

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

Note that the approximation holds uniformly for all $\sigma^t \in \hat{S}$ because (i) the bound on $\mathcal{I}_1$ does not depend on σ^t, and (ii) convergence of $\mathcal{I}_2$ is uniform because c_0^i and $I(p^i)$ are continuous functions of i over the compact set A_0. ▢

PROOF OF COROLLARY 1.

$$\beta_j^t \left(\int_{A_k} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\, di \right)^{\gamma_k} \overset{\text{By Lem. 2}}{=} e^{t\ln\beta_j + \ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{\gamma_k \text{MAR}}{2}\ln t + O(1)}$$

$$= e^{\ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k\text{MAR}}{2}\ln t + O(1)} e^{t\ln\beta_j - (\frac{\gamma_k\text{MAR} - k\text{MAR}}{2})\ln t + O(1)}$$

$$\overset{\text{By BIC}}{=} \left(\int_{A_k} p^i(\sigma^t)\, di \right) e^{t\ln\beta_j - (\frac{\gamma_k\text{MAR} - k\text{MAR}}{2})\ln t + O(1)}.$$ ▢

PROOF OF LEMMA 3. Let A_γ be a positive mass of Bayesian traders with regular priors, $g^i(\theta)$, on the same k-dimensional parameter space Θ. I have to show that their risk-adjusted aggregate beliefs are equivalent to the beliefs of a cluster of Bayesian traders with an identical regular prior f on $\Theta : p^B(\sigma^t)$. Let $\bar{g} = \sup_{i,\theta \in \text{int}\,\Theta} g^i(\theta)$ and $\underline{g} = \inf_{i,\theta \in \text{int}\,\Theta} g^i(\theta)$. Note that $\underline{g} > 0$, because the prior distribution of every trader in A_γ is strictly positive, and note that $\bar{g} < \infty$ because all of the g^is are continuous in the simplex and thus bounded in its (strict) interior. Because the convergence result of Lemma 5 is uniform, it follows that

$$\left(\int_{A_\gamma} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}}\, di \right)^\gamma$$

$$= \left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta) g^i(\theta)\, d\theta \right)^{\frac{1}{\gamma}} di \right)^\gamma$$

$$\in \left[\left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta) \underline{g} \, d\theta \right)^{\frac{1}{\gamma}} di \right)^\gamma, \left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta) \bar{g} \, d\theta \right)^{\frac{1}{\gamma}} di \right)^\gamma \right]$$

$$= \left[\int_\Theta p(\sigma^t|\theta) \underline{g} \, d\theta \left(\int_{A_\gamma} c_0^i \, di \right)^\gamma, \int_\Theta p(\sigma^t|\theta) \bar{g} \, d\theta \left(\int_{A_\gamma} c_0^i \, di \right)^\gamma \right]$$

$$\overset{\text{By Lem. 3}}{=} e^{\ln p^{\hat\theta(\sigma^t)}(\sigma^t) - \frac{k}{2} \ln t + O(1)}$$

$$\overset{\text{By BIC}}{=} O(1) \int_\Theta p(\sigma^t|\theta) f(\theta) \, d\theta$$

$$= O(1) p^B(\sigma^t). \qquad \qquad \square$$

Proof of Proposition 4. Let us focus WLOG on the Bernoulli case: $p^B(\sigma^t) = \int_0^1 p^i(\sigma^t) g^i \, di$.

For the most part, Proposition 4 coincides with Proposition 3. I need to show only two additional things: (a) the MSH fails with a probability arbitrarily close to 1, i.e., $\forall \epsilon > 0$, $p^B(\hat{S}) > 1 - \epsilon$, and (b) lucky traders dominate.

