Market Selection in Large Economies: A Matter of Luck*

Filippo Massari†

September 22, 2017

Abstract

In a general equilibrium model with a continuum of traders and bounded aggregate endowment, we investigate the Market Selection Hypothesis that markets favor traders with accurate beliefs. Contrary to known results for economies with (only) finitely many traders, we find that risk attitudes affect traders' survival and that markets can favor "lucky" traders with incorrect beliefs over "skilled" traders with accurate beliefs. Our 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

*This paper is part of my dissertation presented at Washington University in St Louis. I would like to express my gratitude to my advisors, Werner Ploberger and John Nachbar, for their constant encouragement and many valuable suggestions. I also want to thank Larry Blume, John Geanakoplos, Peter Grünwald, Philip Dybvig, Pietro Dindo, David Easley, David Feldman, Stefano Giglio, Stephanie Heger, Richard Holden, Leonid Kogan, Semyon Malamud, Jonathan Newton, Nick Papageorge, Luciano Pomatto, Alvaro Sandroni, Rabee Tourky, Aleks Yankelevich, Nicholas Yannelis and seminar audiences at SFI (2014), NFA (2014), Northern Econometric Society Meeting, Minneapolis (2014); 10th Annual Cowles GE Conference, Yale (2014); FIRN (2013);NBER/NSF/CEME, Columbia, NYC, USA (2013); University of Sydney (2013); UNSW (2013); S.A.E.T., France (2013); EWGET, Austria (2013); Northwestern MEDS (2012).

†[Corresponding author] Mailing Address: ASB, University of New South Wales, Sydney NSW 2052 Australia, Tel:+61 0414 746 015, e-mail: f.massari@unsw.edu.au

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 supporting 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 — Bewley economy — (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 non-recursive preferences (Borovička, 2013; Dindo, 2016).

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 that vanish can have an everlasting effect on equilibrium prices. Provided that inaccurate traders are overconfident on assets that pay on extremely unlikely events, markets can remain inefficient even if all inaccurate traders vanish.

In our paper, we prove that the MSH is not necessary for market efficiency. We present a model in which accurate traders lose all their wealth, luck — which we properly define in Section 2.2 — is the only force dictating which trader survives, and markets become efficient in the sense that asymptotic prices of short-lived assets reflect correct beliefs.

Our model maintains all the assumptions of Sandroni (2000) and Blume and Easley (2006) and generalizes their setting by allowing for a continuum of traders who are indexed by their beliefs. 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, only with a continuum of agents luck and risk attitudes play a role in traders' survival, and the MSH can fail. In what follows, we use *small economies* to refer to economies with finitely many traders and *large economies* to indicate economies with a continuum of traders.

We begin our paper by presenting an example in which the market selects against traders with correct beliefs, luck is the only determinant of traders' survival, and asymptotic equilibrium prices reflect accurate beliefs. Two conditions are necessary for our result. First, luck can determine survival only if there is a sufficiently rich heterogeneity in 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 diverse enough so that for every path, there is a (lucky) trader that allocates consumption exactly along this path. Ultimately, it is always one trader from that group that 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, in order 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 CRRA parameter smaller than one increases his chances of survival because it gives him enough incentives to save. Contrary to Sandroni (2000) and Blume and Easley (2006), the curvature of preferences matters because although aggregate consumption is bounded, consumption of an infinitesimal trader can become unbounded — and so, individually, we are in the unbounded setting of Yan (2008).[1]

[1] We are thankful to an anonymous referee for providing this interpretation of our result.

In the rest of the paper, we develop the formal theory needed to discuss selection in *large economies* and to understand our leading example. In the tradition of the selection literature, we 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*. Our 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 whose traders have heterogeneous beliefs. Everything else equal, a cluster whose traders have a higher CRRA parameter vanishes against a cluster whose traders have a lower CRRA parameter because it has a lower aggregate savings rate. In Section 6, we provide our main result. Markets become efficient in the sense that asymptotic prices of short-lived assets reflect correct beliefs whenever there is a positive mass of traders with correct beliefs. Even in those cases in which the MSH fails, and traders with correct beliefs vanish against inaccurate traders. There are two appendices. In Appendix A we reconcile the apparent contrast between the selection results in *small* and *large economies*. Proofs are in Appendix B.

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: $\mathcal{S} = \{1, ..., S\}$, 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 $\sigma = (\sigma_1, ...)$. $\sigma^t = (\sigma_1, ..., \sigma_t)$ 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, we 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 $p(\sigma^t) = p(\mathbf{C}(\sigma^t)) = p(\{\sigma_1\} \times ... \times \{\sigma_t\} \times S \times S \times ...)$. 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 iid Bernoulli ($\forall t, p^i(\sigma_t = 1) = i$), 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, we rank traders' accuracy according to the relative likelihood of their beliefs.[2,3]

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

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

Definition 2. *Trader i*

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

Our definition of skills 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,

[2]Focusing on beliefs' likelihood is in the tradition of the selection literature; however, unlike Sandroni (2000) and Blume and Easley (2006), we cannot rely on approximate measures of it. Our result captures $O(\log t)$ differences between traders' log likelihood. Sandroni's definition (average accuracy) is too coarse to capture these differences because the averaging factor, $\frac{1}{t}$, 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 $\lim_{t\to\infty} \frac{p^i(\sigma^t)}{p^j(\sigma^t)}$ does not exist. This case is left unspecified because it doesn't play a role in our results and is potentially distracting.

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). Definition 2 draws the line between skills and no skills between these two notions. 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. We are interested in skilled traders 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. Our next definition refines the notion of skills by putting emphasis on empirical evidence.

Definition 3. *Trader i is*

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

We 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:*

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

Our 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. Consider the following:

Example 1. Suppose we toss a fair coin t times and that we have 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, it corresponds a trader who believes that σ^t will occur for sure. That is, for every $\hat{\sigma}^t$ we have a (lucky) trader, $\hat{i}(\hat{\sigma}^t)$, for which the probability of obtaining a favorable realization is extremely low, $P\left\{\sigma^t \in S^t : \frac{p^{\hat{i}}(\sigma^t)}{P(\sigma^t)} > 0\right\} = \frac{1}{2^t}$; but whose beliefs attach more likelihood to $\hat{\sigma}^t$ than the true probability does, $\frac{p^{\hat{i}}(\hat{\sigma}^t)}{P(\hat{\sigma}^t)} = 2^t$. With $t = \infty$, this belief structure illustrates a setting in which we have a lucky trader in every sequence.[4]

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.

