Price Probabilities: a class of Bayesian and non-Bayesian prediction rules

Filippo Massari*

University of New South Wales

November 7, 2019

Abstract

This paper examines the implications of the market selection hypothesis on the accuracy of the probabilities implied by equilibrium prices and on the "learning" mechanism of markets. I use the standard machinery of dynamic general equilibrium models to generate a rich class of probabilities, *price probabilities*, and discuss their properties. This class includes the Bayes' rule and known non-Bayesian rules. If the prior support is well-specified, I prove that all members of this class perform as well as Bayes' rule in terms of likelihood. If the prior support is misspecified in that the bayesian prior does not converge, I demonstrate that some members of price probabilities significantly outperform Bayes' and characterize them. Because these members are never worse and sometimes better than Bayes, my result challenges the prevailing opinion that Bayes' rule is the only rational way to learn.

Keywords: Non-Bayesian Learning, NML, Safe Bayesian, Prediction Market.

JEL Classification: C53,D70, D81, D83

1 Introduction

It has long been argued that financial markets aggregate the different opinions of their participants efficiently. One explanation is that the market "learns" over time because selection

*I am grateful for the comments and encouragement of Nabil Al-Najjar, Aloisio Araujo, Ken Binmore, Itzhak Gilboa, Simon Grant, Peter Grünwald, Massimo Marinacci, Alvaro Sandroni, three anonymous referees and the seminar participants at WUSTL, CWI, the 2015 World Congress, IMPA, the 2015 European Econometric Society Meeting at U.Bocconi, UNSW, U.Sydney and UTAS.

forces push equilibrium prices to reflect the beliefs of the most accurate trader in the market (*Market Selection Hypothesis*, Friedman, 1953; Sandroni, 2000; Blume and Easley, 2006). This view agrees with the observation that in general equilibrium models with complete markets and equally patient traders having log-utilities, the equilibrium dynamics of the state price densities coincide with the dynamic of the Bayesian posterior calculated from a prior on the set of trader beliefs (Rubinstein, 1974; Blume and Easley, 1993). In this paper, I relax the log-utility assumption and focus on the accuracy of the resulting state price densities.

Price Probability is the class of all probabilities that can be represented as state-price densities of an economy with complete markets, no aggregate risk, and in which the market selection hypothesis holds. This class is rich: Corollary 1, and 2 show that it includes Bayes' rule (BMA)[1] as well as known non-Bayesian rules such as the Normalized Maximum Likelihood (NML) (Rissanen, 1986; Shtar'kov, 1987; Grünwald, 2007) and the Sequential Normalized Maximum Likelihood (SNML) (Roos and Rissanen, 2008).

I find that most members of price probabilities are fundamentally not Bayesian because they do not make convex predictions — i.e., the predictive probability might not be a convex combination of the models in the support (see section 6) — and furthermore, some of its members are time-inconsistent.

Given the overwhelming experimental evidence showing that most agents are not Bayesian (Rabin et al., 2000; Kahneman, 2011), it is natural to ask if non-Bayesian members of the price probability class constitute a "rational" alternative to Bayes' rule. However, what does it mean to be "rational"? Contrary to the axiomatic approach to learning in the economics literature, according to which a prediction rule is rational if consistent with a set of axioms (Ghirardato, 2002; Gilboa and Marinacci, 2011); I take a point of view closer to machine learning (Breiman et al., 2001; Sutton and Barto, 1998) and propose a pragmatic notion of rationality. A prediction rule is pragmatically rational if it guarantees good predictions irrespective of the true probability. I consider these points of view as complementary. The former is appropriate in situations in which an agent is not subject to an external criterion of performance. In this case, a set of axioms can jointly determine an agent's preferences and beliefs. The latter is appropriate for cases in which agent decisions are evaluated according to an external criterion of performance (e.g., Sharpe ratio for portfolio managers, calibration for weather forecasters). Because the criterion pins down agent preferences, a pragmatic agent should internalize this constraint in his decision problem and choose a prediction rule that is optimal for his preferences. These two points of view coincide in well-specified learning problems, but might and often differ in misspecified learning problems.

[1]Bayesian Model Averaging: Hoeting et al. (1999)

Taking advantage of the almost universal predominance of Bayes' rule and its sound axiomatic foundation, I use BMA as a benchmark and propose an accuracy criterion for prediction rules based on likelihood comparisons. I find that if the learning problem is correctly specified all members of price probability are as good as Bayes' because they converge to the truth almost surely at a comparable rate (merge). I call this property $\mathcal{P}$-efficiency.

However, facing a well-specified learning problem is the exception, rather than the rule, for many practical prediction problems (predicting stock market returns, weather, outcomes of sport events...). So, it is pragmatically relevant to compare the performance of prediction rules also under model misspecification. I identify the following categories.

- A probability mixture, p, is *super-efficient* if the log-likelihood ratio between BMA and p is bounded above in every sequence, but there are probabilities, $\hat{P}$, such that it diverges to negative infinity $\hat{P}$-almost surely. That is, p is *super-efficient* if p and BMA use the same prior information, there are no sequences in which BMA is overwhelmingly more accurate than p, and there are cases of misspecification, in which p is overwhelmingly more accurate than BMA.
- A probability mixture, p is *universal-efficient* if the log-likelihood ratio between BMA and p is bounded above and below, in every sequence — that is, if p is qualitatively as accurate as BMA in every sequence.
- A probability mixture, p is *sub-efficient* if the log-likelihood ratio between BMA and p is bounded above almost surely when the model is well-specified, but there cases of misspecification such that it diverges to infinity.

The universal-efficient members of price probabilities are time-inconsistent (except for BMA), thus rational, in all settings in which time inconsistency cannot be used to construct arbitrages.

Time-consistent members of price probability (except for BMA) can either be sub-efficient or super-efficient, depending on the risk attitudes of the agents in the generating economy. The super-efficient members of price probability are generated by economies in which all agents have CRRA utility with parameter $\gamma^i > 1$. In these economies, consumption shares move slower than in log economies which determine a slower convergence toward the model with the maximum likelihood probability. Although non-convex, they share important similarities with known robust algorithms in Computer Science and Game Theory (HEDGE algorithm by Freund and Schapire (1997); Safe Bayesian by Grünwald (2012); and Smooth Fictitious Player by Fudenberg and Levine (1998)). These algorithms show that if the loss (utility) function differs from log-likelihood, an agent can be better off abandoning Bayes' rule for a rule that underreacts to information by giving less weight to past realizations. My result indicates that Baye's rule can be improved also in terms of log-likelihood.

Conversely, the sub-efficient members of price probability are generated by economies in which all agents have CRRA utility with parameter $\gamma^i < 1$. In these economies, consumption shares move more slowly than in log economies which determine a slower convergence toward the model with the maximum likelihood probability. Although non-convex, they share similarities with the algorithms above-mentioned, albeit with parameters that would make them sub-optimally overreact to information.

Section 2 introduces the notation and known probability mixture models. Sections 3 and 4 are about the economic derivation of price probabilities. A reader who is not interested in the economic derivation of Price Probabilities can skip these sections, limit his attention to members of price probability with analytical form and consider Propositions 2 and 3 as definitions. Section 5 discusses the relative accuracy of members of the price probability, while Section 6 shows that most price probabilities are not Bayesian.

2 Environment

Time is discrete and begins at date 0. At each date, a random variable (the economy) can be in S mutually exclusive states, $\mathcal{S} := \{1, ..., S\}$, with a Cartesian product $\mathcal{S}^t = \times^t \mathcal{S}$. The set of all infinite sequences of states is $\mathcal{S}^\infty := \times^\infty \mathcal{S}$, with a representative path, $\sigma = (\sigma_1, ...)$. $\sigma^t = (\sigma_1, ..., \sigma_t)$ denotes the partial history until period t and (σ^{t-1}, σ_t) is the concatenation of σ^{t-1} and σ_t, i.e. the sequence whose first t-1 realizations coincide with σ^{t-1} and last element is σ_t. $\mathcal{C}(\sigma^t)$ is the cylinder set with base σ^t, $\mathcal{C}(\sigma^t) = \{\sigma \in S^\infty | \, \sigma = (\sigma^t, \ldots)\}$, $\mathcal{F}_t$ the σ-algebra generated by the cylinders, $\mathcal{F}_t = \sigma\left(\mathcal{C}(\sigma^t), \forall \sigma^t \in S^t\}\right)$, and $\mathcal{F}$ is the σ-algebra generated by their union, $\mathcal{F} = \sigma\left(\cup^\infty \mathcal{F}_t\right)$. By construction $\{\mathcal{F}_t\}$ is a filtration. For the sake of notation, I assume that past realizations constitute all of the relevant information, i.e. $\mathcal{F}_t := \sigma^t$. In what follows, all variables with index t are assumed to be measurable according to the natural filtration $\mathcal{F}_t$.

2.1 Probability mixture models

This section gives a brief overview of the definition probability mixture and describes known probability mixture models that belong to the price probability class. These mixture models have been derived independently and with different objectives in mind. My framework is the first one to encompass all of them at once. I refer the reader to Foster and Vohra (1999) and Grünwald (2007) for a more comprehensive discussion. Let us start with the instrumental definition of

Log-regret: Given a partial history σ^t and a reference set of probabilities $\mathcal{P}$, the log-regret is the log-likelihood ratio between the model in $\mathcal{P}$ with the highest likelihood on σ^t (i.e. the most accurate model in $\mathcal{P}$ with hindsight) and the probability mixture adopted: given σ^t, $R(p;\sigma^t) = \max_{i\in\mathcal{P}}\{\ln\frac{p^i(\sigma^t)}{p(\sigma^t)}\}$. Log-regret is a measure of how well a probability mixture performs vis a vis the most accurate model in $\mathcal{P}$ with hindsight of the realized sequence. Different sequences have different log-regrets. To avoid this dependence, it is customary to focus on the worst-case log-regret — which means on the log-regret calculated on the least favorable sequence of realizations: $\mathcal{R}(p;t) = \max_{\sigma^t} R(p;\sigma^t)$. A probability mixture with a small worst-case log-regret is desirable because in every sequence it is almost as accurate as the most accurate model in $\mathcal{P}$ with hindsight.

Here, as well as in the rest of the paper, I assume

A0: The set of probabilistic models $\mathcal{P} := \{p^1, ..., p^I\}$ is finite.

- **Probability mixture.** Given a reference set of probability measures $\mathcal{P}$ on $\mathcal{F}$, a probability mixture is any function that combines members of $\mathcal{P}$ to deliver a sequence of probabilities $\{p_t\}_{t=1}^{\infty}$. If the probability mixture can be calculated recursively, its definition coincides with Dawid (1984)'s definition of a statistical forecasting scheme. Otherwise, it represents a sequence of probability assessments.

- **BMA**: Bayesian Model Averaging is considered the "gold standard" among all probability mixtures. Given a Bayesian prior distribution C_0 on a set of probabilities $\mathcal{P}$, BMA directly follows from Bayes' rule:

$$\forall\sigma^t,\ \ p^{BMA}(\sigma^t) = \sum_{i\in\mathcal{P}} p^i(\sigma^t)c_0^i \quad ; \quad p^{BMA}(\sigma_t|\sigma^{t-1}) = \sum_{i\in\mathcal{P}} p^i(\sigma_t|\sigma^{t-1})c_{t-1}^i(\sigma) \tag{1}$$

  Where $c_{t-1}^i(\sigma) = \frac{p^i(\sigma^{t-1})c_0^i}{\sum_{i\in\mathcal{P}} p^i(\sigma^{t-1})c_0^i}$ are the weights of the prior distribution obtained via Bayes' rule from C_0.[2] The prominence of BMA is due to its sound axiomatic foundation, its good predictive performance, and its tractability. BMA is directly implied by Kolmogorov (1933)'s axioms (adopting the standard definition of conditional probability $p^{BMA}(\sigma_t|\sigma^{t-1}) := \frac{p^{BMA}(\sigma^t)}{p^{BMA}(\sigma^{t-1})}$), and it is compatible with Savage (1954)'s axioms (Ghirardato, 2002). Moreover, BMA is consistent — if the true probability belongs to $\mathcal{P}$, BMA's predictions converge to it—, it has bounded worst-case log-regret (if $|\mathcal{P}|$ is finite), and it can be calculated recursively.

