Market Selection in Large Economies: a Matter of Luck*

Filippo Massari

School of Banking and Finance, Australian School of Business

December 5, 2014

Abstract

Within a general equilibrium model with a continuum of traders, we investigate the Market Selection Hypothesis (MSH) - markets select traders with more accurate beliefs. We find, contrary to known results for economies with (only) finitely many traders, that risk attitudes affect survival and that markets might select against traders with accurate beliefs. Remarkably, even in these cases, asymptotic equilibrium prices reflect accurate beliefs. Thus, unlike known violations of MSH, we corroborate Freedman’s conjecture that market selection forces induce rational expectations.

JEL Classification: D51, D01, G1

1 Introduction

According to the market selection hypothesis, henceforth MSH, the market selects for the traders with the most accurate beliefs. This hypothesis, first articulated by Alchian (1950) and Friedman (1953), is one of the key arguments supporting rational expectations: as the consumption-share of traders with correct beliefs converges to one,

*I wish to thank Werner Ploberger, John Nachbar, Larry Blume, David Easley, Pietro Dindo, Alvaro Sandroni, Luciano Pomatto, Philip Dybvig, David Feldman, Aleks Yankelevich, Nick Papageorge and Stephanie Heger.
financial markets can be understood, to a large extent, using models in which traders’ beliefs are correct.

Previous literature has shown that the MSH can fail in partial equilibrium models or if the market contains some inefficiencies (De Long et al (1990, 1991), Shleifer-Summers (1990) and Blume-Easley (1992)). While, in general equilibrium models with finitely many traders, small economies henceforth, the MSH holds true (Sandroni (2000) and Blume-Easley (2006)), albeit mild assumptions on preferences (Kogan et al. (2006)) and on the aggregate endowment process (Kogan et al. (2011), Yan (2008)).

This paper asks whether this fundamental result also applies to general equilibrium models that satisfy the regularity conditions of Sandroni (2000) and Blume-Easley (2006) but are populated by a continuum of traders, henceforth large economies.

There are three main reasons to focus on large economies. First, large economies are better description than small economies of the small sample properties of financial markets. A well developed financial market is populated by a large number of traders and (has) run for finitely many periods. Small economies describe the ideal situation in which the number of trading periods is large enough (traders live forever) for the trader with the most effective investment strategy to dominate the market. Large economies describe the more realistic situation in which the number of trading periods is not large enough to discriminate between traders with investment strategies that are similar but not identical: even if traders live forever, the most accurate trader cannot be uniquely identified after finitely many trading periods. Second, in a large economy it is safe to assume that traders are price-takers (i.e. the competitiveness of walrasian equilibrium Aumann (1965)). It is often argued that price-taking behavior is not at odds with the small economy setting because of the equivalence between a small economy with \( N \) traders and a large economy with \( N \) groups of identical traders. Nevertheless, this level of homogeneity is hardly, if ever, met in financial markets. The large economy setting does not impose this restriction. Third, small economies are special cases of large economies, thus focusing on the large comports no loss of generality.

Contrary to the established results for small economies, we find that, in the large, risk attitudes affect survival and that markets might select against traders with accurate
beliefs.

Risk attitudes affect survival through their effect on aggregate saving. In the CRRA utility specification, the parameter that captures traders’ attitudes toward risk also capture their attitudes toward inter-temporal consumption. Given two groups of traders with equivalent heterogeneous beliefs, the two groups have equivalent aggregate investment decisions but the group whose traders are less risk averse have stronger speculative incentives to invest, thus save more and come to dominate (see Section 4). The same phenomenon is also present in small economies (see Section 6.1), nevertheless in this setting, the best trader in the economy is selected faster and beliefs heterogeneity effectively disappear in finitely many periods eliminating the speculative incentives. Thus, in small economies, risk attitudes affect the asymptotic consumption-share distribution but their effect is not strong enough to drive the consumption-share of a trader all the way to 0 (1) and be captured by the standard, coarse notion, of trader survival.

The effect of risk attitudes on survival suggest that there are large economies in which the MSH fails. Section ?? provide a sufficient condition for these cases to happens, thus showing that having correct beliefs is neither necessary nor sufficient condition for a positive mass of trader to survive. But who survive then? And what happens to equilibrium prices? Our analysis shows that, if the MSH fails, it is impossible to predict which one of the remaining traders will survive. In other words, luck is the only determinant of survival. Specifically, the failure of the MSH we identify occurs if i) the true probability is such that the maximum likelihood parameter is a random variable with continuum support; ii) there is always a (positive mass) of traders whose beliefs are, by luck, arbitrarily close to the empirical maximum likelihood parameter and iii) traders with incorrect beliefs are investing aggressively enough for luck to pay off. The first condition is met, for example, if the true data generating process is a mixture of iid processes (aka, exchangeable process, see Appendix A). These processes are a natural generalization of iid processes, in Kreps (1988) words: “...exchangeability is the same as “independent and identically distributed with a prior unknown distribution function”...”. The last two conditions simply require that there is a trader that gets lucky and that the money in the market moves fast enough for luck to pay off.
Remarkably, even if it is impossible to predict which trader will survive, we are able to show that asymptotic equilibrium prices reflect beliefs that are, ex-post, accurate. This seemingly counterintuitive result depends on the fact that a trader with incorrect beliefs dominates if and only if his beliefs are, on the realized sequence, as good as the correct one. Thus, unlike the other known violations of the market selection hypothesis, our result corroborates Freedman’s conjecture that the selection forces in the market induce rational expectations.

The paper proceeds as follows. Section 2 describes the setting and the assumptions. Section 3 studies economies in which all traders have identical utilities. This setting is already reach enough to illustrate the role played by risk attitudes on survival (Section 5); to show a case in which the MSH fails (Section ??); to highlight that, if the MSH fails, luck is the only way for a trader to survive (Section 5.2) and that asymptotic equilibrium prices reflect accurate beliefs even if the MSH fails (Section 5.3). Section 4 extends our finding to economies with heterogeneous risk attitudes, characterizes equilibrium prices and provides a general sufficient condition for a group of traders to vanish that covers both the small and the large economy setting. In Section 6.1, we discuss the relationship between the small and the large setting.

2 The model

2.1 The probabilistic environment and beliefs accuracy

The model is an infinite horizon Arrow-Debreu exchange economy with complete markets with a unique perishable consumption good. Time is discrete and begins at date 0. At each date there is a finite set of states \( S \equiv \{1, \ldots, S\} \) with cardinality \(|S|=S\). The set of all infinite sequences of states is \( S^\infty \) with representative sequence (path) \( \sigma = (\sigma_1, \ldots) \). We use \( \sigma^t = (\sigma_1, \ldots, \sigma_t) \) to indicate a finite sequences of realizations of length \( t \) and \( \Sigma^t \) for the algebra that consists of all finite unions of sequences of length \( t \). \( \Sigma \) is the smallest \( \sigma \)-algebra on \( \cup_{t=1}^{\infty} \Sigma^t \). In the next paragraphs we introduce a number of economic variable that depends on \( \sigma^t \), all of them are assumed to be \( \Sigma^t \) measurable. The true probability measure on \( \Sigma \) is \( P \), while each trader has subjective, possibly
incorrect, probabilistic view $p^i$ on $\Sigma$. For any probability measure $p^i$, $p^i(\sigma^t)$ is the marginal probability of $\sigma^t$, that is $p^i(\sigma^t) = p^i(\{\sigma_1\} \times ... \times \{\sigma_t\} \times S \times S \times ... )$. With an abuse of notation, $p^i(\sigma^t)$ also indicates the likelihood of $p^i$ on $\sigma^t$. For example, if $p^i$ is an iid Bernoulli model with $p^i(\sigma_t = 1) = i$, then $p(\sigma^t) = t_1(1 - i)^{t_0}$, where $t_1$ and $t_0$ denote, respectively, the number of realizations of state 1 and of state 0 on $\sigma^t$.

Following the tradition in the market selection literature, we assume trader’s beliefs to be dogmatic: traders “agree to disagree” and trade for speculative reasons. We will use two distinct notion of accuracy, both based on the likelihood.$^1$

**Definition 1.** Given a true probability $P$, $p^i$ is more accurate than $p^j$ if $\frac{p^j(\sigma^t)}{p^i(\sigma^t)} \to P$-a.s. $0$.