[4]The belief structure of example 1 is incompatible with the existence of the competitive equilibrium because it requires too many distinct beliefs — to ensure the existence of the competitive equilibrium we need the number of beliefs to be an order of magnitudes smaller than the number of sequences (Ostroy, 1984). The beliefs' structure we use in the rest of the paper only requires one trader per frequency — a number that grows polynomially in t — rather than one trader per sequence — a number that grows exponentially in t.

There are two sets of traders with positive masses (clusters): A_U and A_S. Individual traders, i, can have different beliefs, p^i, but share an identical CRRA utility function ($u^i(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$) with parameter $\gamma < 1$ and discount factor β. As usual, every individual trader in the economy aims to solve:

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

Where E_{p^i} is the expectation according to trader i's beliefs, $c_t^i(\sigma), e_t^i(\sigma)$, 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 cluster A_j's period t aggregate consumption on path σ. In the tradition of the selection literature, the asymptotic fate of a cluster is coarsely characterized by the distinction between those clusters that disappear and those that do not.

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, we randomly select a ball from the urn to determine the state of the economy. The selected ball is then returned to the urn along with one new ball of the same color.

Traders in A_S, *skilled* traders, are allowed to observe the composition of the urn before every draw. They have correct beliefs, $\forall i \in A_S, p^i = P_{Polya}$, 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: $\forall t, \forall i, p^i(R_t) = i$. The *unskilled* cluster collects the different opinions of those traders who, not having access to private information, never change their beliefs. Because the 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 $\left(c_t^i(\sigma)\right)^\gamma = (c_0^i)^\gamma \frac{\beta^t p^i(\sigma^t)}{q_t(\sigma)}$, 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 (Riemann) summing over traders of the same cluster:

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

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

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

By De-Finetti's Theorem (Corollary 2) $\forall \sigma, \forall t, P_{Polya}(\sigma^t) = \int_0^1 p^i(\sigma^t) di$. Thus, Lemma 2 implies that Equation 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, we give an informal preview of the results. These intuitions are demonstrated and further discussed in the remainder of the paper.

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 Equation 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

[5]Because the aggregate endowment is bounded $\left(\frac{C_t^S(\sigma)}{C_t^U(\sigma)}\right)^\gamma \to 0 \Rightarrow C_t^S(\sigma) \to 0$.

because it is more accurate than the *unskilled cluster*.

However, because *unskilled traders* do not have identical beliefs, the right-hand side of Equation 2 does not represent the ratio between two probabilities. Specifically, Corollary 1 and a standard implication of De Finetti's Theorem (Corollary 2) allow us to express Equation 2 in this more informative way:

$$\frac{C_t^S(\sigma)^\gamma}{C_t^U(\sigma)^\gamma} = \frac{P_{Polya}(\sigma^t)\left(\int_{A_S} c_0^i di\right)^\gamma}{\left(\int_{A_U} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i di\right)^\gamma} = \frac{P_{Polya}(\sigma^t)\beta^t\left(\int_{A_S} c_0^i di\right)^\gamma}{P_{Polya}(\sigma^t) * \beta^t e^{-\frac{\gamma-1}{2}\ln t + O(1)}}. \tag{3}$$

Ignoring constant quantities, the numerator is proportional to the product of the true probability of σ^t and the discount factor β^t. The denominator has an aggregate probability term, $P_{Polya}(\sigma^t)$, 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 a Wisdom of the Crowd effect has emerged. Although all unskilled traders have incorrect beliefs, they collectively behave as if they had correct beliefs. Accordingly, the asymptotic fate of the two clusters is uniquely determined by their aggregate discount factor. With $\gamma < 1$, Equation 3 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 3 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 we 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:

It is known (De Finetti, 1937, 's Theorem) that $\forall \sigma, \forall t, P_{Polya}(\sigma^t) = \int_0^1 p^i(\sigma^t) di$. This means that the Pólya urn process produces probabilities that are equivalent, in distribution, to the probabilities obtained via this two-step process. 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 — $\lim_{t\to\infty} \frac{P_{Polya}(\sigma^t)}{P_{Polya}(\sigma^t)} = 1$ 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 $p(R)$ — $\forall i \in A_U, P_{1^{st}step}(\{p(R) = i\}) = 0 \Rightarrow \lim_{t\to\infty} \frac{p^i(\sigma^t)}{P_{Polya}(\sigma^t)} = 0$ 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, $\hat{\imath}$, whose belief coincides with $p(R)$. Among unskilled traders, the market selects for $\hat{\imath}$ because, conditional on $p(R) = \hat{\imath}$, $\hat{\imath}$ is the only accurate trader in the economy.[6]

3.3 Do markets become asymptotically efficient?

Markets do become asymptotically efficient: the asymptotic equilibrium prices of the short-lived asset 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 our leading example, the intuition goes 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).

[6]Assuming 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.

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 we show that violations of the MSH can occur only if the economy has a large number of traders, preferences are elastic enough, 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. We prove that, in all these cases, the next period beliefs of the lucky trader and the truth becomes indistinguishable.

This result does not apply to the setting of example 1 — which violates assumptions C3 and A4 of sections 4.1 and 4.2, respectively. For instance, it is easy to verify that the lucky trader's beliefs never converge to the true probability: $\forall t, ||P(\sigma_t) - p^{\hat{i}}(\sigma_t)||_2 = \sqrt{(\frac{1}{2}-1)^2 + (\frac{1}{2}-0)^2} = \sqrt{\frac{1}{2}} > 0$. 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 our interpretation, our 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.

4 The general model

4.1 The traders in the economy

In this section, we formally describe the space of traders. The economy is characterized by the aggregate preferences $\succ_j$ and by the aggregate time 0 consumption C_0^j of N measurable sets of traders, clusters A_j, $j = 1, ..., N$. Where $\succ_j$ and C_0^j are constructed, respectively, by aggregating the preferences and the initial consumptions of a positive mass of atomless traders i with beliefs p^i, utilities u^i, endowment processes $e_t^i(\sigma)$, and infinitesimal time 0 consumption c_0^i. We assume that trader beliefs are parametric. The set of cluster j's beliefs and the set of parameters that characterize it are $\mathcal{M}_j :=$

$\{p^i : i \in A_j\}$ and $\Theta_j := \{i \in \Theta\}$, respectively.