- **NML**: Normalized Maximum Likelihood is the probability mixture constructed to minimize the maximal worst-case log-regret at any horizon: $p^{NML}(.) := \arg\min \mathcal{R}(p;t)$. Ris-

[2]The unusual notation "$c_{t-1}^i(\sigma)$" for the weights of the prior distribution is to ease the comparison between consumption shares and probabilistic mass. In log-economies, they coincide (Section 4.2).

sanen (1986) and Shtar'kov (1987) independently showed that:

$$\forall\sigma^t, \quad p^{NML}(\sigma^t) = \frac{\max_{i\in\mathcal{P}} p^i(\sigma^t)}{\sum_{\hat{\sigma}^t} \max_{i\in\mathcal{P}} p^i(\hat{\sigma}^t)} \quad ; \quad p^{NML}(\sigma_t|\sigma^{t-1}): \text{ not defined} \tag{2}$$

NML has bounded worst-case log-regret (if $|\mathcal{P}|$ is finite), which makes it desirable on data compression tasks. However, NML is hardly used in prediction tasks because it cannot be calculated recursively since p^{NML} is time-inconsistent across periods: $\sum_{\sigma_t} p^{NML}(\sigma^{t-1}, \sigma_t) \neq p^{NML}(\sigma^{t-1})$. p^{NML} is time-inconsistent because it defines a sequence of unconditional probabilities that do not satisfy the chain-rule. Thus, it does not uniquely define a set of conditional probabilities.

- **SNML**: Sequential Normalized Maximum Likelihood is the probability mixture that, in every period, prescribes using the model in $\mathcal{P}$ that had the highest likelihood in the past. SNML was derived by Roos and Rissanen (2008) to obtain a recursive version of NML, and later applied to the problem of optimal portfolio allocation (Follow the Leader strategy, De Rooij et al., 2014; Massari, 2017). SNML's period t predictions coincide with the conditional probabilities that NML gives to σ_t, assuming that t is the final horizon:

$$\forall\sigma^t,\ p^{SNML}(\sigma^t) = \prod_{\tau=1}^{t} p^{SNML}(\sigma_t|\sigma^{t-1}) \ \ ; \ \ p^{SNML}(\sigma_t|\sigma^{t-1}) = \frac{p^{NML}(\sigma^t)}{\sum_{\sigma_t} p^{NML}(\sigma^{t-1}, \sigma_t)} \tag{3}$$

SNML is consistent and it can be calculated recursively. However, unlike NML, SNML's worst-case regret is unbounded even if the cardinality of $\mathcal{P}$ is finite.

3 Price probabilities

In this section, I introduce the economic setting I use to define price probabilities. Consider an Arrow-Debreu exchange economy with complete markets. The economy contains a finite set of traders $\mathcal{I}$. Each trader, i, has consumption set $\mathbb{R}_+$. A consumption plan $c : S^\infty \to \prod_{t=0}^{\infty} \mathbb{R}_+$ is a sequence of $\mathbb{R}_+$- valued functions $\{c_t(\sigma)\}_{t=0}^{\infty}$. Each trader i is characterized by a payoff function $u^i : \mathbb{R}_+ \to \mathbb{R}$ over consumption, a discount factor $\beta_i \in (0,1)$, and an endowment stream $\{e_t^i(\sigma)\}_{t=0}^{\infty}$. Each trader has a subjective probability p^i on $\mathcal{F}$, his beliefs. I denote the set of trader beliefs by $\mathcal{P} := \{p^i : i \in \mathcal{I}\}$. Beliefs are orthogonal; for example, $\mathcal{P}$ can be a set of distinct i.i.d. measures. Each trader, i, aims to solve:

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

where $q(\sigma^t)$ is the price of a claim that pays a unit of consumption on the last realization of σ^t, in terms of consumption at time zero. Let $q(\sigma_t|\sigma^{t-1})$ be the price of a claim that pays a unit of consumption at period/event σ_t, in terms of consumption at period/event σ^{t-1}. It is worth noting the analogy between the equilibrium relation of time-zero and next-period prices, $q(\sigma_t|\sigma^{t-1}) = \frac{q(\sigma^t)}{q(\sigma^{t-1})}$ (Ljungqvist and Sargent, 2004), and the way unconditional and conditional probabilities are linked, $p(\sigma_t|\sigma^{t-1}) = \frac{p(\sigma^t)}{p(\sigma^{t-1})}$. If the sum of next-period prices were 1, equilibrium prices would define a standard probability measure.

3.1 Assumptions

A competitive equilibrium is a sequence of prices and, for each trader, a consumption plan that is affordable, preference maximal on the budget set, and mutually feasible. Assumptions **A0-A5** are sufficient for the existence of the competitive equilibrium (Peleg and Yaari, 1970) and for the market selection hypothesis to hold (Sandroni, 2000; Blume and Easley, 2006).

- **A0** : The set of traders is finite.
- **A1** : The payoff functions $u^i : \mathbb{R}_+ \to [-\infty, +\infty]$ are C^1, concave, strictly increasing, and satisfy the Inada condition at 0 — that is, $u^i(c)' \to \infty$ as $c \searrow 0$.
- **A2** : For all traders i, j, and for all finite sequences σ^t, $p^i(\sigma^t) > 0 \Leftrightarrow p^j(\sigma^t) > 0$.
- **A3** : The aggregate endowment equals 1 in every period: $\forall(t, \sigma), \sum_{i\in\mathcal{I}} e^i_t(\sigma) = 1$.
- **A4** : All traders have an identical discount factor: $\forall i, \beta^i = \beta$.

Because the second welfare theorem applies, I assume that the initial optimal consumption choices are known and given by $C_0 = [c^1_0 \; ... \; c^I_0] >> 0$. By **A3**, $\sum_{i\in\mathcal{I}} c^i_0 = 1$, which allows us to interpret time-zero consumption shares as the weights that a hypothetical Bayesian prior gives to probabilities in $\mathcal{P}$. The absence of aggregate risk is needed to eliminate biases on risk-neutral probabilities due to aggregate consumption fluctuations.

3.2 The price probability class

Members of price probabilities are obtained by interpreting equilibrium prices of the Arrow securities as representing relative likelihoods and then using these relative likelihoods to construct probabilities via normalization. Given the set of trader beliefs ($\mathcal{P}$), different initial consumption-share distributions (C_0), preferences ($\{u^i\}_{i=1}^I$) and normalization methods determine different probability measures. I call the class of all such probability measures price probabilities:

Definition 1. Price probabilities, *$\mathcal{M}(\mathcal{P})$, is the class of all the probabilities that can be represented as normalized equilibrium prices of an economy that satisfies* **A0-A5**.[3]

[3]This definition can be extended naturally to the continuum by using the model and assumptions of Massari (2019). However, the sub-efficiency of members of $\mathcal{P}$ might not hold in the large as Massari (2019) showed that the

In the rest of the paper, I focus on two normalization methods: p^{NNL}, in which time-zero prices are normalized at every horizon; and p^{SNNL}, in which next-period prices are normalized sequentially.

Definition 2. Normalized Normed Likelihood (NNL)*:*

$$\forall \sigma^t, \quad p^{NNL}(\sigma^t) = \frac{q(\sigma^t)}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)} \quad ; \quad p^{NNL}(\sigma_t|\sigma^{t-1}) : \text{ not defined.}$$

NNL is the only probability measure that preserves the relative likelihoods of time-zero prices at every horizon (a new normalization is done at every horizon). In economic terms, p^{NNL} is the cost of moving a unit of consumption in period/event σ^t in terms of time-zero consumption, divided by the cost of moving a unit of consumption from time-zero to time t for sure. Because all normalizations are conducted with respect to time-zero prices, a set of conditional probabilities such that $p^{NNL}(\sigma^t) = \prod_{\tau=1}^{t} p^{NNL}(\sigma_\tau|\sigma^{\tau-1}) \forall t$ is not guaranteed to exist. More generally,

Proposition 1. *p^{NNL} satisfies the following time-consistency property if and only if all agents in the generating economy have log utility*

$$\forall \sigma^{t-1}, \quad \sum_{\hat{\sigma}^t \in S^t} p(\hat{\sigma}^t \cap \sigma^{t-1}) = p\left(\left(\cup \hat{\sigma}^t \in S^t\right) \cap \sigma^{t-1}\right) = p(\sigma^{t-1}).$$

Thus, p^{NNL} generically defines a sequence of probability measures on $\mathcal{S}^t$, which is not a forecasting scheme.

Definition 3. Sequential Normalized Normed Likelihood (SNNL)*:*

$$\forall \sigma^t, \quad p^{SNNL}(\sigma^t) = \prod_{\tau=1}^{t} p^{SNNL}(\sigma_t|\sigma^{t-1}) \quad ; \quad p^{SNNL}(\sigma_t|\sigma^{t-1}) = \frac{q(\sigma_t|\sigma^{t-1})}{\sum_{\hat{\sigma}_t} q(\hat{\sigma}_t|\sigma^{t-1})}$$

SNNL is the only probability measure that preserves the relative likelihoods of next-period prices. It is the cost of moving a unit of consumption from period/event σ^{t-1} one period ahead in state σ_t, divided by the cost of moving a unit of consumption for sure. Unlike p^{NNL}, p^{SNNL} is a forecasting scheme because it is constructed recursively.

The following Lemma highlights that the relation between p^{NNL} and p^{SNNL} mimics that between NML and SNML: p^{SNNL}'s period t predictions coincide with the conditional probabilities that p^{NNL} gives to σ_t, assuming that t is the final horizon:

MSH can fail in the large.

Lemma 1. *In an economy that satisfies* **A0-A5***,*

$$\forall \sigma^t,\ p^{SNNL}(\sigma_t|\sigma^{t-1}) = \frac{p^{NNL}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NNL}(\sigma^{t-1}, \hat{\sigma}_t)}.$$

4 Price probabilities in identical CRRA economies

If all traders have an identical CRRA utility function, members of price probabilities can be analytically characterized. This setting is flexible enough to show that BMA, NML, and SNML belong to price probabilities (Corollaries 1 and 2). In what follows, I use the notation:

Definition 4. p^{NNL}_{γ} *and* p^{SNNL}_{γ} *denote the* p^{NNL} *and the* p^{SNNL} *probabilities obtained from an economy that satisfies A2-A4 and in which all traders have an identical CRRA utility function with parameter* γ*,* $\forall i \in \mathcal{I}, u^i(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$*.*[4]

4.1 NNL in identical CRRA economies: p^{NNL}_{γ}

Proposition 2. *Given beliefs set* $\mathcal{P}$*, prior* C_0*, and parameter* γ*,* p^{NNL}_{γ} *is given by:*

$$\forall \sigma^t, \quad p^{NNL}_{\gamma}(\sigma^t) = \frac{\left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c^i_0\right)^{\gamma}}{\sum_{\hat{\sigma}^t\in\mathcal{S}^t}\left(\sum_{i\in\mathcal{I}} p^i(\hat{\sigma}^t)^{\frac{1}{\gamma}} c^i_0\right)^{\gamma}}. \tag{4}$$

Equation 4 shows that p^{NNL}_{γ} coincides with the normalized $\frac{1}{\gamma}$ norm of the likelihoods of members of $\mathcal{P}$ according to the measure C_0. Because BMA and NML are the normalized L_1 and L_∞ norms, respectively, they both belong to price probabilities.

Corollary 1. *Given beliefs set* $\mathcal{P}$ *and prior* C_0*,*
i) $\gamma = 1$ *(log)* $\Rightarrow \forall\sigma^t,\ p^{NNL}_1(\sigma^t) = p^{BMA}(\sigma^t)$*;*
ii) $\gamma = 0$ *(linear)* $\Rightarrow \forall\sigma^t, p^{NNL}_0(\sigma^t) = p^{NML}(\sigma^t)$*.*

Proof. *i)* Notice that if $\gamma = 1$, the denominator of Eq.4 equals 1, and compare Eq.4 with Eq.1.
ii) Notice that $lim_{\gamma\to 0}\left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c^i_0\right)^{\gamma} = ||p^i(\sigma^t)||_\infty$: the sup norm; and compare Eq.4 with Eq.2 □

Taking Bayes' rule as a reference point, the effect of γ on p^{NNL}_{γ} is qualitatively as follows. In a log-economy ($\gamma = 1$) p^{NNL}_{γ} coincides with BMA and the interaction between prior information (C_0) and empirical evidence (σ^t) is regulated by Bayes' rule. For $\gamma = 0$, p^{NNL}_{γ} coincides with NML (i.e., it is the optimal probability with respect to worst-case log-regret). Given the

[4] As is customary, I define $\ln 0 = -\infty$. Moreover, I use $\gamma = 0$ as a short notation for the limit equilibrium quantities of an identical CRRA economy in which $\gamma \to 0$ after the equilibrium quantities are calculated.

explosive nature of the log-likelihood on sequences whose frequencies are close to the boundary of the simplex, NML ignores the information of the prior (C_0 plays no role), and it assigns a relatively higher probability to sequences whose frequency lies close to the boundary of the simplex. For values of $\gamma \neq 1$, p_γ^{NNL} represents a compromise between the minimum log-regret approach behind NML and the Bayesian attempt to make the most out of the information in the prior. Compared with a BMA with the same Uniform prior on $\mathcal{P}$, p_γ^{NNL} with $\gamma < (>)1$ assigns more probability to those sequences whose frequency lies close to the boundary (center) of the simplex and penalizes those sequences whose frequency lies close to the center (boundary) of the simplex.