**Definition 2.** Given a path $\sigma$, $p^i$ is empirically more accurate than $p^j$ if $\frac{p^j(\sigma^t)}{p^i(\sigma^t)} \to 0$.

The two definitions are closely related but capture two different type of accuracy. The first definition is an ex-ante notion of accuracy: $P$ is used to ex-ante identify the set of sequences that will occur with probability 1. The second definition is an ex-post notion of accuracy and only depends on the sequence we are given. In a small economies, the most accurate belief according to the first definition is also the most accurate belief according to the second definition. Nevertheless, in large economies, this is not necessarily the case. Our result shows that, if the two notion identifies different beliefs, risk attitudes determine the type of accuracy that is rewarded by the market: if money move fast, the market rewards for being empirically accurate (luck), otherwise it reward for being (probabilistically) accurate (skill).

### 2.2 The traders in the economy

The measure space of traders is $(A, A, i)$ where $A$ is the unit interval, $A$ its Borel subsets, and $i$ is the Lebesgue measure. The economy is characterized by the aggregate

$^1$Focusing on beliefs’ likelihood is in the tradition of the selection literature, but, unlike Sandroni (2000) and Blume-Easley (2006), we cannot rely on approximate measures. The reason is twofold: first, Sandroni’s definition (average accuracy), is to coarse to distinguish between the two type of accuracy I identify (the averaging factor, $\frac{1}{t}$, renders the $O(\log t)$ component of my sufficient condition to vanish (Theorem 1) dimensionality component of the BIC approximation mute (see Section 3); second, Blume-Easley’s definition cannot be applied because, as showed in Massari (2013), it can lead to incorrect results.
preferences, \( \succ_{\gamma_j} \), and by the aggregate time 0 consumption, \( C_0^{\gamma_j} \), of \( N \) measurable sets of traders \( A_{\gamma_j}, j = 1, \ldots, N \). \( \succ_{\gamma_j} \) and \( C_0^{\gamma_j} \) are constructed, respectively, by aggregating the preferences, the initial consumptions of groups of individual traders, \( i \), with believes \( p^i \), utilities \( u^i \), endowment process \( e^i(\sigma^t) \) and infinitesimal time 0 consumption \( c^i_0 \). With an abuse of notation, \( A_{\gamma_j} \) is used to represent a set of traders, \( A_{\gamma_j} = \{ i \in A_{\gamma_j} \} \), a set of probabilities, trader’s beliefs, \( A_{\gamma_j} = \{ p^i : i \in A_{\gamma_j} \} \), and the set of parameters describing traders beliefs \( A_{\gamma_j} = \{ i \in \Theta_{\gamma_j} \} \). For example, if all traders in \( A_{\gamma_j} \) have iid Bernoulli beliefs and the union of their beliefs covers the simplex, \( A_{\gamma_j} \) indicates: \( \{ i \in A_{\gamma_j} \}, \{ p^i : i \in A_{\gamma_j} \} \) and \( \{ i \in (0, 1) \} \).

**Definition 3.** A cluster, \( A_{\gamma_j} \), is a measurable subset of \( A \) such that:

- cluster \( A_{\gamma_j} \) has strictly positive time 0 consumption: \( C_0^{\gamma_j} = \int_{A_{\gamma_j}} c^i_0 \, di > 0 \)
- traders in \( A_{\gamma_j} \) have identical CRRA utility function \( u(c) = \frac{c^{1-\gamma_j} - 1}{1-\gamma_j} \) and identical discount factor \( \beta_j \)
- either all traders in \( A_{\gamma_j} \) have identical beliefs or \( \Theta_{\gamma_j} \) is an open subset of the parameter space of a member of the exponential family.

The definition of cluster ensures tractable aggregate investment strategies for sets of traders with positive aggregate time 0 consumption. Homogeneity of preferences and discount factors allows to express cluster’s optimal consumption choices as a function of his discount factor, risk attitudes and aggregate beliefs. The last condition ensure tractability of aggregate beliefs. If all traders in \( A_{\gamma_j} \) have identical beliefs, aggregation is not problematic and \( A_{\gamma_j} \) can be treated as a trader with positive mass, thus a small economy with \( n \) traders is equivalent to a large economy with \( n \) clusters of traders with identical beliefs. Otherwise, the second part of the condition guaranties that an asymptotic approximation of clusters aggregate risk adjusted beliefs can be done. The exponential family restriction is mild and allows for both iid and non-iid models (for example Markov models with finitely many lags).

Let \( C^{\gamma_j}(\sigma^t) = \int_{A_{\gamma_j}} c^i(\sigma^t) \, di \) be cluster \( A_{\gamma_j} \)'s period \( t \) consumption on \( \sigma \). In the tradition of the selection literature, the asymptotic fate of a cluster is coarsely characterize by the distinction between those clusters who disappear and those who do not.
Definition 4. Cluster $A_{\gamma_j}$ vanishes on $\sigma^t$ if its consumption converges to 0: $\lim_{t \to \infty} C_{\gamma_j}(\sigma^t) = 0$. Cluster $A_{\gamma_j}$ survives on $\sigma^t$ if: $\limsup_{t \to \infty} C_{\gamma_j}(\sigma^t) > 0$. Cluster $A_{\gamma_j}$ dominates on $\sigma^t$ if: $\lim_{t \to \infty} C_{\gamma_j}(\sigma^t) = 1$.

2.3 The assumptions

Throughout the paper we refer to these assumptions.

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

A2: The aggregate endowment is bounded above and below.

A3: For all traders $i, j$, all dates $t$ and all paths $\sigma$, $p_i(\sigma^t) > 0 \iff p_j(\sigma^t) > 0$.

A4: The competitive equilibrium exists.

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

A6: All traders have identical discount factor $\beta$.

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

In this section we present the reference economy and a series of technical result that will be repeatedly used in the rest of the paper.

The economy is a discrete time Arrow-Debreu exchange economy with complete markets, constant aggregate endowment and S states and N clusters of clusters with CRRA parameters \( \gamma_j \) and discount factors \( \beta_j \). Every individual trader in the economy aims to solve:

\[
\max_{\{c(\sigma^t)\}_{t=0}^{\infty}} E_p \sum_{t=0}^{\infty} \beta^t u^1(\sigma^t) \quad \text{s.t.} \quad \sum_{t=0}^{\infty} \sum_{\sigma^t \in \Sigma^t} q(\sigma^t) \left( c^i(\sigma^t) - e^i(\sigma^t) \right) \leq 0.
\]

The true probability, \( P \), is, for now, left unspecified.

Traders’ first order conditions of the maximization problem are sufficient for the pareto optimum and, in every path \( \sigma^t \), can be expressed as:

\[
\left( c^i(\sigma^t) \right)^{\gamma_j} = \left( c^i_0 \right)^{\gamma_j} \beta^t \frac{\int_{A_{\gamma_j}} p^i(\sigma^t) \frac{1}{\gamma_j} \ c^i_0 \ di}{q(\sigma^t)^{\frac{1}{\gamma_j}}}. \tag{1}
\]

Exponentiating by the CRRA parameters and taking ratio, prices simplify out:

\[
\frac{(C^{\gamma_j}(\sigma^t))^{\gamma_j}}{(C^{\gamma_k}(\sigma^t))^{\gamma_k}} = \frac{\beta^t_j \left( \int_{A_{\gamma_j}} c^i_0 p^i(\sigma^t) \frac{1}{\gamma_j} \ di \right)^{\gamma_j}}{\beta^t_k \left( \int_{A_{\gamma_k}} c^i_0 p^i(\sigma^t) \frac{1}{\gamma_k} \ di \right)^{\gamma_k}}. \tag{1}
\]

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

**Lemma 1.** Under A1-A6, \( A_{\gamma_j} \) vanishes on \( \sigma \) if exists \( A_{\gamma_k} \):

\[
\frac{\beta^t_j \left( \int_{A_{\gamma_j}} c^i_0 p^i(\sigma^t) \frac{1}{\gamma_j} \ di \right)^{\gamma_j}}{\beta^t_k \left( \int_{A_{\gamma_k}} c^i_0 p^i(\sigma^t) \frac{1}{\gamma_k} \ di \right)^{\gamma_k}} \rightarrow 0
\]

**Proof.** By A2, \( (C^{\gamma_k}(\sigma^t))^{\gamma_k} \) variates in \( \sigma, j, k \). Thus by Equation 1, \( \forall \gamma_j, \gamma_k \in (0, \infty) \)

\[
\frac{(C^{\gamma_j}(\sigma^t))^{\gamma_j}}{(C^{\gamma_k}(\sigma^t))^{\gamma_k}} \rightarrow 0 \iff C^{\gamma_j}(\sigma^t) \rightarrow 0
\]

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

Definition 5. Given a set of probabilities parametrized by $A_{\gamma j}$ and a path $\sigma^t$, $\hat{p}_{\gamma j}^{(\sigma^t)}$ is the empirically most accurate probability in $A_{\gamma j}$.