Definition 6. *A* cluster*, A_j, is a measurable set of traders such that:*

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 an identical CRRA utility function with parameter γ_j: $\forall i \in A_j, u^i(c) = \frac{c^{1-\gamma_j}-1}{1-\gamma_j}$;*

C3: *either (i) all traders in A_j have identical beliefs, or (ii) all beliefs in $\mathcal{M}_j$ are Multinomial with the same inter-temporal structure (either iid or Markov with finitely many lags) and Θ_j is the whole simplex.*

C1 requires the initial consumption of each cluster to be strictly positive. We ignore subsets of traders whose initial consumption is zero because they cannot affect any equilibrium quantity. **C2** delivers an analytical form for clusters' optimal consumption decisions as a function of its discount factor, risk attitudes, and aggregate beliefs. **C3** disciplines clusters' aggregate beliefs. Either all traders in a cluster are identical and can be treated as a representative trader with positive mass — by (i) and **C2** — ; or, by (ii), it provides enough structure for constructing an asymptotic approximation of clusters' optimal investment strategies (Lemma 2).[7]

The heterogeneity of opinion among traders of the same cluster is captured by the Lebesgue dimensionality of the cluster parameter set, Θ_j. Lemma 2 shows that the dimensionality of Θ_j interacts with cluster risk attitudes to determine survival. In the rest of the paper we use cluster dimensionality to refer to the dimensionality of Θ_j. This use is justified by the fact that the topological properties of a cluster of traders $(A_j, \mathcal{A}_j, i)$ and those of the space of its parameters $(\Theta_j, \mathcal{B}_j, i)$ coincide: each trader is uniquely identified by his beliefs which are uniquely identified by a vector of parameters. This observation allows us to use $\mathcal{A}_j$ and Θ_j interchangeably as the domain of integration.[8]

[7] This assumption can be generalized to include the case of Bayesian traders with identical support but different priors. We discuss this case in Appendix **??**.

[8] To familiarize with our construction, consider the following examples. Let $S = \{W, R\}$. A cluster A_B of traders with iid Bernoulli beliefs has Lebesgue dimensionality one because the set characterizing its trader beliefs is uniquely described by parameters in the one dimensional simplex: $\mathcal{M}_B = \{p^i : \forall t, p^i(W_t) = i,\ i \in \Theta_j\}$ with $\Theta_B = \{i \in (0, 1)\}$. A cluster A_M of traders with Markov (1) Bernoulli beliefs has Lebesgue

Finally, a special role in our condition for a cluster to vanish will be 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, we maintain the following assumptions:

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

A2: The aggregate endowment is, in every period, strictly positive and bounded above.

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, satisfying **C3** (ii), c_0^i is a differentiable and strictly positive function of i, for every i in the interior of Θ_j.

Assumptions **A1-A3** are standard in the selection literature. If the traders in the economy can be organized into finitely many clusters with identical beliefs, the economy is formally equivalent to a small economy and Assumptions **A1-A3** are implied by Sandroni (2000) and Blume and Easley (2006). Assumption **A4** is a smoothness assumption needed for technical reasons.[9] 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, preference 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 (1984) existence theorem. We omit the proof because it is notationally intensive and tangential to the main contribution of our paper.

dimensionality two because the set characterizing its trader beliefs is uniquely described by parameters in a two dimensional simplex: $\mathcal{M}_M = \{p^{\mathbf{i}} : \forall t, p^{\mathbf{i}}(W_t|W_{t-1}) = i', p^{\mathbf{i}}(W_t|R_{t-1}) = i'', \mathbf{i} \in \Theta_M\}$, with $\Theta_M = \{\mathbf{i} := [i'; i''] \in (0,1)\times(0,1)\}$.

[9] Because the second welfare theorem applies, we can think of **A4** as being an assumption on the Pareto weight distribution of the Social Planner (exogenous quantities), rather than an assumption on time 0 equilibrium consumption shares (endogenous quantities). Because for every Social Planner Problems satisfying **A4** there is a sequence of endowments such that the equilibrium time 0 consumption shares satisfy **A4**, there are many competitive equilibria satisfying this property.

4.3 The reference economy

The economy is a discrete time Arrow-Debreu exchange economy with complete markets, bounded aggregate endowment, S states, and N clusters A_j, $j = 1, ..., N$. Every individual trader in the economy aims to solve:

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

Traders' first-order conditions of the maximization problem are sufficient for the Pareto Optimum and, in every path σ^t, can be expressed as $\left(c_t^i(\sigma)\right)^{\gamma_j} = (c_0^i)^{\gamma_j} \frac{\beta^t p^i(\sigma^t)}{q_t(\sigma)}$. Rearranging and (Riemann) summing over traders of the same cluster, $\int_{A_j} c_t^i(\sigma) di = \beta_j^{\frac{t}{\gamma_j}} \frac{\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di}{q_t(\sigma)^{\frac{1}{\gamma_j}}}$. Exponentiating both sides by the CRRA parameter and taking the ratio of clusters' risk-adjusted consumption, prices simplify out:

$$\frac{C_t^j(\sigma)^{\gamma_j}}{C_t^k(\sigma)^{\gamma_k}} = \frac{\beta_j^t \left(\int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di\right)^{\gamma_j}}{\beta_k^t \left(\int_{A_k} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_k}} di\right)^{\gamma_k}}. \tag{4}$$

The following Lemma uses standard arguments in the selection literature to show that Equation 4 is the fundamental quantity to determine which cluster vanishes.

Lemma 1. *Under **A1-A4**, A_j vanishes on σ if exists A_k:* $\frac{\beta_j^t \left(\int_{A_j} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_j}} di\right)^{\gamma_j}}{\beta_k^t \left(\int_{A_k} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma_k}} di\right)^{\gamma_k}} \to 0.$

Proof. By **A2**, $C_t^k(\sigma)^{\gamma_k} < \infty$ $\forall k, \forall \sigma^t$. Thus by Equation 4,$\forall \gamma_j, \gamma_k \in (0, \infty)$

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

□

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. The distinction between risk-adjusted and not risk-adjusted aggregate beliefs is nihil in *small economies* because for clusters with homogeneous beliefs, the common belief can be factored out of the integral.

4.4 Technical contribution

Our main technical contribution is to provide an approximation of Equation 4. We show that the distinction between risk-adjusted and non-risk-adjusted aggregate beliefs does affect cluster survival in *large economies*. Our approximation can be seen as a generalization of a fundamental result about Bayesian accuracy, the BIC approximation (Schwarz, 1978; Clarke and Barron, 1990; Ploberger and Phillips, 2003; Grünwald, 2007). To formally state this result we need some definitions.