4.2 SNNL in identical CRRA economies: p_γ^{SNNL}

Proposition 3. *Given beliefs set $\mathcal{P}$, prior C_0, and parameter γ, p_γ^{SNNL} is given by:*

$$\forall \sigma^t, \quad p_\gamma^{SNNL}(\sigma_t|\sigma^{t-1}) = \frac{\left(\sum_{i\in\mathcal{P}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma}{\sum_{\hat{\sigma}_t\in\mathcal{S}}\left(\sum_{i\in\mathcal{P}} p^i(\hat{\sigma}_t)^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma}; \tag{5}$$

with $c^i_{\gamma,t-1}(\sigma) =^{by\ Eq.10} \frac{p^i(\sigma^{t-1})^{\frac{1}{\gamma}} c^i_0}{\sum_{i\in\mathcal{P}} p^i(\sigma^{t-1})^{\frac{1}{\gamma}} c^i_0}$.

By construction, $\sum_{i\in\mathcal{I}} c^i_{\gamma,t-1}(\sigma) = 1$, thus each $c^i_{\gamma,t-1}(\sigma)$ can be interpreted as being the weight attached to model p^i by a prior distribution $C_{\gamma,t-1}(\sigma)$. Equation 5 shows that p_γ^{SNNL} coincides with the sequentially normalized $\frac{1}{\gamma}$ norm of the next period probabilities members of $\mathcal{P}$ according to the distribution $C_{\gamma,t-1}(\sigma)$. It is easy to verify that SNML belongs to price probabilities.

Corollary 2. *Given beliefs set, $\mathcal{P}$, and prior, C_0, $\forall\sigma^t, p^{SNNL}_{\gamma=0}(\sigma^t) = p^{SNML}(\sigma^t)$.*

Proof. $\forall \sigma^{t-1}$, $p_0^{SNNL}(\sigma_t|\sigma^{t-1}) =^{\text{Lem.1}} \frac{p_0^{NNL}(\sigma^t)}{\sum_{\hat{\sigma}_t} p_0^{NNL}(\sigma^{t-1},\hat{\sigma}_t)} =^{\text{Cor.1}} \frac{p^{NML}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NML}(\sigma^{t-1},\hat{\sigma}_t)} :=^{\text{Eq.3}} p^{SNML}(\sigma_t|\sigma^{t-1})$. ◻

The p_γ^{SNNL} prediction scheme is closely related to other robust prediction schemes as the $C_{\gamma,t-1}(\sigma)$ prior is the core of many known algorithms. $C^i_{\gamma,t-1}(\sigma)$ is a special case of the "Generalized Bayes' rule" introduced by Vovk (1990). The gamma parameter is often called the learning rate as it determines the convergence rate of the posterior. The choice of this parameter plays a fundamental role in both the HEDGE algorithm (Freund and Schapire, 1997) and the Safe Bayesian approach (Grünwald, 2012). p_γ^{SNNL} differs from these algorithms because instead of relying on the generalized prior to mix the probabilities in the support (or the actions) it directly relies on a generalized posterior that treats equally past and future performance. This symmetry of treatment comes with the need to normalize the predictive generalized distribution into a probability. This normalization makes p_γ^{SNNL} prediction non-convex in the sense that its prediction might fall out of the convex combination of models in the support (see Section 6). While the (generalized) prior evolution of the cited algorithm can be directly compared to standard Bayes and values of $\gamma < (>)1$ mapped into predictions that over(under)-react to information with respect to Bayes' rule, this is no longer possible for SNNL, which can deliver non-convex predictions (Section 6).[5] The algorithms mentioned above utilize values of $\gamma \geq 1$, indicating that under-reacting to information is the way to robustify standard Bayes. Similarly, I find that p_γ^{SNNL} can outperforms Bayes' only for values of gamma greater than 1, and only in misspecified learning problems in which the Bayesian posterior does not eventually concentrate on a unique parameter.

5 Asymptotic performance of price probabilities

5.1 The criterion

In this section, I introduce the efficiency criterion I use to characterize the performance price probabilities. Following an established tradition across fields, the criterion I propose is based on *prequential likelihood ratios* (Dawid, 1984; Ploberger and Phillips, 2003).

Definition 6. *Let $p^{BMA}(\sigma^t)$ be the likelihood of a BMA with a full-support prior on $\mathcal{P}$,*

[5] Following Epstein et al. (2008) approach,

Definition 5. *A prediction rule (NB)* ***under-reacts*** *to information with respect to Bayes (B) if for every non-degenerate distribution on the support $C^{NB}(\sigma^t)$ we have that*

$$C^{NB}(\sigma^t) = C^B(\sigma^t) \Rightarrow \forall \sigma_{t+1}, C^{NB}(\theta|\sigma_{t+1}\sigma^t) \in Conv(C^B(\theta|\sigma^t), C^B(\theta|\sigma_{t+1}\sigma^t)).$$

A prediction rule (NB) ***over-reacts*** *to information with respect to Bayes (B) if for every non-degenerate distribution on the support $C^{NB}(\sigma^t)$ we have that*

$$C^{NB}(\sigma^t) = C^B(\sigma^t) \Rightarrow \forall \sigma_{t+1}, C^B(\theta|\sigma_{t+1}\sigma^t) \in Conv(C^{NB}(\theta|\sigma^t), C^{NB}(\theta|\sigma_{t+1}\sigma^t)).$$

For intuition of the effect of γ on the evolution of $C^i_{\gamma,t-1}(\sigma)$, suppose that $S := \{a, b\}$ and every probability in $\mathcal{P} := \{p^i : i \in \mathcal{I}\}$ is i.i.d. Bernoulli $(\forall i, \forall t, p^i(a_t) = i)$. With t_a and t_b representing the number of a, b observations until period t-1, respectively, we obtain:

$$c^i_{\gamma,t-1}(\sigma) = \frac{p^i(\sigma^t)^{\frac{1}{\gamma}} c^i_0}{\sum_{j\in\mathcal{P}} p^j(\sigma^t)^{\frac{1}{\gamma}} c^j_0} = \frac{i^{\frac{t_a}{\gamma}}(1-i)^{\frac{t_b}{\gamma}}}{\sum_{j\in\mathcal{P}} j^{\frac{t_a}{\gamma}}(1-j)^{\frac{t_b}{\gamma}}}.$$

So, if $\gamma < 1$, the model overreacts to empirical evidence: e.g.,$\gamma = \frac{1}{2}$ is equivalent to updating using Bayes' rule "counting every past realization twice." If $\gamma > 1$, the model underreacts to empirical evidence: e.g.,$\gamma = 2$ is equivalent to updating using Bayes' rule "counting every past realization as half."

- *a probability mixture, p, with belief set, $\mathcal{P}$, is* $\mathcal{P}$**-efficient** *if*

$$\forall P \in \mathcal{P}, \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \asymp^{P\text{-}a.s.} 1;$$

- *a probability mixture, p, with belief set, $\mathcal{P}$, is* **universal-efficient** *if*

$$\forall \sigma \in S^\infty, \ \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \asymp 1;$$

- *a probability mixture p with belief set $\mathcal{P}$ is* **super-efficient** *if is $\mathcal{P}$-efficient and*

$$\begin{cases} \forall \sigma \in S^\infty, & \limsup \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} <^{\hat{P}\text{-}a.s.} +\infty \\ \exists \hat{P}: & \lim \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} =^{\hat{P}\text{-}a.s.} -\infty \end{cases};$$

- *a probability mixture, p, with beliefs set, $\mathcal{P}$, is* **sub-efficient** *if is $\mathcal{P}$-efficient and*

$$\begin{cases} \exists \hat{P}: & \lim \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} =^{\hat{P}\text{-}a.s.} +\infty \\ \forall \sigma \in S^\infty, & \liminf \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} >^{\hat{P}\text{-}a.s.} -\infty \end{cases};$$

where the notation $f(x) \asymp g(x)$ abbreviates $\limsup \frac{f(x)}{g(x)} < +\infty$ *and* $\liminf \frac{f(x)}{g(x)} > -\infty$.

So, p, is $\mathcal{P}$-efficient if it is as good as Bayes when the learning problem is well-specified. The rest of the definition characterizes the relative performance of p and p^{BMA} in misspecified setting. I say that p is universal-efficient if it is as accurate as Bayes' in every sequence. A probability mixture p is super-efficient if it does at least as well as Bayes' in every sequence and there are probabilities $\hat{P}$ for which it outperforms Bayes' $\hat{P}$-a.s. — that is, if it guarantees to do at least as well as using Bayes' rule and there are cases (when the model is misspecified) in which it does infinitely better. A probability mixture, p, is sub-efficient if there are no sequences in which it outperforms Bayes', and there are cases of misspecification in which it is infinitely worse.

5.2 Discussion

Objectively comparing different learning rules is not trivial. My criterion has been chosen to satisfy the following desiderata:

- **D1:** The comparison must be performed in every sequence because in most cases in which we need to make predictions, we do not know the true probability. If we knew the true probability, we would not need to find the best mixture of members of $\mathcal{P}$.

- **D2:** The benchmark must be appropriate. The BMA is chosen because it is widely known, utilized, and has a sound axiomatic foundation.[6]

- **D3:** The probability mixture, p, and BMA must use the same support and empirical evidence to be comparable. Otherwise, the comparison would be about the quality of the information, rather than the way to use it.

- **D4:** The criterion must be asymptotic to eliminate the small sample effect of the priors. Furthermore, small sample accuracy criteria should be avoided because they are potentially misleading (Massari, 2013).[7]

The possibility of super-efficient mixture models is, in my experience, often received with skepticism. Here are some responses to concerns raised at conferences and by referees.

- *The criterion is weak. For example, it would be satisfied by using Bayes' rule in an enlarged prior support.*

  This observation is correct. However, it violates **D3**: a different prior support implies different information on the set of possible models. Changing the support alters the intrinsic nature of the learning problem. What we want to achieve is to use the same information more efficiently, not show that a larger prior support can explain more sequences. A super-efficient mixture "beats" Bayes' using the same information.

- *Price probabilities are Bayesian in disguise.*

  This statement is false. In Section 6, I prove that unless all agents in the generating economy have log-utility, p^{NNL} and p^{SNNL} are not Bayesian because they might make predictions that are not convex combinations of models in $\mathcal{P}$.

- *Bayesian updating is almost a tautology if we think of probabilities as empirical frequencies. Why should we abandon it?*

  Bayes rule is not defined when updating from sets of measure 0. When the model is misspecified, the Bayesian measure attaches 0 probability to all tail events that occur P-a.s., thus its application is far from natural. In these cases, and if our ultimate goal is making predictions, it seems natural to compare Bayes' rule against alternative rules on the basis of accuracy, rather than internal consistency. As Dawid (1982) eloquently said: *"If a subjective distribution P attaches probability zero to a non-ignorable event, and if this event happens, then P must be treated with suspicion, and modified or replaced."*

[6] Moreover, BMA has finite worst-case log-regret (if $|\mathcal{P}|$ is finite). Thus a likelihood comparison against BMA is also a way to verify if a probability mixture possesses this fundamental property.

[7] Massari (2013) shows that given two probabilities $\{p^a\}, \{p^b\}$, it is not true that if p^a's next-period predictions are infinitely often more accurate than p^b and never less accurate, then p^a's predictions are more accurate than p^b on long sequences.

- *The super-efficiency result must be incorrect because it is in contrast with Wald (1947)'s Complete Class Theorem (CCT).*

  My result is orthogonal to the CCT. CCT is a result about the optimality of the Bayesian procedure for decision in a static setting. Therefore, CCT is moot about the efficiency of Bayes' rule to incorporate empirical evidence in a prior distribution. More generally, there is no tension between my super-efficiency result and the known optimality of Bayesian decision criteria. If the model is well-specified, all members of price probability are asymptotically indistinguishable from Bayes. If the model is misspecified, while Bayesian predictions are guaranteed to be exactly as accurate as the most accurate model in the support (Berk, 1966), p^{SNNL} can deliver predictions that are even more accurate than that.