Part (a). By assumption, g^i is regular and thus continuous on $(0, 1)$. Therefore, the probability that the g^i gives to the set of parameters in the strict interior of the prior support is arbitrarily close to 1: $\forall \epsilon > 0$, $\exists \epsilon_1 > 0 : p^g(i \in (\epsilon_1, 1 - \epsilon_1)) > 1 - \epsilon$. By the strong law of large numbers, $i \in (\epsilon_1, 1 - \epsilon_1) \Rightarrow \hat{i}(\sigma) \in \hat{S}$ p^i-a.s. so that $\forall \epsilon_1 > 0$, $p^B(\hat{S}) \geq p^g(i \in (\epsilon_1, 1 - \epsilon_1))$. Thus, $\forall \epsilon > 0$, $\exists \epsilon_1 > 0 : p^B(\hat{S}) \geq p^g(i \in (\epsilon_1, 1 - \epsilon_1)) > 1 - \epsilon$.

Part (b). Let $\hat{i}(\sigma^t)$ be the beliefs of the maximum-likelihood trader in the cluster that dominates (A) and let $\{B_t(\hat{i})\}_{t=1}^\infty$ be the following sequence of shrinking subclusters of A: $B_t(\hat{i}) = \{i \in [\hat{i}(\sigma^t) - t^{-\frac{1}{2}+\alpha}, \hat{i}(\sigma^t) + t^{-\frac{1}{2}+\alpha}]\}$ for $0 < \alpha < \frac{1}{2}$. Rearranging (5) and using † and ‡ from the proof of Lemma 5 yields

$$\lim_{t \to \infty} \frac{\displaystyle\int_{i \in A \setminus B_t(\hat{i})} c_t^i(\sigma) \, di}{\displaystyle\int_{i \in B_t(\hat{i})} c_t^i(\sigma) \, di} = \lim_{t \to \infty} \frac{\displaystyle\int_{i \in A \setminus B_t(\hat{i})} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0 \, di}{\displaystyle\int_{i \in B_t(\hat{i})} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0 \, di} = 0.$$

Thus, by Lemma 1, consumption shares concentrate in the shrinking interval $B_t(\hat{i})$ around $p^{\hat{i}(\sigma^t)}$. The market selects for luck because

- $\lim_{t \to \infty} \sup_{i \in B_t(\hat{i})} \|i - \hat{i}\| = 0$, i.e., the market rewards an empirically accurate trader

- $\int_A \lim_{t \to \infty} I_{B_t} g^i \, di = 0$, i.e., trader $\hat{i}$ is not a priori accurate. $\square$

Proof of Proposition 5. Let A_S be a skilled cluster.

Part (a) is necessary for A_S to vanish P-a.s. By contradiction, in a small economy, A_S survives P-a.s. (Sandroni 2000).

Part (b) is necessary for A_S to vanish P-a.s. By contradiction, if $\gamma_j > 1$ for all clusters in the economy, then $s_j - s_s \to -\infty$ $\forall j \neq s$ and A_s dominates by Proposition 1.

Part (c) is necessary for A_S to vanish P-a.s. By contradiction, if $k_P^{\text{BIC}} = 0$, then A_S has the maximal survival index because $k_S^{\text{MAR}} = 0$ and it survives by Proposition 1. $\square$

PROOF OF PROPOSITION 6. If the skilled cluster dominates, the convergence follows from standard economic arguments (Sandroni 2000). Otherwise, the result follows by proving the following four claims.

CLAIM 1. *A cluster that vanishes does not affect next-period equilibrium prices.*

CLAIM 2. *Among traders of the dominating cluster, consumption shares concentrate around the lucky trader.*

CLAIM 3. *The beliefs of non-lucky traders do not affect equilibrium prices.*

CLAIM 4. *The beliefs of the lucky trader are eventually accurate, because they need to beat a skilled cluster.*

Let $\bar{C}$, $\bar{\beta}$, $\bar{\gamma}$, and $\bar{A}$ be the aggregate consumption, discount factor, CRRA parameter, and belief set of the cluster with the highest survival index, $\bar{j}$, respectively.