Definition 6. $\hat{S}$ is the set of sequences such that, for every cluster in the economy, $\hat{p}_{\gamma j}^{(\sigma^t)}$ exists and is strictly bounded away from the boundary of the parameter space.

Focusing on paths in $\hat{S}$ is more general than assuming a specific true distribution and comports almost no loss of generality. For example, if all traders have exchangeable beliefs, $\hat{S}$ is the set of all sequences whose frequency exists and is bounded away from 0 and 1. It is easy to show that for every probabilistic model $P$ whose parameters belong to the strict interior of the parameter space $P(\hat{S}) = 1$.

BIC approximation. Let $\mathcal{M}$ be a member of the exponential family and $p^B(\sigma^t)$ be the bayesian likelihood obtained from a continuous, strictly positive prior distribution, $g$, on a $k_{BIC}$-dimensional non-empty open subset, $\Theta_0$, of the parameter space $\Theta$, then,

$$\forall \sigma \in \hat{S}, \quad p^B(\sigma^t) := \int_{\Theta} p^i(\sigma^t) g^i \, di \approx e^{\ln p^i(\sigma^t) - \frac{k_{BIC}}{2} \ln t}.$$  

Moreover, if $P \in \Theta_0$, then $\hat{p}^{(\sigma^t)}(\sigma_t|\sigma^{t-1}) \rightarrow P$-a.s. $P(\sigma_t|\sigma^{t-1})$.

The BIC approximation shows that the accuracy of the probabilities obtained via Bayes rule depends on the dimensionality of the prior support ($k$). It formalizes the intuition that there is an accuracy cost in using models with redundant parameters because some of the information of the sample is “wasted” to learn that their true value is 0.\footnote{A classical example is the following.} Importantly, the first line holds for every $\sigma \in \hat{S}$, thus independently
from the true probability: probabilities obtained via Bayes’ rule are empirically as accurate as \( p^i(\sigma^t) \), albeit a penalty term that depends on dimensionality of the set of parameters. The second line tells us that, if the true probability is in the prior support, the empirically most accurate probability converges to the true probability.

Lemma 2 obtains a similar approximation for risk adjusted aggregate beliefs.

**Lemma 2.** Under A1-A5, cluster \( A_{\gamma_j} \)'s risk adjusted aggregate beliefs satisfies:

\[
\forall \sigma \in \hat{S}, \quad \left( \int_{A_{\gamma_j}} p^i(\sigma^t)^{\gamma_j} c_0^i d\sigma \right)^{\gamma_j} \approx e^{\ln p^i(\sigma^t)(\sigma^t) - \frac{\gamma_j k_{MAK}^j}{2} \ln t}
\]

Where \( k_{MAK}^j \) is the dimensionality of \( A_{\gamma_j} \) (interpreted as a set of parameters).

Moreover, if \( P \in A_{\gamma_j} \), then \( p^i(\sigma^t)(\sigma_t|\sigma_t-1) \rightarrow P \)-a.s. \( P(\sigma_t|\sigma_t-1) \).

**Proof.** See Appendix

Lemma 2 shows that risk attitudes interact with clusters dimensionality, thus can affect survival. Note that, in small economies \( k_{MAK} = 0 \), thus risk attitudes do not affect survival.

## 4 The role of risk attitudes

To highlight the effect of risk attitudes on clusters’ survival lets start with the simple case in which clusters only differs in their risk attitudes.

**Proposition 1.** In an economy that satisfy A1-A6 with \( N \) clusters with identical beliefs sets with dimensionality \( k_{MAK} > 0 \), the least risk averse cluster dominates on \( \hat{S} \).

**Proof.** Application of Theorem 1

Suppose the true probability is Bernoulli with parameter \( P \). There are two Bayesians traders \( (B^1, B^2) \): \( B^1 \) has a smooth prior on the Bernoulli family (1 parameter: \( k^1 = 1 \)) and \( B^2 \) has a smooth prior on the Markov (1) family (2 parameters: \( k^2 = 2 \)). Since every iid model is also Markov 1, the next period forecasts of both traders converge to the true probability. Nevertheless, application of the BIC approximation reveals that the beliefs of \( B^1 \) are more accurate than the beliefs of \( B^2 \).
**Example 1:** Consider an Arrow-Debreu exchange economy with two states $S = \{W, R\}$. The economy contains two clusters: $A_{\gamma}$ and $A_{\eta}$ with identical discount factor $\beta$ but different risk attitudes $\gamma < \eta$. All traders have iid beliefs and $A_{\eta} = \{i \in (0, 1)\} = A_{\gamma}$. Rearranging the FOC as for Equation 1 and using Lemma 2 we find that the most risk averse cluster vanishes on $\hat{S}$:

$$\lim_{t \to \infty} \left( \frac{C_{A_{\eta}}(\sigma^t)}{C_{A_{\gamma}}(\sigma^t)} \right)^{\eta} = \lim_{t \to \infty} \frac{\beta^t \left( \int_{A_{\eta}} c_{i}^{*} p_{i}^{*}(\sigma^t) \frac{1}{\gamma} di \right)^{\eta}}{\beta^t \left( \int_{A_{\gamma}} c_{i}^{*} p_{i}^{*}(\sigma^t) \frac{1}{\gamma} di \right)^{\gamma}} = \lim_{t \to \infty} \frac{e^{\ln p_{i}^{*}(\sigma^t) - \frac{\gamma}{2} \ln t}}{e^{\ln p_{i}^{*}(\sigma^t) - \frac{\eta}{2} \ln t}} = 0.$$

Example 1 highlights that risk attitudes affects survival through their effect on clusters’ optimal saving rate: at an individual level, every trader has a dogmatic beliefs that the data generating process is iid. Because we have a unique price and heterogeneous opinions, most traders subjectively believe that prices are incorrect and trade for speculative reasons. In the CRRA utility specification, the CRRA parameter captures, at the same time, traders’ attitudes toward risk and their attitudes toward inter-temporal consumption. In particular, if $\gamma < \eta$ each trader in $A_{\gamma}$ optimally decide to postpone more consumption than the trader with the same beliefs in $A_{\eta}$ does. Thus, even if the two cluster have equivalent beliefs, $A_{\gamma}$ saves more and dominates.

Rearranging cluster’s risk adjusted beliefs we can explicitly separate the belief component from the speculative saving component in clusters optimal consumption plan: $e^{\ln p_{i}^{*}(\sigma^t) - \frac{\gamma}{2} \ln t} = e^{\ln p_{i}^{*}(\sigma^t) - \frac{\eta}{2} \ln t} \times e^{-\left(\frac{\gamma}{2} - \frac{1}{2}\right) \ln t}$. The first exponent represent cluster beliefs (it is equivalent to the probability obtained via Bayes’ rule from a strictly positive, continuous prior on $A_{\gamma}$); the second exponent represents the effect of risk attitudes on optimal saving and is the. In the example, the two clusters have identical belief and survival is only determined by differences in the second exponents.

**5 The role of heterogeneity of opinions**

In the previous Section we found that for clusters with heterogeneous beliefs, risk attitudes interact with cluster dimensionality to determine the saving rate. Here I study the case in which the economy has some clusters with heterogenous beliefs and some clusters of identical Bayesian traders. We find that the endogenous saving rate
\((\gamma_j k_j^{MAK})\) and the rate at which the Bayesian learn \((k_j^{BIC})\) can offset each-other.

