Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Classical Subjective Expected Utility

Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci, Luigi Montrucchio

Working Paper n. 400

This Version: April 2012

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Classical Subjective Expected Utility*

Simone Cerreia-Vioglio[a] Fabio Maccheroni[a] Massimo Marinacci[a] Luigi Montrucchio[b]

[a]Department of Decision Sciences, Università Bocconi

[b]Collegio Carlo Alberto, Università di Torino

April 2012

Abstract

We consider decision makers who know that payoff relevant observations are generated by a process that belongs to a given class M, as postulated in Wald [36]. We incorporate this Waldean piece of objective information within an otherwise subjective setting a la Savage [33] and show that this leads to a two-stage subjective expected utility model that accounts for both state and model uncertainty.

1 Introduction

Consider a decision maker who is evaluating acts whose outcomes depend on some verifiable states, that is, on observations (workers' outputs, urns' drawings, rates of inflation, and the like). If the decision maker believes that observations are generated by some probability model, two sources of uncertainty affect his evaluation: model uncertainty and state uncertainty. The former is about the probability model that generates observations, the latter is about the state that obtains (and that determines acts' outcomes).

State uncertainty is payoff relevant and, as such, it is directly relevant for decision maker's decisions. Model uncertainty, in contrast, is not payoff relevant and its role is instrumental relative to state uncertainty. Moreover, models cannot be observed and, while in some cases they have a simple physical description (e.g., urns' compositions), often they do not have it (e.g., fair coins).[1] For all these reasons, the purely subjective choice models a la Savage [33] focus on the verifiable and payoff relevant state uncertainty. They posit an observation space S over which subjective probabilities are derived via betting behavior.

In contrast, classical statistical decision theory a la Wald [36] supposes that decision makers know that observations are generated by a probability model that belongs to a given subset M, whose elements are regarded as alternative random devices that Nature may select to generate

*First preliminary draft: May 2011. We thank Pierpaolo Battigalli and a seminar audience at U. Paris 1 for useful comments. The financial support of the European Research Council (advanced grant, BRSCDP-TEA) is gratefully acknowledged.

[1]See Diaconis, Holmes, and Montgomery [8] for some physical analysis of coin tossing.

observations.[2] In other words, Wald's approach posits a model space M in addition to the observation space S. In so doing, Wald adopted a key tenet of classical statistics, that is, to posit a set of possible data generating processes (e.g., Normal distributions with some possible means and variances), whose relative performance is assessed via available evidence (often collected with i.i.d. trials) through maximum likelihood methods, hypothesis testing, and the like. Though models cannot be observed, in Wald's approach their study is key to better understand state uncertainty.

Is it possible to incorporate this Waldean key piece of objective information within Savage's framework? Our work addresses this question and tries to embed this classical datum within an otherwise subjective setting. Besides its theoretical interest, this question is relevant since some important economic applications assume, at least as a working hypothesis, this Waldean piece of information. For example, M may be the set of equilibrium distributions for observations (e.g., prices in a cobweb model a la Muth).

Our approach takes the objective information M as a primitive and enriches the standard Savage framework with this datum: decision makers know that the true model m that generates data belongs to M. Behaviorally, this translates into the requirement that their betting behavior (and so their beliefs) be consistent with datum M:

$$m(F) \geq m(E) \quad \forall m \in M \Longrightarrow xFy \succsim xEy$$

where xFy and xEy are bets on events F and E, with $x \succ y$. We do not, instead, consider bets on models and, as a result, we do not elicit prior probabilities on models through hypothetical (since models are not observable) betting behavior on models. Nevertheless, our basic representation result, Proposition 4, shows that, under Savage's axioms P.1-P.6 and the above consistency condition, acts are ranked according to the criterion

$$V(f) = \int_{\Delta} \left(\int_S u(f(s))\, dm(s) \right) d\mu(m) \tag{1}$$

where μ is a prior subjective probability on models, whose support is included in M. We call this representation *Classical Subjective Expected Utility* because of the classical Waldean tenet on which it relies.

The prior μ is a subjective probability that may also reflect some personal information on models that decision makers may have, in addition to the objective information M (Proposition 5 behaviorally characterizes its support). Uniqueness of μ corresponds to the linear independence of the set M. For example, M is linearly independent when its members are pairwise orthogonal. Remarkably, Section 5 shows that in intertemporal problems this condition is often satisfied in applications. Specifically, consider a standard intertemporal observation space $\mathcal{Z}^\infty$ whose points are infinite histories $(z_1, ..., z_t, ...)$ of observations. Suppose that $\mathcal{Z}$ is at most countable, and endow

[2] As Wald [37] writes "A characteristic feature of any statistical decision problem is that F is unknown. It is merely assumed to be known that F is a member of a given class Ω of distributions functions. The class Ω is to be regarded as a datum of the decision problem."

$\mathcal{Z}^\infty$ with the filtration $\{\mathcal{B}_t\}$ generated by the elementary cylinder sets $z^t = \{z_1, ..., z_t\}$, that is, histories of observations. By Proposition 6, the conditional version of (1) at z^t is

$$V_{z^t}(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} u(f(z)) \ dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right)$$

where $m\left(z \mid z^t\right)$ and $\mu\left(m \mid z^t\right)$ are, respectively, the conditional model and the posterior probability given the observation history z^t. In this intertemporal setting, pairwise orthogonality holds for collections of i.i.d. models (Proposition 14) and for collections of Markov chains (Proposition 15). Models that are widely used in applications thus satisfy the condition of linear independence that ensures the uniqueness of prior μ.

Moreover, Section 6 shows that under this orthogonality condition there is full learning. That is, decision makers eventually behave as Expected Utility decision makers that know the true model that generates observations. They thus behave consistently with the observations they make, that is, they are long run empiricists. Classical Subjective Expected Utility thus provides a proper decision theoretic framework where to set a common justification of rational expectations that, "with a long enough historical data record, statistical learning will equate objective and subjective probability distributions."[3]

A final, more technical but equally noteworthy, feature of orthogonal data M is that for them the celebrated Lyapunov Convexity Theorem holds also in the infinite case, something that in general is altogether false. In our Savagean setting this key nonatomic property allows to establish in Proposition 10 the representation (1) for countably infinite orthogonal data.

As we detail in the paper, each prior μ induces a *predictive probability* $\bar{\mu}$ on the sample space S through averaging:

$$\bar{\mu}(E) = \int_\Delta m(E)\, d\mu(m)$$

In particular,

$$V(f) = \int_S u(f(s))\, d\bar{\mu}(s) \tag{2}$$

is the reduced form of V, its Savage Subjective Expected Utility representation. When M is a singleton $\{m\}$, we have $\bar{\mu} = m$ for all priors μ and we thus get the von Neumann-Morgenstern Expected Utility representation

$$V(f) = \int_S u(f(s))\, dm(s) \tag{3}$$

where subjective probabilities do not play any role.[4] Classical Subjective Expected Utility thus encompasses both the Savage and von Neumann-Morgenstern representations.

[3] Sargent and Williams [32, p. 361].

[4] Lucas [24, p. 15] writes that "Muth [30] ... [identifies] ... agents' subjective probabilities ... with 'true' probabilities, calling the assumed coincidence of subjective and 'true' probabilities *rational expectations*." (Italics in the original). In our setting, this coincidence is modelled by singleton M and results in the Expected Utility criterion (3). Later in the paper we will consider Muth's cobweb model and what form the rational expectation hypothesis may take in our approach.

In particular, the Savage criterion (2) is what an outside observer, unaware of datum M, would be able to elicit from decision maker's behavior. It is a much weaker representation than the "structural" one (1), which is the criterion that, instead, an outside observer aware of M would be able to elicit. For, this informed observer would be able to focus on the map $\mu \to \bar{\mu}$ from priors with support included in datum M to predictive probabilities. Under the linear independence of datum M, by inverting this map the observer would be able to recover prior μ from the predictive probability $\bar{\mu}$, which can be elicited through standard methods. The richer Waldean representation (1) is thus summarized by a triple (u, M, μ), with $\operatorname{supp} \mu \subseteq M$, while for the usual Savagean representation (2) is enough a pair (u, P).

Summing up, though the work of Savage [33] was inspired by the seminal decision theoretic approach of Wald [36], his purely subjective setup and the ensuing large literature[5] did not consider Wald's classical datum, central in Wald's approach. In this paper we show how to embed this datum in a Savage setting and how to derive the richer Waldean representation (1) by only considering choice behavior based on observables. Battigalli, Cerreia-Vioglio, Maccheroni, and Marinacci (2011) use the Wald-Savage setup of the present paper to study selfconfirming equilibria.

The paper is organized as follows. Section 2 introduces the basic decision theoretic setting and some mathematical preliminaries. Section 3 presents the basic representation result, that Section 4 extends to infinite data. Section 5 illustrates our representation in an important intertemporal setup, in which Section 6 shows that learning occurs. A few important issues, best discussed after the development of the paper analysis, are collected in the Concluding Remarks.

2 Preliminaries

2.1 Setting

We consider a standard Savage setting, where S is a nonempty state space and X is an outcome space. An *act* is a map $f : S \to X$ that produces outcome $f(s)$ in state S. Denote by $\mathcal{F}$ the set of all simple (i.e., finitely valued) acts available to the decision maker.

We consider a binary relation $\succsim$ over $\mathcal{F}$ that represents the decision maker's preferences. We assume that $\succsim$ satisfies the classic Savage's axioms P.1-P.6. By the famous Savage Representation Theorem, due to Savage [33], the preference $\succsim$ satisfies P.1-P.6 if and only if there is a utility function $u : X \to \mathbb{R}$ and a convex-ranged[6] finitely additive probability $P : \Sigma \to [0,1]$ such that $V(f) = \int_S u(f(s))\, dP(s)$ represents $\succsim$. This representation is called Subjective Expected Utility (SEU).

Given any $f, g \in \mathcal{F}$ and $E \in \Sigma$, we denote by fEg the act equal to f on E and to g otherwise, that is,

$$fEg = \begin{cases} f(s) & \text{if } s \in E \\ g(s) & \text{if } s \notin E \end{cases}$$

[5] See Fishburn [12], Kreps [22], and Gilboa [13]. See Jaffray [15] for a different "objective" approach.

[6] That is, for each $E \in \Sigma$ such that $P(E) > 0$ and each $\alpha \in [0,1]$ there exists $F \subseteq E$ such that $P(F) = \alpha P(E)$.

Given any $f, g \in \mathcal{F}$ and $E \in \Sigma$, the conditional preference $\succsim_E$ is a binary relation on $\mathcal{F}$ such that $f \succsim_E g$ whenever $fEh \succsim gEh$ for all $h \in \mathcal{F}$. By P.2, the Sure Thing Principle, $\succsim_E$ is well defined. In particular, an event $E \in \Sigma$ is *null* if, for each $f, g \in \mathcal{F}$, we have $f \sim_E g$ (see Savage [33, p. 24]).

The conditional preference $\succsim_E$ satisfies P.1-P.6 if the primitive preference does (see, e.g., Kreps [22, Chapter 10]). Hence, Savage's Theorem can be stated in conditional form by saying that $\succsim$ satisfies P.1-P.6 if and only if there is a utility function $u : X \to \mathbb{R}$ and a convex-ranged finitely additive probability $P : \Sigma \to [0,1]$ such that, for each nonnull events E,

$$V_E(f) = \int_S u(f(s))\, dP(s \mid E) \tag{4}$$

represents $\succsim_E$. Here $P(\cdot \mid E) : \Sigma \to [0,1]$ is the conditional probability

$$P(F \mid E) = \begin{cases} \frac{P(F \cap E)}{P(E)} & \text{if } P(E) > 0 \\ 0 & \text{else} \end{cases}$$

2.2 Mathematics

We denote by Δ the collection of all (countably additive) probability measures on Σ. Unless otherwise stated, in the paper all probability measures are countably additive.

In the sequel we will often consider subsets M of Δ. Given $M \subseteq \Delta$, we consider M endowed with the σ-algebra

$$\mathcal{M} = \sigma\{m \mapsto m(E) : E \in \Sigma\},$$

that is, with the smallest σ-algebra that makes the real valued and bounded functions on M, of the form $m \mapsto m(E)$, measurable for all $E \in \Sigma$.

In the important special case $M = \Delta$ we write $\mathcal{D}$ in place of $\mathcal{M}$. If $M \subseteq \Delta$ then $\mathcal{M} = M \cap \mathcal{D}$. Throughout the paper we assume that the σ-algebra $\mathcal{D}$ contains all singletons. This is the case if either Σ is countably generated or if for each $\tilde{m} \in M$ there exists $\tilde{E} \in \Sigma$ such that $\tilde{m}\left(\tilde{E}\right) \neq m\left(\tilde{E}\right)$ for all other $m \in M$. This property of $\mathcal{D}$ implies that all finite or countable sets M belong to $\mathcal{D}$.