PROOF OF CLAIM 1. We have

$$\forall \sigma \in \hat{S}, \quad q(\sigma_t|\sigma^{t-1}) = \frac{\bar{C}_{t-1}(\sigma)^{\bar{\gamma}} + o(1)}{\bar{C}_t(\sigma)^{\bar{\gamma}} + o(1)} \left(\frac{\bar{\beta} \left(\left(\int_{\bar{A}} \frac{p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} c_0^i \, di}{\int_{\bar{A}} p^i(\sigma^{t-1})^{\frac{1}{\bar{\gamma}}} c_0^i \, di} \right)^{\bar{\gamma}} + o(1) \right)}{1 + o(1)} \right).$$

In equilibrium

$$C_t^j(\sigma) \stackrel{\text{By Eq. (1)}}{=} \beta^{\frac{t}{\gamma_j}} \frac{\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di}{q_t(\sigma)^{\frac{1}{\gamma_j}}} \quad \Rightarrow \quad q_t(\sigma) = \frac{\sum_j \left(\beta^{\frac{t}{\gamma_j}} \int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j}}{\sum_j C_t^j(\sigma)^{\gamma_j}},$$

so that

$$q(\sigma_t|\sigma^{t-1}) = \frac{q_t(\sigma)}{q_{t-1}(\sigma)} = \frac{\sum_j C_{t-1}^j(\sigma)^{\gamma_j}}{\sum_j C_t^j(\sigma)^{\gamma_j}} \frac{\sum_j \left(\beta^{\frac{t}{\gamma_j}} \int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j}}{\sum_j \left(\beta^{\frac{t-1}{\gamma_j}} \int_{A_j} p^i(\sigma^{t-1})^{\frac{1}{\gamma_j}} c_0^i \, di \right)^{\gamma_j}}. \tag{7}$$

Equation (7) obeys the asymptotic

$$
q(\sigma_t|\sigma^{t-1}) \;=\; \frac{\displaystyle\sum_j C_{t-1}^j(\sigma)^{\gamma_j}}{\displaystyle\sum_j C_t^j(\sigma)^{\gamma_j}} \;\frac{\displaystyle\sum_j \left(\beta^{\frac{t}{\gamma_j}}\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}{\displaystyle\sum_j \left(\beta^{\frac{t-1}{\gamma_j}}\int_{A_j} p^i(\sigma^{t-1})^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}
$$

$$
\overset{(a)}{=} \frac{\bar C_{t-1}(\sigma)^{\bar\gamma}+o(1)}{\bar C_t(\sigma)^{\bar\gamma}+o(1)}
$$

$$
\times \left(\frac{\bar\beta^t\left(\int_{\bar A} c_0^i p^i(\sigma^t)^{\frac{1}{\bar\gamma}}\,di\right)^{\bar\gamma} + \displaystyle\sum_{j\neq \bar j}\left(\beta^{\frac{t}{\gamma_j}}\int_{A_j}p^i(\sigma^t)^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}{\bar\beta^{t-1}\left(\int_{\bar A} c_0^i p^i(\sigma^{t-1})^{\frac{1}{\bar\gamma}}\,di\right)^{\bar\gamma} + \displaystyle\sum_{j\neq \bar j}\left(\beta^{\frac{t-1}{\gamma_j}}\int_{A_j}p^i(\sigma^{t-1})^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}} \right)
$$

$$
= \frac{\bar C_{t-1}(\sigma)^{\bar\gamma}+o(1)}{\bar C_t(\sigma)^{\bar\gamma}+o(1)}
$$

$$
\times \left(\frac{\displaystyle\bar\beta\left(\int_{\bar A}\frac{p^i(\sigma^t)^{\frac{1}{\bar\gamma}}c_0^i\,di}{\int_{\bar A}p^i(\sigma^{t-1})^{\frac{1}{\bar\gamma}}c_0^i\,di}\right)^{\bar\gamma} + \displaystyle\sum_{j\neq\bar j}\frac{\left(\beta^{\frac{t}{\gamma_j}}\int_{A_j}p^i(\sigma^t)^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}{\bar\beta^{t-1}\left(\int_{\bar A}c_0^i p^i(\sigma^{t-1})^{\frac{1}{\bar\gamma}}\,di\right)^{\bar\gamma}}}{1+\dfrac{\displaystyle\sum_{j\neq\bar j}\left(\beta^{\frac{t-1}{\gamma_j}}\int_{A_j}p^i(\sigma^{t-1})^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}{\bar\beta^{t-1}\left(\int_{\bar A}c_0^i p^i(\sigma^{t-1})^{\frac{1}{\bar\gamma}}\,di\right)^{\bar\gamma}}} \right)
$$