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

- *Let $\Theta_0 \subset \Theta$. We 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, is unique, and it belongs to the regular set Θ_0 for all large t.*
- *$\hat{S}$ is the set of all* Θ_0-regular *sequences.*
- *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, Θ_0 be a regular subset of Θ and $p^B(\sigma^t)$ be the Bayesian likelihood obtained from a Θ_0-regular prior; then,*[10]

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

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

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

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

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.[11]

Lemma 2 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 $\Theta_{j,0}$ be a regular subset of Θ_j, then cluster j's risk-adjusted beliefs satisfies:*

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

Where k_j^{MAR} and γ_j are cluster j's dimensionality and CRRA parameter, respectively.

Proof. See Appendix B □

When $\gamma = 1$ (log), Lemma 2 coincides with the BIC approximation. However, for $\gamma \neq 1$ and $k^{MAR} > 0$, 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 diverges 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 (*ii*), 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.

[11] 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 (1 parameter: $k_1^{BIC} = 1$) and B^2 has a smooth prior on the Markov (1) family (2 parameters: $k_2^{BIC} = 2$). 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.

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 $k_j^{MAR} = 0$, 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.

5 Main result

We are now ready to present a general condition for a cluster to vanish that only depends on exogenous quantities. In the tradition of the selection literature, we 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 *is*

$$s_j = t \ln \beta_j + \left[\ln p_j^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k_j^{BIC}}{2} \ln t \right] - \frac{\gamma_j k_j^{MAR}}{2} \ln t.$$

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. $\left[\ln p_j^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{k_j^{BIC}}{2} \ln t \right]$ 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 — it equals zero unless traders in A_j are Bayesian traders with k^{BIC}-dimensional prior support of positive Lebesgue measure. The last term, $\frac{\gamma_j k_j^{MAR}}{2} \ln t$, is new and only appears in the *large economy* setting. It captures the effect of risk attitudes and dimensionality on each cluster saving decisions. This term is absent in *small economies* because k_j^{MAR} is zero in homogeneous belief clusters. 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 $s_j - s_k \to -\infty$.*

Proof. Application of Lemma.1 using Lem.2 and BIC to approximate the RHS of Eq.4. $\square$

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). And differences in the last component indicate that the cluster with the lowest $\gamma_j k$ 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 $O(1) \ln t$. Thus, differences in the first two components always dominate differences in the other components.[12] Therefore, if all traders have an identical discount factor, the leading term of the survival indexes is the one capturing 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, our 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 and last components (Section 5.2). This interaction can be responsible for failures of the MSH.

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

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

[12] Differences in the second two components would disappear if we were to use an average measure of accuracy as in Sandroni (2000) because they would be dominated by the $\frac{1}{t}$ term.

5.1 The role of risk attitudes

To highlight the effect of risk attitudes on cluster survival, we 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, with $k^{MAR} > 0$, the least risk-averse cluster dominates on $\hat{S}$.*[13]

Proof. $\gamma_j < \gamma_k \;\Rightarrow\; s_k - s_j \to -\infty \Rightarrow^{By Th.1} k$ 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 $k_\eta^{MAR} = k_\gamma^{MAR} = 1$. It follows that: $s_\eta - s_\gamma = \frac{\gamma}{2}\ln t - \frac{\eta}{2}\ln t \to -\infty$, 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 inter-temporal consumption. Everything else equal, a low CRRA parameter increases the survival chances of a cluster because it gives to its empirically most accurate traders higher incentives to save.

Corollary 1 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 satisfies:*

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

Proof. See Appendix B. ☐

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

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

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 law of one price, 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 cluster aggregate savings rate. Cluster's optimal choices can be equivalently modeled as those of a representative trader with positive mass, 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 they 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 although aggregate consumption is bounded, consumption of an infinitesimal trader can become unbounded. At an infinitesimal level, a large economy can be thought of as a growing economy, and our results are qualitatively similar to Yan (2008)'s findings in large economies.

5.2 The role of heterogeneity of opinions

In this section, we analyze survival in economies which contain some clusters with heterogeneous beliefs and some clusters with identical Bayesian traders. We show that cluster beliefs dimensionality and risk attitudes ,$\gamma_j k_j^{MAR}$, have an effect on cluster survival which is of the same order as that of belief dimensionality for a Bayesian cluster, k_B^{BIC}. Therefore they can offset each other.

Proposition 3. *Under **A1-A4**, if the economy is HDF and only contains 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:*

$$i)\ \gamma_U \in (0,1) \Leftrightarrow \textit{cluster } A_B \textit{ vanishes}, \ \forall \sigma \in \hat{S}$$
$$ii)\ \gamma_U = 1 \Leftrightarrow \textit{cluster } A_B \textit{ survives but does not dominate}, \ \forall \sigma \in \hat{S}$$
$$iii)\ \gamma_U \in (1,\infty) \Leftrightarrow \textit{cluster } A_B \textit{ dominates}, \ \forall \sigma \in \hat{S}.$$

Proof. Application of Proposition 1 by noticing that $s_U - s_B = \frac{1}{2}(k_B^{BIC} - \gamma_U k_U^{MAR}) \ln t$. $\square$

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^{MAR}=1, k_U^{BIC}=0), while traders in A_B have identical beliefs p^B which are obtained via Bayes' rule from a regular prior distribution on $\Theta_B = (0,1)$ (i.e., k_U^{MAR}=0, k_U^{BIC}=1). The result follows as an application of Proposition 1: $\forall \sigma \in \hat{S}, s_B - s_U = \frac{\gamma k_U^{MAR}}{2} \ln t - \frac{k_B^{BIC}}{2} \ln t \to \begin{cases} 0 \ if \ \gamma \in (0,1) \\ r \in (0,\infty) \ if \ \gamma = 1 \\ \infty \ if \ \gamma \in (1,\infty) \end{cases}$.

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. On the other hand, traders in A_U never learn the truth and have speculative incentives to trade. For $\gamma < 1$, the lucky trader — and a shrinking measure of traders with parameters around him ($B_t \in A_U$) — saves more on the realized path than a log trader would — as Yan (2008) result for growing economies. While those traders whose beliefs are outside B_t save less than what a log trader would save, because they are investing more consumption on paths that do not realize — as Yan (2008) result for shrinking economies. 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.

Our motivation for studying *large economies* is to depart from the strong homogeneity assumptions implicit in the *small economy* setting. It could be argued that requiring all Bayesian traders to share a common prior is inconsistent with the stated goal. The next lemma allows to generalize Proposition 3 to clusters in which traders have different (regular) priors over the same support. In the spirit of Aumann (1976) and Geanakoplos and Polemarchakis (1982), Lemma 3 shows that that the dimensionality term of a positive mass of Bayesian traders with a heterogeneous regular priors on the same support equals 0 ($k^{MAR}_{Bayesian} = 0$) because Bayesian traders with the same prior support do not disagree forever.