Probability measures $\mu : \mathcal{D} \to [0,1]$ will be interpreted as prior probabilities. If M is finite or countable, each μ induces a posterior probability measure $\mu(\cdot \mid E) : \mathcal{D} \to [0,1]$ given by

$$\mu(m \mid E) = \begin{cases} \frac{m(E)\mu(m)}{\sum_{m \in \operatorname{supp}\mu} m(E)\mu(m)} & \text{if } m(E) > 0 \text{ for some } m \in \operatorname{supp}\mu \\ 0 & \text{else} \end{cases}$$

for all $m \in M$.

A subset M of Δ is said to be *measure independent* if, given any bounded measure $\gamma : \mathcal{M} \to \mathbb{R}$,

$$\int_M m(E)\, d\gamma(m) = 0 \quad \forall E \in \Sigma \quad \Longrightarrow \quad \gamma = 0.$$

If M is finite, measure independence reduces to usual notion of linear independence. In other words, by setting $M = \{m_1, ..., m_n\}$, given any collection of scalars $\{\alpha_i\}_{i=1}^n$,

$$\sum_{i=1}^n \alpha_i m_i(E) = 0 \quad \forall E \in \Sigma \Longrightarrow \alpha_1 = \cdots = \alpha_n = 0.$$

When the state space S is finite, this condition further reduces to: given any collection of scalars $\{\alpha_i\}_{i=1}^{n}$,

$$\sum_{i=1}^{n} \alpha_i m_i(s) = 0 \quad \forall s \in S \Longrightarrow \alpha_1 = \cdots = \alpha_n = 0.$$

This is a condition of linearly independence of the $|M|$ vectors $(m(s) : s \in S) \in \mathbb{R}^{|S|}$, which amounts to require that the rank of the associated $|S| \times |M|$ matrix is full.

Two measures m and $\tilde{m}$ in Δ are *orthogonal (or singular)*, written $m \perp \tilde{m}$, if there exists $E \in \Sigma$ such that $m(E) = 0 = \tilde{m}(E^c)$. A collection of models $M \subseteq \Delta$ is *(pairwise) orthogonal* if all its elements are pairwise orthogonal.

If $m(E) = 0$ implies $\tilde{m}(E) = 0$ for each $E \in \Sigma$, then we say that $\tilde{m}$ is *absolutely continuous* with respect to m and we write $\tilde{m} \ll m$. In this case $d\tilde{m}/dm$ denotes the *Radon-Nikodym derivative* of $\tilde{m}$ with respect to m. If we have both $\tilde{m} \ll m$ and $m \ll \tilde{m}$, we write $m \equiv \tilde{m}$ and we say that m and $\tilde{m}$ are *equivalent*.

In applications, probability models are often assumed to have a density. In our setting this means that there is a probability $\lambda \in \Delta$ such that $m \ll \lambda$ for all $m \in M$.[7] In this case we say that M is *dominated*. Finite collections of models are trivially dominated, as well as countable ones: for, it is enough to set $\lambda = \sum_{k=1}^{\infty} 2^{-k} m_k$. Using a result of Halmos and Savage [14], next we show that for orthogonal subsets this is the only case.

Lemma 1 *An orthogonal subset M of Δ is dominated if and only if it is finite or countable.*

Finally, a measure $m \in \Delta$ is *nonatomic* if, for each $E \in \Sigma$ such that $m(E) > 0$, there exists $F \subseteq E$ such that $0 < m(F) < m(E)$. We denote by $\Delta_{na}(S)$ the collection of all nonatomic probability measures. The main property of nonatomic measures is the classic Lyapunov Theorem that says that the range $\{(m_1(E), ..., m_n(E)) : E \in \Sigma\}$ of a finite collection $\{m_i\}_{i=1}^{n}$ of nonatomic measures is a convex subset of $\mathbb{R}^n$. In particular, a single probability measure is nonatomic if and only if it is convex-ranged.

3 Finite representation

3.1 Basic result

The first issue to consider in our normative approach is how decision makers' behavior should reflect the fact that they regard M as a datum of the decision problem. To this end, given a subset M of Δ say that an event E is *unanimous* if $0 < m(E) = m'(E) < 1$ for all $m, m' \in M$. In other words, all probability models in M assign the same probability to event E.

Definition 2 *A preference $\succsim$ is consistent with a subset M of Δ if, for some outcomes $x \succ y$,*

$$m(F) = m(E) \quad \forall m \in M \Longrightarrow xFy \sim xEy \tag{5}$$

[7] The apparently weaker requirement that λ be σ-finite is actually equivalent to λ be a probability, as Halmos and Savage [14, p. 322] observed.

for all $F \in \Sigma$ and all unanimous $E \in \Sigma$.

Consistency requires that the decision maker is indifferent among bets on events that all probability models in M classify as equally likely. The next stronger consistency property requires that decision makers prefer to bet on events that are more likely according to all models.

Definition 3 *A preference $\succsim$ is order consistent with a subset M of Δ if, for some outcomes $x \succ y$,*

$$m(F) \geq m(E) \quad \forall m \in M \Longrightarrow xFy \succsim xEy \tag{6}$$

for all $F \in \Sigma$ and all unanimous $E \in \Sigma$.

Both these notions are minimal consistency requirements among information and preference that behaviorally reveal that decision makers consider M as a datum of the decision problem. To an outside observer, aware of datum M, these consistency notions are the behavioral markers that reveal that the decision makers actually regard M as datum of the decision problem.

We can now state our basic representation results, which considers finite sets M of nonatomic models.

Proposition 4 *Let M be a finite subset of $\Delta_{na}(S)$ and $\succsim$ a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies P.1-P.6 and it is order consistent with M;

(ii) there exist a non-constant utility function $u: X \to \mathbb{R}$ and a prior $\mu: \mathcal{D} \to [0,1]$, with $\operatorname{supp}\mu \subseteq M$, such that

$$V(f) = \int_{\Delta} \left(\int_S u(f(s))\, dm(s) \right) d\mu(m) \tag{7}$$

represents $\succsim$.

Moreover, u is cardinally unique, while μ is unique for each such $\succsim$ if and only if M is linearly independent.

Uniqueness of the prior μ is an important feature of this result. In fact, it pins down μ even though its domain is made of unobservable probability models. Because of the structure of Δ, it is the linear independence of M – not just its affine independence – that turns out to be equivalent to this uniqueness property. This simple, but useful, fact is well known (see, e.g., Teicher [35]).

Each prior $\mu: \mathcal{D} \to [0,1]$ induces a *predictive probability* $\bar{\mu}: \Sigma \to [0,1]$ on the sample space through reduction:

$$\bar{\mu}(E) = \int_{\Delta} m(E)\, d\mu(m) \qquad \forall E \in \Sigma. \tag{8}$$

The reduction map $\mu \mapsto \bar{\mu}$ relates subjective probabilities on the sample space to subjective probabilities on space of models, that is, prior and predictive probabilities.[8] Clearly, (7) implies that

$$V(f) = \int_S u(f(s))\, d\bar{\mu}(s) \qquad \forall f \in \mathcal{F}, \tag{9}$$

which is the reduced form of V, its Savage's SEU form. As observed in the Introduction, this is the criterion that an outside observer, unaware of datum M, would be able to elicit from decision maker's behavior. It is a much weaker representation than the "structural" one (7), which can be equivalently written as

$$V(f) = \int_M \left(\int_S u(f(s))\, dm(s) \right) d\mu(m)$$

since $\operatorname{supp} \mu \subseteq M$ and $M \in \mathcal{D}$. This is the criterion that, instead, an outside observer aware of M would be able to elicit. For, denote by $\Delta(M)$ the collection of all priors μ such that $\operatorname{supp} \mu \subseteq M$. The informed observer would be able to focus on the restriction $\rho_M : \Delta(M) \to \Delta$ of the reduction map on $\Delta(M)$. If M is linear independent, the restriction ρ_M is one-to-one and thus allows prior identification from a behaviorally elicited Savagean probability $P \in \Delta$ through the inverse $\rho_M^{-1}(P)$, at least in principle.[9]

The structural representation (7) is a version of Savage's representation that may be called *Classical Subjective Expected Utility* since it takes into account Waldean information, with its classical flavor.[10] In place of the usual SEU pair (u, P) the representation is now characterized by a triple (u, M, μ), with $\operatorname{supp} \mu \subseteq M$. According to the Bayesian paradigm, the prior μ quantifies probabilistically the decision maker's uncertainty about which model in M is the true one. This kind of uncertainty is sometimes called *(probabilistic) model uncertainty* or *parametric uncertainty.*

In the Introduction we observed that when datum M is a singleton $\{m\}$ the Classical SEU criterion (7) reduces to

$$V(f) = \int_S u(f(s))\, dm(s) \tag{10}$$

In this case it trivially holds $\bar{\mu} = m$ and so subjective beliefs do not play any role. For this reason (10) is a von Neumann-Morgenstern Expected Utility criterion, which is thus the special case of Classical SEU that corresponds to singleton data.

In contrast, when M is nonsingleton but the support of some prior μ is a singleton, say $\operatorname{supp} \mu = \{\tilde{m}\}$ with $\tilde{m} \in M$, then it is the decision maker's personal information that prior μ reflects that leads him to the Dirac predictive probability $\mu = \delta_{\tilde{m}}$. In this case,

$$V(f) = \int_\Delta \left(\int_S u(f(s))\, dm(s) \right) d\delta_{\tilde{m}}(m) = \int_S u(f(s))\, d\tilde{m}(s)$$

is a Savage's SEU criterion.

[8] Notice that probability measures on S can play two conceptually altogether different roles: predictive probabilities and probability models.

[9] In fact, computing inverses can be highly nontrivial, as the trapdoor functions of Diffie and Hellman [9] famously show.

[10] Diaconis and Freedman [7] call "classical Bayesianism" the Bayesian approach that considers as a datum of the statistical problem the collection of all possibe data generating mechanisms.

3.2 Support

In Proposition 4 the support of the prior is included in M, i.e., $\operatorname{supp}\mu \subseteq M$. For, because of consistency models are assigned positive probability only if they belong to datum M. But, the decision maker may well disregard some models in M because of some personal information that his subjective belief μ may reflect.[11] In this case the inclusion is strict and $\mu(m) = 0$ for some $m \in M$.

Next we behaviorally characterize – through a consistency condition – the models in M that belong to the prior's support. These are the models that the decision maker believes to carry significant probabilistic information for his decision problem. We consider linearly independent data M in view of the uniqueness result in Proposition 4.[12]

Proposition 5 *Let M be linearly independent. In Proposition 4, a model $m \in M$ belongs to* $\operatorname{supp}\mu$ *if and only if, for all $E \subseteq F$,*

$$m(E) < m(F) \Longrightarrow xEy \prec xFy \tag{11}$$

for some $x \succ y$.

The significance of a model m is thus revealed by the rankings of nested events $E \subseteq F$. Since they are nested, all models agree that $m(E) \leq m(F)$. This agreement is what turns out to make it possible the behavioral identification, through (11), of the pivotal role of a model m, and so of whether it belongs to the support of the prior.

3.3 Variations

We close by establishing the conditional and orthogonal versions of Proposition 4. We begin with the conditional version, that is, with the counterpart of representation (4) under Waldean information.

Proposition 6 *Let M be a finite subset of $\Delta_{na}(S)$ and $\succsim$ a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies P.1-P.6 and it is order consistent with M;

(ii) there exist a non-constant utility function $u : X \to \mathbb{R}$ and a prior $\mu : \mathcal{D} \to [0,1]$, with $\operatorname{supp}\mu \subseteq M$*, such that, for all nonnull events E,*

$$V_E(f) = \int_\Delta \left(\int_S u(f(s))\, dm(s \mid E) \right) d\mu(m \mid E) \tag{12}$$

represents $\succsim_E$.

Moreover, u is cardinally unique, while μ is unique for each such $\succsim$ if and only if M is linearly independent.

[11] In fact, the interpretation of μ is purely subjective, not at all logical/objective a la Carnap and Keynes.

[12] The "only if" part actually holds even without linear independence.