$$
\overset{(b)}{=} \frac{\bar C_{t-1}(\sigma)^{\bar\gamma}+o(1)}{\bar C_t(\sigma)^{\bar\gamma}+o(1)}\left(\frac{\displaystyle\bar\beta\left(\int_{\bar A}\frac{p^i(\sigma^t)^{\frac{1}{\bar\gamma}}c_0^i\,di}{\int_{\bar A}p^i(\sigma^{t-1})^{\frac{1}{\bar\gamma}}c_0^i\,di}\right)^{\bar\gamma} + o(1)}{1+o(1)} \right).
$$

By Proposition 1, $j\neq\bar j \Rightarrow C_t^j(\sigma)^{\gamma_j}=o(1)$. By Lemma 5,

$$
j\neq\bar j \quad\Rightarrow\quad \frac{\beta_j^t\left(\int_{A_j}p^i(\sigma^t)^{\frac{1}{\gamma_j}}c_0^i\,di\right)^{\gamma_j}}{\bar\beta^t\left(\int_{\bar A}c_0^i p^i(\sigma^t)^{\frac{1}{\bar\gamma}}\,di\right)^{\bar\gamma}} = o(1).
$$

◁

PROOF OF CLAIM 2. We have

$$\forall \sigma \in \hat{S}, \quad \sup_{i \in \bar{C}_{t-1}} \left\| \frac{c_{t-1}^i(\sigma)^{\bar{\gamma}}}{c_t^i(\sigma)^{\bar{\gamma}}} - \frac{c_{t-1}^{\hat{i}(\sigma^t)}(\sigma)}{c_t^{\hat{i}(\sigma^t)}(\sigma)} \right\| \to 0.$$

Let $\{\bar{B}_T\}_{T=1}^{\infty}$ be a sequence of subsets of $\bar{A}$ centered around $\hat{i}$ as in the proof of Lemma 4 but with $T = o(t)$ and $\bar{B}_T^c = \bar{A} \setminus \bar{B}_T$ its complement, and let $\bar{C}_{\bar{B}_T}(\sigma)$ and $\bar{C}_{\bar{B}_T^c}(\sigma)$ be the aggregate consumption of traders in B_T and B_T^c, respectively. By Lemma 5, †, and ‡, $\bar{C}_{\bar{B}_T^c}(\sigma)/\bar{C}_{\bar{B}_T}(\sigma) \to 0$ for every T. Thus,

$$\frac{\bar{C}_{t-1}(\sigma)^{\bar{\gamma}}}{\bar{C}_t(\sigma)^{\bar{\gamma}}} = \frac{\left(\bar{C}_{\bar{B}_{T,t}}(\sigma) + C_{\bar{B}_{T,t}^c}(\sigma)\right)^{\bar{\gamma}}}{\left(\bar{C}_{\bar{B}_{T,t-1}}(\sigma) + C_{\bar{B}_{T,t-1}^c}(\sigma)\right)^{\bar{\gamma}}}$$

$$\in \left\{ \frac{\min\{c \in \bar{C}_{\bar{B}_{T,t}}(\sigma)\}}{\max\{c \in \bar{C}_{\bar{B}_{T,t-1}}(\sigma)\}} - o(1), \frac{\max\{c \in \bar{C}_{\bar{B}_{T,t}}(\sigma)\}}{\min\{c \in \bar{C}_{\bar{B}_{T,t-1}(\sigma)}\}} + o(1) \right\}$$

$$\to^a \frac{c_t^{\hat{i}(\sigma^t)}(\sigma)}{c_{t-1}^{\hat{i}(\sigma^t)}(\sigma)}.$$

The limit follows because $\sup_{i \in \bar{B}_{T,t-1}} \|\hat{i}(\sigma^{t-1}) - i\| \to^{t,T \to \infty} 0$ uniformly and c^i is differentiable in i. ◁