**Proposition 2.** Under A1-A6, if the economy contains two clusters: \(A_{\gamma_U}\), whose traders have heterogeneous beliefs; and \(A_{\gamma_B}\), whose traders are Bayesian with identical, continuous, strictly positive prior on \(A_U\), then

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

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

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

**Proof.** Application of Theorem 1

**Example 2:** 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, \(\gamma\) and discount factor, \(\beta\). Traders in \(A_U\) have heterogeneous iid beliefs \(p^i\) such that \(A_U = \{i \in (0, 1)\}\), while traders in \(A_B\) have identical beliefs \(p^B\) which are obtained via Bayes’ rule from a continuous, strictly positive prior, \(g^i\), on \((0, 1)\):

\[
p^B(\sigma_t) = \int_0^1 g(\theta)p(\sigma_t | \theta)d\theta.
\]

By construction, all traders in \(A_B\) have identical beliefs, thus \(p^B(\sigma_t)^{\frac{1}{\gamma}}\) can be taken out of the integral and Equation 1 becomes:

\[
\frac{(C^B(\sigma_t))^\gamma}{(C^U(\sigma_t))^\gamma} = \frac{p^B(\sigma_t) \left(\int_{A_B} c^i \gamma^i \right)^\gamma}{\left(\int_{A_U} c^i \gamma^i \right)^\gamma}.
\]

Applying the BIC and Lemma 2 to Equation 2 an taking limits we find that

\[
\lim_{t \to \infty} \frac{(C^B(\sigma_t))^\gamma}{(C^U(\sigma_t))^\gamma} = \lim_{t \to \infty} \frac{p^B(\sigma_t) \left(\int_{A_B} c^i \gamma^i \right)^\gamma}{\left(\int_{A_U} c^i \gamma^i \right)^\gamma} = \lim_{t \to \infty} \frac{e^{\ln p^B(\sigma_t) - \frac{1}{2} \ln t}}{e^{\ln p^U(\sigma_t) - \frac{1}{2} \ln t}} = \begin{cases} 0 & \text{if } \gamma \in (0, 1) \\ 1 & \text{if } \gamma = 1 \\ \infty & \text{if } \gamma \in (1, \infty) \end{cases}
\]

and the result follows from Lemma 1.

It is interesting to note that i) risk attitudes of the Bayesian cluster do not affect survival (because \(k_{\gamma_U}^{MAK} = 0\); ii) if the traders in \(A_U\) have log utility \((\gamma = 1)\), risk attitudes have no effect on aggregation. This is because log utility is the knife edge case in which the speculative incentives do not distort the optimal saving rule of individual

\[^{4}k_{\gamma}=1\text{ because for the Bernoulli model, we only need to estimate one parameter}\]
traders (because the income effect and the substitution effect cancels out). Thus, in the spirit of Rubinstein’s (1974), we find that a cluster of traders with log utility and heterogeneous beliefs can be equivalently modeled as a cluster of traders with log utility and identical beliefs that correspond to the consumption-share weighted average of the beliefs of the original cluster. In our case, \( A_U = A_B \), thus the representative agents of the two clusters have equally accurate beliefs and both survives.

5.1 A Violation of the MSH

Proposition 2 holds on \( \hat{S} \), thus it leaves us almost complete freedom to chose the data generating process. If we assume that true probability coincides with the beliefs of the Bayesian cluster, all traders in \( A_B \) have correct beliefs, yet for \( \gamma < 1 \) cluster \( A_B \) vanish, i.e. the MSH fails.

**Proposition 3.** Under A1-A6, if the economy contains a clusters: \( A_{\gamma U} \), whose traders have heterogeneous beliefs and cluster \( A_{\gamma B} \), whose traders are Bayesian with identical, continuous, strictly positive prior on \( A_{\gamma U} \), then the MSH fails on \( \hat{S} \).

But what does it means that the true probability coincides with the probability obtain via Bayes rule? In Appendix A we maintain that this type of probabilities are well defined, constitute a natural generalization of the iid case and, for the case in which we are ignorant of the true distribution, are the only probabilities that are logically consistent with our ignorance. Here we present an example in which \( P = p_B \).

**Example:** The economy is a discrete time Arrow-Debreu exchange economy with two states \( S = \{W, R\} \). The true probability \( P \) evolves according to this (Polya urn) process: the process starts with an urn containing one White ball (\( W \)) and one Red ball (\( R \)). At 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 new ball of the same color.

There are two clusters, \( A_U \) and \( A_B \), with identical risk attitude, \( \gamma \), and discount factor, \( \beta \). Traders in \( A_U \), *unskilled* traders, have heterogeneous iid beliefs \( p^i \) such that \( A_U = \{i \in (0,1)\} \). Since \( P \) is not iid, all traders in \( A_U \) have incorrect beliefs. Traders
in $A_B$, skilled traders, are allowed in every period to observe the composition of the urn before the extraction is made, thus they share correct beliefs, $P$. By construction, all traders in $A_B$ have identical correct beliefs, thus $P(\sigma^t)$ can be taken out of the integral and Equation 1 becomes:

$$\frac{(C^B(\sigma^t))^\gamma}{(C^U(\sigma^t))^\gamma} = \frac{P(\sigma^t)}{\left(\frac{\int_{A_B} c^0_i \, di}{\int_{A_U} c^0_i \, p^i(\sigma^t)^\frac{1}{2} \, di}\right)^\gamma}.$$  

The result follows from Proposition 2 by noticing that $P(R_{t+1}|\sigma^t) = p^B(R_{t+1}|\sigma^t) = \frac{1 + \sum_{t=0}^t \sum_{i=0}^{\tau} \sigma_i = R}{t+2}$: the true probabilities$^5$ and the conditional probabilities obtained via Bayes’ rule from the Uniform prior on $A_U$ $^6$ coincide on $\hat{S}$.

The interested reader will find a direct proof of this result in Appendix C.

The example also highlight a case in which the most accurate beliefs differs from the empirically most accurate beliefs: ex-ante $\forall i \in A_U$, $p^i(\sigma^t) \rightarrow P$-a.s 0 because each $p^i$ has infinitesimal probability to be the asymptotic frequency. Nevertheless, the beliefs of traders in $A_U$ covers all the possible frequencies, thus on each sequence, there is a trader in $A_U$ that happens to have the model with the highest likelihood. Thus, ex-post, $\frac{P(\sigma^t)}{p^i(\sigma^t)} \rightarrow 0$.

### 5.2 If you’re so rich, why aren’t you smart?

In Proposition 3 we present a case in which a cluster whose members have correct beliefs is driven out of the market by a cluster whose members have incorrect beliefs. But, if skilled traders vanish, who dominates then? It turns out that, among unskilled traders, the market selects for traders that are empirically accurate. Given our definition of

$^5$In every period, the true probability of selecting a Red is the fraction of Red balls in the urn:

$P(R_{t+1}|\sigma^t) = \frac{1 + \sum_{t=0}^t \sum_{i=0}^{\tau} \sigma_i = R}{t+2}$.

Denominator: the urn has the 2 initial balls plus a ball for every extraction.

Numerator: the urn has 1 initial Red ball, plus a Red ball for every time a Red ball has been selected.

$^6$Let $t_R = \sum_{t=0}^t \sum_{i=0}^{\tau} \sigma_i = R$. By the definition of conditional probability and of the Beta function:

$$p^B(R_{t+1}|\sigma^t) = \frac{p^B((\sigma_1) \times \cdots \times (\sigma_t) \times (R_{t+1}))}{p^B((\sigma_1) \times \cdots \times (\sigma_t))} = \frac{\int_0^1 \tau_{R+1}^{t+1} (1-x)^{t+1-t_R} \, dx}{\int_0^1 \tau_R^{t+1} (1-x)^{t+1-t_R} \, dx} = \frac{B(t_R+2, t+1-t_R+1)}{B(t_R+1, t-t_R+1)} = \frac{(t_R+1)(t+1)}{t_R(t+1)} = t_R+1$$
luck, these traders cannot have any other merit but to have made a lucky guess. Thus, if the MSH fails, luck and not skill is the only determinant of survival.