- *Where are the tricks/hidden assumptions?*

  The crucial assumption needed to ensure super-efficiency is the non-convex prior support. In Proposition 4, I show that super-efficiency occurs only on those sequences on which the Bayesian posterior does not concentrate fast enough on a unique model. By concavity of the log-likelihood function, this event can happen only if the support contains two orthogonal models with similar likelihood, but no intermediate model (i.e. if the prior support is not convex.[8]

- *These results are practically irrelevant.*

  Regarding relevance, the super-efficiency properties of the parameters I identify apply verbatim to all standard prediction problems. I choose to work in a parametric setting with finitely many parameters only for ease of exposition and to maintain the state price interpretation. For applications of similar results, I refer the reader to Grünwald and van Ommen (2014), which shows that underreacting rules outperform the BMA with prior on a finite set of linear regression models. Furthermore, consistent with our results,Timmermann (2006) brings evidence that forecasting combinations of statistical models with weights evolving slower than BMA outperform BMA in many cases of misspecification. Avramov (2002); Cremers (2002) have found that BMA guarantees better out-of-sample prediction than that obtained using model selection criteria — which are qualitatively equivalent to p_0^{SNNL}— in the context of forecasting U.S. stock market indices.

- *Asymptotic criteria are not relevant for investment decisions on a finite horizon (Samuelson, 1971, 1979).*

  Samuelson's critique is based on the argument that, given beliefs and prices, different pref-

[8] To convexify the prior support is hardly a solution. First, to convexify the prior support violates **D3**. Second, it is often difficult to do in a non-parametric context. Last, increasing the dimensionality of the prior support its undesirable as it reduces the learning rate (Schwarz, 1978; Clarke and Barron, 1990).

erences determine different optimal investment strategies. His critique does not apply here because preferences play no role in my accuracy criterion.

5.3 Asymptotic performance of p^{SNNL}

Theorem 1. *p^{SNNL} is $\mathcal{P}$-efficient in any economy that satisfies* **A0-A5**. *Furthermore,*
i) $p^{SNNL}_{\gamma^i>1}$ is super-efficient;
ii) $p^{SNNL}_{\gamma^i=1}$ is universal-efficient;
iii) $p^{SNNL}_{\gamma^i<1}$ is sub-efficient;
where $p^{SNNL}_{\gamma^i>1}, p^{SNNL}_{\gamma^i=1}$ and $p^{SNNL}_{\gamma^i<1}$ denote the p^{SNNL} probabilities obtained from an economy that satisfies A2-A4 and in which all traders have CRRA utility function with parameter $\gamma^i > 1, \gamma^i = 1$, and $\gamma^i < 1$, respectively.

So, p^{SNNL}, is as good as Bayes' when the learning problem is well-specified. Furthermore, Theorem 1 specifies how risk attitudes of agents in the generating economy affects the accuracy of p^{SNNL} in misspecified learning problems.

Intuition:

- p^{SNNL} is $\mathcal{P}$-efficient irrespective of preferences because when there is a unique agent that knows the truth, P, he dominates P-a.s.. Furthermore, inspection of the FOC shows that his consumption-share converges to one at an exponential rate, which guarantees that $p^{SNNL} \to P$ at an exponential rate (i.e. p^{SNNL} merges to P).
- $p^{SNNL}_{\gamma^i=1}$ is universal-efficient because it coincides with Bayesian updating.
- $p^{SNNL}_{\gamma^i>(<)1}$ is super-efficient (sub-efficient). Let us focus on the (more interesting) case of $\gamma^i > 1$ and make use of the economy generating $p^{SNNL}_{\gamma^i>1}$ to gain intuition. From the proof of Theorem 1, we see that

$$\begin{aligned}
\ln \frac{p^{BMA}(\sigma^t)}{p^{SNNL}(\sigma^t)} &= \ln \frac{p^{BMA}(\sigma^t)}{q(\sigma^t)} + \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \\
&\underbrace{=}_{\substack{\text{by Massari (2017)} \\ \text{Corollary 1}}} O(1) + \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right)
\end{aligned}$$

So, $p^{SNNL}_{\gamma^i>1}$ is super-efficient if there are paths on which $\sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \to -\infty$ and no path on which $\sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \to +\infty$. In economic terms, for all $(t-1, \sigma)$, $\sum_{\tilde{\sigma}_t} q_t(\tilde{\sigma}_t|\sigma^{t-1})$ is the cost of moving a unit of consumption for sure a period ahead, i.e., the reciprocal of the risk-free rate. The effect of risk attitudes on the risk-free rate follows this intuition. In every period most agents subjectively believe that assets are mispriced and trade for speculative reasons because they disagree. When agents have log utility ($\gamma = 1$), prices (and thus interest rates) do not affect optimal saving choices (the substitution effect equals the income effect) and the reciprocal of the risk-free rate is given by the discount factor: for all $(t, \sigma), \beta = \sum_{\tilde{\sigma}_t} q_t(\tilde{\sigma}_t|\sigma^{t-1})$. However, if $\gamma > 1$, the substitution effect is stronger than the income effect; so, agents optimally invest less than if they had log utility, and a higher risk-free rate arise $\sum_{\tilde{\sigma}_\tau} q_\tau(\tilde{\sigma}_\tau|\sigma^{\tau-1}) < \beta$ for all (t, σ) in which the consumption-share is not degenerate. When the consumption share distribution is non-degenerate a positive fraction of periods, this effect cumulates to $-\infty$.

The argument above suggests that non-concentration of the Bayesian prior plays a special role in determining the (sub)super-efficient condition. This is indeed the case:

Proposition 4.
i) In every path in which p^{BMA}'s posterior does not concentrate on a unique model,

$$\lim \ln \frac{p^{SNNL}_{\gamma^i>1}(\sigma^t)}{p^{BMA}(\sigma^t)} = +\infty \quad \textit{and} \quad \lim \ln \frac{p^{SNNL}_{\gamma^i<1}(\sigma^t)}{p^{BMA}(\sigma^t)} = -\infty$$

ii) In every path in which p^{BMA}'s posterior concentrates exponentially fast on a unique model,

$$\ln \frac{p^{SNNL}_{\gamma^i>1}(\sigma^t)}{p^{BMA}(\sigma^t)} \asymp 1 \quad \textit{in any economy that satisfies } \textbf{A0-A5}$$

Proposition 4 tells us that $\gamma^i < 1$ is detrimental while $\gamma^i > 1$ is desirable in cases in which the Bayesian prior does not concentrate on a unique model in the support. If the Bayesian posterior does not concentrate, then the true model must be somewhere in the middle because the data supports more than one model. In this case, the generalized prior, by giving more (less) weight to empirical evidence produces forecasts that are closer (further) to the truth than the Bayesian.

Known asymptotic results in Bayesian statistics[9] make Proposition 4 useful in recognizing the probabilities that determine the (sub)super-efficiency condition.[10]

Next, we make use of the analytical form of p^{SNNL}_γ to present three cases showing that p^{SNNL}_γ can significantly outperform but never underperform BMA if $\gamma > 1$; whereas p^{SNNL}_γ can significantly underperform but never outperform BMA with if $\gamma < 1$. The generality of the example rests in the choice of the sequences in Cases 1 and 2. Case 1 utilizes the sequence on

[9] If $|\mathcal{P}| < \infty$, in most standard settings (if members of $\mathcal{P}$ are either i.i.d. or conditionally iid), the Bayesian posterior does not concentrate if and only if there is more than one model with the same K-L divergence. Otherwise, it concentrates exponentially fast.

[10] Proposition 4 enormously simplifies this task. Even if traders' beliefs and the true measure are iid, p^{SNNL}'s dynamic is path-dependent.

which the Bayesian posterior concentrates the least; case 2 utilizes the sequence in which the convergence rate of the Bayesian posterior is the fastest. Case 3 illustrates Proposition 3.

Example 1: Let $S = \{a, b\}, C_0 = [.5\ .5]$, and $\mathcal{P} = \{p^1, p^2\}$, with p^1, p^2 i.i.d. measures: $\forall t, p^1(a_t) = \frac{1}{3} = p^2(b_t)$. Consider three $p^{SNNL}_{\gamma_j}$s with parameters $\gamma_0 = 0$, $\gamma_1 = 1$ and $\gamma_2 = 2$, respectively.

Case a: The true probability, P, is degenerate. It gives probability 1 to the alternating sequence $\{a, b, a, ...\}$. This is the most favourable case for probability mixtures with slow concentration rate of the prior. Because both models are equally (in)accurate, the best predictor is the one giving equal weight to p^1 and p^2 in every period (as C_0 does). By Equation 5:

$$p_0^{SNNL}(a_t|\sigma^{t-1}) = \frac{p^{NML}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NML}(\sigma^{t-1}, \hat{\sigma}_t)} = \begin{cases} \frac{1}{2}\ if\ t\ odd \\ \frac{2}{3}\ if\ t\ even \end{cases}$$

$$p_1^{SNNL}(a_t|\sigma^{t-1}) = \sum_{i\in\mathcal{I}} p^i(a) \frac{p^i(\sigma^{t-1}) c_0^i}{\sum_{i\in\mathcal{I}} p^i(\sigma^{t-1})\, c_0^i} = \begin{cases} \frac{1}{2}\ if\ t\ odd \\ \frac{5}{9}\ if\ t\ even \end{cases}$$

$$p_2^{SNNL}(a_t|\sigma^{t-1}) = \frac{\left(\sum_{i\in\mathcal{I}} p^i(a)^{\frac{1}{2}} \frac{p^i(\sigma^{t-1})^{\frac{1}{2}} c_0^i}{\sum_{i\in\mathcal{I}} p^i(\sigma^{t-1})^{\frac{1}{2}} c_0^i}\right)^2}{\sum_{\hat{\sigma}_t}\left(\sum_{i\in\mathcal{I}} p^i(\hat{\sigma}_t)^{\frac{1}{2}} \frac{p^i(\sigma^{t-1})^{\frac{1}{2}} c_0^i}{\sum_{i\in\mathcal{I}} p^i(\sigma^{t-1})^{\frac{1}{2}} c_0^i}\right)^2} = \begin{cases} \frac{1}{2}\ if\ t\ odd \\ \frac{9}{17}\ if\ t\ even \end{cases}$$

Thus, on $\{a, b, a, ...\}, \forall\mu \in (0, 1)$, $$\begin{cases} \frac{p_\mu^{BMA}(\sigma^t)}{p_0^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{1}{3}\right)^{\frac{t}{2}}\left(\frac{2}{3}\right)^{\frac{t}{2}} + (1-\mu)\left(\frac{2}{3}\right)^{\frac{t}{2}}\left(\frac{1}{3}\right)^{\frac{t}{2}}}{\left(\frac{1}{2}\right)^{\frac{t}{2}}\left(\frac{1}{3}\right)^{\frac{t}{2}}} \to +\infty \\ \frac{p_\mu^{BMA}(\sigma^t)}{p_1^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{1}{3}\right)^{\frac{t}{2}}\left(\frac{2}{3}\right)^{\frac{t}{2}} + (1-\mu)\left(\frac{2}{3}\right)^{\frac{t}{2}}\left(\frac{1}{3}\right)^{\frac{t}{2}}}{\left(\frac{1}{2}\right)^{\frac{t}{2}}\left(\frac{4}{9}\right)^{\frac{t}{2}}} = O(1) \\ \frac{p_\mu^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{1}{3}\right)^{\frac{t}{2}}\left(\frac{2}{3}\right)^{\frac{t}{2}} + (1-\mu)\left(\frac{2}{3}\right)^{\frac{t}{2}}\left(\frac{1}{3}\right)^{\frac{t}{2}}}{\left(\frac{1}{2}\right)^{\frac{t}{2}}\left(\frac{9}{17}\right)^{\frac{t}{2}}} \to 0 \end{cases}.$$

Case a shows that, p_2^{SNNL} produces predictions that are closer to the empirical frequency than p_1^{SNNL} and p_0^{SNNL} and thus more accurate.

Case b: The true probability, P, is degenerate. It gives probability 1 to the sequence $\{a, a, a, ...\}$. Because p^2 is clearly the best model, case b is the most favourable sequence for probability mixtures that overreact to empirical evidence.