Lemma 3. *A positive mass 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.*

Proof. See Appendix B. □

5.3 The role of the true probability

In Proposition 3, we make a comparison between a cluster of Bayesian traders and a cluster of traders with heterogeneous iid beliefs. The result holds in every path in $\hat{S}$. Therefore, it does not depend on the true data-generating process. If we assume the true probability coincides with the beliefs of the Bayesian cluster, we have an economy in which all traders in A_B have correct beliefs and yet for $\gamma < 1$ cluster A_B vanishes: the MSH fails, and luck is the sole determinant of trader survival.

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 is the sole determinant of trader survival.*

Proof. See Appendix B. □

But what does it mean that the true probability coincides with the probability obtained via Bayes' rule? It is well known that the Pólya urn process in the leading example of page 8 satisfies this requirement (De Finetti, 1937):

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:

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

Proof. Standard application of De Finetti's Theorem (e.g., Mahmoud, 2008, pg.30). $\square$

Similar examples can be constructed as long as the true data-generating process is exchangeable but not iid. Where exchangeable means that the true data-generating process is such that the probability of finite sequences does not depend on the order of the realizations.

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

From the Definition 11 it follows that every sequence of iid random variables, conditional on some underlying distributional form, is exchangeable. De Finetti (1937)'s 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 variable has a representation of the form: $p(\sigma^t) = \int_\Theta p^i(\sigma^t)g^i di$; where the p^i represents 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) it corresponds an 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 we used for our leading example. There are two clusters, A_S and A_U, with am identical discount factor β. Traders in A_S, *skilled* traders, are Bayesian with Uniform prior on

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

In Example 4, a smaller value of γ is needed than in the leading example 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 ($k_{M1}^{MAR} = 2$) rather than an iid cluster ($k_{IID}^{MAR} = 1$). 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

We presented two examples in which the MSH fails. The examples have three elements in common: a large number of traders, preferences that are elastic enough, 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:*

a) *at least one cluster, A_j, has heterogeneous traders;*

b) *cluster A_j's trader preferences are elastic enough: $\gamma_j \leq 1$;*

c) *the true maximum-likelihood parameter is a random variable with continuum support. That is, $k_P^{BIC} > 0$.*

Proof. See Appendix B. □

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 elastic enough 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 skilled cluster's, A_S. If the maximum-likelihood parameters of the true data-generating process were either constants or random variables with finite support, then A_S must have maximal survival index because both its BIC and MAR components equal 0.

6 Markets are asymptotically efficient

In this section, we give formal proof for what is stated in the leading example. The prices of short-period assets 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.*

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

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

Proof. See Appendix B. ☐

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 equilibrium prices. And 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.[14]

7 Conclusions

This paper extends the work started by Sandroni (2000) and Blume and Easley (2006) on market selection to the *large economy* setting. Our generalization alters some of the basic implications of their model: in *large economies*, risk attitudes do affect trader survival and the MSH can fail. We 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, we 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.

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

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

Proof. Application of Proposition 1, noticing that in a small economy $\forall_{j=1,\ldots,N}, k_j^{MAR} = 0$. □

[14]Proposition 6, together with Example 1 (page 7), could foster the incorrect conjecture that our result implies that the market can achieve perfect foresight on iid coin tosses. This conjecture is incorrect and not consistent with our result for at least two reasons: first, the competitive equilibrium does not exist with a belief structure like the one in Example 1 (Ostroy, 1984) — it violates our 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.

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 non-zero 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. $\forall \gamma_U, \gamma_B \in (0, \infty), \forall \epsilon > 0$, *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* $A_U = \{p^1, ..., p^n\}$ *and* n *Bayesian traders,* A_B*, with prior* g *on* A_U*; the following inequalities hold* $\forall \sigma \in \hat{S}^*$*:*

$$\begin{aligned} &i)\ \gamma_U \in (0,1) \Leftrightarrow \lim_{t\to\infty} C_B(\sigma^t) < \epsilon; \\ &ii)\ \gamma_U = 1 \Leftrightarrow \lim_{t\to\infty} C_B(\sigma^t) \in (\epsilon, 1-\epsilon); \\ &iii)\ \gamma_U \in (1,\infty) \Leftrightarrow \lim_{t\to\infty} C_B(\sigma^t) > 1-\epsilon. \end{aligned}$$

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

Proof. Proven by Example 5. □

Example 5: (Small sample analog of Example 3).

Consider a small economy with two states $\mathcal{S} = \{W, R\}$ and $2n$ traders. Traders 1,...,n, group A_U, have a CRRA parameter γ_U and heterogeneous iid Bernoulli beliefs: $\{p^1(w), ..., p^{n-1}(w)\} = \{\frac{1}{n}, ..., \frac{n-1}{n}, 1\}$. Traders $n+1$,...,$2n$ are Bayesian traders, group A_B, with prior G on $A_U = \{\cup_{i=1}^n p^i\}$ $\left(i.e., p^B(\sigma^t) = \sum_{i=1}^n p^i(\sigma^t) g^i\right)$ and CRRA parameter γ_B.

For ease of exposition, let $\forall i, c_0^i = \frac{1}{2n}$, G be the Uniform prior, and p^1 be the model on which

p^B concentrates. Rearranging the FOC as for Equation 4 and working through the notation:[15]

$$\frac{C_B(\sigma^t)^{\frac{1}{\gamma_U}}}{C_U(\sigma^t)^{\frac{1}{\gamma_B}}} = \frac{(\frac{1}{2})^{\gamma}_{B}\left(\frac{1}{n}+\frac{1}{n}\sum_{i=2}^{n}\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)}{\left(\frac{1}{2n}+\frac{1}{2n}\sum_{i=2}^{n}\left(\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}} \to^{t\to\infty} \left(2^{\gamma_U-\gamma_B}\right) n^{\gamma_U-1} \quad \forall\sigma \in \hat{S}^*.^{16} \tag{5}$$