The representation of the conditional preference $\succsim_E$ thus depends on the conditional model $m(\cdot \mid E) : \Sigma \to [0,1]$ and on the posterior probability $\mu(\cdot \mid E) : \mathcal{D} \to [0,1]$ that, respectively, update in light of E the model m and prior μ. Criterion (12) shows how decision makers currently plan to use the information they may gather through observations to update their inference on the actual model that generates data.[13]

The conditional predictive probability $\bar{\mu}(\cdot \mid E) : \Sigma \to [0,1]$ is given by

$$\bar{\mu}(F \mid E) = \int_{\Delta} m(F \mid E)\, d\mu(m \mid E) \qquad \forall F \in \Sigma. \tag{13}$$

The reduced form of (12) is thus given by

$$V_E(f) = \int_S u(f(s))\, d\bar{\mu}(s \mid E) \tag{14}$$

The conditional representations (12) and (14) are, respectively, induced by the primitive representations (7) and (9) via conditioning.

Orthogonality is a simple, but important, sufficient condition for linear independence.

Lemma 7 *Orthogonal subsets M are linearly independent.*

Section 5 will show that some fundamental classes of models satisfy this convenient condition. Because of its importance, the following result shows what form the Classical SEU representation of Proposition 4 takes in this case.

Proposition 8 *Let M be a finite and orthogonal subset of $\Delta_{na}(S)$ and $\succsim$ a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) *$\succsim$ satisfies P.1-P.6 and it is consistent with M;*

(ii) *there exist a non-constant utility function $u : X \to \mathbb{R}$ and a prior $\mu : \mathcal{D} \to [0,1]$, with $\operatorname{supp} \mu \subseteq M$, such that*

$$V(f) = \int_{\Delta} \left(\int_S u(f(s))\, dm(s) \right) d\mu(m)$$

represents $\succsim$.

Moreover, μ is unique and u is cardinally unique.

Notice that here consistency suffices and that the prior μ is automatically unique because of the orthogonality of M.

[13] As Marschak [27, p. 109] remarked "to be an 'economic man' implies being a 'statistical man'"

4 Infinite representation

We now consider infinite data, that is, collections M that may have an infinite number of elements. The extension of the previous representations to the infinite case is nontrivial because the Lyapunov Theorem, which plays a key role in these Savagean results, in general fails for infinite collections of nonatomic measures.[14] However, for the fundamental orthogonal case this classical theorem holds and we are thus able to establish an infinite version of Proposition 8. To this end, we need the following stronger version of consistency.

Definition 9 *A preference $\succsim$ is strongly consistent with a subset M of Δ if there are outcomes $x \succ y$ such that*

$$m(F) = m(E) \quad \forall m \in M \Longrightarrow xFy \sim xEy \tag{15}$$

for all $E, F \in \Sigma$.

In other words, decision makers are indifferent among bets on events that, model by model, have the same probability. It is no longer enough to consider only unanimous events. Though stronger than consistency, condition (15) is still a natural consistency condition.

We can now state our main representation result, the infinite version of Proposition 8. Recall that, by Lemma 1, dominated orthogonal subsets are finite or countable.

Proposition 10 *Let M be a dominated orthogonal subset of $\Delta_{na}(S)$ and $\succsim$ a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies P.1-P.6 and is strongly consistent with M;

(ii) there exist a non-constant utility function $u : X \to \mathbb{R}$ and a prior $\mu : \mathcal{D} \to [0,1]$, with $\operatorname{supp} \mu \subseteq M$, such that

$$V(f) = \int_{\Delta} \left(\int_S u(f(s)) \; dm(s) \right) d\mu(m) \tag{16}$$

represents $\succsim$.

Moreover, μ is unique and u is cardinally unique.

Since finite collections M are trivially dominated, this result generalizes Proposition 8 to infinite M, modulo the stronger version of consistency assumed. More importantly, the examples considered in the intertemporal setting show that orthogonal sets M of models are often used in applications.

The reduction map ρ_M between prior and predictive probabilities is easily seen to preserve convexity: for all $\alpha \in [0,1]$ it holds $\alpha\mu + (1-\alpha)\mu' \mapsto \alpha\bar{\mu} + (1-\alpha)\bar{\mu}$, and viceversa. More interestingly, in the orthogonal case it also preserves both orthogonality and equivalence, as next we show.

[14] As well known, Savage-type results critically rely on the range convexity of the involved subjective probabilities. The importance of the Lyapunov Theorem in our setting is an example of this well know methodological feature of Savage's approach.

Proposition 11 *Under the hypotheses of Proposition 10, two priors μ and μ' are orthogonal (resp., equivalent) if and only if their predictive probabilities $\bar{\mu}$ and $\bar{\mu}'$ are orthogonal (resp., equivalent).*

Notice that the "if" part for orthogonality and the "only if" part for equivalence hold in general, even if M is not orthogonal.

5 Intertemporal illustration

We illustrate the previous results through a standard intertemporal decision problem where observations are generated over time. We first introduce the intertemporal setting and we then give a few important examples of orthogonal collections of probability models.

5.1 Setting

Consider an intertemporal decision problem where information builds up through observations generated by a sequence of random variables $\{Z_t\}$ defined on some (possibly unverifiable, except to Laplace's demon) underlying space and taking values on observation spaces $\mathcal{Z}_t$ that, for ease of exposition, we assume to be at most countable (all results in this section actually hold in Polish spaces). For example, the sequence $\{Z_t\}$ can model subsequent draws of balls from a sequence of (possibly identical) urns; in this case $\mathcal{Z}_t$ consists of the possible colors of the balls that can be drawn in urn t.

Suppose, for convenience, that all observation spaces are identical – each denoted by $\mathcal{Z}$ and endowed with the σ-algebra $\mathcal{B} = 2^{\mathcal{Z}}$ – and that the relevant state space S for the decision problem is the overall sample space $\mathcal{Z}^\infty = \prod_{t=1}^{\infty} \mathcal{Z}$. Its points $z = (z_1, ..., z_t, ...)$ are the possible observation paths generated by the sequence $\{Z_t\}$. Without loss of generality, we identify $\{Z_t\}$ with the coordinate process such that $Z_t(z) = z_t$.

Endow $\mathcal{Z}^\infty$ with the product σ-algebra $\mathcal{B}^\infty$ generated by the elementary cylinder sets

$$z^t = \{z_1, ..., z_t\} = \{z_1\} \times \cdots \times \{z_t\} \times \mathcal{Z} \times \cdots$$

The elementary cylinder sets are the observables in this intertemporal setting. In particular, the filtration $\{\mathcal{B}_t\}$, where $\mathcal{B}_0 \equiv \{S, \emptyset\}$ and $\mathcal{B}_t$ is the algebra generated by the cylinders $\{z_1, ..., z_t\}$, records the building up of observations. Clearly, $\mathcal{B}^\infty$ is the σ-algebra generated by the filtration $\{\mathcal{B}_t\}$, that is, $\mathcal{B}^\infty = \sigma\left(\bigcup_t \mathcal{B}_t\right)$.

Since elementary cylinder sets z^t are observable, conditioning relative to them is especially important. In particular,[15] the probability measure $m\left(\cdot \mid z^t\right) : \Sigma \to [0,1]$ given by

$$m\left(E \mid z^t\right) = \begin{cases} \frac{m\left(E \cap z^t\right)}{m\left(z^t\right)} & \text{if } m\left(z^t\right) > 0 \\ 0 & \text{else} \end{cases}$$

[15] To ease notation we write $m(z_1, ..., z_n)$ in place of $m(\{z_1, ..., z_n\})$.

is the conditional distribution of model m given observations $z^t = \{z_1, ..., z_t\}$, while the probability measure $\mu : \mathcal{D} \to [0,1]$ given by

$$\mu\left(A \mid z^t\right) = \begin{cases} \frac{\int_A m(z^t) d\mu}{\int_\Delta m(z^t) d\mu} & \text{if } \int_\Delta m\left(z^t\right) d\mu > 0 \\ 0 & \text{else} \end{cases}$$

is the *posterior distribution* of prior μ given observations $z^t = \{z_1, ..., z_t\}$.

In this intertemporal setting the pair (S, Σ) is thus given by $(\mathcal{Z}^\infty, \mathcal{B}^\infty)$. The space of models Δ consists of all probability measures $m : \mathcal{B}^\infty \to [0,1]$. Acts are adapted outcome processes $f = \{f_t\} : \mathcal{Z}^\infty \to X$, which we often call *plans*. The outcome space X has also a product structure $X = \mathcal{C}^\infty$, where $\mathcal{C}$ is a common instant outcome space. We consider Classical SEU representations

$$V(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} u(f(z)) \ dm(z) \right) d\mu(m) \tag{17}$$

Its conditional version relative to cylinder sets z^t is:

$$V_{z^t}(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} u(f(z)) \ dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \tag{18}$$

The function $u : \mathcal{C}^\infty \to \mathbb{R}$ in (17) and (18) is an intertemporal utility $u(c) = u(c_1, ..., c_t, ...)$ that, under standard additional conditions, has a classic discounted form

$$u(c_1, ..., c_t, ...) = \sum_{t=1}^{\infty} \beta^{t-1} \upsilon(c_t) \tag{19}$$

with subjective discount factor $\beta \in [0,1]$ and instantaneous utility function $\upsilon : \mathcal{C} \to \mathbb{R}$. For instance, the discounted version of (18) is:

$$V_{z^t}(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} \sum_{\tau=1}^{\infty} \beta^{\tau-1} \upsilon(f_\tau(z)) \ dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \tag{20}$$

Throughout the section we assume that utility is bounded, that is, $\sup_{c \in \mathcal{C}^\infty} |u(c)| < \infty$. For example, in the discounting case (19) this condition holds provided instantaneous utility functions are bounded, that is, $\sup_{c \in \mathcal{C}} |\upsilon(c)| < \infty$.

5.2 A general singularity condition

A general characterization of the orthogonality of two measures in a filtration setting is due to Kabanov, Liptser, and Shiryaev [17], who generalized an earlier classic results of Kakutani [18]. To present it, denote by m_t the restriction on $\mathcal{B}_t$ of a model $m \in \Delta$. Given two models $m, \tilde{m} \in \Delta$ such that $m_t \ll \tilde{m}_t$ for each t, let $\lambda_t : \mathcal{Z}^\infty \to \mathbb{R}$ be their *likelihood ratio* given by

$$\lambda_t\left(z^t\right) = \frac{dm_t}{d\tilde{m}_t}\left(z^t\right) = \begin{cases} \frac{m(z^t)}{\tilde{m}(z^t)} & \text{if } \tilde{m}\left(z^t\right) > 0 \\ 0 & \text{else} \end{cases} \tag{21}$$

Given the two models m and $\tilde{m}$, the likelihood ratio process $\{\lambda_t\}$ can be constructed from observations; we will say more about it in Section 6. Here define the conditional likelihood $l_t : \mathcal{Z}^\infty \to \mathbb{R}$ by

$$l_t\left(z^t\right) = \begin{cases} \frac{\lambda_t\left(z^t\right)}{\lambda_{t-1}\left(z^{t-1}\right)} & \text{if } \lambda_{t-1}\left(z^{t-1}\right) > 0 \\ 0 & \text{else} \end{cases}$$

for each $t \geq 2$.

The next fundamental lemma of Kabanov et al [17] characterizes orthogonal measures in a general filtration setup $\{\Sigma_t\}_{t=0}^{\infty}$ through a predictive property of the conditional likelihood l_t.

Lemma 12 *Two probability measures $m, \tilde{m} \in \Delta$, with $m_t \ll \tilde{m}_t$ for each t, are orthogonal if and only if*

$$\sum_{t=1}^{\infty} \left[1 - E_{\tilde{m}}\left(\sqrt{l_t} \mid \Sigma_{t-1}\right)\right] = \infty \qquad m\text{-a.e.} \tag{22}$$

The series in (22) has positive terms. For, as observed by [17, p. 213], it holds

$$E_{\tilde{m}}\left(\sqrt{l_t} \mid \Sigma_{t-1}\right) \leq 1 \qquad m\text{-a.e.}$$

In particular, a simple sufficient condition for (22) is

$$\limsup_t E_{\tilde{m}}\left(\sqrt{l_t} \mid \Sigma_{t-1}\right) < 1 \qquad m\text{-a.e.}$$

Condition (22) can be easily stated for collections M of probabilities. For later reference, next we state this version of Lemma 12.

Proposition 13 *A collection M of models, with equivalent restrictions on each $\mathcal{B}_t$, is orthogonal provided condition (22) holds for all $m, \tilde{m} \in M$.*

In view of all this, to ease the derivation in the rest of the paper we will often consider models m that are *strictly positive* on each $\mathcal{B}_t$, that is, $m_t\left(z^t\right) > 0$ for each elementary cylinder $\left\{z^t\right\}$. Clearly, any two such models have equivalent restrictions on $\mathcal{B}_t$. We denote by Δ_+ the set of all models $m : \mathcal{B}^\infty \to [0,1]$ that are strictly positive on each $\mathcal{B}_t$.

