# Price Probabilities: A class of Bayesian and non-Bayesian prediction rules

Filippo Massari*

*University of New South Wales*

February 6, 2018

**Abstract**

I use the standard machinery of dynamic general equilibrium models to generate a rich class of probabilities and discuss their properties. This class includes 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, I demonstrate how rules that underreact to new information can significantly outperform Bayes'. Because underreaction is never worse and sometimes better than Bayesian predictions, 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

*I am thankful 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.

# 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 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 is confirmed by 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: 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).

Non-Bayesian members of the price probability class are either time-inconsistent or (over)underreact to empirical evidence. 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"? I believe the answer to this question should be pragmatic. A rule is rational if an agent cannot be persuaded by a logical argument to use an alternative rule (Gilboa, 2015).

Taking advantage of the almost universal predominance of Bayes' rule and its sound axiomatic foundation, it is natural to use BMA as a benchmark for rationality. A prediction rule is rational if, in every sequence, its predictions are qualitatively at least as accurate

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

as the one obtained via Bayes' rule using the same information. To compare the relative performance of members of price probability and BMA, I propose an accuracy criterion. Following an established tradition across fields, my criterion relies on asymptotic likelihood comparisons. In every sequence, I compare the likelihood of a probability mixture of models in a certain (support) set $\mathcal{P}$, against the likelihood of BMA with a regular prior on the same support.

- 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 ($\hat{P} \notin \mathcal{P}$), in which $p$ is overwhelmingly more accurate than BMA $\hat{P}$-a.s..

- 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 ($P \in \mathcal{P}$), but there are probabilities, $\hat{P} \notin \mathcal{P}$, such that it diverges to infinity $\hat{P}$-almost surely.

I adopt this accuracy criterion as my criterion of pragmatic rationality. Underreacting members of price probability are super-efficient and thus rational. For example, a situation in which an underreacting rule outperforms Bayes' is as follows. Suppose you observe repeated tosses of a fair coin but erroneously believe the probability of Heads to be either $\frac{1}{3}$ or $\frac{2}{3}$ with equal prior probability. Bayes' rule gives predictions that most of the time are arbitrarily close to either $\frac{1}{3}$ or $\frac{2}{3}$. An underreacting rule gives predictions that are closer to $\frac{1}{2}$ than Bayes' and therefore more accurate. Time-inconsistent members are universal-efficient, thus rational in all settings in which time inconsistency cannot be used to construct arbitrages. Finally, overreacting members are not rational because they are sub-efficient.

Contrary to the axiomatic approach to learning in the economics literature (Ghirardato, 2002; Gilboa and Marinacci, 2011), my approach is purely pragmatic (closer to Machine Learning and Reinforcement Learning, Breiman et al., 2001; Sutton and Barto, 1998). A rule is desirable not for the axioms it satisfies but for its practical performance. 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 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 with respect to his preferences.

My model of (over)underreaction is consistent with the axiomatization of Epstein (2006). It provides a general class of predictions in which (over)underreaction is path-dependent. If there is a unique best model in the support, the posterior concentrates on it and (over)underreaction vanishes; otherwise, it persists. Moreover, I extend the analysis of Epstein et al. (2008) characterizing the performance of (over)underreacting models under misspecification. Notably, I show that the underreacting rule can be more accurate than Bayes' under misspecification.

My results are inspired by known 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) differs from log-likelihood, an agent can be better off abandoning Bayes' rule for a rule that gives less weight to past realizations. My contribution is to show that, in certain cases of model misspecification, underreacting rules can improve on Bayes' even if their performance is measured according to the log-likelihood criterion.

Section 2 introduces the notation and known probability mixture models. Section 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 this section and consider Propositions 1 and 2 as definitions. Sections 5, 6, and 7 discuss the accuracy of members of the price

probability class, while Section 8 compares them with Bayes' rule. Proofs are in the online Appendix.

# 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 $(\sigma^{t-1}, \sigma_t)$ is the concatenation of $\sigma^{t-1}$ and $\sigma_t$, i.e. the sequence whose first $t$-1 realizations coincide with $\sigma^{t-1}$ and last element is $\sigma_t$. $\mathcal{C}(\sigma^t)$ is the cylinder set with base $\sigma^t$, $\mathcal{C}(\sigma^t) = \{\sigma \in S^\infty | \, \sigma = (\sigma^t, \ldots)\}$, $\mathcal{F}_t$ the $\sigma$-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 $\sigma$-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 standard definition of probability mixture models, log-regret, and known probability mixture models which belong to the price probability class. I refer to Foster and Vohra (1999) and Grünwald (2007) for a more comprehensive discussion. 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.

- **Probability mixture.** Given a reference set of orthogonal probability measures $\mathcal{P} = \{p^1, ..., p^I\}$ 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.

- **Log-regret**. Given a partial history $\sigma^t$ 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 $\sigma^t$ (i.e. the most accurate model in $\mathcal{P}$ with hindsight) and the probability mixture adopted: given $\sigma^t$, $R(p;\sigma^t) = \sup_{i\in\mathcal{P}}\{\ln \frac{p^i(\sigma^t)}{p(\sigma^t)}\}$. 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 favourable sequence of realizations: $\mathcal{R}(p;t) = \sup_{\sigma^t} R(p;\sigma^t)$. A probability mixture with 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.

- **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:

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

  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

[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).

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:$p^{NML}(.) := \arg\min \mathcal{R}(p;\sigma^t)$. Rissanen (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 $\sigma_t$, 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. A reader who is mostly interested in the properties and performance of members of price probabilities can skip this section and consider Propositions 1 and 2 to be the definitions of the non-Bayesian rules I propose.

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 iid 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 $\sigma^t$, 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 $\sigma_t$, in terms of consumption at period/event $\sigma^{t-1}$.

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 A1-A4 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):

**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 $\sigma^t$, $p^i(\sigma^t) > 0 \Leftrightarrow p^j(\sigma^t) > 0$.

**A3** : The aggregate endowment equals 1 in every period: $\forall \sigma, \forall t, \sum_{i \in \mathcal{I}} e_t^i(\sigma) = 1$.