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 we further assume $P = p^B$, $\max_i g^i < \epsilon$, and $\gamma_U < 1$; then:*

$$\begin{aligned}
&i) \quad \forall i \in A_U, P\{\sigma : \lim_{t\to\infty} \tfrac{p^i(\sigma^t)}{P(\sigma^t)} > 0\} < \epsilon;\\
&ii) \quad \lim_{t\to\infty} C_B(\sigma^t) < \epsilon \ P\text{-}a.s.;\\
&iii) \ \exists\hat{i}(\sigma) \in A_U: \ \lim_{t\to\infty} c_t^{\hat{i}(\sigma)}(\sigma) > 1-\epsilon \ P a.s..
\end{aligned}$$

Proof.
i) $\forall i \in \{1,...,n\}, P\{\sigma : \lim_{t\to\infty} \frac{p^i(\sigma^t)}{P(\sigma^t)} > 0\} = g^i < \epsilon$. Because, by assumption, $\forall i$, the prior attaches probability $g^1 < \epsilon$ to those sequences to which p^i gives probability 1.
ii) $P = p^B \Rightarrow P(\{\hat{S}^*\}) = 1$. Thus, rearranging Eq.5, $\bar{n} > \left(\frac{\epsilon^{\frac{1}{\gamma_B}}}{(1-\epsilon)^{\frac{1}{\gamma_U}}}2^{\gamma_B-\gamma_U}\right)^{\frac{1}{\gamma_U-1}} \Rightarrow \lim_{t\to\infty} C_B(\sigma^t) < \epsilon$ P-a.s..
iii) $P = p^B \Rightarrow P(\{\hat{S}^*\}) = 1$. Thus, by Massari (2017)'s necessary and sufficient condition for a trader to vanish, only one trader, i, in A_U survives P-a.s.. The result follows noticing that $\lim_{t\to\infty} c^i(\sigma^t) = \lim_{t\to\infty} C_U(\sigma^t) = \lim_{t\to\infty} 1 - C_B(\sigma^t) = 1-\epsilon$ for $n > \bar{n}$. □

The intuition mimics the one we presented in Section 3.2. The data-generating process can be understood as describing this two-steps 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 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-ϵ).

[15] $\frac{C_B(\sigma^t)^{\frac{1}{\gamma_B}}}{C_U(\sigma^t)^{\frac{1}{\gamma_U}}} = \frac{\left(\sum_{i\in A_B} c_t^i(\sigma)\right)^{\gamma_B}}{\left(\sum_{i\in A_U} c_t^i(\sigma)\right)^{\gamma_U}} = \frac{\left(\sum_{i=n}^{2n}\frac{1}{2n}p^B(\sigma^t)^{\frac{1}{\gamma_B}}\right)^{\gamma_B}}{\left(\sum_{i=1}^{n}\frac{1}{2n}p^i(\sigma^t)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}} = \frac{(\frac{1}{2})^{\gamma}_{B}p^B(\sigma^t)}{\left(\sum_{i=1}^{n}\frac{1}{2n}p^i(\sigma^t)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}} = \frac{(\frac{1}{2})^{\gamma}_{B}\left(\frac{1}{n}+\frac{1}{n}\sum_{i=2}^{n}\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)}{\left(\frac{1}{2n}+\frac{1}{2n}\sum_{i=2}^{n}\left(\frac{p^i(\sigma^t)}{p^1(\sigma^t)}\right)^{\frac{1}{\gamma_U}}\right)^{\gamma_U}}.$

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

Examples 3 and 5 highlights 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 gamma determine a faster consumption-shares concentration rate, thus a 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 $O\left(\frac{1}{t^{\frac{k^{BIC}}{2}}}\right)$ and $O\left(\frac{1}{t^{\frac{k^{MAR}}{2\gamma}}}\right)$), the two measures are not mutually absolutely continuous, and asymptotic consumption-shares depend on γ, k^{BIC} and k^{MAC}.

B Appendix: Proofs

We make use of the notations o(.) and $O(.)$ with the following meanings: $f(x) = o(g(x))$ abbreviates $\lim_{x\to\infty} \frac{f(x)}{g(x)} \to 0$, $f(x) = O(g(x))$, abbreviates $\limsup \frac{f(x)}{g(x)} < +\infty$, and $f(x) \sim g(x)$ abbreviates $\lim \frac{f(x)}{g(x)} = 1$.

Proof of Lemma 2

Proof. $\left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di\right)^{\gamma_j} = e^{\ln p^{\hat{i}(\sigma^t)}(\sigma^t) - \frac{\gamma_j k_j^{MAR}}{2} \ln t + O(1)}$. It 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, pg. 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 c_0^i be a function that satisfies **A4**; then:*

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

Where $D(p^{\hat{i}(\sigma^t)}||p^i) = E_{p^{\hat{i}(\sigma^t)}} \ln \frac{p^{\hat{i}(\sigma^t)}}{p^i}$ is the Kullback-Leibler divergence between $p^{\hat{i}(\sigma^t)}$ and p^i.

Proof.

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

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

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

□

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

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

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

Proof. By Lemma 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 focus on the case in which $\mathcal{M}$ is the iid Bernoulli family, so that $p^i = i$.[17] 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 $\frac{1}{\sqrt{t}}$. Because $\sigma \in \hat{S}$, B_t concentrates around $\hat{i}$ and because c_0^i is continuous, 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 $c_0^i > 0$. We always assume $t > T$.

By additivity of the integral:

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

[17] The generalization to the Multinomial case and non-iid inter-temporal structures is straightforward.

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

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

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

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

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

By Equation 6 and the definition of B_t

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

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

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

for $a = \frac{1}{2\gamma} \min\limits_{i \in int(A)} I(p^i) > 0$ and $k = \int_{A \setminus B_t} c_0^i di < \int_A c_0^i di < \infty$.

‡ **Second integral:** $\mathcal{I}_2 = \int_{B_t} e^{-\frac{t}{\gamma} D(p^{\hat{i}(\sigma^t)}||p^i)} c_0^i di \sim \frac{\sqrt{2\pi} c_0^{\hat{i}}}{\sqrt{t \frac{I(p^{\hat{i}})}{\gamma}}}.$
Let $I_t^- = \inf\limits_{i' \in B_t} I(p^{i'})$, $I_t^+ = \sup\limits_{i' \in B_t} I(p^{i'})$, $c_t^- = \inf\limits_{i' \in B_t} c_0^{i'}$, $c_t^+ = \sup\limits_{i' \in B_t} c_0^{i'}$,
by Equation 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, we get

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

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

$i)\sqrt{t\frac{I_t^-}{\gamma}}$ on the right integral, to get

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

We now recognize both integrals as standard Gaussian. 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 $\frac{c_0^{\hat{i}}}{\sqrt{t\frac{I(p^{\hat{i}})}{\gamma}}}$ and we get $\mathcal{I}_2 \sim \frac{\sqrt{2\pi}c_0^{\hat{i}}}{\sqrt{t\frac{I(p^{\hat{i}})}{\gamma}}}$.