5.3 Independence

An important special case of Proposition 13 is the i.i.d. case originally studied by Kakutani [18]. Consider a model $m \in \Delta$ that makes the coordinate process $\{Z_t\}$ i.i.d., with marginal distribution $\pi : \mathcal{B} \to [0,1]$. In this case, m is a product probability on $\mathcal{B}^\infty$ uniquely determined by the marginal π. In particular, it holds $m\left(z^t\right) = \prod_{i=1}^{t} \pi\left(z_i\right)$ on each elementary cylinder $z^t = \{z_1, ..., z_t\}$.

Proposition 14 *A collection $M \subseteq \Delta_+$ of models that make the coordinate process $\{Z_t\}$ i.i.d. is orthogonal.*

For example, in the i.i.d. binomial case, with $\mathcal{Z} = \{0,1\}$ and $E_m(Z_1) \in (0,1)$, we can parametrize M with the open unit interval $(0,1)$. By Proposition 14, the set $M = \{m_\alpha\}_{\alpha\in(0,1)}$ is orthogonal. As observed by Kakutani [18, p. 223], this would also follow from the strong law of large numbers, which implies $m\left(\lim t^{-1}\sum_{i=1}^{t} Z_i = E_m(Z_1)\right) = 1$ for each $m \in M$. Kakutani's remark is easily generalized, via the Pointwise Ergodic Theorem, to show that models that make the coordinate process $\{Z_t\}$ stationary and ergodic are orthogonal. Proposition 14 thus holds more generally for them.

By Proposition 4, if $\succsim$ satisfies P.1-P.6 and is strongly consistent with a countable collection M of i.i.d. models, then there is a cardinally unique utility function u and a unique prior μ, with $\operatorname{supp}\mu \subseteq M$, such that

$$V(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} u(f(z))\ dm(z)\right) d\mu(m) \tag{23}$$

represents $\succsim$. As already observed, the reduced form

$$V(f) = \int_{\mathcal{Z}^\infty} u(f(z))\ d\bar{\mu}(z),$$

is what can be elicited from behavior without the knowledge of M. The predictive probability $\bar{\mu}$ is exchangeable. In the purely subjective approach a la de Finetti that Savage adopted, from this reduced exchangeable form – via de Finetti Theorem-type arguments – it is inferred a candidate collection M of models, a "subjective datum," for which the structural form (23) holds. Here we follow an opposite Waldean path where M is an "objective datum" of the problem. The relations with the de Finetti subjective approach will be further discussed in the Concluding Remarks.

5.4 Gaussian and Markov cases

Consider a model m that makes the coordinate process $\{Z_t\}$ a Markov chain with transition functions $\pi_t : \mathcal{Z} \times \mathcal{B} \to [0,1]$ for $t \geq 1$, where $\pi_t(z_t, \cdot) : \mathcal{B} \to [0,1]$ is strictly positive[16] for each $z_t \in \mathcal{Z}$, and $\pi_t(\cdot, z_{t+1}) : \mathcal{Z} \to [0,1]$ is a function for each $z_{t+1} \in \mathcal{B}$. Given an initial probability distribution π_0 on $\mathcal{B}$, the model m is uniquely determined by π as follows:

$$m\left(z^t\right) = \pi_0(z_1) \prod_{i=1}^{t-1} \pi_i(z_i, z_{i+1})$$

for each cylinder set $z^t = \{z_1, ..., z_t\}$. Denote by $\|\cdot\|$ the Euclidean norm of $\mathbb{R}^{|\mathcal{Z}|}$.

Proposition 15 *A collection $M \subseteq \Delta_+$ of models that make the coordinate process $\{Z_t\}$ a Markov chain is orthogonal if*

$$\liminf_t \left\| \sqrt{\tilde{\pi}_t(z_t, \cdot)} - \sqrt{\pi_t(z_t, \cdot)} \right\| > 0 \qquad m\text{-a.e.} \tag{24}$$

for all $m, \tilde{m} \in M$.

[16] That is, $\pi_n(z_n, z_{n+1}) > 0$ for all $z_{n+1} \in \mathcal{Z}$.

In the homogeneous case, when there is a transition function π such that $\pi_t = \pi$ for all t, it is easy to see that condition (24) is always satisfied. We thus have the following generalization of Proposition 14.

Corollary 16 *A collection $M \subseteq \Delta_+$ of models that make the coordinate process $\{Z_t\}$ a homogeneous Markov chain is orthogonal.*

We close the study of the Markov case by observing that $2^{-1}\left\|\sqrt{\tilde{\pi}_t\left(z_t,\cdot\right)}-\sqrt{\pi_t\left(z_t,\cdot\right)}\right\|$ is the Hellinger distance between the probability measures $\tilde{\pi}_t\left(z_t,\cdot\right)$ and $\pi_t\left(z_t,\cdot\right)$ on $\mathcal{B}$. Condition (24) can thus be stated in terms of this distance.

For ease of exposition so far we considered at most countable $\mathcal{Z}$, though the results of this section hold for Polish spaces. In this final part of this section we relax this assumption and suppose that $\mathcal{Z} = \mathbb{R}$ endowed with its Borel σ-algebra. Consider a model $m \in \Delta$ that makes the coordinate process $\{Z_t\}$ independent with Gaussian marginal distribution $\pi_t : \mathcal{B} \to [0,1]$ with parameters $\left(a_t, \sigma_t^2\right)$. The next result shows that a collection of independent Gaussian models is orthogonal under mild conditions on the parameters.

Proposition 17 *A collection M of models that make the coordinate process $\{Z_t\}$ independent, with equivalent Gaussian marginals, is orthogonal if, for all $m, \tilde{m} \in M$, either* $\lim_t \tilde{\sigma}_t^2/\sigma_t^2 \neq 1$ *or* $\lim_t a_t \neq \lim_t \tilde{a}_t$ *and* $\limsup_t \tilde{\sigma}_t^2 < \infty$.

6 Learning

6.1 Beliefs

In intertemporal decision problems, where conditional preferences are contingent upon more and more observations, it is important to see if decision makers eventually learn from observations the true model among those in the support of their prior probabilities (that is, among those that ex ante they regarded as possible data generating processes). Decision makers' long run behavior is consistent with observations, which asymptotically determine their behavior by "swamping" any subjective beliefs they may have.

The likelihood ratio process $\{\lambda_t\}$ plays a central role in the study of this issue. As well known, it is a martingale with respect to the filtration $\{\mathcal{B}_t\}$ and so, by the Martingale Convergence Theorem, it converges. If it converges to zero $\tilde{m}$-a.e., it means that asymptotically the data will reveal that $\tilde{m}$ is the true model. The next known result[17] shows that orthogonality is necessary and sufficient for this key limit behavior of the likelihood ratio.

Lemma 18 *Given any two probability measures $m, \tilde{m} \in \Delta$, with $m_t \ll \tilde{m}_t$ for each t, it holds*

$$\lambda_t\left(z^t\right) \to 0 \qquad \tilde{m}\text{-}a.e. \tag{25}$$

if and only if m and $\tilde{m}$ are orthogonal.

[17] See, e.g., Billingsley [4, p. 494], and Stroock, [34, p. 287] (who credits Borge Jessen). Since we could not find a complete proof of this result, in the Appendix we give a simple proof.

We can thus distinguish asymptotically a true model $\tilde{m}$ relative to all orthogonal alternative models, and only relative to them. Without orthogonality, there can be only imperfect distinguishability among alternative models (i.e., among alternative simple hypotheses in the Statistics terminology).

The orthogonality condition (22) is thus equivalent to the limit condition (25) of Lemma 18. In particular, the important examples studied earlier in this section all satisfy condition (25). This is the case for collections of i.i.d. models (Proposition 14), as well as collections of Markov chains under some mild conditions (Proposition 15). Remarkably, the models that are most widely used in applications are thus within the scope of Lemma 18.

When M is finite and orthogonal, Lemma 18 ensures that asymptotically the true model in M will be detected almost surely. This key asymptotic property of the likelihood ratio easily translates in a consistency property of the prior μ. For, suppose $\tilde{m} \in \operatorname{supp} \mu$ and let $\lambda_t^m\left(z^t\right) = m\left(z^t\right)/\tilde{m}\left(z^t\right)$ be the likelihood ratio of m with respect to $\tilde{m}$. Then

$$\mu\left(\tilde{m} \mid z_1, ..., z_t\right) = \frac{\mu\left(\tilde{m}\right)\tilde{m}\left(z^t\right)}{\sum_{m \in \operatorname{supp}\mu} \mu\left(m\right) m\left(z^t\right)} = \frac{\mu\left(\tilde{m}\right)}{\mu\left(\tilde{m}\right) + \sum_{\tilde{m} \neq m \in \operatorname{supp}\mu} \mu\left(m\right)\lambda_t^m\left(z^t\right)}$$

and so

$$\mu\left(\tilde{m} \mid z_1, ..., z_t\right) \to 1 \qquad \tilde{m}\text{-a.e.} \tag{26}$$

if and only if, for each $m \in \operatorname{supp}\mu$ distinct from $\tilde{m}$ it holds

$$\lambda_t^m\left(z_1, ... z_t\right) \to 0 \qquad \tilde{m}\text{-a.e.} \tag{27}$$

that is, if and only if (25) holds for each such m. The prior thus asymptotically concentrates on the true model if and only if the likelihood ratios of each alternative model with respect to the true one vanish asymptotically (under the true model). The next lemma builds on this simple observation.[18]

Lemma 19 *Let $M \subseteq \Delta_+$ be a finite collection of models and $\mu : \mathcal{D} \to [0,1]$ a prior with* $\operatorname{supp}\mu \subseteq M$*. If $\tilde{m} \in \operatorname{supp}\mu$ is the true model, then*

$$\mu\left(m \mid z_1, ..., z_t\right) \to \delta_{\tilde{m}} \qquad \tilde{m}\text{-a.e.}$$

if and only if $\tilde{m}$ is orthogonal with respect to all other models $m \in \operatorname{supp}\mu$.

The orthogonality of the true model with respect to all its possible alternative models is thus a necessary and sufficient condition for full learning, that is, for data to "swamp" the prior. Without this condition, there exist priors that, though they contain the true model in their support, will never learn the true model regardless of the available amount of data.

Propositions 14 and 15 show that this apparently strong orthogonality requirement is satisfied in some fundamental cases that are widely used in applications. For them, Lemma 19 ensures full learning.

[18] In the statement we have to require that the restrictions on each $\mathcal{B}_n$ are equivalent if, instead of $M \subseteq \Delta_+\left(\mathcal{Z}^\infty\right)$, we only assume that $M \subseteq \Delta\left(\mathcal{Z}^\infty\right)$.

Finally, in terms of predictive probabilities it is easy to see that, for all $E \in \Sigma$, it holds

$$\left|\bar{\mu}\left(E \mid z_1, ..., z_t\right) - \tilde{m}\left(E \mid z_1, ..., z_t\right)\right| \to 1 \qquad \tilde{m}\text{-a.e.} \tag{28}$$

under the hypotheses of the previous lemma. The predictive probability thus converges to the true model.

6.2 Dynamic Choices

The previous lemmas lead to a decision theoretic learning result that shows that Classical SEU decision makers will eventually learn the true model and behave accordingly. To establish this learning result we need to introduce a dynamic version of our representation. To this end, consider the nodes $z^t = \{z_1, ..., z_t\}$ identified by histories of observations z^t up to t. In the dynamic setting they are decision nodes and, for this reason, at each of them there is a preference $\succeq_{z^t}$ over continuation plans from this node onwards, so that the family $\{\succeq_{z^t}\}$ characterizes a decision maker at all possible nodes that he can reach. In particular, $\succsim = \succeq_{\emptyset}$. That is, the primitive static preference studied so far may be regarded as the empty history preference in the family $\succeq_{z^t}$.[19]

The domain of each preference $\succeq_{z^t}$ is the set $\mathcal{F}$ of plans. However, to capture the idea that only continuation plans matter we require *Consequentialism*, that is,

$$f\left(z^t, z_{t+1}, ...\right) = g\left(z^t, z_{t+1}, ...\right) \quad \forall \left(z_{t+1}, ...\right) \in \mathcal{Z}^{\infty} \Longrightarrow f \sim_{z^t} g$$

for all $f, g \in \mathcal{F}$. Another classical property of the family $\{\succeq_{z^t}\}$ of preferences is *Dynamic Consistency*, that is, given any two plans f and g that are identical up to node z^t, it holds

$$f \succsim_{z^t} g \Longrightarrow f \succeq_{z^t} g$$

That is, the original ranking of plans at node z^n is not reversed once the node is reached.