**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 $C_0 = [c_0^1 \; ... \; c_0^I]' >> 0$. By A3, $\sum_{i \in \mathcal{I}} c_0^i = 1$, 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 traders 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 A1-A4.*

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):*

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

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 $\sigma^t$ 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. In Section 8, we show that a set well behaved conditional probabilities exists if and only if all traders have log-utility. Thus, $p^{NNL}$ typically 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 $\sigma^{t-1}$ one period ahead in state $\sigma_t$, 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 $\sigma_t$, assuming that $t$ is the final horizon:

**Lemma 1.** *In an economy that satisfies A1-A4,*

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

# 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 and to discuss relevant deviations from Bayes' rule.
In what follows, I use the notation:

**Definition 4.** *$p_\gamma^{NNL}$ and $p_\gamma^{SNNL}$ 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 $\gamma$, $\forall i \in \mathcal{I}, u^i(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$.*[3]

## 4.1 NNL in identical CRRA economies: $p_\gamma^{NNL}$

**Proposition 1.** *Given beliefs set $\mathcal{P}$, prior $C_0$, and parameter $\gamma$, $p_\gamma^{NNL}$ is given by:*

$$\forall \sigma^t, \quad p_\gamma^{NNL}(\sigma^t) = \frac{\left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\right)^\gamma}{\sum_{\hat{\sigma}^t\in\mathcal{S}^t}\left(\sum_{i\in\mathcal{I}} p^i(\hat{\sigma}^t)^{\frac{1}{\gamma}} c_0^i\right)^\gamma}. \tag{4}$$

Equation 4 shows that $p_\gamma^{NNL}$ 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 $L_\infty$ norms, respectively, they both belong to price probabilities.

**Corollary 1.** *Given beliefs set $\mathcal{P}$ and prior $C_0$,*
*i) $\gamma = 1$ (log) $\Rightarrow \forall \sigma^t,\ p_1^{NNL}(\sigma^t) = p^{BMA}(\sigma^t)$;*
*ii) $\gamma = 0$ (linear) $\Rightarrow \forall \sigma^t, p_0^{NNL}(\sigma^t) = p^{NML}(\sigma^t)$.*

*Proof.* *i)* Notice that if $\gamma = 1$, the denominator of Eq.4 equals 1, and compare Eq.4 with Eq.1.
*ii)* Notice that $lim_{\gamma\to 0}\left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\right)^\gamma = ||p^i(\sigma^t)||_\infty$: the sup norm; and compare Eq.4 with Eq.2 □

Taking Bayes' rule as a reference point, the effect of gamma on $p^{NNL}$ is qualitatively as follows. In a log-economy ($\gamma = 1$) $p^{NNL}$ coincides with BMA and the interaction between

[3] 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.

prior information ($C_0$) and empirical evidence ($\sigma^t$) is regulated by Bayes' rule. For $\gamma = 0$, $p^{NNL}$ coincides with NML (i.e., it is the optimal probability with respect to worst-case log-regret). Given the 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 $\gamma \neq 1$, $p^{NNL}$ 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 $\mathcal{P}$, $p^{NNL}$ 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: $p_\gamma^{SNNL}$

**Proposition 2.** *Given beliefs set $\mathcal{P}$, prior $C_0$, and parameter $\gamma$, $p_\gamma^{SNNL}$ 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* $c^i_{\gamma,t-1}(\sigma) =^{by\ Eq.11} \frac{p^i(\sigma^{t-1})^{\frac{1}{\gamma}} c^i_0}{\sum_{i\in\mathcal{P}} p^i(\sigma^{t-1})^{\frac{1}{\gamma}} c^i_0}$.

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 $C_{\gamma,t-1}(\sigma)$. The gamma parameter affects the evolution of this distribution. If gamma equals 1 (log), $C^i_{\gamma,t-1}(\sigma)$ coincides with the Bayesian prior distribution obtained from $C_0$ on $\sigma^{t-1}$ (compare with Equation 1). Taking Bayes' rule as my reference point, the effect of gamma on $C_{\gamma,t-1}(\sigma)$ is qualitatively as follows. If gamma is greater than 1, $C^i_{\gamma,t-1}(\sigma)$ gives less weight to empirical evidence than Bayes' because it is less concentrated around the model with the highest likelihood. Conversely, if gamma is lower than 1, $C^i_{\gamma,t-1}(\sigma)$ gives more weight to empirical evidence than Bayes'. The normalizing component and the use of the $\frac{1}{\gamma}$ norm only mitigates this effect.

**Proposition 3.** *Given a belief set $\mathcal{P}$ and a prior $C_0$,*

*i)* $\gamma > 1 \Leftrightarrow p^{SNNL}_{\gamma}$ *underreacts to empirical evidence;*

*ii)* $\gamma = 1$ *(log)* $\Leftrightarrow p^{SNNL}_{\gamma}$ *coincides with Bayesian updating;*

*iii)* $\gamma < 1 \Leftrightarrow p^{SNNL}_{\gamma}$ *overreacts to empirical evidence.*

For intuition, suppose $S := \{a, b\}$ and every probability in $\mathcal{P} := \{p^i : i \in \mathcal{I}\}$ is iid 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:

$$c^i_{\gamma,t-1}(\sigma) = \frac{p^i(\sigma^t)^{\frac{1}{\gamma}} c^i_0}{\sum_{j\in\mathcal{P}} p^j(\sigma^t)^{\frac{1}{\gamma}} c^j_0} = \frac{i^{\frac{t_a}{\gamma}}(1-i)^{\frac{t_b}{\gamma}}}{\sum_{j\in\mathcal{P}} j^{\frac{t_a}{\gamma}}(1-j)^{\frac{t_b}{\gamma}}}. \tag{6}$$

Equation 6 highlights the effect of gamma on the evolution of $C^i_{\gamma,t-1}(\sigma)$. If gamma is smaller than 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 is greater than 1, the model underreacts to empirical evidence: e.g.,$\gamma = 2$ is equivalent to updating using Bayes' rule "counting every past realization as half."[4]

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.* $\forall \sigma^{t-1}$, $p^{SNNL}_0(\sigma_t|\sigma^{t-1}) =^{\text{Lem.1}} \frac{p^{NNL}_0(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NNL}_0(\sigma^{t-1},\hat{\sigma}_t)} =^{\text{Cor.1}} \frac{p^{NML}(\sigma^t)}{\sum_{\hat{\sigma}_t} p^{NML}(\sigma^{t-1},\hat{\sigma}_t)} :=^{\text{Eq.3}} p^{SNML}(\sigma_t|\sigma^{t-1})$.