Putting † and ‡ together:

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

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

Proof.

$$\begin{aligned}\beta_j^t\left(\int_{A_k} p^i(\sigma^t)^{\frac{1}{\gamma}}c_0^i di\right)^{\gamma_k} &=^{By\ Lem.2}\ e^{t\ln\beta_j+\ln p^{\hat{i}(\sigma^t)}(\sigma^t)-\frac{\gamma_k MAR}{2}\ln t+O(1)}\\ &= e^{\ln p^{\hat{i}(\sigma^t)}(\sigma^t)-\frac{k^{MAR}}{2}\ln t+O(1)} * e^{t\ln\beta_j-\left(\frac{\gamma k^{MAR}-k^{MAR}}{2}\right)\ln t+O(1)}\\ &=^{By\ BIC}\int_{A_k} p^i(\sigma^t)di * e^{t\ln\beta_j-\left(\frac{\gamma k^{MAR}-k^{MAR}}{2}\right)\ln t+O(1)}\end{aligned}$$

□

Proof of Lemma 3

Proof. Let A_γ be a positive mass of Bayesian traders with regular priors, $g^i(\theta)$ on the same k-dimensional parameter space Θ. We 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 $\bar{g} = \sup_{i,\theta\in int\Theta} g^i(\theta)$ and $\underline{g} = \inf_{i,\theta\in int\Theta} g^i(\theta)$. Note that $\underline{g} > 0$, because the prior distribution of every trader in A_γ is strictly positive and that $\bar{g} < \infty$ because all of the g^i's are continuous in the simplex, thus bounded in its (strict) interior. Because the convergence result of Lemma 5 is uniform, it follows that

$$
\begin{aligned}
\left(\int_{A_\gamma} c_0^i p^i(\sigma^t)^{\frac{1}{\gamma}} di\right)^\gamma &= \left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta) g^i(\theta)\ d\theta\right)^{\frac{1}{\gamma}} di\right)^\gamma \\
&\in \left[\left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta)\underline{g}\ d\theta\right)^{\frac{1}{\gamma}} di\right)^\gamma, \left(\int_{A_\gamma} c_0^i \left(\int_\Theta p(\sigma^t|\theta)\bar{g}\ d\theta\right)^{\frac{1}{\gamma}} di\right)^\gamma\right] \\
&= \left[\int_\Theta p(\sigma^t|\theta)\underline{g}\ d\theta \left(\int_{A_\gamma} c_0^i di\right)^\gamma, \int_\Theta p(\sigma^t|\theta)\bar{g}\ d\theta \left(\int_{A_\gamma} c_0^i di\right)^\gamma\right] \\
&=^{By\ Lem.3}\ e^{\ln p^{\hat{\theta}(\sigma^t)}(\sigma^t) - \frac{k}{2}\ln t + O(1)} \\
&=^{By\ BIC}\ O(1)\int_\Theta p(\sigma^t|\theta) f(\theta) d\theta \\
&= O(1) p^B(\sigma^t).
\end{aligned}
$$

□

Proof of Proposition 4

Proof. Lets 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. We only need to show 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, 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 Equation 4 and using † and ‡ from the proof of Lemma 5,

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

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:

- $\lim_{t\to\infty} \sup_{i\in B_t(\hat{i})} ||i-\hat{i}|| = 0$: the market rewards an empirically accurate trader.
- $\int_A \lim_{t\to\infty} I_{B_t} g^i di = 0$, trader $\hat{i}$ is not a priori accurate.

□

Proof of Proposition 5

Proof. Let A_S be a skilled cluster.

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);

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.

c) is necessary for A_S to vanish P-a.s..

By contradiction: if $k_P^{BIC} = 0$, then A_S has the maximal survival index because $k_S^{MAR} = 0$ and survives by Proposition 1. □

Proof of Proposition 6

Proof. If the skilled cluster dominates, the convergence follows from standard economic arguments (Sandroni, 2000). Otherwise, the result follows proving these 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.

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

In equilibrium (as per Equation 1), $C_t^j(\sigma) = \beta^{\frac{t}{\gamma_j}} \frac{\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di}{q_t(\sigma)^{\frac{1}{\gamma_j}}}$. Exponentiating by γ_j and summing over clusters, we obtain $q_t(\sigma) = \frac{\sum_j \left(\beta^{\frac{t}{\gamma_j}} \int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di\right)^{\gamma_j}}{\sum_j C_t^j(\sigma)^{\gamma_j}}$, so that:

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

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

By Lemma 5, $j \neq \bar{j} \Rightarrow \frac{\beta_j^t \left(\int_{A_j} p^i(\sigma^t)^{\frac{1}{\gamma_j}} c_0^i di\right)^{\gamma_j}}{\bar{\beta}^t \left(\int_{\bar{A}} c_0^i p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} di\right)^{\bar{\gamma}}} = o(1)$.

Therefore, Equation 7 obeys the following asymptotic.

$$\begin{aligned}
q(\sigma_t|\sigma^{t-1}) &= \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}} \\
&= \frac{\bar{C}_{t-1}(\sigma)^{\bar{\gamma}} + o(1)}{\bar{C}_t(\sigma)^{\bar{\gamma}} + o(1)} \left(\frac{\bar{\beta}^t \left(\int_{\bar{A}} c_0^i p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} di\right)^{\bar{\gamma}} + \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}} + \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)} \left(\frac{\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}} + \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 + \frac{\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) \\
&= \frac{\bar{C}_{t-1}(\sigma)^{\bar{\gamma}} + o(1)}{\bar{C}_t(\sigma)^{\bar{\gamma}} + o(1)} \left(\frac{\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)
\end{aligned}$$

Claim 2: $\forall \sigma \in \hat{S}, \quad \sup_{i \in \bar{C}_{t-1}} \left\| \frac{c_{t-1}^i(\sigma)^{\bar{\gamma}}}{c_t^i(\sigma)^{\bar{\gamma}}} - \frac{c_{t-1}^{\hat{i}(\sigma^t)}(\sigma)}{c_t^{\hat{i}(\sigma^t)}(\sigma)} \right\| \to 0.$