A Classical SEU preference representation for $\succeq_{z^t}$ that satisfies Consequentialism and Dynamic Consistency is the preference functional $W_{z^t} : \mathcal{F} \to \mathbb{R}$ given by

$$W_{z^t}(f) = \int_{\Delta} \left(\int_{\mathcal{Z}^{\infty}} u_t\left(f_t(z), ...\right) \, dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \tag{29}$$

where $\operatorname{supp} \mu$ is included in some set $M \subseteq \Delta$ of models, $m\left(z \mid z^t\right)$ is the conditional distribution of future observations given the past ones z^t, and $u_t : \mathcal{C}^{\infty} \to \mathbb{R}$ are recursive intertemporal utilities that satisfy equations

$$u_t\left(c_t, ...\right) = \phi\left(c_t, u_{t+1}\left(c_{t+1}, ...\right)\right)$$

for some suitable aggregator $\phi : \mathcal{C} \times \mathbb{R} \to \mathbb{R}$ (cf. Marinacci and Montrucchio [26, p. 1790]).

[19] Notice that $\succeq_{z^n}$ and $\succsim_{z^n}$ denote altogether different notions: the former is a primitive preference at node z^n, the latter is the conditional preference of the primitive static preference $\succsim$ in light of evidence z^n. It is also important to observe that $\mathcal{Z}^{\infty} = \times_{k>n} \mathcal{Z}$ and $\mathcal{C}^{\infty} = \times_{k>n} \mathcal{C}$, that is, the possible future observations at node z^n are identical as at the initial node.

We call the family $\{W_{z^t}\}$ given by (29) a *recursive Classical SEU* representation. It is summarized by the triple $(\{u_t\}, M, \mu)$, with $\operatorname{supp}\mu \subseteq M$.[20] In particular, $W_\emptyset$ is the Classical SEU criterion (17).

For example, if $\phi(c, y) = \upsilon(c) + \beta y$, we have the discounted form

$$W_{z^t}(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \upsilon(f_\tau(z)) \; dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \tag{30}$$

whose recursive form is[21]

$$W_{z^t}(f) = u(f_t(z)) + \beta \int_\Delta \left(\int_{\mathcal{Z}} W_{\{z^t, z_{t+1}\}}(f) \; dm\left(z_{t+1} \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \tag{31}$$

6.3 Long run empiricists

We can now state our learning result.

Proposition 20 *Suppose $(\{u_t\}, M, \mu)$ is a recursive Classical SEU decision maker, with finite and orthogonal M. If $\tilde{m} \in \operatorname{supp}\mu$ is the true model, then*

$$\left| W_{z^t}(f) - \int_{\mathcal{Z}^\infty} u_t(f_t(z), ...) \; d\tilde{m}\left(z \mid z^t\right) \right| \to 1 \qquad \tilde{m}\text{-a.e.} \tag{32}$$

for all plans $f \in \mathcal{F}$.

The focus on orthogonal sets M is natural in view of the previous two lemmas. As observations build up, a Classical SEU decision maker will thus behave more and more like a decision maker that knows the true model. That is, like a SEU decision maker

$$\int_{\mathcal{Z}^\infty} u_t(f_t(z), ...) \; d\tilde{m}\left(z \mid z^t\right)$$

that uses the correct conditional distribution of future observations. Classical SEU decision makers are thus long run empiricists.

The reduced form of (29) is given by

$$W_{z^t}(f) = \int_{\mathcal{Z}^\infty} u_t(f_t(z), ...) \; d\bar{\mu}\left(z \mid z^t\right)$$

where $\bar{\mu}\left(z \mid z^t\right)$ is the conditional predictive distribution of future observations given history z^t. From (28) it follows that, under the hypotheses of Proposition 20, it holds:

$$\left| \int_{\mathcal{Z}^\infty} u_t(f_t(z), ...) \; d\bar{\mu}\left(z \mid z^t\right) - \int_{\mathcal{Z}^\infty} u_t(f_t(z), ...) \; d\tilde{m}\left(z \mid z^t\right) \right| \to 1 \qquad \tilde{m}\text{-a.e.}$$

The reduced form thus converges to the correct Expected Utility model.

[20] An axiomatic derivation, along the lines of Johnsen and Donaldson [16], of the recursive criterion (29) is beyond the scope of the present paper.

[21] See the Appendix for a short derivation.

7 Concluding remarks

Perspectives Our "classical" approach is very different, in a sense opposite, to the purely subjective derivations – through de Finetti Theorem-type arguments – of priors μ over collections M of models. In a nutshell, while in our approach datum M is a primitive notion upon which the analysis relies, in the de Finetti approach it is a subjective construct, a "subjective datum/parameterization", inferred from the large sample betting behavior peculiar to the de Finetti Theorem. As we remarked in Section 5.3, these arguments – widely discussed in the Bayesian literature – show when subjective predictive probabilities can be viewed as derived within a subjective parametric setup (see Al-Najjar and De Castro [1], Epstein and Seo [11], Klibanoff, Mukerji, and Seo [21], and Cerreia-Vioglio et al [5] for recent decision theoretic analyses of parametric models along these lines).

In contrast, here the class M is an element, a datum, of the problem and our purpose is to investigate how to embed it in an otherwise subjective setting. In applications where it is natural to assume the existence of a datum M, a de Finettian perspective would be a straitjacket. This is why here we take a different approach, in which classical and subjective features coexist.

Relatedly, a version of the orthogonal representations of Propositions 8 and 11 (but not of our basic result, Proposition 4) can be derived in an Ascombe-Aumann setting using the techniques of [5].[22] However, the original motivation of [5] is in a de Finetti perspective that, as just noticed, is different from our Waldean one. As a result, our paper is set in a Savage setting (with consistency notions purely based on betting behavior) and, more generally, its analysis develops along altogether different lines than that of [5], both conceptually and mathematically.

Marschak Some works of Jacob Marschak have been a source of inspiration of our exercise, in particular his [27] and [28] articles. The former paper discusses a version of criterion (1), nicely summarized by the sentence "to be an 'economic man' implies being a 'statistical man'" that we mentioned after Proposition 6.

More importantly, our work addresses the issue that he raised in the latter paper, in which he asked how to pin down subjective beliefs on models from observables. In so doing, our analysis also shows that to study general data M, possibly linearly dependent, it is necessary to go beyond betting behavior on observables. For example, the study of second order acts in Klibanoff, Marinacci, and Mukerji [20] goes in that direction, though the classic work of Marschak and Radner [29] already considered acts – and so bets – on models in a Savage framework.[23]

[22] In [5] we would need the following lemma (we use its terminology): Let M be a finite subset of Δ and $\{E_n\}_{n\in\mathbb{N}}$ be a partition such that $m_n(E_n) = 1$ and $m_n(E_k) = 0$ if $k \neq n$; then both (S, Σ, M) and $(S, \Sigma, \mathrm{co}M)$ are Dynkin spaces with respect to $\sigma(E_n : n \in \mathbb{N})$ and the regular conditional probability $p : S \to \Delta$ given by $p(\omega) = m_n$ if $\omega \in E_n$.

[23] See, for example, Table 2.1 on page 48, which in our setup corresponds to a state space $S \times M$ (see also Radner [31] for a similar state space). In the same page they write "... in the formulation of a decision problem, the states of the environment must be described in sufficient detail to cover not only those aspects relevant to the payoff function, but also those aspects relevant to the type of information on which the decisions may be based."

Experimentation In our Savagean setting uncertainty is exogenous in that it is not affected by decision makers' choices. As a result, there is no room for experimentation, that is, choices whose purpose is to acquire information on how uncertainty may resolve. The extension of Classical SEU to include endogenous uncertainty is a natural next step in our analysis. This would allow to relate our decision theoretic framework to control models a la Easley and Kiefer [10], which used versions of the functional form (1) to study learning with endogenous uncertainty.

Model uncertainty Throughout the paper the datum M was supposed to be known to the decision maker, there is no uncertainty about it. In other words, the decision maker has enough information to identify the collection of all models that can generate the observations. If this is not the case we have a further source of uncertainty, which is often called *statistical model uncertainty* (see, e.g., Claeskens and Hjort, [6]). This would add a further layer in the representation (7) with a "meta-prior" over 2^M.

Principal Principle By our consistency conditions, for each $\alpha \in [0,1]$ it holds

$$m(E) = \alpha \quad \forall m \in M \Longrightarrow \bar{\mu}(E) = \alpha$$

for all priors μ with $\operatorname{supp} \mu \subseteq M$. In other words, if all models in M agree that the probability of some event E is α, then its predictive probability is also α, regardless of any personal information that a prior μ might reflect.

Mutatis mutandis, this property can be seen as a form of the Principal Principle of Lewis [23], an important notion in the Philosophy of Probability that requires that degrees of beliefs be consistent with objective chances, and only with them (any other possible information on events becomes irrelevant once objective chances are available).

Axiom P.6 Since $M \subseteq \Delta_{na}(S)$, it is easy to see that axiom P.6 is no longer needed in the first part of the proof of Savage's Representation Theorem that derives the subjective probability P. However, it is still needed later in the proof of Savage's Theorem (see his Theorem 5.2.2) and this is why it appears in our Proposition 4.

8 Appendix: proofs and related analysis

Denote by $\Delta(M)$ the collection of all probability measures $\mu : \mathcal{M} \to [0,1]$. Given a set $M \subseteq \Delta$, among priors and predictive distributions there is a two ways relation. For example, given a predictive probability P denote by $\Gamma_M(P)$ the (possibly empty) set of priors $\mu \in \Delta(M)$ that induce P, that is, $\Gamma_M(P) = \{\mu \in \Delta(M) : P = \bar{\mu}\}$. On the other hand, each prior μ induces a predictive probability $\bar{\mu}$.

It is easy to see that the correspondence $\Gamma_M : \Delta \to 2^{\Delta(M)}$ is convex valued and with disjoint images, i.e., $\Gamma_M(P) \cap \Gamma_M(P') = \emptyset$ if $P \neq P'$. Given Γ_M, the effective domain of Γ_M is defined to be

$$\operatorname{dom} \Gamma_M = \{P \in \Delta : \Gamma_M(P) \neq \emptyset\}$$

In general $\operatorname{dom}\Gamma_M$, is the collection, $\{\bar{\mu} : \mu \in \Delta(M)\}$, of all predictive probabilities that are induced by priors in $\Delta(M)$. If M is finite then $\operatorname{dom}\Gamma_M = \operatorname{co}(M)$. We say that P is M-representable if $\Gamma_M(P) \neq \emptyset$ and if, in addition, $\Gamma_M(P)$ is a singleton, we say that P is M-identifiable.[24]

Lemma 21 *Given $M \subseteq \Delta$, each predictive probability $P \in \operatorname{dom}\Gamma_M$ is M-identifiable if and only if M consists of measure independent models.*

Proof We first prove sufficiency. Consider $P \in \operatorname{dom}\Gamma_M$. Assume that M consists of measure independent models. Since $P \in \operatorname{dom}\Gamma_M$, there exists $\mu \in \Delta(M)$ such that $P = \bar{\mu}$. Next, consider $\mu_1, \mu_2 \in \Delta(M)$ such that $\bar{\mu}_1 = \bar{\mu}_2 = P$. Define $\gamma = \mu_1 - \mu_2$. We have that γ is a bounded measure on $\mathcal{M}$. It follows that for each $A \in \mathcal{F}$

$$\int_M m(A)\, d\gamma(m) = \int_M m(A)\, d\mu_1(m) - \int_M m(A)\, d\mu_2(m) = P(A) - P(A) = 0.$$

Since M consists of measure independent models, it follows that $\mu_1 - \mu_2 = \gamma = 0$, that is, $\mu_1 = \mu_2$. This proves the uniqueness of μ and the M-identifiability of P. As to the converse, assume that each $P \in \operatorname{dom}\Gamma_M$ is M-identifiable. By contradiction, assume that M does not consist of measure independent models. Thus, there exists a bounded measure $\gamma : \mathcal{M} \to \mathbb{R}$ such that

$$\gamma \neq 0 \text{ and } \int_{\mathcal{M}} m(A)\, d\gamma(m) = 0 \qquad \forall A \in \mathcal{F}. \tag{33}$$

Since γ is a bounded signed measure, γ admits a decomposition in terms of two finite measures, in other words, $\gamma = \gamma^+ - \gamma^-$ where γ^+ and γ^- are, respectively,the positive and negative part of γ. By (33), we have that

$$0 = \int_M m(S)\, d\gamma(m) = \int_M 1_M d\gamma = \gamma(M) = \gamma^+(M) - \gamma^-(M).$$

Since $\gamma \neq 0$, this implies that $\gamma^+(M) = \gamma^-(M) > 0$. Set

$$\mu_1 = \frac{\gamma^+}{\gamma^+(M)}, \quad \mu_2 = \frac{\gamma^-}{\gamma^-(M)}, \text{ and } \quad \eta = \gamma/\gamma^+(M).$$

It follows that $\mu_1, \mu_2 \in \Delta(M)$, $\mu_1 \neq \mu_2$, and $\eta = \mu_1 - \mu_2$. By (33), we have that for each $A \in \Sigma$

$$\begin{aligned} 0 &= \frac{1}{\gamma^+(S)} \left(\int_M m(A)\, d\gamma(m) \right) = \int_M m(A)\, d\eta(m) \\ &= \int_M m(A)\, d\mu_1(m) - \int_M m(A)\, d\mu_2(m) = \bar{\mu}_1(A) - \bar{\mu}_2(A). \end{aligned}$$

If we define P by $\bar{\mu}_1 = P = \bar{\mu}_2$ then we have that $P \in \operatorname{dom}\Gamma_M$ but $\mu_1 \neq \mu_2$, a contradiction with each $P \in \operatorname{dom}\Gamma_M$ being M-identifiable. ■

Proof of Lemma 1 By Lemma 7 of Halmos and Savage [14], there is a countable subset $\tilde{M} = \{m_1, ..., m_n, ...\} \subseteq M$ such that

$$m(E) = 0 \quad \forall m \in \tilde{M} \quad \Longrightarrow \quad m(E) = 0 \quad \forall m \in M \tag{34}$$

[24] The notion of identifiability is based on Teicher [35].