Thus,on $\{a, a, a, ...\}, \forall\mu \in (0, 1)$, $$\begin{cases} \frac{p_\mu^{BMA}(\sigma^t)}{p_0^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{2}{3}\right)^t + (1-\mu)\left(\frac{1}{3}\right)^t}{\frac{1}{2}\left(\frac{2}{3}\right)^{t-1}} = O(1) \\ \frac{p_\mu^{BMA}(\sigma^t)}{p_1^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{2}{3}\right)^t + (1-\mu)\left(\frac{1}{3}\right)^t}{\frac{1}{2}\left(\frac{2}{3}\right)^t + \frac{1}{2}\left(\frac{1}{3}\right)^t} = O(1) \\ \frac{p_\mu^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} = \frac{\mu\left(\frac{2}{3}\right)^t + (1-\mu)\left(\frac{1}{3}\right)^t}{\left(\frac{1}{2}\left(\frac{2}{3}\right)^{\frac{t}{\gamma}} + \frac{1}{2}\left(\frac{1}{3}\right)^{\frac{t}{\gamma}}\right)^{\gamma} * e^{-\sum_{\tau=1}^{t} \ln\left(q(a|\sigma^{\tau-1}) + q(b|\sigma^{\tau-1})\right)}} = O(1) \end{cases}.$$ [11]

[11] By Proposition 4, ii)

Case b shows that, although p_0^{SNNL} only takes one observation to correctly identify the most accurate model, p_0^{SNNL} and p^{BMA} converge to p^2 fast enough not to compromise their asymptotic likelihood performance.

Case c: Draws are i.i.d. with true probability P ($\forall t, P(a_t) = \frac{1}{2}$). Because p^1 and p^2 are equally (in)accurate, it is easy to show that the Bayesian posterior does not concentrate (Massari, 2013) and Proposition 4 implies, $\begin{cases} \frac{P^{BMA}(\sigma^t)}{p_0^{SNNL}(\sigma^t)} \to^{P\text{-a.s.}} +\infty \\ \frac{P^{BMA}(\sigma^t)}{p_1^{SNNL}(\sigma^t)} \asymp^{P\text{-a.s.}} 1 \\ \frac{P^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} \to^{P\text{-a.s.}} 0 \end{cases}$.

Remark: The relationship between p^{NNL} and p^{SNNL} mimics that between NML and SNML. Each next-period forecast of the p^{SNNL} corresponds to the last period conditional distribution of the corresponding p^{NNL} probability. Thus, p^{SNNL} can be thought of as a compromise to make p^{NNL} recursive. This interpretation makes the super-efficiency part of Theorem 1 even more surprising. It shows that a forecaster can perform significantly better by using a recursive method even when he knows the final horizon of his prediction task. Because a recursive method does not use the length of the sequence he is forecasting as an input, this result illustrates a case in which ignoring some relevant information increases prediction accuracy.

5.4 Asymptotic performance of p^{NNL}

Theorem 2. *p^{NNL} is universal-efficient in any economy that satisfies* **A0-A5**.

Theorem 2 tells us that, although non-convex and time-inconsistent, p^{NNL} performs qualitatively as well as BMA in terms of likelihood in any sequence (and thus P-a.s.). If we are only concerned about accuracy, there is no reason to consider non-convexity or time-consistency to be a fundamental property of rational forecasts.

Intuition: universal efficiency of p^{NNL} implies that the prediction of p^{NNL} and those of p^{BMA} must be asymptotically equivalent. For intuition, consider Equation 16 in the proof of Theorem 2 and the unconditional Bayesian probabilities from a full support prior C_0:

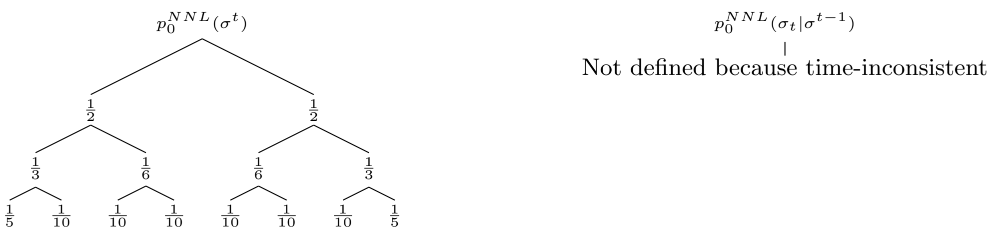
$$p^{NNL}(\sigma^t) \propto \frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{\sum_{j\in\mathcal{I}}\frac{1}{u_j'(c^j(\sigma^t))}}$$

$$p^{BMA}(\sigma^t) = \sum_{i\in\mathcal{I}} p^i(\sigma^t) c_0^i.$$

If the term $\sum_{j\in\mathcal{I}}\frac{1}{u_j'(c^j(\sigma^t))}$ were time and state-independent, the proportionality sign would incorporate it with $\frac{1}{u_i'(c_0^i)}$ into a prior so that p^{NNL} would map into a p^{BMA} with different prior. However, the time-dependence of $\sum_{j\in\mathcal{I}}\frac{1}{u_j'(c^j(\sigma^t))}$ makes p^{NNL} time-inconsistent (to act as if it had a prior distribution that depends on the predictions horizon), while its dependence on states makes p^{NNL} non-convex. Asymptotically, p^{NNL} and p^{BMA} are equivalent because the time and state dependence of this term becomes negligible since differences on prior distributions have negligible effect compared to that of models' likelihoods.

Theorem 2 does not justify the systematic use of time-inconsistent probabilities in every decision problem. Time-inconsistent members of $\mathcal{M}(\mathcal{P})$ are undesirable in many economic settings because they do not rule out dynamic arbitrage (Lehrer and Teper, 2016).

Example 2: Let $S = \{a, b\}, C_0 = [\frac{1}{2}\ \frac{1}{2}]$, and $\mathcal{P} = \{p^1, p^2\}$, with p^1, p^2 i.i.d. measures: $\forall t, p^1(a_t) = \frac{1}{3} = p^2(b_t)$. It can be easily calculated that $p_0^{NNL}(\sigma^t) = \frac{\max_i p^i(\sigma^t)}{\sum_{\hat{\sigma}^t} \max_i p^i(\hat{\sigma}^t)}$ attaches the following three of unconditional probabilities.



So, a risk-neutral agent with p_0^{NML} beliefs who does not discount the future is at time-zero indifferent between:

- $\$\frac{1}{3}$ and a lottery, L1, that pays \$1 if $\{R, L\}$ realizes, \$0 otherwise;
- $\$\frac{2}{10}$ and a lottery, L2, that pays \$1 if $\{R, L, L\}$ realizes, \$0 otherwise;
- $\$\frac{1}{10}$ and a lottery, L3, that pays \$1 if $\{R, L, R\}$ realizes, \$0 otherwise.

Selling L1 to him for $\$\frac{1}{3}$ and buying from him L2 and L3 for a total of $\$\frac{3}{10}$ constitutes an arbitrage: if $\{R, L\}$ does not realize, I make a profit $\frac{1}{3}$-$\frac{3}{10} > 0$. If $\{R, L\}$ does realize, I make the same profit because I can use the market to pay the dollar I lose in $t = 2$ with the dollar I win for sure in t=3 (because either $\{R, L, R\}$ or $\{R, L, L\}$ will happen for sure).

However, this arbitrage opportunity can be generated only if p^{NNL} is used in markets that allow for both time-zero and sequential trading. An arbitrage can be constructed against an agent with p^{NNL} beliefs only because his beliefs correspond to a state of mind in which trade can only occur at time-zero. If he knew his final horizon t and he was given the possibility to trade sequentially, then he could use his p^{NNL} at t to construct a set of prequential conditional probabilities via backward induction to avoid arbitrages.

6 Price probabilities are not Bayesian

In this section, I show that members of price probability are generically not Bayesian because p^{NNL} and p^{SNNL} produce convex predictions if and only if all agents in the generating economy have log utility ($\gamma = 1$).

Definition 7. *A probability mixture p has* ***convex predictions*** *if in every application (i.e. for every state space S, for every prior support $\mathcal{P}$, prior distribution C_0, and path) its predictions belongs to $Conv(\mathcal{P})$.*

The Bayesian predictive distribution has convex predictions because in every application the predictive distribution is a prior-weighted average of the models in the support. So, probability mixtures that have non-convex predictions are fundamentally non-Bayesian. $\mathcal{P}$-efficiency of probability mixture with non-convex predictions shows that consistency is orthogonal to convex prediction, and thus not a fundamental property of rational decisions.

Proposition 5. *p^{SNNL} and p^{NNL} are convex if and only if all agents in the generating economy have log utility.*

Example 3 shows that p^{SNNL}_{γ} is not convex for $\gamma \neq 1$.

Example 3: Let $S = \{a, b, c\}, C_0 = [\frac{1}{2}\ \frac{1}{2}]$, and $\mathcal{P} = \{p^1, p^2\}$, with p^1, p^2 i.i.d. measures: $\forall t, [p^1(a_t)\ p^1(b_t)\ p^1(c_t)] = [.2\ .6\ .2]$ and $[p^2(a_t)\ p^2(b_t)\ p^2(c_t)] = [.2\ .2\ .6]$, . Consider p^{SNNL}_{γ} obtained with $\gamma_0 = 0$ and $\gamma_1 = 1$ and $\gamma_2 = 2$, respectively.

$$
\begin{aligned}
p^{SNNL}_{\gamma_0}(a_1) &= \frac{\max_{i=1,2} p^i(a_1)}{\sum_{\sigma_1} \max_{i=1,2} p^i(\sigma_1)} = \frac{.2}{.2 + .6 + .6} = \frac{1}{7} \neq .2 && \Rightarrow p^{SNNL}_{\gamma_0} \notin Conv\{p^1, p^2\} \\
p^{SNNL}_{\gamma_1}(a_1) &= \frac{p^1(a_1)c_0^1 + p^2(a_2)c_0^2}{\sum_{\sigma_1} p^i(\sigma_1)c_0^i} = .2 && \Rightarrow p^{SNNL}_{\gamma_0} \in Conv\{p^1, p^2\} \\
p^{SNNL}_{\gamma_2}(a_1) &= \frac{\left(p^1(a_1)^{.5}c_0^1 + p^2(a_2)^{.5}c_0^2\right)^2}{\sum_{\sigma_1} \left(p^1(\sigma_1)^{.5}c_0^1 + p^2(\sigma_2)^{.5}c_0^2\right)^2} \\
&= \frac{.2}{.2 + 2\left(\sqrt{.6}.5 + \sqrt{.2}.5\right)^2} \approx .21 \neq .2 && \Rightarrow p^{SNNL}_{\gamma_0} \notin Conv\{p^1, p^2\}
\end{aligned}
$$

7 Conclusion

I use the standard machinery of dynamic general equilibrium models to generate a rich class of probabilities and to discuss their properties. All members of price probability do as well as Bayes' when the learning problem is well-specified while some of them can significantly improve on Bayes in misspecified settings. My result challenges the prevailing opinion that Bayes' rule is the only rational way to learn.

7.1 Applications and future developments

My results go beyond the simple parametric setting I have adopted in this paper, which was chosen exclusively for illustrative purposes. The proofs of Theorem 1 and 2 apply, with minor notational changes, to the case in which models in $\mathcal{P}$ are parametric models whose parameters get learned over time. For example, suppose you are an investor. To choose how to invest in the market, you would like to know what is the true data-generating process of the market's returns. Unfortunately, nobody is currently able to tell you what THE true model is, and the best you can do is to recognize that there is a set of candidate modes: $\mathcal{P}$. Given your prior opinion on the merits of these models, the problem you face is to decide how to change the weights of your prior as a function their likelihood performance. Theorem 1 tells you that, unless you are sure that one of the models in $\mathcal{P}$ is correct, you should pragmatically prefer to use p^{SNNL} with a large gamma over a Bayesian Model Average ($\gamma = 1$) and over standard criteria of model selection that uniquely identify a "true" model in $\mathcal{P}$ such as BIC ($\gamma = 0$).

The only problem left is how to choose gamma. I have shown that if the posterior does not concentrate, larger values of gamma deliver more accurate forecasts. However, this improvement in accuracy comes with a cost. Although, for every $1 < \gamma < \infty$ the asymptotic log-likelihood ratio between p^{BMA} and p^{SNNL}_{γ} is bounded above, this ratio increases monotonically in gamma when the model is correctly specified (slower learning rate implies slower convergence rate). If the model is misspecified and the posterior does converge to a unique model, it is the relative position of the projection of P on $\mathcal{P}$ that determines whether larger values of gamma improve or deteriorate p^{SNNL}_{γ}'s accuracy. In a prediction context, it is natural to use past data to determine the value of gamma on-line. This line of reasoning is inspired by the Safe Bayesian approach (Grünwald and van Ommen (2014)) and the Flip-Flop algorithm (De Rooij et al. (2014)). The implementation of this intuition into the equations governing p^{SNNL}_{γ} is left to future research.