PROOF OF CLAIM 3. We have

$$\forall \sigma \in \hat{S}, \quad \left\| \left(\int_{\bar{A}} \frac{p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} c_0^i \, di}{\int_{\bar{A}} p^i(\sigma^{t-1})^{\frac{1}{\bar{\gamma}}} c_0^i \, di} \right)^{\bar{\gamma}} - \bar{p}^{\hat{i}(\sigma^t)}(\sigma_t | \sigma^{t-1}) \right\|_{\infty} \to 0,$$

$$\left(\int_{\bar{A}} \frac{p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} c_0^i \, di}{\int_{\bar{A}} p^i(\sigma^{t-1})^{\frac{1}{\bar{\gamma}}} c_0^i \, di} \right)^{\bar{\gamma}}$$

$$= \left(\frac{\int_{\bar{B}_t} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di + \int_{\bar{A} \setminus \bar{B}_t} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di}{\int_{\bar{B}_{t-1}} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di + \int_{\bar{A} \setminus \bar{B}_{t-1}} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di} \right)^{\gamma}$$

$$= \left(\frac{\int_{\bar{B}_t} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di + o\left(\int_{\bar{B}_{t-1}} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)} \| p^i)} c_0^i \, di \right)}{\int_{\bar{B}_{t-1}} e^{-\frac{t-1}{\gamma} D(p^{\hat{i}(\sigma^{t-1})} \| p^i)} c_0^i \, di + o\left(\int_{\bar{B}_{t-1}} e^{-\frac{t-1}{\gamma} D(p^{\hat{i}(\sigma^{t-1})} \| p^i)} c_0^i \, di \right)} \right)^{\gamma}$$

$$\in \left\{ \frac{\displaystyle \inf_{i\in\bar{B}_{t-1}} p^i(\sigma^t)\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{t\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}}{\displaystyle \sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{(t-1)\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}} - o(1); \right.$$

$$\left. \frac{\displaystyle \sup_{i\in\bar{B}_{t-1}} p^i(\sigma^t)\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{t\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}}{\displaystyle \inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{(t-1)\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}} + o(1) \right\}$$

$$\in \left\{ \inf_{i\in\bar{B}_{t-1}} p^i(\sigma_t|\sigma^{t-1}) \frac{\displaystyle \inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{t\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}}{\displaystyle \sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{(t-1)\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}} - o(1); \right.$$

$$\left. \sup_{i\in\bar{B}_{t-1}} p^i(\sigma_t|\sigma^{t-1}) \frac{\displaystyle \sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{t\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}}{\displaystyle \inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1})\left(\frac{\sqrt{2\pi}c_0^i}{\sqrt{(t-1)\dfrac{I(p^i)}{\gamma}}}\right)^{\gamma}} + o(1) \right\}$$

$$\to^a \bar{p}^{\hat{i}(\sigma^{t-1})}(\sigma_t|\sigma^{t-1}).$$

Because $\exists T < \infty : \forall t > T, \forall i \in B_{t-1}$, $p^i(\sigma^t)$, $p^i(\sigma^{t-1})$, $I(p^i)$, and c_0^i are all differentiable and strictly positive functions of i, and $\sup_{i \in \bar{B}_{t-1}} \|\hat{i}(\sigma^{t-1}) - i\| \to 0$ uniformly.   ◁

PROOF OF CLAIM 4. We have

$$\forall \sigma \in \hat{S}, \quad \left\| P\big(\sigma_t|\sigma^{t-1}\big) - p^{\hat{i}(\sigma^{t-1})}\big(\sigma_t|\sigma^{t-1}\big) \right\|_\infty \to 0.$$