**Definition 7.** Trader $i$ is lucky if these conditions are met:
- the maximum likelihood parameters of the true probability are a random variables
- he believes them to be deterministic constants
- their realized value coincides with trader $i$’s beliefs.

The definition of luck is stringent but unambiguous. It requires the true maximum likelihood parameters to be RVs, instead of constants. The reason is that if the maximum likelihood parameters are random variables, it is impossible for a trader to know their value before making investment decisions. Thus ruling out the possible confusion between those traders that know the true parameter because of their informational advantage and those who happens to use the true parameters by chance.

**Proposition 4.** Under A1-A6, if the MSH fails, luck is the sole determinant of survival.

*Proof.* See Appendix

In example 3, the intuition goes as follows. For the Polya urn process described, the empirical maximum likelihood parameter is the realized frequency. If $\gamma < 1$ skilled traders vanish, thus a trader with correct beliefs vanishes for sure. Among the unskilled traders, the market selects for empirically accurate traders, which is to say for the traders whose beliefs are in a shrinking interval around the empirical frequency. Since the empirical frequency is a random variable (the composition of the urn changes stochastically), these traders cannot have any particular merit beside the fact the they made, by luck, the correct guess. Thus, luck is the only determinant of survival.

### 5.3 Asymptotic equilibrium prices reflect accurate beliefs

If the MSH holds, convergence to rational expectations follows from standard economic arguments, but what happens when the market does not select for the traders with correct beliefs? Here we show that, contrary to the other failure of the MSH identified
in the literature, in our setting selection induce rational expectation, albeit second order term due to the effect of risk attitudes: by selecting for lucky traders, the market brings equilibrium prices to reflect beliefs that are, ex post, as accurate as the beliefs of the most accurate trader in the economy.

**Proposition 5.** Under A1-A6, asymptotic prices reflect the, ex post, most accurate beliefs among traders.

*Proof. See Appendix B*

The reason why equilibrium prices reflect accurate beliefs is that in homogeneous discount factor economies, the leading term in the survival index is the accuracy of the most accurate beliefs in the cluster. Thus consumption-shares concentrates on traders whose beliefs lies in a shrinking interval around the most accurate trader in the economy and prices reflect his beliefs, albeit a second order effect due to the long lasting heterogeneity.

In example 3, the intuition goes as follows. The Polya urn process can be equivalently thought of as representing the case in which Nature randomizes at time 0 to decide which iid model to use (i.e. the asymptotic frequency $\hat{p}$). Ex-ante, skilled traders’ beliefs are correct because they know that Nature is choosing the parameter at random ($\lim_{t \to \infty} \frac{p_B}{P} = 1$ P-a.s.); while each unskilled traders incorrectly believe that there is a unique possible parameter ($\forall i \in A_U \lim_{t \to \infty} \frac{p_i}{P} = 0$ P-a.s.). Ex-post, the market selects for the lucky iid traders whose beliefs are close to the realized parameter, i.e. for the traders that have rational expectations, conditionally on the realized value of $\hat{p}$. Thus, even if the MSH fails, asymptotic equilibrium prices reflect beliefs that are empirically correct.

### 5.4 Risk attitudes, heterogeneity and equilibrium prices

**Proposition 6.** Under A1-A5, $\forall \sigma \in \hat{S}$ equilibrium prices satisfies:

$$q(\sigma^t) \approx \max_{j \in N} e^{t \ln \beta_j + \left(\frac{\gamma_j k_{MAK}}{2} - \frac{k_{MAK}^j}{2}\right) \ln t} \times e^{\ln p_j^j(\sigma^t) \ln(\sigma^t) - \frac{k_{BIC}^j}{2} \ln t - \frac{k_{MAK}^j}{2} \ln t}$$

*Proof. See Appendix 16*
If the cluster that dominates has homogeneous beliefs, \( k^{MAK} = 0 \), or log utility \( \gamma = 1 \), then equilibrium prices are asymptotically as accurate as the discounted probabilities of the most accurate cluster. If the cluster that dominates has heterogeneous beliefs and \( \gamma \neq 1 \), then heterogeneity does not vanish fast enough and risk attitudes have a second order effect on the asymptotic discount factor:

\[
e^{-\left(\frac{\gamma_j k^{MAK}}{2} - \frac{k^{MAK}}{2}\right) \ln t}
\]

If the cluster that dominates is less (more) risk averse than log, the inter-temporal discount factor is diminished (increased) by an endogenous component that captures the interaction between the subjective arbitrage opportunity due to heterogeneity of opinions, \( k^{MAK} \), and the effect of \( \gamma \) on inter-temporal consumption choices. This asymptotic effect is consistent with the small sample effect of beliefs heterogeneity on equilibrium prices identified by the growing literature that explains price anomalies via beliefs heterogeneity (among others Cvitanic-Jouini et al. (2012), Jouini-Napp (2007)). My result complement theirs by showing that heterogeneity can indeed survive for long time and have ever lasting effects on equilibrium prices.

6 A general overview

We are now ready to present a general sufficient 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 8.** Cluster’s \( A_{\gamma_j} \) survival index is

\[
s_{\gamma_j} = t \ln \beta_j + \ln \hat{p}_j^{(\sigma^t)}(\sigma^t) - \frac{k^{BIC}_j}{2} \ln t - \frac{\gamma_j k^{MAK}_j}{2} \ln t
\]

The survival index has four terms: \( t \ln \beta_j \) is the discount factor: more patient clusters have higher survival chances because attach higher value future consumption. \( \ln \hat{p}_j^{(\sigma^t)}(\sigma^t) - \frac{k^{BIC}_j}{2} \ln t \) represent the likelihood of the belief of the most accurate trader in the cluster, \( k^{BIC}_j \) is the BIC dimensionality term and appears if traders in \( A_{\gamma_j} \) are identical Bayesian traders with \( k \)-dimensional prior support of positive Lebesgue measure (or learn according to an asymptotically equivalent rule). \( \gamma_j k^{MAK}_j \) is the
dimensionality component found in Lemma 2 and capture the speculative incentive of $A_\gamma_j$. This term is 0 whenever all traders have common beliefs, thus is not present in small economies.

**Theorem 1.** Under A1-A5, cluster $j$ vanishes in $\hat{S}$ if there is a cluster $k$ such that:

\[ s_j - s_k \to -\infty \]

**Proof.** By the FOC:

\[ \frac{(C^{A_\gamma}(\sigma^i))}{(C^{A_\gamma}(\sigma^j))} = \frac{\beta^i \left( \int_{A_\gamma} c_0^i p^i(\sigma^i)^{1/2} \, d\theta \right)^{\eta_i}}{\beta^j \left( \int_{A_\gamma} c_0^j p^j(\sigma^j)^{1/2} \, d\theta \right)^{\eta_j}}. \]

The result follows from Lemma 1 after approximating the RHS using Lemma 2 and the BIC. \qed

Theorem 1 highlights that the survival of a cluster depends on four exogenous components. Keeping the other three components equal, differences in the first components indicate that the least patient cluster vanishes; differences in the second components indicate that a cluster vanishes if its most accurate trader is less accurate than the most accurate trader of another cluster; differences in the third components indicate that among Bayesian learners whose support contain the true probability, the one that has to estimate the larger number of parameter vanishes in probability \(^7\) (as for Blume-Easley (2006), Theorem 6) and differences in the last components indicate that the cluster with the lowest $\gamma_j k$ term dominates, because he decide to optimally save more.

These four components have difference intensity: The first two component diverge at rate $t$, while the last two components diverge at rate $\log t$. Thus differences in the first two components always dominate differences in the second two components.

It follows that risk attitudes can affect survival not only via direct comparison of the last term of the survival indexes (Proposition 1) but also via the interaction between its third and the last component. This is exactly what happens in Section 3: the first two components of the survival indexes are identical ($\beta_U = \beta_B$, $p_U^{i(\sigma)} = p_B^{i(\sigma)}$) and differences in the third and fourth components determines survival: $k_B^{BIC} = 1 \neq 0 = k_U^{BIC}$ and $\gamma k_B^{MAK} = 0 \neq \gamma 1 = \gamma k_U^{MAK}$.