□

[4] $C^i_{\gamma,t-1}(\sigma)$ 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). $p^{SNNL}$ differs from these algorithms because instead of relying on the generalized prior it directly depends on the sequential normalization of the $\frac{1}{\gamma}$ norm.

# 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). In every sequence, I compare the likelihood of a probability mixture with belief set, $\mathcal{P}$, against the likelihood of BMA with regular prior[5] on the same support.

**Definition 5.** *Let $p^{BMA}(\sigma^t)$ be the likelihood of a BMA with a regular prior on $\mathcal{P}$,*

- *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*

$$\forall P \in \mathcal{P}, \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \asymp^{P\text{-a.s.}} 1, \quad \textit{and} \quad \begin{cases} \nexists \hat{P} : \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \to^{\hat{P}\text{-a.s.}} +\infty; \\ \exists \hat{P} : \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \to^{\hat{P}\text{-a.s.}} -\infty; \end{cases}$$

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

$$\forall P \in \mathcal{P}, \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \asymp^{P\text{-a.s.}} 1, \quad \textit{and} \quad \begin{cases} \exists \hat{P} : \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \to^{\hat{P}\text{-a.s.}} +\infty; \\ \nexists \hat{P} : \lim_{t\to\infty} \ln \frac{p^{BMA}(\sigma^t)}{p(\sigma^t)} \to^{\hat{P}\text{-a.s.}} -\infty; \end{cases}$$

*Where the notation $f(x) \asymp g(x)$ abbreviates $\limsup \frac{f(x)}{g(x)} < +\infty$ and $\liminf \frac{f(x)}{g(x)} > 0$.*

In other words, $p$ is universal-efficient if it is as accurate as the prediction obtained using Bayes' rule in every sequence. A probability mixture $p$ is super-efficient if it does as well

[5] A prior is regular if it attaches positive mass on every element of the prior support.

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 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

Comparing different learning rules in an objective way 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. BMA is chosen because it is widely known and applied 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 on 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.

[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 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 8, I prove that, if the prior support only contains iid models, the only member of price probability that admits a Bayesian interpretation is $\gamma = 1$.

- *Bayesian updating is almost a tautology if we think of probabilities as empirical frequencies. Why should I 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."*

- *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, underreacting rules are not dominated because they are asymptotically Bayesian. If the model is misspecified, while Bayesian predictions can only be as accurate as the most accurate model in the support (Berk, 1966), underreacting rules can deliver predictions that are even more accurate than that.

- *Where are the trick/hidden assumptions?*

  The crucial assumption needed to ensure super-efficiency is a non-convex prior support. In Proposition 4, I show that underreacting rules are more accurate than Bayes', whenever 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 underreacting rules apply verbatim to all standard prediction problems. I chose to work in a parametric setting with finitely many parameters only for ease of exposition and to maintain the state price interpretation. For applications, we refer the reader to Grünwald and van Ommen (2014), which shows that underreacting rules outperform the BMA with prior, $g$, 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. While Avramov (2002); Cremers (2002) have found that BMA guaranties better out-of-sample prediction than prediction obtained using model selection criteria — which are qualitatively equivalent to $p_0^{SNNL}$— in the context of forecasting U.S. stock market indices.

[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 and is hardly a solution even in simple parametric settings because the increase in the dimensionality of the prior support reduces the learning rate (Schwarz, 1978; Clarke and Barron, 1990).

- *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 preferences determine different optimal investment strategies. His critique does not apply here because my accuracy criterion (preferences) is kept fixed in my comparison.

# 6 Asymptotic performance of $p_\gamma^{SNNL}$

**Theorem 1.** *If the cardinality of $\mathcal{P}$ is finite,*
*i)* $\gamma > 1 \Rightarrow p_\gamma^{SNNL}$ *is super-efficient;*
*ii)* $\gamma = 1 \Rightarrow p_\gamma^{SNNL}$ *is universal-efficient;*
*iii)* $\gamma < 1 \Rightarrow p_\gamma^{SNNL}$ *is sub-efficient.*

Theorem 2 shows that $p_\gamma^{SNNL}$ is qualitatively as accurate as BMA whenever the model is correctly specified ($p_\gamma^{SNNL}$ is, at least, sub-efficient). Furthermore, it shows that there are cases in which a rule that underreacts to empirical evidence can significantly outperform Bayes' rule. Moreover, because underreacting members of $\mathcal{M}(\mathcal{P})$ never underperform, and can even outperform Bayes' by an infinite amount, a Pascal (1668) wager's argument suggests that underreacting rules should be pragmatically preferred to Bayes' unless we are certain that our model is correctly specified.

Next, we present two cases showing that underreacting rules ($\gamma > 1$) can significantly outperform but never underperform BMA; whereas overreacting rules ($\gamma < 1$) can significantly underperform but never outperform BMA.

**Example 1:** Let $S = \{a, b\}, C_0 = [\frac{1}{2}\ \frac{1}{2}]'$, and $\mathcal{P} = \{p^1, p^2\}$, with $p^1, p^2$ iid measures: $\forall t, p^1(a_t) = \frac{1}{3} = p^2(b_t)$. Consider three $p_{\gamma_j}^{SNNL}$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 favorable case for probability mixtures that underreact to empirical evidence. 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:

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

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

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

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

Case $a$ shows that, by underreacting, $p_2^{SNNL}$ 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 favorable sequence for forecasting systems that overreact to empirical evidence.