I can have a violation of the MSH involving a positive mass of traders only if there is a cluster j such that $P = p^j$. Moreover, by Proposition 5, this can happen only if $P = \int_{\Theta_j} p^i g^i \, di$ with Θ_j with a positive Lebesgue measure. Finally, the empirical accuracy term dominates the cluster dimensionality and the BIC dimensionality term, so only a cluster with an empirically accurate trader can dominate a cluster with correct beliefs. Therefore, it must be the case that the maximum-likelihood parameter according to P, $p_P^{\hat{i}(\sigma^t)}$, and the maximum-likelihood trader of the competing unskilled cluster coincide: $p_P^{\hat{i}(\sigma^t)} = \bar{p}^{\hat{i}(\sigma^t)}$. The result follows by applying the proof of Claim 3 (with $\gamma = 1$) to the true probability, $P(\sigma_t|\sigma^{t-1})$:

$$\lim_{t \to \infty} \left\| P\big(\sigma_t|\sigma^{t-1}\big) - \bar{p}^{\hat{i}(\sigma^t)}\big(\sigma_t|\sigma^{t-1}\big) \right\|_\infty$$

$$:= \lim_{t \to \infty} \left\| \int_{\Theta_J} \frac{p^i(\sigma^t) g_0^i \, di}{\int_{\Theta_J} p^i(\sigma^{t-1}) g_0^i \, di} - \bar{p}^{\hat{i}(\sigma^t)}\big(\sigma_t|\sigma^{t-1}\big) \right\|_\infty$$

$$= \lim_{t \to \infty} \left\| p_P^{\hat{i}(\sigma^t)}\big(\sigma_t|\sigma^{t-1}\big) - \bar{p}^{\hat{i}(\sigma^t)}\big(\sigma_t|\sigma^{t-1}\big) \right\|_\infty = 0.$$

Because the convergence results in Claims 1–4 are all uniform, I obtain the desired result:

$$\forall \sigma \in \hat{S}, \quad \left\| q\big(\sigma_t|\sigma^{t-1}\big) - \frac{u^{\hat{i}(\sigma^{t-1})}\big(c_{t-1}^{\hat{i}(\sigma^t)}(\sigma)\big)'}{u^{\hat{i}(\sigma^{t-1})}\big(c_t^{\hat{i}(\sigma^t)}(\sigma)\big)'} \beta P\big(\sigma_t|\sigma^{t-1}\big) \right\|_\infty \to 0.$$ □

PROOF OF PROPOSITION 8. Let WLOG p^1 be the model on which p^B concentrates. Rearranging the FOC as for (5) and working through the notation yields

$$\frac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} = \frac{\left(\sum_{i \in A_B} c_0^i p^B(\sigma^t)^{\frac{1}{\gamma_B}} \right)^{\gamma_B}}{\left(\sum_{i \in A_U} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_U}} \right)^{\gamma_U}}$$

$$= \frac{\left(\sum_{i=n+1}^{2n} c_0^i \right)^{\gamma_B} p^B(\sigma^t)}{\left(\sum_{i=1}^{n} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_U}} \right)^{\gamma_U}}$$

$$= \frac{\left(\sum\limits_{i=n+1}^{2n} c_0^i\right)^{\gamma_B} \sum\limits_{i=1}^{n} g^i p^i(\sigma^t)}{\left(\sum\limits_{i=1}^{n} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}}$$

$$= \frac{\left(\sum\limits_{i=n+1}^{2n} c_0^i\right)^{\gamma_B} \left(g^1 + \sum\limits_{i=2}^{n} g^i \frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)}{\left(c^1 + \sum\limits_{i=2}^{n} c^i \frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)^{\gamma_U}}$$

$$= O(1) \frac{g^1 + o_t(1)}{(c^1 + o_t(1))^{\gamma_U}} = O(n^{\gamma_U - 1}) \quad \text{for all large } t, \text{ for all } \sigma \in \hat{S}^*.$$

Thus, for all large t, for all $\sigma \in \hat{S}^*$

$$\lim_{n \to \infty} \frac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} = \begin{cases} 0 & \text{if } \gamma_U \in (0,1), \\ K \in (0,\infty) & \text{if } \gamma_U = 1, \\ \infty & \text{if } \gamma_U \in (1,\infty). \end{cases}$$

The result follows because for all large t, for all $\sigma \in \hat{S}^*$,

$$C_B(\sigma^t) + C_U(\sigma^t) \in (0,\infty)$$

$$\Rightarrow \quad \begin{cases} \dfrac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} \to 0 & \Leftrightarrow \quad \dfrac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} \to 0; \\ \dfrac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} \to K \in (0,\infty) & \Leftrightarrow \quad \dfrac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} \to K \in (0,\infty); \\ \dfrac{C_B(\sigma^t)^{\gamma_B}}{C_U(\sigma^t)^{\gamma_U}} \to \infty & \Leftrightarrow \quad \dfrac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} \to 1. \end{cases} \qquad \square$$