It holds $\tilde{M} = M$. Suppose there is $m_0 \in M$ such that $m_0 \notin \tilde{M}$. Consider the countable family $\tilde{M} \cup \{m_0\}$ of pairwise orthogonal probability measures. By Lemma 23-(i), there is a partition $\{E_n\}_{n \geq 0}$ such that, for each $n \geq 0$, $m_n(E_n) = 1$ and $m_n(E_k) = 0$ if $k \neq n$. This contradicts (34) since $m_0(E_0) = 1$ and $m(E_0) = 0$ for all $m \in \tilde{M}$. ■

Proof of Proposition 4 (i) implies (ii) By Savage's Theorem, there exist a non-constant function $u : X \to \mathbb{R}$ and a unique convex-ranged $P : \Sigma \to [0,1]$ such that for each f and g in $\mathcal{F}$

$$V(f) = \int_S u(f(s))\, dP(s) \text{ and } V(f) \geq V(g) \Leftrightarrow f \succsim g.$$

By order consistency, there exists a unanimous event $E \in \Sigma$ such that for each $F \in \Sigma$

$$m(F) = m(E) \quad \forall m \in M \quad \Longrightarrow \quad P(F) = P(E)$$

and

$$m(F) \geq m(E) \quad \forall m \in M \quad \Longrightarrow \quad P(F) \geq P(E).$$

By assumption, each m is convex-ranged. By Theorem 20 of Marinacci and Montrucchio [25] and since it is immediate to check that E is a radial set, we have that $P \in \operatorname{cone} M$. Since $P \in \Delta$, it follows that $P \in \operatorname{co} M$. In turn, this implies that P is countably additive and that there is a probability $\mu : 2^M \to [0,1]$ such that $P(E) = \sum_{m \in M} m(E)\, \mu(m)$ for all $E \in \Sigma$. Hence, we can conclude that

$$V(f) = \int_S u(f(s))\, dP(s) = \int_M \left(\int_S u(f(s))\, dm(s) \right) d\mu(m) \qquad \forall f \in \mathcal{F}.$$

Since the support of μ is a subset of M, it is immediate to see that it consists of nonatomic elements.

(ii) implies (i) Define $P = \bar{\mu}$. Since each $m \in M$ is a nonatomic probability measure, we next show that P is a nonatomic probability measure as well. Indeed, it is enough to prove that P is convex ranged. Consider $E \in \Sigma$ such that $P(E) > 0$ and $\alpha \in (0,1)$. Since the collection $\{m_i\}_{i=1}^n$ is a collection of nonatomic probability measures, by the Lyapunov Theorem there exists $F_\alpha \in \Sigma$ such that $m(F_\alpha) = \alpha m(E)$ for all $m \in M$. This implies that $\bar{\mu}(F_\alpha) = \sum_{m \in M} m(F_\alpha)\, \mu(m) = \sum_{m \in M} \alpha m(E)\, \mu(m) = \alpha \bar{\mu}(E)$. We can conclude that $V : \mathcal{F} \to \mathbb{R}$ defined by

$$V(f) = \int_S u(f(s))\, dP(s) = \int_M \left(\int_S u(f(s))\, dm(s) \right) d\mu(m) \qquad \forall f \in \mathcal{F},$$

represents $\succsim$ where P is nonatomic. By Savage's Theorem, it follows that $\succsim$ satisfies P.1-P.6.

At last, we show that $\succsim$ is order consistent with M. Let E be an unanimous event and let $x \succ y$. Without loss of generality assume that $u(x) = 1$ and $u(0) = 0$. If $F \in \Sigma$ is such that $m(F) \geq m(E)$ (resp., $m(F) = m(E)$) for each $m \in M$ then $V(xFy) = \bar{\mu}(F) \geq \bar{\mu}(E) = V(xEy)$ (resp., $V(xFy) = \bar{\mu}(F) = \bar{\mu}(E) = V(xEy)$), and so $xFy \succsim xEy$ (resp., $xFy \sim xEy$).

Finally, the cardinal uniqueness of u is obvious. The uniqueness of μ follows from Lemma 21 and the uniqueness of the predictive probability P derived in (i) implies (ii). ■

Proof of Proposition 5 Set $M = \{m_1, ..., m_n\}$. By Proposition 4,

$$P(E) = \sum_{i=1}^{n} \chi_i m_i(E) \qquad \forall E \in \Sigma \tag{35}$$

with each $\chi_i \geq 0$ and $\sum_{i=1}^{n} \chi_i = 1$. To prove the "only if" part, wlog suppose that $m_1 \in \operatorname{supp} \mu$, so that $\chi_1 > 0$. Let $E \subseteq F$. Suppose $m_1(E) < m_1(F)$. Then, $P(E) = \sum_{i=1}^{n} \chi_i m_i(E) < \sum_{i=1}^{n} \chi_i m_i(F) = P(F)$ since $\chi_1 > 0$.

As to the converse, suppose $\tilde{m} \in M$ is such that (11) holds for all $E \subseteq F$. We want to show that $\tilde{m} \in \operatorname{supp} \mu$. Set $\tilde{M} = \{m \in M : m \neq \tilde{m}\}$. Let

$$\Lambda = \left\{E : m(E) = \frac{1}{2} \text{ for all } m \in M\right\} \quad \text{and} \quad \tilde{\Lambda} = \left\{E : m(E) = \frac{1}{2} \text{ for all } m \in \tilde{M}\right\}.$$

Clearly, $\Lambda \subseteq \tilde{\Lambda}$. We want to show that $\Lambda \neq \tilde{\Lambda}$. Suppose, per contra, that $\Lambda = \tilde{\Lambda}$. This implies that, for all E,

$$m(E) = \frac{1}{2} \quad \forall m \in \tilde{M} \quad \Longrightarrow \quad \tilde{m}(E) = \frac{1}{2}$$

Since each $m \in M$ is convex-ranged, from Marinacci and Montrucchio [25, Theorem 20] it follows that $\tilde{m} \in \operatorname{span} \tilde{M}$, which contradicts the linear independence of M. We conclude that $\Lambda \neq \tilde{\Lambda}$.

Let $\tilde{E} \in \tilde{\Lambda}$ and $\tilde{E} \notin \Lambda$. Then, $1/2 = m\left(\tilde{E}\right) \neq \tilde{m}\left(\tilde{E}\right)$ for all $m \in \tilde{M}$. If $P = \tilde{m}$, we trivially have $\tilde{m} \in \operatorname{supp} \mu$. If $P \neq \tilde{m}$, there is some $m \in \tilde{M}$ such that $\mu(m) > 0$. Since $m\left(\tilde{E}\right) = 1/2$, by (35) this implies $P\left(\tilde{E}\right) > 0$. By the Lyapunov Theorem, there is $F \subseteq \tilde{E}$ such that $P(F) = 2^{-1} P\left(\tilde{E}\right)$ and $m(F) = 2^{-1} m\left(\tilde{E}\right)$ for all $m \in M$. Then,

$$\begin{aligned} P(F) &= \sum_{m \in M} m(F)\mu(m) = \frac{1}{4}(1 - \mu(\tilde{m})) + \frac{1}{2}\tilde{m}\left(\tilde{E}\right)\mu(\tilde{m}) \\ &< P\left(\tilde{E}\right) = \sum_{m \in M} m\left(\tilde{E}\right)\mu(m) = \frac{1}{2}(1 - \mu(\tilde{m})) + \tilde{m}\left(\tilde{E}\right)\mu(\tilde{m}) \end{aligned}$$

and so $\mu(\tilde{m}) > 0$. We conclude that $\tilde{m} \in \operatorname{supp} \mu$. ■

Proof of Proposition 6 Let $P(E) > 0$. By (4),

$$V_E(f) = \int_S u(f(s))\, d\bar{\mu}(s \mid E)$$

Set $\operatorname{supp}_E \mu = \{m \in \operatorname{supp} \mu : m(E) > 0\}$. Moreover,

$$\begin{aligned} \int_\Delta m(F \mid E)\, d\mu(m \mid E) &= \sum_{m \in \operatorname{supp}_E \mu} m(F \mid E)\mu(m \mid E) = \sum_{m \in \operatorname{supp}_E \mu} \frac{m(F \cap E)}{m(E)} \frac{m(E)\mu(m)}{\sum_{m \in \operatorname{supp}_E \mu} m(E)\mu(m)} \\ &= \frac{1}{\sum_{m \in \operatorname{supp}_E \mu} m(E)\mu(m)} \left(\sum_{m \in \operatorname{supp}_E \mu} \frac{m(F \cap E)}{m(E)} m(E)\mu(m)\right) \\ &= \frac{\sum_{m \in \operatorname{supp}_E \mu} m(F \cap E)\mu(m)}{\sum_{m \in \operatorname{supp}_E \mu} m(E)\mu(m)} = \frac{\bar{\mu}(F \cap E)}{\bar{\mu}(E)} = \bar{\mu}(F \mid E). \end{aligned}$$

Hence,

$$V_E(f) = \int_S u(f(s))\, d\bar{\mu}(s \mid E) = \int_\Delta \left(\int_S u(f(s))\, dm(s \mid E) \right) d\mu(m \mid E)$$

as desired. ■

Proof of Proposition 8 By Savage's Theorem, there exist a non-constant function $u : X \to \mathbb{R}$ and a unique convex-ranged $P : \Sigma \to [0,1]$ such that

$$V(f) = \int_S u(f(s))\ dP(s)$$

By (5),

$$m(F) = m(E) \quad \forall m \in M \quad \Longrightarrow \quad P(F) = P(E) \tag{36}$$

for all $F \in \Sigma$. Since each m is convex-ranged, from Theorem 20 of Marinacci and Montrucchio [25] it follows that $P \in \operatorname{span} M$, i.e., there is a collection $\{\chi_i\}_{i=1}^n$ of scalars such that $P(E) = \sum_{i=1}^n \chi_i m_i(E)$ for each $E \in \Sigma$. From $P(S) = \sum_{i=1}^n \chi_i m_i(S)$ it follows that $\sum_{i=1}^n \chi_i = 1$. By Lemma 7, there exists a partition $\{E_i\}_{i=1}^n$ such that $m_i(E_i) = 1$ for each $i = 1, ..., n$. Hence, for each i it holds $P(E_i) = \chi_i$, and so $\chi_i \geq 0$. We conclude that $P \in \operatorname{co} M$. The rest of the proof is similar to that of Proposition 4. ■

Proof of Lemma 7 See Lemma 23. ■

Proof of Proposition 11 By Lemma 1, M is countable. We first consider orthogonality and then equivalence. (i) Suppose $\mu \perp \mu'$, i.e., there is $A \in \mathcal{D}$ such that $\mu(A) = 1 = \mu'(A^c)$. By Lemma 23-(i), there exists a countable partition $\{E_m\}$ such that $m(E_m) = 1$ and $m'(E_m) = 0$ if $m' \neq m$. Set $E = \bigcup \{E_m : m \in A\}$. Clearly, $E \in \Sigma$. Moreover, $m(E) = 1$ for all $m \in A$ and $m(E) = 0$ for all $m \in A^c$. Then,

$$\bar{\mu}(E) = \sum_{m \in M} m(E)\mu(m) = \sum_{m \in A} m(E)\mu(m) = \sum_{m \in A} \mu(m) = \mu(A) = 1$$

and

$$\bar{\mu}(E) = \sum_{m \in M} m(E)\mu'(m) = \sum_{m \in A^c} m(E)\mu'(m) = 0,$$

which implies $\bar{\mu} \perp \bar{\mu}'$.