Thus,on $\{a,a,a,...\}, \forall\alpha\in(0,1)$, $\begin{cases} \frac{p_\alpha^{BMA}(\sigma^t)}{p_0^{SNNL}(\sigma^t)} = \frac{\alpha\left(\frac{2}{3}\right)^t+(1-\alpha)\left(\frac{1}{3}\right)^t}{\frac{1}{2}\left(\frac{2}{3}\right)^{t-1}} \asymp 1 \\ \frac{p_\alpha^{BMA}(\sigma^t)}{p_1^{SNNL}(\sigma^t)} = \frac{\alpha\left(\frac{2}{3}\right)^t+(1-\alpha)\left(\frac{1}{3}\right)^t}{\frac{1}{2}\left(\frac{2}{3}\right)^t+\frac{1}{2}\left(\frac{1}{3}\right)^t} \asymp 1 \\ \frac{p_\alpha^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} = \frac{\alpha\left(\frac{2}{3}\right)^t+(1-\alpha)\left(\frac{1}{3}\right)^t}{\left(\frac{1}{2}\left(\frac{2}{3}\right)^{\frac{t}{\gamma}}+\frac{1}{2}\left(\frac{1}{3}\right)^{\frac{t}{\gamma}}\right)^\gamma * e^{-\sum_{\tau=1}^t \ln\left(q(a|\sigma^{\tau-1})+q(b|\sigma^{\tau-1})\right)}} \asymp 1 \end{cases}$.[9]

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.

Cases $a$ and $b$ suggest that non-concentration of the Bayesian posterior plays a special

[9]By Proposition 4 $\frac{p_\alpha^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} = \frac{\alpha\left(\frac{2}{3}\right)^t+(1-\alpha)\left(\frac{1}{3}\right)^t}{\left(\frac{1}{2}\left(\frac{2}{3}\right)^{\frac{t}{\gamma}}+\frac{1}{2}\left(\frac{1}{3}\right)^{\frac{t}{\gamma}}\right)^\gamma * e^{-\sum_{\tau=1}^t \ln\left(q(a|\sigma^{\tau-1})+q(b|\sigma^{\tau-1})\right)}} \asymp 1$

role in determining the (sub)super-efficient condition. This is indeed the case:

**Proposition 4.** *For every regular prior, $C_0$, on a finite support, $\mathcal{P}$, and $\gamma \in (0,\infty) \setminus 1$,*
*i)* $\lim \ln \frac{p^{SNNL}(\sigma^t)}{p^{BMA}(\sigma^t)} = \pm\infty$ *in every path in which $p^{BMA}$'s posterior does not concentrate on a unique model;*
*ii)* $\lim \ln \frac{p^{SNNL}(\sigma^t)}{p^{BMA}(\sigma^t)} \asymp 1$ *in every path in which $p^{BMA}$'s posterior concentrates exponentially fast on a unique model.*

Proposition 4 tells us that overreaction is detrimental while underreaction is desirable in cases in which the posterior does not concentrate on a model in the support. If the 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, giving more (less) weight to the prior produces forecasts that are closer (further) to the truth than Bayesian.

Known asymptotic results in Bayesian statistics[10] make Proposition 4 useful in recognizing the probabilities that determine the (sub)super-efficiency condition.[11] For example, Proposition 4 can be used to analyze:

**Case c:** The true probability is iid: $\forall t, P(a_t) = \frac{1}{2}$. 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, $\begin{cases} \frac{P^{BMA}(\sigma^t)}{p_0^{SNNL}(\sigma^t)} \to^{P\text{-a.s.}} +\infty \\ \frac{P^{BMA}(\sigma^t)}{p_1^{SNNL}(\sigma^t)} \asymp^{P\text{-a.s.}} 1 \\ \frac{P^{BMA}(\sigma^t)}{p_2^{SNNL}(\sigma^t)} \to^{P\text{-a.s.}} 0 \end{cases}$. The intuition goes as follows.

- $p_1^{SNNL}(\sigma_t|\sigma^{t-1})$ coincides, in every period, with $P^{BMA}_{\alpha=.5}(\sigma_t|\sigma^{t-1})$; hence, it does as well as a Bayesian with regular prior on $\mathcal{P}$;

- $p_2^{SNNL}(\sigma_t|\sigma^{t-1})$ smoothly oscillates between $p^1$ and $p^2$, but it spends more time "close to the middle" than $P^{BMA}_{\alpha=.5}(\sigma_t|\sigma^{t-1})$ because it underreacts to empirical evidence. Because the true distribution lies between $p^1$ and $p^2$, spending "more time close to the middle" makes $p_2^{SNNL}(\sigma_t|\sigma^{t-1})$'s forecasts more accurate than Bayes';

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

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

- $p_0^{SNNL}(\sigma_t|\sigma^{t-1})$ changes his forecasts discontinuously every time the model that performed best in the past changes. Thus, it spends less time "close to the middle" than $P^{BMA}$. Because the true distribution is between $p^1$ and $p^2$ spending "less time close to the middle" makes $p_0^{SNNL}(\sigma_t|\sigma^{t-1})$'s forecasts less accurate than Bayes'.

**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.

# 7 Asymptotic performance of $p^{NNL}$

**Theorem 2.** *If the cardinality of $\mathcal{P}$ is finite, $\forall\gamma \in [0,\infty), p_\gamma^{NNL}$ is universal-efficient.*[12]

Theorem 2 tells us that, although time-inconsistent, $p^{NNL}$ performs qualitatively as well as BMA in terms of likelihood. If we are only concerned about accuracy, there is no reason to consider time-consistency to be a fundamental property of rational forecasts.

This result 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:** a risk-neutral agent who does not discount the future and whose beliefs are $p_0^{NML}(R,L) = \frac{1}{3}$, $p_0^{NML}(R,L,L) = \frac{2}{10}$ and $p_0^{NML}(R,L,R) = \frac{1}{10}$ (as in Example 3, Section 8), is at time-zero indifferent between:

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

[12]More generally, $p^{NNL}$ is universal-efficient in any economy that satisfies A1-A4 with $|\mathcal{P}| < \infty$.

- $\$\frac{2}{10}$ 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 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 which 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.

This procedure is equivalent to the "massaging" process described in Savage (1954) and Binmore (2008) as sufficient to deduce a Bayesian prior from the set of subjective relative likelihoods of an agent. Incidentally, the beliefs of Example 3 can be used to show that "massaging" is not sufficient to imply Bayes' rule. "Massaging" $p_0^{NNL}$ forward, I obtain $p^{SNNL}$ which is prequential but not Bayesian (because it is not-exchangeable). Fixing the final horizon and "massaging" $p_0^{NNL}$ backward, I obtain a measure that is prequential and exchangeable but still not Bayesian because there is no prior that can generate that set of conditional beliefs.