PROOF OF PROPOSITION 9.   (i) For all $i \in A_U$, $P\{\sigma : \lim_{t\to\infty} p^i(\sigma^t)/P(\sigma^t) > 0\} = g^i < \epsilon$. Because, by assumption, $\forall i$, the prior attaches probability $g^i < \epsilon$ to those sequences to which p^i gives probability 1.

(ii) We have $P = p^B \Rightarrow P(\{\hat{S}^*\}) = 1$. Thus, Proposition 8(i) becomes

$$\frac{C_B(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} < \epsilon, \quad P\text{-a.s.}$$

(iii) We have $P = p^B \Rightarrow P(\{\hat{S}^*\}) = 1$. Thus, Proposition 8(i) implies

$$\lim_{t \to \infty} \frac{C_U(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} > 1 - \epsilon, \quad P\text{-a.s.}$$

Moreover, Massari's (2017) necessary and sufficient condition for a trader to vanish implies that only one trader, i, in A_U survives P-a.s., and the result follows by noticing that

$$\lim_{t \to \infty} \frac{c^i(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} = \lim_{t \to \infty} \frac{C_U(\sigma^t)}{C_B(\sigma^t) + C_U(\sigma^t)} > 1 - \epsilon, \quad P\text{-a.s.} \qquad \square$$

APPENDIX C: EXISTENCE OF THE COMPETITIVE EQUILIBRIUM

To prove the existence of the competitive equilibrium, I will show that my assumptions guarantee that the economy satisfies Ostroy's (1984) sufficient conditions for the existence of a competitive equilibrium. According to Ostroy (1984), a large economy is a pair $(\succ, W_0)$, where $\succ$ describes preferences and W_0 defines an initial allocation of commodities to a *group* of traders. Ostroy's (1984) theorem shows that the competitive equilibrium exists in economies that satisfy three sets of assumptions: on the commodity space (Y.1), on the aggregate preferences (S.1–S.7), and the initial allocation of commodities (T.1–T.2).

- Assumption Y.1 requires that the commodity space has an "order compatible" basis. This assumption is satisfied if the commodity space contains only countably many elements. Assumption C3 ensures that this is indeed the case:

LEMMA 6. *For all C3$\Rightarrow$ the commodity space is countable.*

PROOF. If all traders have exchangeable beliefs (either Bayesian or iid)—conditionally exchangeable if the beliefs have a Markov structure—they can only distinguish consumption by the length of the sequence and the average number of realizations (conditional average number of realizations if Markov). Because the length of the sequence and and the average number of realizations belong to countable sets, $\mathbb{N}$ and $\mathbb{Q}$, respectively, the equilibrium consumption allocations need only to span a countable space and the orthonormal basis exists.[17] $\qquad \square$

- Assumptions S1–S7 are regularity assumptions on preferences. They are implied by C2 and A1, which require common CRRA preferences among traders of the same cluster.

- Assumptions T1 and T2 are regularity conditions on the measurability of the initial allocation. They are satisfied by assuming that the initial allocations are described by a Lebesgue measurable function (C1). This function represents the initial equilibrium consumption shares (i.e., the inverse of the risk-adjusted Pareto weight).

Assumption A3 is necessary to rule out difficulties that can arise if agents have orthogonal beliefs on finite horizons. Assumptions A2 and A4 are ancillary to the existence of the competitive equilibrium, but are useful for its characterization.

---

[17]The belief structure of Example 1 violates this requirement: assuming Dirac measures on every sequence requires the commodity space to be large enough to distinguish between every sequence. Therefore, in Example 1, the commodity space is the space of all binary sequences with the sup norm, and the equilibrium does not exist because this space does not have an orthonormal basis.