As to the converse, suppose $\bar{\mu} \perp \bar{\mu}'$. There exists $E \in \Sigma$ such that $\bar{\mu}(E) = 1 = \bar{\mu}'(E^c)$. Set $A = \{m \in M : m(E) > 0\}$. We have $A \in \mathcal{D}$ since A is countable. It holds

$$1 = \bar{\mu}(E) = \sum_{m \in M} m(E)\mu(m) = \sum_{m \in A} m(E)\mu(m) \leq \sum_{m \in A} \mu(m) = \mu(A) \leq 1$$

and so $\mu(A) = 1$. Moreover,

$$0 = \bar{\mu}'(E) = \sum_{m \in M} m(E)\mu'(m) = \sum_{m \in A} m(E)\mu'(m). \tag{37}$$

If $\mu'(\tilde{m}) > 0$ for some $\tilde{m} \in A$, then $\sum_{m \in A} m(E)\mu'(m) \geq \tilde{m}(E)\mu'(\tilde{m}) > 0$, which contradicts (37). Hence, $\mu'(m) = 0$ for all $m \in A$, and so $\mu'(A) = 0$. We conclude that $\mu \perp \mu'$.

(ii) Suppose $\mu \equiv \mu'$. Let $\bar{\mu}(E) = 0$, so that $\bar{\mu}(E) = \sum_{m\in M} m(E)\mu(m) = 0$. Then, $\mu(\{m : m(E) > 0\}) = 0$, and so $\mu'(\{m : m(E) > 0\}) = 0$. In turn this implies $\bar{\mu}'(E) = \sum_{m\in M} m(E)\mu'(m) = 0$. Hence, $\bar{\mu}' \ll \bar{\mu}$. A similar argument shows that $\bar{\mu} \ll \bar{\mu}'$.

Conversely, suppose $\bar{\mu} \equiv \bar{\mu}'$. Let $A \in \mathcal{D}$ be such that $\mu(A) = 0$. By proceeding as in the proof of Lemma 24, we can construct pairwise disjoint events $\{E_m\}_{m\in A}$ such that $m(E_m) = 1/2$ and $m'(E_m) = 0$ for all $m' \in A$ distinct from m. Set $E = \bigcup_{m\in A} E_m$. Then, $m(E) = 1/2$ and $m'(E) = 0$ for all $m' \notin A$, so that

$$\bar{\mu}(E) = \sum_{m\in M} m(E)\mu(m) = \frac{1}{2}\mu(A) = 0.$$

This implies

$$0 = \bar{\mu}'(E) = \sum_{m\in M} m(E)\mu'(m) = \frac{1}{2}\mu'(A)$$

and so $\mu'(A) = 0$. We conclude that $\mu' \ll \mu$. A similar argument shows that $\mu \ll \mu'$. ∎

Proof of Proposition 14 Given any two such models m and $\tilde{m}$, if they are equivalent it holds $l_n = d\pi/d\tilde{\pi}$. Hence, condition (22) becomes

$$E_m\left(\sqrt{l_n} \mid \Sigma_{n-1}\right) = \int_{\mathcal{Z}} \sqrt{\frac{d\pi}{d\tilde{\pi}}}d\tilde{\pi} \tag{38}$$

and so

$$\sum_n \left[1 - E_{\tilde{m}}\left(\sqrt{l_n} \mid \Sigma_{n-1}\right)\right] = \sum_n \left[1 - \int_{\mathcal{Z}} \sqrt{\frac{d\pi}{d\tilde{\pi}}}d\tilde{\pi}\right] = \infty \Longleftrightarrow \int_{\mathcal{Z}} \sqrt{\frac{d\pi}{d\tilde{\pi}}}d\tilde{\pi} \neq 1.$$

Since $d\pi/d\tilde{\pi} \geq 0$, by the Jensen inequality it holds $0 \leq \int_{\mathcal{Z}} \sqrt{d\pi/d\tilde{\pi}}d\tilde{\pi} \leq \sqrt{\int_{\mathcal{Z}} (d\pi/d\tilde{\pi})\, d\pi} = 1$. Hence,

$$\int_{\mathcal{Z}} \sqrt{\frac{d\pi}{d\tilde{\pi}}}d\tilde{\pi} = 1 \Longleftrightarrow \frac{d\pi}{d\tilde{\pi}} = 1 \quad \pi\text{-a.e.}$$

that is, if and only if $\pi = \tilde{\pi}$. We conclude that condition (38) holds if and only if m and $\tilde{m}$ are distinct. Hence, M consists of pairwise orthogonal probability measure. ∎

The next result implies, as a special case, Proposition 17.

Proposition 22 *Two probability measures $m, \tilde{m} \in \Delta$ that make the coordinate process $\{Z_n\}$ independent, with equivalent Gaussian marginals, are orthogonal provided*

$$\sum_{n=1}^{\infty} \left[\left(\frac{\tilde{\sigma}_n^2}{\sigma_n^2} - 1\right) + \left(\frac{a_n - \tilde{a}_n}{\sigma_n}\right)^2\right] = \infty \tag{39}$$

Proof It follows from Kabanov et al [17] and Lemma 7. ∎

Proof of Proposition 15 Let $\tilde{m}, m \in M$. It holds

$$\frac{dm_n}{d\tilde{m}_n}(z_1, ..., z_n) = \frac{\pi_0(z_1)}{\tilde{\pi}_0(z_1)} \frac{\prod_{i=1}^{n-1} \pi_i(z_i, z_{i+1})}{\prod_{i=1}^{n-1} \tilde{\pi}_i(z_i, z_{i+1})}$$

and so

$$l_n\left(z_1,...,z_n\right)=\frac{\frac{\pi_0(z_1)\prod_{i=1}^{n-1}\pi_i(z_i,z_{i+1})}{\tilde{\pi}_0(z_1)\prod_{i=1}^{n-1}\tilde{\pi}_i(z_i,z_{i+1})}}{\frac{\pi_0(z_1)\prod_{i=1}^{n-2}\pi_i(z_i,z_{i+1})}{\tilde{\pi}_0(z_1)\prod_{i=1}^{n-2}\tilde{\pi}_i(z_i,z_{i+1})}}=\frac{\pi_{n-1}\left(z_{n-1},z_n\right)}{\tilde{\pi}_{n-1}\left(z_{n-1},z_n\right)}$$

Hence,

$$E_{\tilde{m}}\left(\sqrt{l_n}\mid\Sigma_{n-1}\right)=\sum_{\zeta\in\mathcal{Z}}\sqrt{\frac{\pi_{n-1}\left(z_{n-1},\zeta\right)}{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}}\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)=\sum_{\zeta\in\mathcal{Z}}\sqrt{\pi_{n-1}\left(z_{n-1},\zeta\right)}\sqrt{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}$$

Since $\sum_{\zeta\in\mathcal{Z}}\pi_{n-1}\left(z_{n-1},\zeta\right)=\sum_{\zeta\in\mathcal{Z}}\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)=1$, it holds

$$\sum_{\zeta\in\mathcal{Z}}\left(\sqrt{\pi_{n-1}\left(z_{n-1},\zeta\right)}-\sqrt{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}\right)^2=2-2\sum_{\zeta\in\mathcal{Z}}\sqrt{\pi_{n-1}\left(z_{n-1},\zeta\right)}\sqrt{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}$$

and so

$$1-E_{\tilde{m}}\left(\sqrt{l_n}\mid\Sigma_{n-1}\right)=\frac{1}{2}\sum_{\zeta\in\mathcal{Z}}\left(\sqrt{\pi_{n-1}\left(z_{n-1},\zeta\right)}-\sqrt{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}\right)^2$$

By Lemma 12, the probability measures $\tilde{m}$ and m are orthogonal if and only if

$$\sum_n\left\|\sqrt{\pi_n\left(z_n,\cdot\right)}-\sqrt{\tilde{\pi}_n\left(z_n,\cdot\right)}\right\|=\sum_n\left[\sum_{\zeta\in\mathcal{Z}}\left(\sqrt{\pi_{n-1}\left(z_{n-1},\zeta\right)}-\sqrt{\tilde{\pi}_{n-1}\left(z_{n-1},\zeta\right)}\right)^2\right]=\infty\qquad m\text{-a.e.}$$

In particular, this is the case if (24) holds. ∎

Proof of Corollary 16 Let $\tilde{m}$ and m be two distinct elements of M, so that $\pi\neq\tilde{\pi}$. Then, there are $\zeta,\zeta'\in\mathcal{Z}$ such that $\pi\left(\zeta,\zeta'\right)\neq\tilde{\pi}\left(\zeta,\zeta'\right)$. Hence, $\left\|\sqrt{\tilde{\pi}\left(z_n,\cdot\right)}-\sqrt{\pi\left(z_n,\cdot\right)}\right\|\geq\left(\sqrt{\pi\left(\zeta,\zeta'\right)}-\sqrt{\tilde{\pi}\left(\zeta,\zeta'\right)}\right)^2>0$, and so (24) holds. ∎

Proof of Lemma 18 Set $p=2^{-1}\left(m+\tilde{m}\right)$. By Lemma 5 of Kabanov et al [17], the process $\{\lambda_n\}$ converges $\tilde{m}$-a.e., with

$$\tilde{m}\left(\lim_n\lambda_n=\frac{dm/dp}{d\tilde{m}/dp}\right)=1$$

Since $\tilde{m}\left(d\tilde{m}/dp=0\right)=0$, it holds

$$\tilde{m}\left(\frac{dm/dp}{d\tilde{m}/dp}=0\right)=1\Longleftrightarrow\tilde{m}\left(dm/dp=0\right)=1.$$

It remains to show that $\tilde{m}\left(dm/dp=0\right)=1$ if and only if $m\perp\tilde{m}$. The "only if" is obvious since

$$m\left(\frac{dm}{dp}=0\right)=\int_{\left\{\frac{dm}{dp}=0\right\}}\frac{dm}{dp}dp=0$$

that implies $m\perp\tilde{m}$. As to the converse, suppose there is $E\in\mathcal{B}^\infty$ such that $m\left(E\right)=0$ and $\tilde{m}\left(E\right)=1$. Then,

$$0=m\left(E\right)=\int_E\frac{dm}{dp}dp\Longrightarrow p\left(E\cap\left(\frac{dm}{dp}>0\right)\right)=0\Longrightarrow\tilde{m}\left(E\cap\left(\frac{dm}{dp}>0\right)\right)=0.$$

But, $\tilde{m}(E) = 1$ implies $\tilde{m}(E \cap (dm/dp > 0)) = \tilde{m}(dm/dp > 0)$, and so $\tilde{m}(dm/dp > 0) = 0$. This completes the proof. ■

Proof of Lemma 19 "If" Let $m \in M$. Since $\tilde{m} \perp m$ for each $\tilde{m} \neq m \in \operatorname{supp}\mu$, by Lemma 18 it holds (27), and so (26) holds. As to the converse, from (26) it follows that (27) holds for each $\tilde{m} \neq m \in \operatorname{supp}\mu$. By Lemma 18, $\tilde{m} \perp m$ for each $\tilde{m} \neq m \in \operatorname{supp}\mu$. ■

Derivation of (31) Using (13) and, via the reduced form, the recursive form of expected utility we can write

$$\begin{aligned}
W_{z^t}(f) &= \int_\Delta \left(\int_{\mathcal{Z}^\infty} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \upsilon(f_\tau(z))\ dm\left(z \mid z^t\right) \right) d\mu\left(m \mid z^t\right) \\
&= \int_{\mathcal{Z}^\infty} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \upsilon(f_\tau(z))\ d\bar{\mu}\left(z \mid z^t\right) = \upsilon(f_t(z)) + \beta \int_\Delta \left(\int_{\mathcal{Z}} W_{\{z^t, z_{t+1}\}}\ (f)\, d\bar{\mu}\left(z^t, z_{t+1} \mid z^t\right) \right) \\
&= \upsilon(f_t(z)) + \beta \int_\Delta \left(\int_{\mathcal{Z}} W_{\{z^t, z_{t+1}\}}(f)\ dm_{t+1}\left(z_{t+1} \mid z^t\right) \right) d\mu\left(m \mid z^t\right)
\end{aligned}$$

as desired. ■

Proof of Proposition 20 Let $f \in \mathcal{F}$ and $\varepsilon > 0$. By Lemma 19, there is n_ε such that, $\tilde{m}$-a.e.,

$$\mu(\tilde{m} \mid z_1, ..., z_n) \geq 1 - \varepsilon \quad \text{and} \quad \mu(m \mid z_1, ..., z_n) \leq \varepsilon \qquad \forall \tilde{m} \neq m \in \operatorname{supp}\mu$$

Then, $\tilde{m}$-a.e., it holds

$$\begin{aligned}
&\left| W_{z^n}(f) - \int_{\mathcal{Z}^\infty} u(f_n(z), ...)\ dm(z \mid z^n) \right| \\
= &\left| \int_{\Delta(\mathcal{Z}^\infty)} \left(\int_{\mathcal{Z}^\infty} u(f_n(z), ...)\ dm(z \mid z_1, ..., z_n) \right) d\left(\mu(m \mid z_1, ..., z_n) - \delta_{\tilde{m}}\right) \right| \\
\leq &\ \varepsilon \sup_{z \in \mathcal{Z}^\infty} |u(z)|
\end{aligned}$$

and so (32) holds. ■

8.1 Proof of Proposition 10

The proof of this proposition relies on few lemmas.