**Proposition 5.** *A prior that makes $p_0^{NNL}(\sigma^3)$ in Example 3 consistent with Bayes' rule does not exist.*

*Proof.* By symmetry of $p_0^{NNL}(\sigma^3)$, if this prior existed it had to give the same weight to model $p^1$ and $p^2$. Unconditional probabilities obtained via Bayes' rule from this prior coincide with $p_1^{NNL}(\sigma^3) \neq p_0^{NNL}(\sigma^3)$, a contradiction. □

# 8 Price probabilities are not Bayesian

In this section, I introduce two characterizing properties of Bayesian updating (on iid sequences) and discuss whether these properties are satisfied by members of price probabilities.

I demonstrate that $p_\gamma^{NNL}$ and $p_\gamma^{SNNL}$ satisfy both properties if and only if gamma equals 1. Members of price probability are Bayesian if and only if $\gamma = 1$.

**Definition 6.** *A probability mixture, p, is prequential if:*

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

Prequentiality (see Grünwald (2007) for details) coincides with Kolmogorov's third axiom (additivity). An agent with non-prequential beliefs believes that the sum of the probabilities of disjoint events differs from the probability of their union. In the economics literature, non-prequential beliefs are called time-inconsistent because a trader with non-prequential beliefs can be put into arbitrage (see example and discussion in Section 7). In the behavioral literature, a violation of this property is labeled conjunction fallacy (Kahneman, 2011).

**Definition 7.** *A probability mixture, p, is exchangeable if $p(\sigma^t) = p(\hat{\sigma}^t)$ whenever two partial histories $\sigma^t, \hat{\sigma}^t$ share the same frequency.*

Exchangeability captures the idea that the probability of a sequence of events does not depend on the order of the realizations. This assumption is deeply connected to Bayes' rule: De Finetti (1931)'s theorem implies that a measure on infinite sequences is exchangeable if and only if it is a Bayesian mixture of iid probabilities. Exchangeability is an appealing criterion whenever all the models in $\mathcal{P}$ are iid. For example, I expect a rational agent facing repeated iid tosses from a coin with an unknown bias to attach the same probability to the sequences of realizations $\{H, H, T\}$ and $\{T, H, H\}$. In terms of conditional forecasts, an agent who attaches less probability to $\{H, H, T\}$ than $\{T, H, H\}$ will appear as either over-weighting or under-weighting (relatively to a Bayesian) the first two realizations.

The next proposition shows that most members of $\mathcal{M}(\mathcal{P})$ are not Bayesian.

**Proposition 6.** *Given a prior, $C_0$, if all probabilities in $\mathcal{P}$ are iid,*

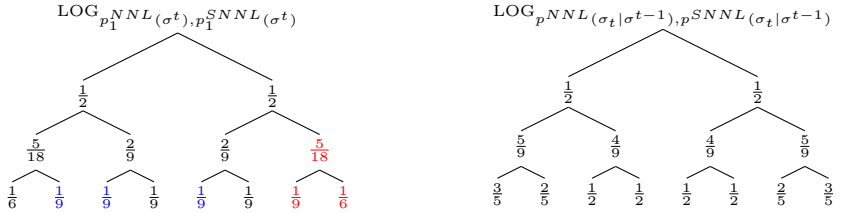
- *a) $\gamma = 1 \Rightarrow p_\gamma^{NNL}, p_\gamma^{SNNL}$ and $p^{BMA}$ coincide and are prequential and exchangeable.*

- *b)* $\gamma \neq 1 \Rightarrow$ *i)* $p_\gamma^{NNL}$ *is exchangeable but not prequential;*
  *ii)* $p_\gamma^{SNNL}$ *is prequential but not exchangeable.*

Example 3 illustrates Proposition 6. It shows how different values of gamma affect $p_\gamma^{NNL}$ and $p_\gamma^{SNNL}$ on sequences of length 3 (unconditional probabilities on the left-hand tree, conditional probabilities on the right-hand tree).

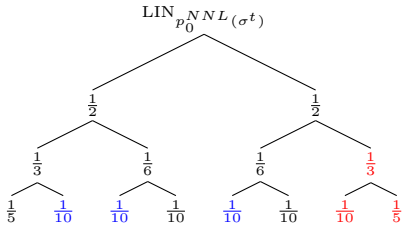
**Example 3:** Let $S = \{a, b\}, C_0 = [\frac{1}{2}\ \frac{1}{2}]'$, and $\mathcal{P} = \{p^1, p^2\}$, with $p^1, p^2$ iid measures: $\forall t, p^1(a_t) = \frac{1}{3} = p^2(b_t)$. Consider $p_\gamma^{NNL}$ and $p_\gamma^{SNNL}$ obtained with $\gamma_0 = 0$ and $\gamma_1 = 1$, respectively.

**Case *a*:** $\gamma = 1$. By Corollary 1 and Proposition 3, $p_1^{NNL}(\sigma^t)$ and $p_1^{SNNL}(\sigma^t)$ coincide with BMA with prior $C_0$, which is *prequential* (e.g., $p(\{R,R,R\}) + p(\{R,R,L\}) = p(\{R,R\})$) and *exchangeable* (e.g., $p(\{L,L,R\}) = p(\{L,R,L\}) = p(\{R,L,L\})$).



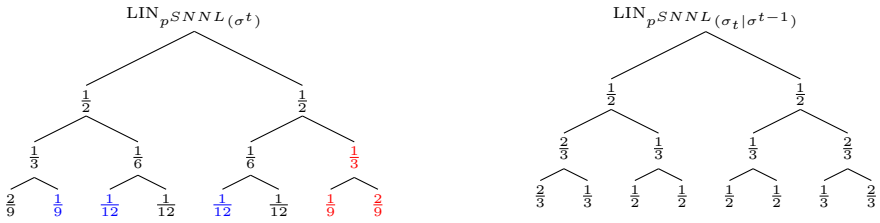

**Case *b*:** $\gamma = 0$.

- Normalized Normed Likelihood $\left(p_0^{NNL}(\sigma^t) = \frac{\max_i p^i(\sigma^t)}{\sum_{\hat{\sigma}^t} \max_i p^i(\hat{\sigma}^t)}\right)$, is *exchangeable* (e.g.,$p_0^{NNL}(\{L,L,R\}) = p_0^{NNL}(\{L,R,L\}) = p_0^{NNL}(\{R,L,L\})$) but not *prequential* (e.g.,$p_0^{NNL}(R,R,R)+p_0^{NNL}(R,R,L) = \frac{3}{10} \neq \frac{1}{3} = p_0^{NNL}(R,R)$. Comparing $p_0^{NNL}$ with $p^{BMA}$ shows that $p_0^{NNL}$ attaches more probability to extreme sequences ($\{L,L,L\}$ and $\{R,R,R\}$) than $p^{BMA}$ does.