A Appendix

Proof of Lemma 1

Proof.

$$p^{SNNL}(\sigma_t|\sigma^{t-1}) =^{By\ Def.3} \frac{q(\sigma_t|\sigma^{t-1})}{\sum_{\hat{\sigma}_t} q(\hat{\sigma}_t|\sigma^{t-1})} = q(\sigma_t|\sigma^{t-1}) * \frac{q(\sigma^{t-1})}{\sum_{\bar{\sigma}^t} q(\bar{\sigma}^t)} * \frac{\sum_{\bar{\sigma}^t} q(\bar{\sigma}^t)}{q(\sigma^{t-1})} * \frac{1}{\sum_{\hat{\sigma}_t} q(\hat{\sigma}_t|\sigma^{t-1})}$$
$$= \frac{q(\sigma^t)}{\sum_{\bar{\sigma}^t} q(\bar{\sigma}^t)} * \frac{1}{\frac{\sum_{\hat{\sigma}_t} q(\sigma^{t-1},\hat{\sigma}_t)}{\sum_{\bar{\sigma}^t} q(\bar{\sigma}^t)}} =^{By\ Def.2} \frac{p^{NNL}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NNL}(\sigma^{t-1},\hat{\sigma}_t)}.$$

□

Lemma 2. *In an economy that satisfies* **A0-A5***, equilibrium prices are given by:*

$$q(\sigma^t) = \frac{\beta^t \sum_{i\in\mathcal{P}} p^i(\sigma^t) \frac{1}{u^i(c_0^i)'}}{\sum_{j\in\mathcal{I}} \frac{1}{u^j(c_t^j(\sigma))'}} \tag{6}$$

Proof. The Lagrangian problem associated with each trader's maximization problem is

$$L_i = E_{p^i} \sum_{t=0}^{\infty} \beta^t u^i(c_t^i(\sigma)) + \lambda_i \left(\sum_{t=0} \sum_{\sigma^t \in S^t} q(\sigma^t) \left(c_t^i(\sigma) - e_t^i(\sigma) \right) \right).$$

By equating the derivatives of this Lagrangian to 0 I get, for all (t, σ),

$$\frac{\partial L_i}{\partial c_t^i(\sigma)} = 0 \Rightarrow \beta^t p^i(\sigma^t) u^i(c_t^i(\sigma))' = \lambda_i q(\sigma^t) \tag{7}$$

Letting $q_0 = 1$ (the price of one unit of consumption at t=0 equals 1) I find that $\lambda_i = u^i(c_0^i)'$, the result follows rearranging summing over traders and rearranging. □

Proof of Proposition 1

Proof. By contradiction, assume H_0 : there exists a non-log economy such that

$$\forall \sigma^{t-1}, \quad \sum_{\hat{\sigma}^t \in S^t} p(\hat{\sigma}^t \cap \sigma^{t-1}) = p\left(\left(\cup \hat{\sigma}^t \in S^t \right) \cap \sigma^{t-1} \right) = p(\sigma^{t-1}).$$

$$\begin{aligned}
\text{Then,} \quad & \forall \sigma^t, \frac{q(\sigma^t)}{q(\sigma^{t-1})} =^{Eq.\ condition} q(\sigma_t|\sigma^{t-1}) \\
\Leftrightarrow \quad & \forall \sigma^t, \frac{p^{NNL}(\sigma^t)}{p^{NNL}(\sigma^{t-1})} = \frac{\frac{q(\sigma^t)}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}}{\frac{q(\sigma^{t-1})}{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})}} = q(\sigma_t|\sigma^{t-1}) \frac{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)} \\
\Leftrightarrow \quad & \forall \sigma^{t-1}, \sum_{\sigma_t} \frac{p^{NNL}(\sigma^{t-1}, \sigma_t)}{p^{NNL}(\sigma^{t-1})} = \sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \frac{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)} \\
\underbrace{\Leftrightarrow}_{\text{If } H_0 \text{ is true}} \quad & \forall \sigma^{t-1}, 1 = \sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \frac{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)} \\
\Leftrightarrow \quad & \forall \sigma^{t-1}, \frac{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})} = \sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \\
\Leftrightarrow \quad & \forall \tilde{\sigma}^{t-1}, \bar{\sigma}^{t-1}, \ \sum_{\sigma_t} q(\sigma_t|\tilde{\sigma}^{t-1}) = \frac{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})} = \sum_{\sigma_t} q(\sigma_t|\bar{\sigma}^{t-1});
\end{aligned}$$

the last equality tells us that the risk-free rate is independent of histories and thus of consumption shares. This property is known to hold only in log-economies. □

Proof of Proposition 2 and 3:

Proof. Substituting $c_t^i(\sigma)^{-\gamma}$ for $u^i(\sigma^t)'$ and $u^i(c_0^i)$ for λ_i in Equation 7,

$$\beta^t p^j(\sigma^t) c_t^j(\sigma)^{-\gamma} = (c_0^j)^{-\gamma} q(\sigma^t) \tag{8}$$

taking the ratio of traders i, j' FOCs: $\frac{\beta^t p^i(\sigma^t) c_t^i(\sigma)^{-\gamma}}{\beta^t p^j(\sigma^t) c_t^j(\sigma)^{-\gamma}} = \frac{(c_0^i)^{-\gamma} q(\sigma^t)}{(c_0^j)^{-\gamma} q(\sigma^t)}$; solving for $c^i(\sigma^t)$:

$$c_t^i(\sigma) = \left(\frac{p^i(\sigma^t)}{p^j(\sigma^t)} \right)^{\frac{1}{\gamma}} \frac{c_0^i}{c_0^j} c_t^j(\sigma). \tag{9}$$

Substituting Equation 9 in the market-clearing condition (which holds with equality because of monotonicity of u^i): $1 = \sum_{i\in\mathcal{I}} c_t^i(\sigma) = c_t^j(\sigma)\frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i}{p^j(\sigma^t)^{\frac{1}{\gamma}} c_0^j}$; solving for $c_t^j(\sigma)$:

$$c_t^j(\sigma) = \frac{p^j(\sigma^t)^{\frac{1}{\gamma}} c_0^j}{\sum_{i\in\mathcal{I}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i}. \tag{10}$$

Substituting $c_t^j(\sigma)$ in Equation 8 and rearranging, I obtain

$$q(\sigma^t) = \beta^t \left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\right)^{\gamma} \tag{11}$$

The result follows substituting Equations 11 in Definition 2 and 3, respectively. □

Lemma 3. *Under **A0-A5**, if agents' utilities are CRRA, for all (t,σ),*

$$\begin{array}{l} \forall i, \gamma^i \geq 1 \Rightarrow \frac{1}{\beta}\sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \leq 1 \\ \forall i, \gamma^i \leq 1 \Rightarrow \frac{1}{\beta}\sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \geq 1 \end{array},$$

with equality if and only if ether the consumption share is degenerate, or $\gamma^i = 1$ for all agents, or all agents have identical beliefs.

Proof. In this proof I omit the conditioning notation for prices and probabilities. That is, for $j \in \mathcal{I} \cup SNNL, p^i(\sigma_t|\sigma^{t-1}) := p^i(\sigma_t|)$ and $q(\sigma_t|\sigma^{t-1}) := q(\sigma_t|)$.
On every equilibrium path $\forall(t,\sigma)$ and for all i,

$$c_t^i(\sigma) = \left(\frac{\beta p^i(\sigma_t|)}{q(\sigma_t|)}\right)^{\frac{1}{\gamma^i}} c_{t-1}^i(\sigma).$$

Multiplying on both sides by $\frac{q(\sigma_t|)}{\beta}$,

$$\frac{q(\sigma_t|)}{\beta} c_t^i(\sigma) = p^i(\sigma_t|)^{\frac{1}{\gamma^i}} \left(\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}} c_{t-1}^i(\sigma).$$

Summing on both sides over all the agents,

$$\frac{q(\sigma_t|)}{\beta} \sum_{i\in\mathcal{I}} c_t^i(\sigma) = \sum_{i\in\mathcal{I}} p^i(\sigma_t|)^{\frac{1}{\gamma^i}} \left(\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}} c_{t-1}^i(\sigma).$$

Dividing on both sides by the aggregate endowment (which is constant over t)

$$\frac{q(\sigma_t|)}{\beta} = \sum_{i\in\mathcal{I}} p^i(\sigma_t|)^{\frac{1}{\gamma^i}} \left(\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}} \phi_{t-1}^i,$$

where $[\phi_{t-1}^1, ..., \phi_{t-1}^I]$ is the consumption shares distribution in $(t-1, \sigma^{t-1})$.
Summing on both sides over the states:

$$\sum_{\sigma_t} \frac{q(\sigma_t|)}{\beta} = \sum_{i\in\mathcal{I}} \sum_{\sigma_t} p^i(\sigma_t|)^{\frac{1}{\gamma^i}} \left(\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}} \phi_{t-1}^i.$$

Multiplying the right-hand side by $\frac{\prod_{k\in\mathcal{I}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^k}}}{\prod_{j\in\mathcal{I}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^j}}}=1$ we can express the left-hand side as a function of p^{SNNL}.

$$\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}=\sum_{i\in\mathcal{I}}\sum_{\sigma_t}p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\phi^i_{t-1}\frac{\prod_{k\in\mathcal{I}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^k}}}{\prod_{j\neq i}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^j}}}. \tag{12}$$

- Let us focus on the case in which $\forall i,\gamma^i\geq 1$.

  Let $i^*:=\arg\max_{i\in\mathcal{I}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}}$, so that $\forall k,i\in\mathcal{I}$, $\frac{\prod_{k\neq i^*}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^k}}}{\prod_{j\neq i}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^j}}}\leq 1$.

  It follows that

$$\begin{aligned}\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}&=\sum_{i\in\mathcal{I}}\sum_{\sigma_t}p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\phi^i_{t-1}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^{i^*}}}\frac{\prod_{k\neq i^*}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}}}{\prod_{j\neq i}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^j}}}\\&\leq\sum_{i\in\mathcal{I}}\sum_{\sigma_t}p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\phi^i_{t-1}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^{i^*}}}.\end{aligned}$$

  Rearranging,

$$\begin{aligned}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{\frac{1}{\gamma^{i^*}}}&\leq\sum_{i\in\mathcal{I}}\sum_{\sigma_t}p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\phi^i_{t-1}\\&\leq^{(a)}\sum_{i\in\mathcal{I}}\sum_{\sigma_t}\left(\frac{1}{\gamma^i}p^i(\sigma_t|)+\left(1-\frac{1}{\gamma^i}\right)p^{SNNL}(\sigma_t|)\right)\phi^i_{t-1}=1\\\Rightarrow\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}&\leq 1.\end{aligned} \tag{13}$$

  $(a):\forall i\in\mathcal{I},\gamma^i\geq 1\Rightarrow\forall\sigma_t,\ p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\leq\frac{1}{\gamma^i}p^i(\sigma_t|)+\left(1-\frac{1}{\gamma^i}\right)p^{SNNL}(\sigma_t|)$, because strict concavity of log ensures that

$$\begin{aligned}\ln\left(p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\right)&=\frac{1}{\gamma^i}\ln p^i(\sigma_t|)+\left(1-\frac{1}{\gamma^i}\right)\ln p^{SNNL}(\sigma_t|)\\&\leq\ln\left(\frac{1}{\gamma^i}p^i(\sigma_t|)+\left(1-\frac{1}{\gamma^i}\right)p^{SNNL}(\sigma_t|)\right).\end{aligned}$$

- Let's focus on the case in which $\forall i,\gamma^i\leq 1$.

  Let $i^{**}:=\arg\min_{i\in\mathcal{I}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^i}}$; thus $\forall k,i\in\mathcal{I}$, $\frac{\prod_{k\neq i^{**}}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^k}}}{\prod_{j\neq i}\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{1-\frac{1}{\gamma^j}}}\geq 1$.