\(^7\)The reason why we can’t get a stronger result here is due to fact that the maximum likelihood probabilities of two nested families cannot be distinguished a.s.
6.1 Small and Large economies

A large economy with finitely many clusters of identical traders is formally equivalent to a small economy, thus the condition of Theorem 1 also applies to this setting. In this case, the beliefs of the most accurate trader in the cluster coincides with trader’s beliefs and the risk/dimensionality component is mute ($k^{MAK}=0$). Thus, consistently with Sandroni (2000) and Blume-Easley (2006) finding, risk attitudes do not play a role on survival:

**Corollary 1.** In a small economy that satisfy A1, A2, A3 and A6, $\forall \sigma \in \hat{S}$, the market selects for the most accurate trader.

*Proof.* By noticing that the first component (by A6) and the last component ($k^{MAK} = 0$ in a small economy) of the survival indexes become mute. 

The different implication of risk attitudes on survival for *large and small economies* can be puzzling. Corollary 1 applies to economies with an arbitrarily large number of traders and yet are not valid in large economies. It turns out that this discontinuity is only apparent, as it is generated by the dichotomic definition of survival, not by a qualitative difference between the two settings. If instead of focusing on 0 Vs positive asymptotic consumption we were focusing on the size of the asymptotic consumption-shares, we would have found no discontinuity between the two settings: in small economies, risk attitudes have an effect on asymptotic consumption-shares that has the same direction as the one found in large economies. The following example illustrates the point.

**Example 4:** Consider an Arrow-Debreu exchange economy with two states $S = \{W, R\}$. Traders 1,...,n have heterogeneous iid beliefs $p^i$, cluster $A_U$, and traders $n+1,...,2n$ are Bayesian traders, cluster $A_B$, with Uniform prior on $A_U = \{\cup_{i=1}^n p^i\}$: $p^b(\sigma^t) := \sum_{i=1}^n \frac{1}{n} p^i(\sigma^t)$. All trader have the same CRRA utility function with parameter $\gamma$ and identical inter-temporal discount factor $\beta$. Let assume that $\forall i, c_0^i = \frac{1}{2n}$ and

---

8According to the definition, $A_U$ is not a cluster because $c_0$ is not continuous. Traders are grouped in this fashion to maintain the parallel with the approach followed in Section 3.
that \( p^1 \) is the most accurate trader. Rearranging the FOC,

\[
\left( \sum_{i \in A_B} c_i(\sigma^t) \right)^\gamma \left( \sum_{i \in A_U} c_i(\sigma^t) \right)^\gamma = \frac{\left( \sum_{i=n+1}^{2n} \frac{1}{2n} p^B(\sigma^t)^{\frac{1}{2}} \right)^\gamma}{\left( \sum_{i=1}^{n} \frac{1}{2n} p^B(\sigma^t)^{\frac{1}{2}} \right)^\gamma} + \frac{\left( \frac{1}{2} \right)^{\gamma} \left( \frac{1}{n^2} + \frac{1}{n^2} \sum_{i=2}^{n} \frac{p^i(\sigma^t)}{p^i(\sigma^t)} \right)^\frac{1}{2}}{\left( \frac{1}{2n^2} + \frac{1}{2n^2} \sum_{i=2}^{n} \frac{p^i(\sigma^t)}{p^i(\sigma^t)} \right)^\frac{1}{2}} \xrightarrow{P-a.s.} n(\gamma - 1),
\]

Which shows that (i) risk attitudes have an effect on the asymptotic consumption share of the Bayesian trader and (ii) for \( \gamma > 0 \) and \( n < \infty \), this effect is not strong enough to be detected by the definition of survival but has the same direction as Proposition 2.\(^{10}\)

The reason is intuitive: the consumption-share of the trader with correct beliefs initially grow faster than the consumption-share of the Bayesian traders because the beliefs of the Bayesian traders are initially incorrect. Nevertheless, as the beliefs of the Bayesian concentrate around the true probability, this difference disappears. Risk attitudes have an effect on asymptotic consumption-shares because determine how fast consumption-shares move: the lower the gamma, the faster consumption-shares move and the lower will be the asymptotic consumption-shares of the Bayesians.

The cardinality of \( A \) has affects survival through its effect on the convergence rates of the Bayesian posterior and of the wealth-shares: if \( |A| < |\mathbb{R}| \), both convergence rates are exponential, thus, in finitely many periods, the Bayesian learns the true probability and heterogeneity disappears from the market. While, if \( |A| = |\mathbb{R}| \), both convergence rates are slower than exponential (they are respectively \( O\left( \frac{1}{\sqrt{k_{\text{BIC}}} t} \right) \) and \( O\left( \frac{1}{\sqrt{k_{MAK}} t} \right) \)); in this case the market tolerates a mild form of long run beliefs heterogeneity and the posterior never exactly concentrates on the true probability.

### 7 Conclusions

This paper extends the project started by Sandroni (2000) and Blume-Easley (2006) on market selection in general equilibrium complete markets to the large economy setting.

\(^{10}\)Letting \( n \to \infty \), the example suggests that an economy with countably many traders can behave like a large economy, albeit complications on the way to properly construct and interpret this limit.
We show that large economies are qualitatively different from small economies in that risk attitudes do play a role on survival. It follows that contrary to the standard result, markets can fail to identify the traders with correct beliefs. This failure of the MSH is qualitatively different from all of the other cases found in the literature in that it does not invalidate Freedman’s conjecture that the selection forces in the market support the adoption of rational expectations. It turns out that equilibrium prices can be asymptotically correct even if the market selects against traders with correct beliefs. Our result shows that risk attitudes and aggregation affect investment decisions in a non-trivial way even when traders optimize on saving decisions and allocations decisions at the same time. Moreover, our setting allows to discuss the role played by luck in financial markets and its relation with risk attitudes. In particular we find cases in which to be lucky is the only way to survive.

A Exchangeability and De Finetti’s theorem

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

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

**Definition 9.** An infinite sequence of realization $\sigma^\infty$ is exchangeable if, for every finite $t$, $P(\sigma_1, ..., \sigma_t) = P(\sigma_{\pi(1)}, ..., \sigma_{\pi(t)})$ for any permutation $\pi$ of the indices.

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

**De Finetti’s Theorem.** A probability distribution $P$ on $\Sigma$ is exchangeable if and only if $P$ is a mixture of iid distributions $(Q)$: $P(A) = \int Q^\infty(A) \mu(dQ)$, for some probability distribution $\mu$ on the space of all probability distributions on $S$.
For an intuition of the relationship between exchangeable and iid processes consider these examples of Polya urn processes. Suppose we have an urn that contains \( N \) balls with a certain composition of Black balls and White balls. (i) Sampling from the urn with replacement is an iid, hence exchangeable process. (ii) Sampling from the urn, replacing each ball extracted with \((n>1)\) balls of the same color is exchangeable, not iid, because the probability of an outcome depends on the previous outcomes, and, by De Finetti’s Theorem, there is a mixture of iid distribution that coincides with this model. (iii) Sampling from the urn without replacement is an exchangeable process that is not iid, but De Finetti’s Theorem does not apply because the process cannot generate infinite sequences.

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

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

B Appendix

In this appendix we make use of the notation \( o(.) \) and \( \approx \) with the following meanings. \( f(x) = o(g(x)) \), abbreviates \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \), while \( f(x) \approx g(x) \), is used, not conventionally, to abbreviate \( \lim_{x \to \infty} \frac{f(x)}{g(x)} \in (0, +\infty) \).

Proof of Lemma 2

Proof. First: \( \left( \int_{A_{\gamma_j}} p_i^{(\sigma)} \hat{c}_j d\sigma \right)^{\gamma_j} \approx e^{\ln p_i^{(\sigma)}(\sigma) - \frac{\gamma_j M A K}{2} \ln \epsilon} \). It follows from Lemma 4 by substituting \( A_{\gamma_j} \) for \( A \), multiplying by \( \gamma_j \), exponentiating and ignoring the constant terms.
Second: if $P \in A_{\gamma}$, then $\hat{p}(\sigma_t) \rightarrow^{P.a.s.} P(\sigma_t | \sigma_{t-1})$ by consistency of the maximum likelihood estimator.