LIN $p_0^{NNL}(\sigma_t|\sigma^{t-1})$
Not defined because not prequential

- Sequential Normalized Normed Likelihood $\left(p_0^{SNNL}(\sigma_t|\sigma^{t-1}) = \frac{p_0^{NNL}(\sigma^t)}{\sum_{\hat{\sigma}_t} p_0^{NNL}(\sigma^{t-1},\hat{\sigma}_t)}\right)$, is *prequential* (e.g.,$p_0^{SNNL}(\{R,R,R\})+p_0^{SNNL}(\{R,R,L\}) = p_0^{SNNL}(\{R,R\})$) but not *exchangeable* (e.g., $p_0^{SNNL}(\{L,L,R\}) = \frac{1}{9} \neq \frac{1}{12} = p_0^{SNNL}(\{L,R,L\}) = p_0^{SNNL}(\{R,L,L\})$). The tree on the right shows that $p_0^{SNNL}$ "overreacts" to empirical evidence. Unlike $p^{BMA}$, a single observation in favor of $p^1$ ($p^2$) suffices to make its conditional probabilities coincide with $p^1$ ($p^2$) (e.g., $p_0^{SNNL}(L|L)=\frac{2}{3}=p^2(L)$ and $p_0^{SNNL}(R|R)=\frac{2}{3}=p^1(R)$).




The example shows that, different members of $\mathcal{P}$ attach different probabilities to small samples. A natural question to ask is which model is the best. In the previous section we showed that, if there is a unique most accurate model in $\mathcal{P}$, members of price probabilities are asymptotically equivalent in terms of likelihood. Thus, if an agent is certain that there is a best model in $\mathcal{P}$ all these probability mixtures are equally valid. However, if this is not the case, Theorem 1 tells us that there are situations in which it is sub-optimal to use Bayes' rule. By underreacting (SNNL with $\gamma > 1$), an agent can produce forecasts that are more accurate than those of an agent that uses Bayes' rule. Because a mild form of underreaction is safe (it never does significantly worse than Bayes' rule) and, in some cases, is much better that standard Bayesian, it seems to us that there is no reason to believe that Bayes' rule is the only rational way to learn.

# 9 Conclusion

I use the standard machinery of dynamic general equilibrium models to generate a rich class of probabilities and to discuss their properties. This class includes probabilities consistent with Bayes' rule and known behavioral biases. I propose an objective criterion of accuracy and use it to discuss the performance of non-Bayesian rules against the Bayesian benchmark. If the prior support is correctly specified, I prove that all members of this class perform as well as Bayes' rule according to my criterion. If the prior support is misspecified, I demonstrate how rules that underreact to new information can significantly outperform Bayes' rule. Because underreaction is never worse and sometimes better than Bayes', I provides a full range of alternative rules to challenge the prevailing opinion that Bayes' rule is the only rational way to learn.

## 9.1 Discussion

My results go beyond the simple parametric setting I 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 market's returns. Unfortunately, nobody is currently able to tell you what is THE true model, 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 are facing 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 that, before any empirical application of $p_\gamma^{SNNL}$, I need to give an objective rule to choose gamma. It is easy to show 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 $p_\gamma^{SNNL}$ 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 S$p_\gamma^{SNNL}$'s accuracy. With hindsight, I would choose a small gamma, when the data clearly suggests a unique best model, and a large gamma otherwise. However, We are interested in prediction, not fitting the data. The best I can do is 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 main difficulty is implementing this intuition on the different

equations governing $p_\gamma^{SNNL}$.

# 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 A1-A4, equilibrium prices are given by:*

$$q(\sigma^t) = \frac{\beta^t \sum_{i\in\mathcal{P}} p^i(\sigma^t) \frac{1}{u^i(c_0^i)'}}{\sum_{j\in\mathcal{I}} \frac{1}{u^j(c_t^j(\sigma))'}} \tag{7}$$

*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, $\forall\sigma, \forall t$,

$$\frac{\partial L_i}{\partial c_t^i(\sigma)} = 0 \Rightarrow \beta^t p^i(\sigma^t) u^i(c_t^i(\sigma))' = \lambda_i q(\sigma^t) \tag{8}$$

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 and 2:**

*Proof.* Substituting $c_t^i(\sigma)^{-\gamma}$ for $u^i(\sigma^t)'$ and $u^i(c_0^i)$ for $\lambda_i$ in Equation 8,

$$\beta^t p^j(\sigma^t) c_t^j(\sigma)^{-\gamma} = (c_0^j)^{-\gamma} q(\sigma^t) \tag{9}$$

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 $c^i(\sigma^t)$:

$$c_t^i(\sigma) = \left( \frac{p^i(\sigma^t)}{p^j(\sigma^t)} \right)^{\frac{1}{\gamma}} \frac{c_0^i}{c_0^j} c_t^j(\sigma). \tag{10}$$

Substituting Equation 10 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 $c_t^j(\sigma)$:

$$c_t^j(\sigma) = \frac{p^j(\sigma^t)^{\frac{1}{\gamma}} c_0^j}{\sum_{i\in\mathcal{I}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i}. \tag{11}$$

Substituting $c_t^j(\sigma)$ in Equation 9 and rearranging, I obtain

$$q(\sigma^t) = \beta^t \left(\sum_{i\in\mathcal{P}} p^i(\sigma^t)^{\frac{1}{\gamma}} c_0^i\right)^{\gamma} \tag{12}$$

The result follows substituting Equations 12 in Definition 2 and 3, respectively. □

**Lemma 3.** *In an identical CRRA economy that satisfies A1-A4:*

$$\forall \sigma \in S^{\infty}, \forall t \in \{1, ..., +\infty\}, \gamma'' \geq 1 \geq \gamma' \Rightarrow \frac{\sum\limits_{\sigma_t\in\mathcal{S}} q_{\gamma'}(\sigma_t|\sigma^{t-1})}{\beta} \geq 1 \geq \frac{\sum\limits_{\sigma_t\in\mathcal{S}} q_{\gamma''}(\sigma_t|\sigma^{t-1})}{\beta}$$