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 $T = o(t)$, $\bar{B}_T^c = \bar{A} \setminus \bar{B}_T$ its complement and $\bar{C}_{\bar{B}_T}(\sigma)$ 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 ‡, $\frac{\bar{C}_{\bar{B}_T^c}(\sigma)}{\bar{C}_{\bar{B}_T}(\sigma)} \to 0$ for every T. Thus

$$\begin{aligned}
\frac{\bar{C}_{t-1}(\sigma)^{\bar{\gamma}}}{\bar{C}_t(\sigma)^{\bar{\gamma}}} &= \frac{\left(\bar{C}_{\bar{B}_{T,t}}(\sigma) + C_{\bar{B}^c_{T,t}}(\sigma)\right)^{\bar{\gamma}}}{\left(\bar{C}_{\bar{B}_{T,t-1}}(\sigma) + C_{\bar{B}^c_{T,t-1}}(\sigma)\right)^{\bar{\gamma}}} \\
&\in \left\{ \frac{\min\{c \in \bar{C}_{\bar{B}_{T,t}}(\sigma)\}}{\max\{c \in \bar{C}_{\bar{B}_{T,t-1}}(\sigma)\}} - o(1), \frac{\max\{c \in \bar{C}_{\bar{B}_{T,t}}(\sigma)\}}{\min\{c \in \bar{C}_{\bar{B}_{T,t-1}(\sigma)}\}} + o(1) \right\} \\
&\to^a \frac{c_t^{\hat{i}(\sigma^t)}(\sigma)}{c_{t-1}^{\hat{i}(\sigma^t)}(\sigma)}.
\end{aligned}$$

(a) 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.

Claim 3: $\forall \sigma \in \hat{S}$, $\left\| \left(\int_{\bar{A}} \frac{p^i(\sigma^t)^{\frac{1}{\bar{\gamma}}} c_0^i di}{\int_{\bar{A}} p^i(\sigma^{t-1})^{\frac{1}{\bar{\gamma}}} c_0^i di} \right)^{\bar{\gamma}} - \bar{p}^{\hat{i}(\sigma^t)}(\sigma_t|\sigma^{t-1}) \right\|_\infty \to 0.$

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

$$\begin{aligned}
&\in \left\{ \frac{\inf_{i\in\bar{B}_{t-1}} p^i(\sigma^t) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{t \frac{I(p^i)}{\gamma}}} \right)^{\gamma}}{\sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{(t-1)} \frac{I(p^i)}{\gamma}} \right)^{\gamma}} - o(1); \frac{\sup_{i\in\bar{B}_{t-1}} p^i(\sigma^t) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{t \frac{I(p^i)}{\gamma}}} \right)^{\gamma}}{\inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{(t-1)} \frac{I(p^i)}{\gamma}} \right)^{\gamma}} + o(1) \right\} \\
&\in \left\{ \inf_{i\in\bar{B}_{t-1}} p^i(\sigma_t|\sigma^{t-1}) \frac{\inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{t \frac{I(p^i)}{\gamma}}} \right)^{\gamma}}{\sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{(t-1)} \frac{I(p^i)}{\gamma}} \right)^{\gamma}} - o(1); \sup_{i\in\bar{B}_{t-1}} p^i(\sigma_t|\sigma^{t-1}) \frac{\sup_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{t \frac{I(p^i)}{\gamma}}} \right)^{\gamma}}{\inf_{i\in\bar{B}_{t-1}} p^i(\sigma^{t-1}) \left(\frac{\sqrt{2\pi} c_0^i}{\sqrt{(t-1)} \frac{I(p^i)}{\gamma}} \right)^{\gamma}} + o(1) \right\} \\
&\to^a \bar{p}^{\hat{i}(\sigma^{t-1})}(\sigma_t|\sigma^{t-1}).
\end{aligned}$$

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

Claim 4: $\forall \sigma \in \hat{S}$, $\left\| P(\sigma_t|\sigma^{t-1}) - p^{\hat{i}(\sigma^{t-1})}(\sigma_t|\sigma^{t-1}) \right\|_\infty \to 0.$

We 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, thus 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 applying the proof of **C3** (with $\gamma = 1$) to the true probability: $P(\sigma_t|\sigma^{t-1}) = \int_{\Theta_J} \frac{p^i(\sigma^t)g_0^i di}{\int_{\Theta_J} p^i(\sigma^{t-1})g_0^i di}$:

$$\begin{aligned}
\lim_{t\to\infty} \left\| P(\sigma_t|\sigma^{t-1}) - \bar{p}^{\hat{i}(\sigma^t)}(\sigma_t|\sigma^{t-1}) \right\|_\infty &= \lim_{t\to\infty} \left\| \int_{\Theta_J} \frac{p^i(\sigma^t)g_0^i di}{\int_{\Theta_J} p^i(\sigma^{t-1})g_0^i di} - \bar{p}^{\hat{i}(\sigma^t)}(\sigma_t|\sigma^{t-1}) \right\|_\infty \\
&= \lim_{t\to\infty} \left\| p_P^{\hat{i}(\sigma^t)}(\sigma_t|\sigma^{t-1}) - \bar{p}^{\hat{i}(\sigma^t)}(\sigma_t|\sigma^{t-1}) \right\|_\infty \\
&= 0.
\end{aligned}$$

Because the convergence results in **Claims 1-4** are all uniform, we obtain the desired:

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

□

References

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

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4:1236–1239.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *The Annals of Mathematical Statistics*, 33(3):882–886.

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

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

Borovička, J. (2013). Survival and long-run dynamics with heterogeneous beliefs under recursive preferences. *Available at SSRN 2023501.*

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

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

Cvitanić, J. and Malamud, S. (2011). Price impact and portfolio impact. *Journal of Financial Economics*, 100(1):201–225.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1991). The survival of noise traders in financial markets. *The Journal of Business*, 64(1):pp. 1–19.

Dindo, P. (2016). Survival in speculative markets. *Available at SSRN 2705337*.

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

Geanakoplos, J. D. and Polemarchakis, H. M. (1982). We can't disagree forever. *Journal of Economic Theory*, 28(1):192–200.

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

Kalai, E. and Lehrer, E. (1994). Weak and strong merging of opinions. *Journal of Mathematical Economics*, 23(1):73–86.

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

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

Mahmoud, H. (2008). *Pólya urn models*. CRC press.

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

Massari, F. (2017). Markets with heterogeneous beliefs: A necessary and sufficient condition for a trader to vanish. *Journal of Economic Dynamics and Control*.

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

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

Pólya, G. (1930). Sur quelques points de la théorie des probabilités. In *Annales de l'institut Henri Poincaré*, volume 1, pages 117–161.

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

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

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

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