The proof of Lemmas 3 and 4 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). We assume that $\mathcal{M}$ is the iid Bernoulli model. The generalization to other members of the exponential family is straightforward.

**Lemma 3.** Let $\mathcal{M}$ be a member of the exponential family parametrized by $A$ and $c_0$ a function that satisfies A5, then:

$$\ln \int_A p^i(\sigma^t)^t c_0^i di = \ln \int_A p^i(\sigma^t)^t c_0^i di + \ln p^i(\sigma^t)(\sigma^t)$$

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

**Proof.**

$$\ln \int_A p^i(\sigma^t)^t c_0^i di = \ln \int_A p^i(\sigma^t)^t c_0^i di + \ln p^i(\sigma^t)(\sigma^t)^t - \ln p^i(\sigma^t)(\sigma^t)^t$$

$$= \ln \int_A \frac{p^i(\sigma^t)^t}{p^i(\sigma^t)(\sigma^t)^t} c_0^i di + \ln p^i(\sigma^t)(\sigma^t)^t$$

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

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

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

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

**Lemma 4.** Let $\mathcal{M}$ be a member of the exponential family parametrized by $A$ and $c_0$ a function that satisfies A5, then:

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

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

**Proof.** By Lemma 3

$$\ln \int_A p^i(\sigma^t)^t c_0^i di = \ln \int_A e^{-1} D(\hat{p}(\sigma^t)||p^i) c_0^i di + \frac{1}{\gamma} \ln p^i(\sigma^t)(\sigma^t)$$
For $0 < \alpha < \frac{1}{2}$ let $B_t = \{ i : p^i \in [p^{i(\sigma^t)} - t^{1+\alpha}, p^{i(\sigma^t)} + t^{1+\alpha}] \}$. By additivity of the integral:

$$
\int_A e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di = \int_{A \setminus B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di + \int_{B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di
$$

Since $c^i_0$ is continuous on $A$ and strictly positive in $int(A)$ there is a $T$, such that $\forall t > T c^i_0 > 0, \forall i \in B_t$. In what follows we always assume $t > T$. The proof is done by performing a second order Taylor expansion of $D(p^{i(\sigma^t)} || p^i)$ to bound the two integrals. $M$ is a member of the exponential family, thus, by the results in Chapter 19 of Grünwald, $D(p^i || P)$ can be well approximated in $B$ as follows

$$
D(p^{i(\sigma^t)} || p^i) = \frac{1}{2} \left( p^{i(\sigma^t)} - p^i \right)^2 I(p^i) \tag{3}
$$

for some $i^* \in B_t$ such that $p^{i^*}$ lies between $p^i$ and $p^{i(\sigma^t)}$.

† **First integral:** $\exists k, a < \infty : \mathcal{I}_1 = \int_{A \setminus B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di < ke^{-at^{2\alpha}} \to 0$

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

$$
0 < \int_{A \setminus B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di < \int_{A \setminus B_t} e^{-\frac{\alpha}{2} \min_{i \in A \setminus B_t} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di
$$

By Equation 3

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

so that, since $I(p^i)$ is continuous and $> 0$ for all $i \in A$, and $\int_{A \setminus B_t} c^i_0 \, di < \infty$,

$$
0 < \int_{A \setminus B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di < \int_{A \setminus B_t} e^{-\frac{\alpha}{2} \left( \frac{1}{2} t^{1+2\alpha} \min_{i \in int(A)} I(p^i) \right) c^i_0 \, di} < ke^{-at^{2\alpha}}
$$

For $a = \min_{i \in A \setminus B_t} \frac{I(p^i)}{2^\gamma} > 0$ and $k = \int_{A \setminus B_t} c^i_0 \, di < \int_A c^i_0 \, di < \infty$.

‡ **Second integral:** $\mathcal{I}_2 = \int_{B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di \approx \frac{2\pi c^i_0}{\sqrt{I(p^i)}}$

Let $I_t^- := \inf_{i \in B_t} I(p^i), I_t^+ := \sup_{i \in B_t} I(p^i), c_t^- := \inf_{i \in B_t} c^i_0, c_t^+ := \sup_{i \in B_t} c^i_0$, by Equation 3

$$
\mathcal{I}_2 = \int_{B_t} e^{-\frac{\alpha}{2} D(p^{i(\sigma^t)} || p^i)} c^i_0 \, di = \int_{B_t} e^{-\frac{\alpha}{2} (p^{i(\sigma^t)} - p^i)^2 I(i)} c^i_0 \, di
$$

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

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

24
We now perform the substitutions $z := (p^i(\sigma^t) - p^i) \sqrt{t \frac{I_t \gamma}{\sigma}}$ on the left integral and $z := (p^i(\sigma^t) - p^i) \sqrt{t \frac{I_t \gamma}{\sigma}}$ on the right integral, to get

\[
\frac{c^i_t}{\sqrt{t \frac{I_t \gamma}{\sigma}}} \int_{|z| < t \sigma \sqrt{I_t}} e^{-\frac{z^2}{2}} dz \leq \mathcal{I}_2 \leq \frac{c^i_t}{\sqrt{t \frac{I_t \gamma}{\sigma}}} \int_{|z| < t \sigma \sqrt{I_t}} e^{-\frac{z^2}{2}} dz
\]

We now recognize both integrals as standard Gaussian.

Since, as $t \to \infty$, $I_t^- \to I(p^i)$ and $I_t^+ \to I(p^i)$, the domain of integration tends to infinity for both integrals, so that they both converge to $\sqrt{2\pi}$. Since $c^i_t \to c_0^i$ and $c_t^- \to c_t^0$, the constant in both integrals converges to $\frac{c_0^i}{\sqrt{t \frac{I_t \gamma}{\sigma}}}$ and we get $\mathcal{I}_2 \approx \frac{\sqrt{2\pi}c_0^i}{\sqrt{t \frac{I_t \gamma}{\sigma}}}$.

Putting † and ‡ together:

\[
\ln \int_A p^i(\sigma^t) e^{c_0^i} di = \ln (\mathcal{I}_1 + \mathcal{I}_2) + \frac{1}{\gamma} \ln p^i(\sigma^t) (\sigma^t) \\
= -\frac{1}{\gamma} \ln p^i(\sigma^t) (\sigma^t) + \ln \sqrt{\gamma} + \ln c_0^i - \frac{1}{2} \ln \frac{t}{2\pi} - \ln \sqrt{\det I(p^i)} + o(1)
\]

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

**Proof of Proposition 4**

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

\[
\lim_{t \to \infty} \int_{i \in A \setminus B_t} c^i(\sigma^t) di = \lim_{t \to \infty} \int_{i \in A \setminus B_t} e^{-\frac{1}{2} D(p^i(\sigma^t)||p^i)} w(i) di \\
= \lim_{t \to \infty} \int_{B_t} e^{-\frac{1}{2} D(p^i(\sigma^t)||p^i)} w(i) di \to 0
\]

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

**Proof of Proposition 5**

Proof. If the MSH holds, the result follows from standard arguments. If it does not hold, by Proposition 4, consumption-shares concentrates on traders whose beliefs are, in every $\sigma^t$, close to the belief with maximal empirical likelihood: $p^i(\sigma^t)$. Thus, by standard arguments, prices reflects beliefs that becomes arbitrarily close to $p^i(\sigma^t)$. \qed
**Proof of Proposition 6**

Proof. We will start by showing that \( \forall \sigma \in \Sigma, q(\sigma') = \max_{j \in \mathbb{N}} \left\{ \beta_j^t \left( \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau} \right)^\gamma \right\} \).

By the FOC, \( \forall j: (C_j^t(\sigma'))^\gamma = \frac{\beta_j^t \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau}}{q(\sigma')}^\gamma \).

The proof is done by contradiction: we have to consider two possible cases:

\( i \) \( \exists j, \exists \sigma^t : \frac{\beta^t \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau}}{q(\sigma')} \to \infty \). By the FOC, \( C_j^t(\sigma') = \frac{\beta^t \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau}}{q(\sigma')} \to \infty \) violating the bounded aggregate endowment assumption (A2).