References

Alchian, Armen A. (1950), "Uncertainty, evolution, and economic theory." *Journal of Political Economy*, 58, 211–221. [437]

Blackwell, David and Lester Dubins (1962), "Merging of opinions with increasing information." *Annals of Mathematical Statistics*, 33, 882–886. [440]

Blume, Lawrence and David Easley (1992), "Evolution and market behavior." *Journal of Economic Theory*, 58, 9–40. [438]

Blume, Lawrence and David Easley (2006), "If you're so smart, why aren't you rich? Belief selection in complete and incomplete markets." *Econometrica*, 74, 929–966. [437, 438, 440, 448, 451, 453, 456, 457, 458]

Borovička, Jaroslav (2018), "Survival and long-run dynamics with heterogeneous beliefs under recursive preferences." Unpublished paper, SSRN 1852047. [438]

Clarke, Bertrand S. and Andrew R. Barron (1990), "Information-theoretic asymptotics of Bayes methods." *Information Theory, IEEE Transactions on*, 36, 453–471. [449]

Cvitanić, Jaksa and Semyon Malamud (2011), "Price impact and portfolio impact." *Journal of Financial Economics*, 100, 201–225. [438]

Cvitanić, Jakša, Elyès Jouini, Semyon Malamud, and Clotilde Napp (2012), "Financial markets equilibrium with heterogeneous agents." *Review of Finance*, 16, 285–321. [438]

de Finetti, Bruno (1937), "La prévision: Ses lois logiques, ses sources subjectives." In *Annales de l'institut Henri Poincaré*, volume 7, 1–68, Gauthier-Villars. [444, 455]

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann (1991), "The survival of noise traders in financial markets." *Journal of Business*, 64, 1–19. [437]

Dindo, Pietro (2018), "Survival in speculative markets." Unpublished paper, SSRN 2705337. [438]

Friedman, Milton (1953), *Essays in Positive Economics*, volume 231. University of Chicago Press. [437]

Grünwald, Peter D. (2007), *The Minimum Description Length Principle*. MIT press. [449, 460, 462]

Kalai, Ehud and Ehud Lehrer (1994), "Weak and strong merging of opinions." *Journal of Mathematical Economics*, 23, 73–86. [440]

Kogan, Leonid, Stephen A. Ross, Jiang Wang, and Mark M. Westerfield (2006), "The price impact and survival of irrational traders." *Journal of Finance*, 61, 195–229. [438]

Kreps, David M. (1988), *Notes on the Theory of Choice*. Westview Press, Boulder. [455]

Mahmoud, Hosam (2008), *Pólya urn Models*. Chapman and Hall/CRC, New York, New York. [442, 455]

Massari, Filippo (2013), "Comment on if you're so smart, why aren't you rich? Belief selection in complete and incomplete markets." *Econometrica*, 81, 849–851. [440]

Massari, Filippo (2017), "Markets with heterogeneous beliefs: A necessary and sufficient condition for a trader to vanish." *Journal of Economic Dynamics and Control*, 78, 190–205. [440, 471]

Ostroy, Joseph M. (1984), "On the existence of Walrasian equilibrium in large-square economies." *Journal of Mathematical Economics*, 13, 143–163. [442, 446, 448, 457, 471]

Ploberger, Werner and Peter C. B. Phillips (2003), "Empirical limits for time series econometric models." *Econometrica*, 71, 627–673. [449]

Pólya, George (1930), "Sur quelques points de la théorie des probabilités." In *Annales de l'institut Henri Poincaré*, volume 1, 117–161, Gauthier-Villars. [442]

Rubinstein, Mark (1974), "An aggregation theorem for securities markets." *Journal of Financial Economics*, 1, 225–244. [453]

Sandroni, Alvaro (2000), "Do markets favor agents able to make accurate predictions?" *Econometrica*, 68, 1303–1341. [437, 438, 440, 445, 448, 451, 453, 456, 457, 458, 464, 465]

Schwarz, Gideon (1978), "Estimating the dimension of a model." *Annals of Statistics*, 6, 461–464. [449, 457]

Yan, Hongjun (2008), "Natural selection in financial markets: Does it work?" *Management Science*, 54, 1935–1950. [438, 447, 453, 454]