  Proceeding as above, we obtain the opposite inequality:

$$\left(\sum_{\sigma_t}\frac{q(\sigma_t|)}{\beta}\right)^{\frac{1}{\gamma^{i^{**}}}}\geq\sum_{i\in\mathcal{I}}\sum_{\sigma_t}p^i(\sigma_t|)^{\frac{1}{\gamma^i}}p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}}\phi^i_{t-1}. \tag{14}$$

The result follows by showing that

$$\gamma^i \leq 1 \ \forall i \Rightarrow \ln \sum_{i\in\mathcal{I}} \sum_{\sigma_t} p^i(\sigma_t|)^{\frac{1}{\gamma^i}} p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}} \phi^i_{t-1} \geq 0.$$

For convenience, let $\forall i, \eta_i := \frac{1}{\gamma^i}$; so that $\forall i, \eta_i \in (1, \infty)$.

$$\begin{aligned}
\ln \sum_{i\in\mathcal{I}} \sum_{\sigma_t} p^i(\sigma_t|)^{\frac{1}{\gamma^i}} p^{SNNL}(\sigma_t|)^{1-\frac{1}{\gamma^i}} \phi^i_{t-1} &= \ln \sum_{i\in\mathcal{I}} \sum_{\sigma_t} \frac{p^i(\sigma_t|)^{\eta_i}}{p^{SNNL}(\sigma_t|)^{\eta_i-1}} \phi^i_{t-1} \\
&\geq^{(a)} \sum_{i\in\mathcal{I}} \phi^i_{t-1} \ln \sum_{\sigma_t} \frac{p^i(\sigma_t|)^{\eta_i}}{p^{SNNL}(\sigma_t|)^{\eta_i-1}} \\
&= \sum_{i\in\mathcal{I}} (\eta_i - 1)\phi^i_{t-1} \left(\frac{1}{\eta_i - 1} \ln \sum_{\sigma_t} \frac{p^i(\sigma_t|)^{\eta_i}}{p^{SNNL}(\sigma_t|)^{\eta_i-1}} \right) \\
&=^{(b)} \sum_{i\in\mathcal{I}} (\eta_i - 1)\phi^i_{t-1} D_{\eta^i}(p^i_t || p^{RN}_t) \\
&\geq^{(c)} 0.
\end{aligned}$$

(a): By concavity of log.
(b): Recognizing the definition of the Rényi divergence ($D_{\eta^i}(p^i_t||p^{RN}_t)$) between p^i_t and p^{RN}_t (Rényi, 1961; Van Erven and Harremos, 2014).
(c): Rény divergence is weakly positive, it equals 0 if and only if $p^i = p^{SNNL}$ (Van Erven and Harremos, 2014).

An inspection of Equation (12) shows that equality holds if and only if
— all agents have identical beliefs because

$$\forall i, p^i_t = p_t = p^{SNNL}_t \Rightarrow \forall i, \left(\sum_{\sigma_t} \frac{q(\sigma_t|)}{\beta} \right)^{1-\frac{1}{\gamma^i}} = \sum_{i\in\mathcal{I}} \sum_{\sigma_t} p_t(\sigma_t|)^{\frac{1}{\gamma^i}} p_t(\sigma_t|)^{1-\frac{1}{\gamma^i}} \phi^i_{t-1} = 1;$$

— or the consumption share distribution is degenerate because

$$\phi^i_{t-1} = 1 \Rightarrow p^i_t = p_t = p^{SNNL}_t \Rightarrow \left(\sum_{\sigma_t} \frac{q(\sigma_t|)}{\beta} \right)^{1-\frac{1}{\gamma^i}} = \sum_{i\in\mathcal{I}} \sum_{\sigma_t} p_t(\sigma_t|)^{\frac{1}{\gamma^i}} p_t(\sigma_t|)^{1-\frac{1}{\gamma^i}} \phi^i_{t-1} = 1;$$

— or $\gamma^i = 1$ for all agents because $\sum_{\sigma_t} \frac{q(\sigma_t|)}{\beta} = \sum\limits_{i\in\mathcal{I}} \sum\limits_{\sigma_t} p^i(\sigma_t|) = 1$.

□

Lemma 4. *In an economy that satisfies* **A0-A5***, for all* $\sigma \in S^\infty$, $\begin{cases} \liminf_t \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i_t(\sigma))} > a > 0 \\ \limsup_t \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i_t(\sigma))} < b < \infty \end{cases}$.

Proof.

- $\forall \sigma \in S^\infty, \liminf \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i_t(\sigma))} \geq \min_{[c^1,\ldots,c^I]} \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i)} > 0$ because $\sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i)} = 0$ if and only if $\forall i,\ u'_i(c^i) = \infty \Leftrightarrow^{A1} \forall i, c^i = 0$, which violates market-clearing $\left(\forall t, \sum_{i\in\mathcal{I}} c^i = \sum_{i\in\mathcal{I}} e^i =^{A3} 1\right)$.
- $\forall \sigma \in S^\infty, \limsup \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i(\sigma))} \leq \max_{[c^1,\ldots,c^I]} \sum_{i\in\mathcal{I}} \frac{1}{u'_i(c^i)} < |\mathcal{I}| \max_i \frac{1}{u'_i(1)} < \infty$ because market clearing and **A3** implies $\max_i c^i = 1$; and **A1** implies $\forall i, \max_{c\leq 1} \frac{1}{u^i(c)'} = \frac{1}{u^i(1)'} < \infty$.

□

Proof of Theorem 1:

Proof. Definition 3 allows to rewrite $\ln \frac{p^{BMA}(\sigma^t)}{p^{SNNL}(\sigma^t)}$ as follows:

$$\ln \frac{p^{BMA}(\sigma^t)}{p^{SNNL}(\sigma^t)} = \sum_{\tau=1}^{t} \sum_{\sigma_\tau} I_{\sigma_\tau} \ln \frac{p^{BMA}(\sigma^\tau|\sigma^{\tau-1})}{\frac{q(\sigma^\tau|\sigma^{\tau-1})}{\sum_{\hat{\sigma}_t} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}} \qquad = \ln \frac{\beta^t p^{BMA}(\sigma^t)}{q(\sigma^t)} + \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right)$$

$$\underbrace{=}_{\substack{\text{by Massari (2017)} \\ \text{Corollary 1}}} \quad O(1) + \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right)$$

- $p^{SNNL}(\sigma^t)$ is $\mathcal{P}$-efficient.
  I need to show that if there is a trader i^* that knows the truth, P, then

$$\sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \asymp^{P\text{-a.s}} 1. \tag{15}$$

  Taking ratio of the FOCs, in every equilibrium path and for every agent $i \neq i^*$ it holds

$$\frac{\frac{1}{u^i(c_t^i(\sigma))'}}{\frac{1}{u^{i^*}(c_t^{i^*}(\sigma))'}} = \frac{p^i(\sigma^t)}{P(\sigma^t)} \frac{\frac{1}{u^i(c_0^i)'}}{\frac{1}{u^{i^*}(c_0^{i^*})'}};$$

  so, $\frac{1}{u^{i^*}(c_t^{i^*}(\sigma))'} \to^{P\text{-a.s.}} \frac{1}{u^{i^*}(1)'}$ at an exponential rate.
  By Equation 6 and the definition of $q(\cdot|)$ we have that for all (t, σ)

$$q(\sigma_t|) = \frac{q(\sigma^t)}{q(\sigma^{t-1})} = \frac{\frac{\beta^t \sum_{i\in\mathcal{P}} p^i(\sigma^t) \frac{1}{u^i(c_0^i)'}}{\sum_{j\in\mathcal{I}} \frac{1}{u^j(c_t^j(\sigma))'}}}{\frac{\beta^{t-1} \sum_{k\in\mathcal{P}} p^k(\sigma^{t-1}) \frac{1}{u^k(c_0^k)'}}{\sum_{l\in\mathcal{I}} \frac{1}{u^l(c_{t-1}^l(\sigma))'}}} = \beta \sum_{i\in\mathcal{P}} p^i(\sigma^t|) \frac{p^i(\sigma^{t-1}) \frac{1}{u^i(c_0^i)'}}{\sum_{k\in\mathcal{P}} p^k(\sigma^{t-1}) \frac{1}{u^k(c_0^k)'}} \frac{\sum_{l\in\mathcal{I}} \frac{1}{u^l(c_{t-1}^l(\sigma))'}}{\sum_{j\in\mathcal{I}} \frac{1}{u^j(c_t^j(\sigma))'}}.$$

  Thus, $\frac{1}{\beta} \sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1}) \to^{P\text{-a.s.}} 1$ at an exponential rate which is sufficient to bound the sum in Equation 15.
- $p^{SNNL}_{\gamma^i=1}$ is universal-efficient. By Lemma 3, $\gamma = 1 \Rightarrow \forall (t, \sigma), \frac{1}{\beta} \sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1}) = 1$, so that

$$\forall (t, \sigma), \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) = 0 \asymp 1.$$

- $p^{SNNL}_{\gamma^i<(>)1}$ is sub-efficient (super-efficient).
  The result follows by showing that $\exists \hat{P} : \gamma < (>) 1 \Rightarrow \sum_{\tau=1}^{t} ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \to +(-)\infty$.
  Proceeding by steps:
  i) $\sum_{\tau=1}^{t} ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \to \pm\infty$ if consumption shares do not concentrate on one trader.
  Proof: By Lemma 3, $\forall i \in \mathcal{I}, \gamma^i < (>) 1 \Leftrightarrow \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right) \geq (\leq) 0$, with equality if and only if the consumption-share distribution is degenerate (or beliefs are identical). Thus, if for all $i \in \mathcal{I}, \gamma^i < (>) 1$, all terms of the sum have the same sign. So, if consumption shares do not concentrate on a unique trader, $\exists \eta > 0 : |\ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right)| > \eta$ infinitely often and the sum diverges.

ii) $p^{BMA}(\sigma^t)$'s prior, $C_{t,\gamma=1}$, does not eventually concentrate on a unique trader if and only if, $\forall \gamma \in (0,+\infty)$, $p^{SNNL}(\sigma^t)$'s generalized prior, $C_{t,\gamma}$, does not concentrate on a unique trader.[12]

Proof:$C_{t,\gamma=1}$ does not concentrate on a unique trader

$$\Leftrightarrow \exists \eta > 0, \exists i,j \in \mathcal{P} : \limsup \frac{p^i(\sigma^t)}{\sum_{i\in\mathcal{P}} p^i(\sigma^t)} > \eta \text{ and } \limsup \frac{p^j(\sigma^t)}{\sum_{i\in\mathcal{P}} p^i(\sigma^t)} > \eta$$

$$\Leftrightarrow \exists \eta_\gamma > 0 : \limsup \frac{p^i(\sigma^t)^{\frac{1}{\gamma}}}{\sum_{i\in\mathcal{P}} p^i\sigma^{t\,\frac{1}{\gamma}}} = c^i_{\gamma,t}(\sigma) > \eta_\gamma \text{ and } \limsup \frac{p^j(\sigma^t)^{\frac{1}{\gamma}}}{\sum_{i\in\mathcal{P}} p^i\sigma^{t\,\frac{1}{\gamma}}} = c^j_{\gamma,t}(\sigma) > \eta_\gamma$$

$\Leftrightarrow C_{t,\gamma}$ does not concentrate on a unique trader.

iii) $\exists \hat{P}$ such that $C_{1,t}$ does not concentrate on a unique trader:
Proof: The proof is constructive. Let $i,j \in \mathcal{P}$ be two orthogonal models, it is easy to verify that the Bayesian posterior does not converges P_{θ_0}-a.s. for P_{θ_0} defined recursively as follows:

$$\forall \sigma^{t-1}, P_{\theta_0}(\sigma_t|\sigma^{t-1}) := \begin{cases} p^i(\sigma_t|\sigma^{t-1}), & \text{if } p^j(\sigma^{t-1}) \in \arg\max_{i\in\mathcal{P}} \\ p^j(\sigma_t|\sigma^{t-1}), & \text{otherwise} \end{cases}.$$

Concentration cannot occur because j has maximal expected likelihood (thus highest expected survival index) whenever his likelihood is comparatively low, however he cannot dominate since he cannot beat agent i, because i becomes the most accurate as soon agent j consumption share passes a threshold (i.e., his likelihood becomes comparative large).