\( ii \) \( \exists \sigma^t : \forall j = 1, \ldots, n; \frac{\beta^t \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau}}{q(\sigma')} \to 0 \). By the FOC, \( \forall j = 1, \ldots, n; C_j^t(\sigma') = \frac{\beta^t \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau}}{q(\sigma')} \to 0 \) violating the positive aggregate endowment assumption (A2).

The result follows using Lemma2 to approximate \( \max_{j \in \mathbb{N}} \left\{ \beta_j^t \left( \int_{A_j} c_0^t p^t(\sigma') \frac{dt}{\tau} \right)^\gamma \right\} \) on \( \hat{S} \) and rearranging. \( \square \)

**C Appendix**

Let \( \hat{p}_{\sigma^t} = \frac{1}{t} \sum_{i=1}^{t} I_{\sigma_i=R} \) be the frequency of Red balls on \( \sigma^t \).

**Lemma 5.** Assume that \( c_0^t \) is uniform in \( A_U \), then for all \( \sigma \in \hat{S} \), \( \gamma < 1 \Rightarrow A_B \) vanishes.

Proof. All traders in \( A_B \) have identical beliefs \( P \), aggregating the FOC as for Eq.2,

\[
\frac{\int_{A_B} c^t(\sigma')d\tau}{\int_{A_U} c^t(\sigma')d\tau} = \frac{P(\sigma^t)^{\frac{1}{\tau}}}{{\int_{A_U} c_0^t p^t(\sigma') \frac{dt}{\tau}}} = \frac{P(\sigma^t)^{\frac{1}{\tau}}}{{\int_{A_U} p^t(\sigma^t) \frac{dt}{\tau}}}.
\]

By Lemma 1, \( A_U \) vanishes if Eq.4 converges to 0. By Lemmas 6 and 7,

\[
\lim_{t \to \infty} \frac{P(\sigma^t)^{\frac{1}{\tau}}}{{\int_{A_U} p^t(\sigma^t) \frac{dt}{\tau}}} = \lim_{t \to \infty} \frac{\left(\hat{p}_{\sigma^t}(1-\hat{p}_{\sigma^t})^{1-\hat{p}_{\sigma^t}}\right)^{\frac{1}{\sqrt{t}}}}{\left(\hat{p}_{\sigma^t}(1-\hat{p}_{\sigma^t})^{1-\hat{p}_{\sigma^t}}\right)^{\frac{1}{\sqrt{t}}}} = \lim_{t \to \infty} \sqrt{t^{1-\frac{1}{\sqrt{t}}}} \to 0 \quad \text{for} \ \gamma < 1
\]

\( \square \)

**Lemma 6.** For the Polya urn process described, for every \( \sigma \in \hat{S} \), \( P(\sigma^t) \approx p^t (1 - \hat{p}_{\sigma^t})^{(1-\hat{p}_{\sigma^t})^t} \frac{1}{\sqrt{t}} \).
Proof.

\[ P(\sigma^t) = \frac{a(\hat{p}_{\sigma^t})!((1 - \hat{p}_{\sigma^t})t)!}{(t+1)*t!} \]

\[ \approx b,c \frac{1}{2} e^{-\hat{p}_{\sigma^t} t} (\hat{p}_{\sigma^t} t)^{\hat{p}_{\sigma^t} t} \frac{1}{2} e^{-((1 - \hat{p}_{\sigma^t})t) t} ((1 - \hat{p}_{\sigma^t} t) t)^{(1 - \hat{p}_{\sigma^t} t) t} + \frac{1}{2} e^{-t} t^t \approx c \frac{1}{\sqrt{t}} \]

a) It follows noticing that:
   i) \( \hat{p}_{\sigma^t} t \) is the number of realizations of \( R \) (integer by construction);
   ii) by exchangeability, permutations of the indices do not change the probability of finite sequences, thus \( \forall \sigma^t, P(\sigma^t) = P(\sigma^t_R) \) where \( \sigma^t \) is a generic sequence of length \( t \) with \( \hat{p}_{\sigma^t} t \) realizations of state \( R \) and \( \sigma^t_R = \{R_1, ..., R_{\hat{p}_{\sigma^t} t}, W_{\hat{p}_{\sigma^t} t+1}, ..., W_t\} \) is the sequence of length \( t \) in which state \( R \) appears only in the first \( \hat{p}_{\sigma^t} t \) entries;
   iii) Applying the iterative formula for the Polya urn composition:

\[ P(\sigma^t) = (\hat{p}_{\sigma^t} t)!((1 - \hat{p}_{\sigma^t} t) t)! \]

b) By Sterling’s approximation, \( \forall x \in \mathbb{N}, x! \in \left[ \sqrt{2\pi x}^{x+\frac{1}{2}} e^{-x}, x^{x+\frac{1}{2}} e^{-x+1} \right] \).

c) Disregarding the finite constants.

Lemma 7. For every \( \sigma \in \hat{S}, \int_0^1 \hat{p}_{\sigma^t} (\sigma^t)^{\frac{1}{2}} di \approx \left( \frac{\hat{p}_{\sigma^t} t}{(1 - \hat{p}_{\sigma^t}) t} \right)^{\frac{1}{2}} \frac{1}{\sqrt{t}}. \)
Proof. By definition, \( \hat{p}_{\sigma_t} t = \sum_{\tau=1}^{t} I_{\sigma_\tau = R} \) and \( p'(\sigma^t) = i\hat{p}_{\sigma_t} (1 - i)^{(1 - \hat{p}_{\sigma_t})t} \), thus

\[
\int_{0}^{1} \hat{p}_{\sigma_t} \cdot (\sigma^t)^{\frac{\gamma}{\gamma}} di = \int_{0}^{1} i \hat{p}_{\sigma_t} \cdot (1 - i)^{(1 - \hat{p}_{\sigma_t})t} di
\]

\[
= B \left( \frac{\hat{p}_{\sigma_t} t}{\gamma} + 1, \frac{(1 - \hat{p}_{\sigma_t})t}{\gamma} + 1 \right) \quad \text{by definition of Beta function}
\]

\[
= \frac{\Gamma \left( \frac{\hat{p}_{\sigma_t} t}{\gamma} + 1 \right)}{\Gamma \left( \frac{1}{\gamma} + 2 \right)} \quad \text{expressing the Beta function as a ratio of Gamma functions}
\]

\[
= a \quad \frac{1}{\frac{t}{\gamma} + 1} \quad \frac{\left( \frac{\hat{p}_{\sigma_t} t}{\gamma} \right)! \cdot \left( \frac{1 - \hat{p}_{\sigma_t} t}{\gamma} \right)!}{\left( \frac{t}{\gamma} \right)!} \quad \Gamma \left( \frac{1}{\gamma} \right) \Gamma \left( \frac{1}{\gamma} - \frac{1}{\gamma} \right) \Gamma \left( \frac{1}{\gamma} - \frac{2}{\gamma} \right) \Gamma \left( \frac{1}{\gamma} - \frac{3}{\gamma} \right) \Gamma \left( \frac{1}{\gamma} - \frac{4}{\gamma} \right)
\]

\[
\approx 1 \quad \frac{\left( \frac{\hat{p}_{\sigma_t} t}{\gamma} \right)! \cdot \left( \frac{1 - \hat{p}_{\sigma_t} t}{\gamma} \right)!}{\left( \frac{t}{\gamma} \right)!} \quad \text{because} \quad \frac{\Gamma \left( \frac{\hat{p}_{\sigma_t} t}{\gamma} - \frac{1}{\gamma} \right)}{\Gamma \left( \frac{t}{\gamma} - \frac{1}{\gamma} \right)} \frac{\Gamma \left( \frac{1 - \hat{p}_{\sigma_t} t}{\gamma} - \frac{1}{\gamma} \right)}{\Gamma \left( \frac{1}{\gamma} - \frac{1}{\gamma} \right)} \text{is finite}
\]

\[
\approx \left( \hat{p}_{\sigma_t} (1 - \hat{p}_{\sigma_t}) t^{\frac{1}{\gamma}} \right) \frac{1}{\sqrt{t}} \quad \text{using Sterlings approximation as in Lem.6}
\]

\( a \) where, for every number \( x \), \( \lfloor x \rfloor \) is the largest integer smaller than \( x \). \( \square \)

References