*With equality if and only if $\gamma = 1$ or the consumption-share distribution is degenerate.*

*Proof.* Using Equation 12 and the definition of $q(\sigma_t|\sigma^{t-1})$,

$$\begin{aligned}
\frac{\sum_{\sigma_t\in\mathcal{S}} q_\gamma(\sigma_t|\sigma^{t-1})}{\beta} &= \sum_{\sigma_t\in\mathcal{S}} \frac{\left(\sum_{i\in\mathcal{I}} p^i(\sigma^{t-1},\sigma_t)^{\frac{1}{\gamma}} c_0^i\right)^{\gamma}}{\left(\sum_{i\in\mathcal{I}} p^i\left(\sigma^{t-1}\right)^{\frac{1}{\gamma}} c_0^i\right)^{\gamma}} \\
&= \sum_{\sigma_t\in\mathcal{S}} \left(\sum_{i\in\mathcal{I}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} \frac{p^i(\sigma^{t-1})^{\frac{1}{\gamma}} c_0^i}{\sum_{i\in\mathcal{I}} p^i\left(\sigma^{t-1}\right)^{\frac{1}{\gamma}} c_0^i}\right)^{\gamma} \\
&= \sum_{\sigma_t\in\mathcal{S}} \left(\sum_{i\in\mathcal{I}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} c_{\gamma,t-1}^i(\sigma)\right)^{\gamma} && \text{by Eq. 11;} \\
&= \sum_{\sigma_t\in\mathcal{S}} f\left(\sum_{i\in\mathcal{I}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} c_{\gamma,t-1}^i(\sigma)\right)
\end{aligned}$$

Note that $\sum_{i\in\mathcal{I}} c_{\gamma,t-1}^i(\sigma) = 1$, and $f(.)$ is strictly concave $\Leftrightarrow \gamma < 1$, $f(.)$ is linear $\Leftrightarrow \gamma = 1$ and $f(.)$ is strictly convex $\Leftrightarrow \gamma > 1$. Let $\gamma > 1$, by Jensen inequality:

$$\begin{aligned}
\sum_{\sigma_t\in\mathcal{S}} f\left(\sum_{i\in\mathcal{I}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} c_{\gamma,t-1}^i(\sigma)\right) &\leq \sum_{\sigma_t\in\mathcal{S}}\sum_{i\in\mathcal{I}} c_{\gamma,t-1}^i(\sigma) f(p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}}) && \text{with equality iff } \exists i : c_{\gamma,t-1}^i(\sigma) = 1; \\
&= \sum_{i\in\mathcal{I}} c_{\gamma,t-1}^i(\sigma) \sum_{\sigma_t\in\mathcal{S}} p^i(\sigma_t|\sigma^{t-1}) \\
&= 1 && \text{because } \forall i, \sum_{\sigma_t\in\mathcal{S}} p^i(\sigma_t|\sigma^{t-1}) = 1.
\end{aligned}$$

The cases $\gamma = 1, \gamma < 1$ can be proven using the same logic. □

**Proof of Proposition 3**

*Proof.* By Eq.11, the consumption of the trader with maximum likelihood on $\sigma^t$, $\hat{i}(\sigma^t)$, is given by $c_t^{\hat{i}}(\sigma) = \frac{p^{\hat{i}(\sigma^t)}(\sigma^t)^{\frac{1}{\gamma}} c_0^{\hat{i}}}{\sum_{j\in\mathcal{I}} p^j(\sigma^t)^{\frac{1}{\gamma}} c_0^j}$. Claims $i, ii, iii$ follow by noticing that $p^{SNNL}_{\gamma=1} = p^{BMA}$ and that to higher gamma corresponds a lower weight to $c_t^{\hat{i}}$: $\frac{\partial c_t(\hat{i}(\sigma))}{\partial\gamma} = -c_0^{\hat{i}} p^{\hat{i}}(\sigma^t)^{\frac{1}{\gamma}} \frac{\sum_{j\neq\hat{i}(\sigma^t)} c_0^j p^j(\sigma^t)^{\frac{1}{\gamma}} \ln \frac{p^{\hat{i}}(\sigma^t)}{p^j(\sigma^t)}}{\gamma^2 \left(\sum_{j\in\mathcal{I}} p^j(\sigma^t)^{\frac{1}{\gamma}} c_0^j\right)^2} < 0.$

□

**Lemma 4.** *In an economy that satisfies A1-A4,* $\forall\sigma \in S^\infty : \sum_{i\in\mathcal{I}} \frac{1}{u_i'(c_t^i(\sigma))} \asymp 1$.

*Proof.*

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

□

**Proof of Theorem 1:**

*Proof.* Let us rewrite $\ln \frac{p^{BMA}(\sigma^t)}{p^{SNNL}(\sigma^t)}$ as follows:

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

The result follows from these two claims:

- Claim 1: $\ln \frac{\beta^t p^{BMA}(\sigma^t)}{q(\sigma^t)} \asymp 1$:

  *Proof*: By Eq.7, $q(\sigma^t) = \beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t) \frac{\frac{1}{u^i(c_0^i)'}}{\sum_{j\in\mathcal{I}} \frac{1}{u_j'(c_t^j(\sigma))}} \asymp^{by\ Lem.4} \beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t) \frac{1}{|\mathcal{I}|} \asymp \beta^t p^{BMA}(\sigma^t)$.

- Claim 2: $\exists \hat{P} : \gamma < (>) 1 \Rightarrow \sum_{\tau=1}^{t} ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \to +(-)\infty$ :

*i)* $\sum_{\tau=1}^{t} ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \to \pm\infty$ iff consumption shares do not concentrate on one trader.