□

Proof of Theorem 2:

Proof. Substituting Equation 6 in the definition of p^{NNL}.

$$p^{NNL}(\sigma^t) = \frac{q(\sigma^t)}{\sum_{\sigma^t} q(\sigma^t)} = \frac{\dfrac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^t))}}}{\beta^t \sum_{\hat{\sigma}^t\in\mathcal{S}^t}\left(\dfrac{\sum_{k\in\mathcal{I}} p^k(\hat{\sigma}^t)\frac{1}{u_k'(c_0^i)}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c(\hat{\sigma}^t))}}\right)} \tag{16}$$

Let $\liminf \sum_{i\in\mathcal{I}} \frac{1}{u_i'(c(\hat{\sigma}^t))} = a$ and $\limsup \sum_{i\in\mathcal{I}} \frac{1}{u_i'(c(\hat{\sigma}^t))} = b$; by Lem.4; $0 < a \le b < \infty$, thus

$$p^{NNL}(\sigma^t) \in [\frac{\dfrac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{b}}{\beta^t \sum_{\sigma\in\mathcal{S}^t}\left(\dfrac{\sum_{k\in\mathcal{I}} p^k(\sigma^t)\frac{1}{u_k'(c_0^k)}}{a}\right)}, \frac{\dfrac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{a}}{\beta^t \sum_{\sigma\in\mathcal{S}^t}\left(\dfrac{\sum_{k\in\mathcal{I}} p^k(\sigma^t)\frac{1}{u_k'(c_0^k)}}{b}\right)}]$$

$$\Rightarrow p^{NNL}(\sigma^t) \in [\frac{a}{b}\sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}, \frac{b}{a}\sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}]$$

$$\Leftrightarrow \ln \frac{p^{BMA}(\sigma^t)}{p^{NNL}(\sigma^t)} \asymp 1.$$

□

[12] The proof slightly differs for $\gamma = 0$ because I need the stronger condition that the model with the highest likelihood changes infinitely to ensure $\sum_{\tau=1}^t \ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \to \pm\infty$ (See Massari (2017)).

Proof of Proposition 4:

Proof. As in the proof of Th.1: $\ln \frac{p^{BMA}(\sigma^t)}{p^{SNNL}(\sigma^t)} = O(1) + \sum_{\tau=1}^{t} \ln \left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta} \right)$.

- Part *i*) mimics the step of Theorem 1, except that non-concentration is now assumed.
- Part *ii*) mimics the proof of $\mathcal{P}$-efficiency in Theorem 1, except that the exponential concentration rate is assumed, rather than deduced from the fact that one agents knows the truth. For an application, consider Example 1:

$$e^{-\sum_{\tau=1}^{t} \ln\left(q(a|\sigma^{\tau-1})+q(b|\sigma^{\tau-1})\right)} = EXP\left[-\sum_{\tau=1}^{t} \ln \frac{\left(\frac{1}{2}\left(\frac{2}{3}\right)^{\frac{\tau}{\gamma}} + \frac{1}{2}\left(\frac{1}{3}\right)^{\frac{\tau}{\gamma}}\right)^{\gamma} + \left(\frac{1}{2}\left(\frac{1}{3}\right)^{\frac{1}{\gamma}}\left(\frac{2}{3}\right)^{\frac{\tau-1}{\gamma}} + \frac{1}{2}\left(\frac{2}{3}\right)^{\frac{1}{\gamma}}\left(\frac{1}{3}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}{\left(\frac{1}{2}\left(\frac{2}{3}\right)^{\frac{\tau-1}{\gamma}} + \frac{1}{2}\left(\frac{1}{3}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}\right]$$

$$= EXP\left[-\sum_{\tau=1}^{t} \ln \frac{\left(\left(\frac{2}{3}\right)^{\frac{1}{\gamma}} + \left(\frac{1}{3}\right)^{\frac{1}{\gamma}}\left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma} + \left(\left(\frac{1}{3}\right)^{\frac{1}{\gamma}} + \left(\frac{2}{3}\right)^{\frac{1}{\gamma}}\left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}{\left(1 + \left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}\right]$$

Taylor expanding the two terms on the numerator around $\frac{2}{3}^{\frac{1}{\gamma}}$ and $\frac{1}{3}^{\frac{1}{\gamma}}$ and the term in the denumerator around 1, respectively, it follows that $\exists \eta \in (0, \frac{1}{2})$:

$$EXP\left[-\sum_{\tau=1}^{t} \ln \frac{\left(\left(\frac{2}{3}\right)^{\frac{1}{\gamma}} + \left(\frac{1}{3}\right)^{\frac{1}{\gamma}}\left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma} + \left(\left(\frac{1}{3}\right)^{\frac{1}{\gamma}} + \left(\frac{2}{3}\right)^{\frac{1}{\gamma}}\left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}{\left(1 + \left(\frac{1}{2}\right)^{\frac{\tau-1}{\gamma}}\right)^{\gamma}}\right] \in \left[e^{-\sum_{\tau=1}^{t}\left(\frac{1}{2}+\eta\right)^{\tau}}; e^{-\sum_{\tau=1}^{t}\left(\frac{1}{2}-\eta\right)^{\tau}}\right] = O(1).$$

□

Proof of Proposition 5:

Proof. $\gamma = 1 \Rightarrow p_\gamma^{NNL} = p_\gamma^{SNNL} = p^{BMA}$ which is convex.

Generically, p^{SNNL} is not convex: For intuition, let start with the case of identical CRRA economies. If the state space have at least 3 states and all models in $\mathcal{P}$ attach the same probability to state $\hat{s}, p(\hat{s})$ then p_γ^{SNNL} is not convex in non-log economies because:

$$\begin{aligned} p^{SNNL}(\hat{s}_t|\sigma^{t-1})_\gamma &= \frac{\left(\sum_{i\in\mathcal{P}} p^i(\hat{s}_t|\sigma^{t-1})^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma}{\sum_{\hat{\sigma}_t\in\mathcal{S}}\left(\sum_{i\in\mathcal{P}} p^i(\hat{\sigma}_t)^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma} \\ &= \frac{p^i(\hat{s}_t)}{\sum_{\hat{\sigma}_t\in\mathcal{S}}\left(\sum_{i\in\mathcal{P}} p^i(\hat{\sigma}_t)^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma} \\ &\neq p^i(\hat{s}_t) \end{aligned}$$

where the last inequality follows because $\sum_{\hat{\sigma}_t\in\mathcal{S}}\left(\sum_{i\in\mathcal{P}} p^i(\hat{\sigma}_t)^{\frac{1}{\gamma}} c^i_{\gamma,t-1}(\sigma)\right)^\gamma \neq 1$ for all non degenerate consumption shares.

For the general case, I first show that there are cases in which the normalizing factor in

p^{SNNL} is different from 1. Then generalize the above example.

$$\forall \sigma^t,\ p^{SNNL}(\sigma_t|\sigma^{t-1}) \underbrace{=}_{\substack{\text{by} \\ \text{Lemma 1}}} \frac{p^{NNL}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NNL}(\sigma^{t-1}, \hat{\sigma}_t)}$$

$$\underbrace{=}_{\substack{\text{by} \\ \text{Equation 16}}} \frac{\frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^t))}}}{\sum_{\hat{\sigma}_t} \frac{\sum_{k\in\mathcal{I}} p^k(\sigma^t)\frac{1}{u_k'(c_0^k)}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c^l(\sigma^{t-1}\hat{\sigma}_t))}}} = \frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)\frac{1}{u_i'(c_0^i)}}{\sum_{\hat{\sigma}_t}\sum_{k\in\mathcal{I}} p^k(\sigma^t)\frac{1}{u_k'(c_0^k)}\frac{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\sigma_t))}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c^l(\sigma^{t-1}\hat{\sigma}_t))}}}$$

$$= \frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t|\sigma^{t-1})g^i(\sigma^{t-1})}{\sum_{\hat{\sigma}_t}\sum_{k\in\mathcal{I}} p^k(\hat{\sigma}_t|\sigma^{t-1})g^k(\sigma^{t-1})\frac{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\sigma_t))}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c^l(\sigma^{t-1}\hat{\sigma}_t))}}};$$

where $\forall i, g^i(\sigma^{t-1}) := \frac{\frac{p^i(\sigma^{t-1})}{u_i'(c_0^i)}}{\sum_{j\in\mathcal{I}}\frac{p^j(\sigma^{t-1})}{u_j'(c_0^j)}}$ can be thought of as the weight of a prior distribution $G(\sigma^{t-1})$. Suppose that the distribution of consumption is not degenerate and such that $\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\sigma_t))} \in \text{argmax}_{\hat{\sigma}} \sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\hat{\sigma}_t))}$. It follows that

$$\sum_{\hat{\sigma}_t}\sum_{k\in\mathcal{I}} p^k(\hat{\sigma}_t|\sigma^{t-1})g^k(\sigma^{t-1})\frac{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\sigma_t))}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c^l(\sigma^{t-1}\hat{\sigma}_t))}} \neq 1.$$

So, if the state space have at least 3 states and all models in $\mathcal{P}$ attach the same probability to state $\hat{s}, p(\hat{s})$ then p^{SNNL} is not convex in non-log economies because:

$$p^{SNNL}(\hat{s}_t|\sigma^{t-1}) = \frac{\sum_{i\in\mathcal{I}} p^i(\hat{s}_t|\sigma^{t-1})g^i(\sigma^{t-1})}{\sum_{\hat{\sigma}_t}\sum_{k\in\mathcal{I}} p^k(\hat{\sigma}_t|\sigma^{t-1})g^k(\sigma^{t-1})\frac{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c^j(\sigma^{t-1}\hat{s}_t))}}{\sum_{l\in\mathcal{I}} \frac{1}{u_l'(c^l(\sigma^{t-1}\hat{\sigma}_t))}}} \neq p(\hat{s}).$$

Generically, p^{NNL} is not convex because $p^{SNNL}(\sigma_1) = p^{NNL}(\sigma_1)$.

□

References

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64(3):423–458.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics*, 37(1):51–58.

Blume, L. and Easley, D. (1993). Economic natural selection. *Economics Letters*, 42(2):281–289.

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.

Breiman, L. et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231.

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

Cremers, K. M. (2002). Stock return predictability: A bayesian model selection perspective. *The Review of Financial Studies*, 15:1223–1249.

Dawid, A. P. (1982). The well-calibrated bayesian. *Journal of the American Statistical Association*, 77(379):605–610.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, pages 278–292.

De Rooij, S., Van Erven, T., Grünwald, P. D., and Koolen, W. M. (2014). Follow the leader if you can, hedge if you must. *The Journal of Machine Learning Research*, 15(1):1281–1316.

Epstein, L. G., Noor, J., and Sandroni, A. (2008). Non-bayesian updating: a theoretical framework. *Theoretical Economics*, 3(2):193–229.

Foster, D. P. and Vohra, R. (1999). Regret in the on-line decision problem. *Games and Economic Behavior*, 29(1):7–35.

Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139.

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

Fudenberg, D. and Levine, D. (1998). Learning in games. *European economic review*, 42(3):631–639.

Ghirardato, P. (2002). Revisiting savage in a conditional world. *Economic Theory*, 20(1):83–92.

Gilboa, I. and Marinacci, M. (2011). Ambiguity and the bayesian paradigm. *Chapter*, 7:179–242.

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

Grünwald, P. (2012). The safe bayesian. In *Algorithmic Learning Theory*, pages 169–183. Springer.

Grünwald, P. and van Ommen, T. (2014). Inconsistency of bayesian inference for misspecified linear models, and a proposal for repairing it. *arXiv preprint arXiv:1412.3730*.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kolmogorov, A. (1933). Grundbegriffe der wahrscheinlichkeitstheorie. *Ergebnisse Mathematische*, 2.

Lehrer, E. and Teper, R. (2016). Who is a bayesian? *Working paper*.

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

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

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

Massari, F. (2019). Market selection in large economies: A matter of luck. *Theoretical Economics*, 14(2):437–473.

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

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

Rabin, M. et al. (2000). *Inference by believers in the law of small numbers.* Institute of Business and Economic Research.

Rényi, A. (1961). On measures of entropy and information. Technical report, HUNGARIAN ACADEMY OF SCIENCES Budapest Hungary.

Rissanen, J. (1986). Stochastic complexity and modeling. *The Annals of Statistics*, pages 1080–1100.

Roos, T. and Rissanen, J. (2008). On sequentially normalized maximum likelihood models. *Workshop on Information Theoretic Methods in Science and Engineering.*

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

Samuelson, P. A. (1971). The ?fallacy? of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of Sciences*, 68(10):2493–2496.

Samuelson, P. A. (1979). Why we should not make mean log of wealth big though years to act are long. *Journal of Banking & Finance*, 3(4):305–307.

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

Savage, L. J. (1954). *The foundations of statistics.* John Wiley & Sons.

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

Shtar'kov, Y. M. (1987). Universal sequential coding of single messages. *Problemy Peredachi Informatsii*, 23(3):3–17.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge.

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting*, 1:135–196.

Van Erven, T. and Harremos, P. (2014). Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.

Vovk, V. G. (1990). Aggregating strategies. In *Proc. Third Workshop on Computational Learning Theory*, pages 371–383. Morgan Kaufmann.

Wald, A. (1947). An essentially complete class of admissible decision functions. *The Annals of Mathematical Statistics*, 18(4):549–555.