*Proof*: By Lemma 3, Jensen inequality $\gamma < (>) 1 \Leftrightarrow \ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \geq (\leq) 0$, with equality iff the consumption-share distribution is degenerate. Thus, given $\gamma$, all terms of the sum have the same sign. 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 posterior, $C_{t,\gamma=1}$, does not concentrate on a unique model iff, $\forall\gamma \in (0,+\infty)$, $p^{SNNL}(\sigma^t)$'s posterior, $C_{t,\gamma}$, does not concentrate on a unique model.[13]

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

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

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

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

*iii)* $\exists\hat{P}$ such that $C_{1,t}$ does not concentrate on a unique model:
*Proof*: Because $\mathcal{P}$ has finitely many models, it exists a sequence such that the two most accurate models in $\mathcal{P}$ have comparable likelihood (it can be constructed recursively by choosing the next realization to favor the model with the lower likelihood). Alternatively, a non-degenerate measure that satisfies this condition can be constructed using Chernoff's bond (Cover and Thomas, 2012). □

**Proof of Theorem 2:**

*Proof.* Substituting Equation 7 in the definition of $p^{NNL}$.

$$p^{NNL}(\sigma^t) = \frac{q(\sigma^t)}{\sum_{\sigma^t} q(\sigma^t)} = \frac{\frac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i}{\sum_{i\in\mathcal{I}} \frac{1}{u_i'(c(\sigma^t))}}}{\beta^t \sum_{\hat{\sigma}^t\in\mathcal{S}^t}\left(\frac{\sum_{i\in\mathcal{I}} p^i(\hat{\sigma}^t)c_0^i}{\sum_{i\in\mathcal{I}} \frac{1}{u_i'(c(\hat{\sigma}^t))}}\right)}$$

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; $0 < a \leq b < \infty$, thus

$$p^{NNL}(\sigma^t) \in [\frac{\frac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i}{b}}{\beta^t \sum_{\sigma\in\mathcal{S}^t}\left(\frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i}{a}\right)}, \frac{\frac{\beta^t \sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i}{a}}{\beta^t \sum_{\sigma\in\mathcal{S}^t}\left(\frac{\sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i}{b}\right)}]$$

$$\Rightarrow p^{NNL}(\sigma^t) \in [\frac{a}{b}\sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i, \frac{b}{a}\sum_{i\in\mathcal{I}} p^i(\sigma^t)c_0^i]$$

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

□

**Proof of Proposition 4:**

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

[13]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)).

By Claim 1 in the proof of Th.1, $\ln \frac{\beta^t p^{BMA}(\sigma^t)}{q(\sigma^t)} \asymp 1$. For the second term I have:

- Part $i$) mimics the step of Th. 1, except that non-concentration is now assumed.
- Part $ii$) follows because if the concentration rate is exponential a Taylor expansion ensures that $\exists 0 < a < b < 1 : ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \in [b^\tau; a^\tau]$, so that $\sum_{\tau=1}^t ln\left(\frac{\sum_{\hat{\sigma}_\tau} q(\hat{\sigma}^\tau|\sigma^{\tau-1})}{\beta}\right) \asymp 1$. For a practical example of the details, 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 $\exists \eta \in (0, \frac{1}{2}) :$

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

□

**Proof of Proposition 6:**

*Proof.* (a): In a log economy $p^{NNL} = p^{SNNL} = p^{BMA}$ which is prequential and exchangeable.

$(b) : i)$ Exchangeable: because the denominator in Eq. 7 is constant at every horizon.

Non-prequential: by contradiction, assume $H_0 : \exists \gamma \neq 1 : p^{NNL}$ is prequential.

$$\forall \sigma^t, \frac{q(\sigma^t)}{q(\sigma^{t-1})} =^{Eq.\ condition} q(\sigma_t|\sigma^{t-1})$$

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

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

$$\Leftrightarrow^{if\ H_0\ is\ true} \forall \sigma^{t-1}, 1 = \sum_{\sigma_t} q(\sigma_t|\sigma^{t-1}) \frac{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})}{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}$$

$$\Leftrightarrow \forall \sigma^{t-1}, \frac{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})} = \sum_{\sigma_t} q(\sigma_t|\sigma^{t-1})$$

$$\Leftrightarrow \forall \tilde{\sigma}^{t-1}, \bar{\sigma}^{t-1}, \ \sum_{\sigma_t} q(\sigma_t|\tilde{\sigma}^{t-1}) = \frac{\sum_{\hat{\sigma}^t} q(\hat{\sigma}^t)}{\sum_{\hat{\sigma}^{t-1}} q(\hat{\sigma}^{t-1})} = \sum_{\sigma_t} q(\sigma_t|\bar{\sigma}^{t-1})$$

$$\Leftrightarrow \forall \tilde{\sigma}^{t-1}, \bar{\sigma}^{t-1}, \ \sum_{\sigma_t} \left( \sum_{i \in \mathcal{P}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} \frac{p^i(\tilde{\sigma}^{t-1})^{\frac{1}{\gamma}}}{\sum_j p^j(\tilde{\sigma}^{t-1})^{\frac{1}{\gamma}}} \right)^{\gamma} = \sum_{\sigma^t} \left( \sum_{i \in \mathcal{P}} p^i(\sigma_t|\sigma^{t-1})^{\frac{1}{\gamma}} \frac{p^i(\bar{\sigma}^{t-1})^{\frac{1}{\gamma}}}{\sum_{i \in \mathcal{P}} p^i(\bar{\sigma}^{t-1})^{\frac{1}{\gamma}}} \right)^{\gamma}.$$

Which is true iff gamma=1 (i.e. all traders have log-utility). A contradiction of $H_0$.

*ii*) Prequential: because the measure is constructed recursively in a forward fashion. Non-exchangeable: by De Finetti (1931)'s theorem, a measure on infinite sequences is exchangeable if and only if it can be represented as a mixture of iid measures if and only if it exists a prior such that it coincides with BMA. The (sub)super-efficiency of $p^{SNNL}$ proved in Theorem 1 implies that there exists no such prior if $\gamma \neq 1$, thus $p^{SNNL}$ is not exchangeable.

□

# 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.

Binmore, K. (2008). *Rational decisions*. Princeton University Press.

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.

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

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 Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. *Atti del Congresso Internazionale dei matematici, Bologna*.

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. (2006). An axiomatic model of non-bayesian updating. *Review of Economic Studies*, pages 413–436.

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. (2015). Rationality and the bayesian paradigm. *Journal of Economic Methodology*, 22(3):312–334.

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.

Pascal, B. (1668). *Pascal's Pensees (English translation by John Walker)*. Available online at http://www.gutenberg.org/files/18269/18269-h/18269-h.htm.

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.

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.

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.