Lemma 23 *Let $M = \{m_1, ..., m_n, ...\}$ be a countable subset of Δ. If the elements of M are pairwise orthogonal, then:*

(i) there exists a countable partition $\{E_n\}$ such that $m_n(E_n) = 1$ and $m_n(E_k) = 0$ if $k \neq n$;

(ii) M is measure independent.

Proof of Lemma 23 (i) Suppose that $M = \{m_1, ..., m_n, ...\}$ consists of pairwise orthogonal elements. Consider m_1. For each $i \neq 1$ there is $E_{1i} \in \Sigma$ such that $m_1(E_{1i}) = 1 = m_i(E_{1i}^c)$. By setting $E_1 = \bigcap_{i\neq 1} E_{1i}$ we then have $m_1(E_1) = 1$ and $m_i(E_1) = 0$ for each $i \neq 1$. Consider m_2. Since $m_2(E_1^c) = 1$, for each $i > 2$ there is an event $E_{2i} \subseteq E_1^c$ such that $m_2(E_{2i}) = 1 = m_i(E_{2i}^c)$. By setting $E_2 = \bigcap_{i\neq 1} E_{2i}$ we then have $m_2(E_2) = 1$ and $m_i(E_2) = 0$ for each $i > 2$. By proceeding in this way we can construct the desired partition.

(ii) Given any collection $\{\alpha_n\}$ of scalars, from $\alpha_n m_n(E_n) + \sum_{k\neq n} \alpha_k m_k(E) = 0$ it follows $\alpha_n = 0$ for each n. Hence, M is measure independent. ■

The next lemma could be proved through the notion of thin set due to Kingman and Robertson [19], but we prefer a simpler direct approach.

Lemma 24 *Let $M = \{m_1, ..., m_n, ...\}$ be a countable subset of nonatomic measures in Δ. If the elements of M are pairwise orthogonal, except at most a finite number of them, then*

$$\{(m_1(E), ..., m_n(E), ...) : E \in \Sigma\}$$

is a convex subset of $\mathbb{R}^\infty$.

Proof Suppose that all elements of M, except one, are pairwise orthogonal. To ease notation, let us write M as $\{p, m_1, ..., m_n, ...\}$, where p is the unique element not pairwise orthogonal. We want to show that $R_M = \{(p(E), m_1(E), ..., m_n(E), ...) : E \in \Sigma\}$ is a convex subset of $\mathbb{R}^\infty$. Let $x, y \in R_M$ and $\alpha \in [0,1]$. By definition, there are events A and B such that $x = (p(A), m_1(A), ..., m_n(A), ...)$ and $y = (p(B), m_1(B), ..., m_n(B), ...)$. By Lemma 23-(i), there exists a countable partition $\{E_n\}$ such that $m_n(E_n) = 1$ and $m_n(E_k) = 0$ if $k \neq n$. Hence,

$$x = (p(A), m_1(A \cap E_1), ..., m_n(A \cap E_n), ...) \quad \text{and} \quad y = (p(B), m_1(B \cap E_1), ..., m_n(B \cap E_n), ...).$$

Fix n. Since p and each m_n are nonatomic, the Lyapunov Theorem implies that for each of pair (m_n, p) there is an event $E_n^\alpha \subseteq E_n$ such that

$$m_n(E_n^\alpha) = \alpha m_n(A \cap E_n) + (1-\alpha) m_n(B \cap E_n) \quad \text{and} \quad p(E_n^\alpha) = \alpha p(A \cap E_n) + (1-\alpha) p(B \cap E_n).$$

Let $E^\alpha = \bigcup_n E_n^\alpha$ be the union of the pairwise disjoint events $\{E_n^\alpha\}$ that we just found. It holds $m_n(E^\alpha) = m(E_n^\alpha)$ for each n, and

$$p(E^\alpha) = \sum_n p(E_n^\alpha) = \alpha \sum_n p(A \cap E_n) + (1-\alpha) \sum_n p(B \cap E_n) = \alpha p(A) + (1-\alpha) p(B)$$

Hence

$$\begin{aligned} & \alpha x + (1-\alpha) y \\ = \; & \alpha (p(A), m_1(A \cap E_1), ..., m_n(A \cap E_n), ...) + (1-\alpha)(p(B), m_1(B \cap E_1), ..., m_n(B \cap E_n), ...) \\ = \; & (p(E^\alpha), m_1(E^\alpha), ..., m_n(E^\alpha), ...) \end{aligned}$$

and so $\alpha x + (1-\alpha) y \in R_M$. This completes the proof when there is a unique element of M that is not pairwise orthogonal. If there is a finite number of them, say $\{p_1, ..., p_k\}$, we can proceed as before by applying the Lyapunov Theorem on each $(m_n, p_1, ..., p_k)$. ■

Lemma 25 *Let M be a subset of Δ such that, for some $p \in \Delta$,*

$$m(E) = 0 \quad \forall m \in M \Longrightarrow p(E) = 0$$

Then, for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$m(E) < \delta \quad \forall m \in M \Longrightarrow p(E) < \varepsilon$$

Proof We generalize the argument in Bhasakara Rao and Bhasakara Rao [3, p. 161] for singleton M. Suppose per contra that there is $\varepsilon > 0$ such for every $k \geq 1$ there is an event E_k such that $m(E_k) < 2^{-k}$ for all $m \in M$ and $p(E_k) \geq \varepsilon$. Then

$$m\left(\bigcap_k \bigcup_{s \geq k} E_s\right) = \lim_k m\left(\bigcup_{s \geq k} E_s\right) \leq \lim_k \sum_{s \geq k} m(E_k) \leq \lim_k \sum_{s \geq k} \frac{1}{2^k} \to 0 \qquad \forall m \in M$$

and

$$p\left(\bigcap_k \bigcup_{s \geq k} E_s\right) = \lim_k p\left(\bigcup_{s \geq k} E_s\right) \geq \limsup_k p(E_k) \geq \varepsilon$$

which contradicts

$$m\left(\bigcap_k \bigcup_{s \geq k} E_s\right) = 0 \quad \forall m \in M \Longrightarrow p\left(\bigcap_k \bigcup_{s \geq k} E_s\right) = 0$$

as desired. ■

Proof of Proposition 10 Let l_∞ be the Banach space of bounded sequences, that is, $l_\infty = \{x \in \mathbb{R}^\infty : \|x\|_\infty < \infty\}$ where $\|x\|_\infty = \sup_n |x_n|$. By Lemma 23,

$$\{(m_1(E), ..., m_n(E), ...) : E \in \Sigma\} = [0,1]^\infty$$

We can then define $F : [0,1]^\infty \to [0,1]$ by

$$F(m_1(E), ..., m_n(E), ...) = P(E) \qquad \forall E \in \Sigma.$$

Let $\alpha \in [0,1]$,

$$x = (m_1(A), ..., m_n(A), ...) \in [0,1]^\infty \quad \text{and} \quad y = (m_1(B), ..., m_n(B), ...) \in [0,1]^\infty.$$

By Lemma 24, there is $E^\alpha \in \Sigma$ such that $P(E^\alpha) = \alpha P(A) + (1-\alpha) P(B)$ and

$$m_n(E^\alpha) = \alpha m_n(A) + (1-\alpha) m_n(B) \qquad \forall n \geq 1.$$

Hence,

$$\begin{aligned} F(\alpha x + (1-\alpha) y) &= F(m_1(E^\alpha), ..., m_n(E^\alpha), ...) = P(E^\alpha) \\ &= \alpha P(A) + (1-\alpha) P(B) = \alpha F(x) + (1-\alpha) F(y) \end{aligned}$$

Therefore, F is a affine on $[0,1]^\infty$. A routine argument shows that there is a linear functional $L : l_\infty \to \mathbb{R}$ such that $L(x) = F(x)$ for all $x \in [0,1]^\infty$. This functional is bounded. For, let $x \in l_\infty$ such that $\|x\|_\infty \leq 1$. Then, $x = x^+ - x^-$, with $x^+, x^- \in [0,1]^\infty$.[25] Hence,

$$|L(x)| = \left|L\left(x^+ - x^-\right)\right| = \left|F\left(x^+\right) - F\left(x^-\right)\right| \leq \left|F\left(x^+\right)\right| + \left|F\left(x^-\right)\right| \leq 2$$

since $F(x) \in [0,1]$ for all $x \in [0,1]^\infty$. We conclude that $\|L\| = \sup\{|L(x)| : \|x\|_\infty \leq 1\} \leq 2$.

Claim Let $\{x^s\} \subseteq l_\infty$ be uniformly bounded (i.e., $\|x^s\| \leq K$ for all s and some $K > 0$) and such that $x_n^s \to x_n$ for each n. Then, $L(x^s) \to L(x)$.

Proof of the claim It is enough to show that, if $\{x^s\} \subseteq [0,1]^\infty$ is such that $x_n^s \to 0$ for each n, then $F(x^s) \to 0$. Let $\{E_s\}$ be such that

$$x^s = \left(m_1(E_s), ..., m_n(E_s), ...\right) \qquad \forall s \geq 1$$

Hence $m(E_s) \to 0$ for all $m \in M$. By Lemma 25, this implies $P(E_s) \to 0$, and so $F(x^s) \to 0$. □

Since, by the Claim, the bounded linear functional $L : l_\infty \to \mathbb{R}$ is bounded pointwise continuous, there exists $\chi \in l_1$ such that $L(x) = \sum_n \chi_n x_n$ for each $x \in l_\infty$. In particular, this implies

$$P(E) = \sum_n \chi_n m_n(E) \qquad \forall E \in \Sigma.$$

Since $P(S) = \sum_n \chi_n m_n(S)$ it follows $\sum_n \chi_n = 1$. By Lemma 23, there is a partition $\{E_n\}$ such that $m_n(E_n) = 1$ for each n. Then, $P(E_n) = \chi_n m_n(E) = \chi_n$ for each n and so $\chi_n \geq 0$ for each n, as desired. ■

References

[1] N. I. Al-Najjar and L. De Castro (2010), Parametric representation of preferences, mimeo.

[2] P. Battigalli, S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci (2011), Selfconfirming equilibrium and model uncertainty, Igier WP 428.

[3] K. P. S. Bhaskara Rao and M. Bhaskara Rao (1983), *Theory of charges*, Academic Press: New York.

[4] P. Billingsley (1986), *Probability and measure*, Wiley: New York.

[5] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio (2011), Ambiguity and robust statistics, Igier WP 382.

[6] G. Claeskens and N. L. Hjort (2008), *Model selection and model averaging.* Cambridge: Cambridge University Press.

[25] Here $x_n^+ = \max\{x_n, 0\}$ and $x_n^- = \min\{x_n, 0\}$.

[7] P. Diaconis and D. Freedman (1986), On the consistency of Bayes estimates, *Annals of Statistics*, 14: 1-26.

[8] P. Diaconis, S. Holmes, and R. Montgomery (2007), Dynamical bias in the coin toss, *SIAM Review*, 49: 211-235.

[9] W. Diffie and M. Hellman (1976), New directions in cryptography, *IEEE Transaction Information Theory*, 22: 644-654.

[10] D. Easley and N. M. Kiefer (1988), Controlling a stochastic process with unknown parameters, *Econometrica*, 56, 1045-1064.

[11] L. G. Epstein and K. Seo (2010), Symmetry of evidence without evidence of symmetry, *Theoretical Economics*, 5: 313–368.

[12] P.C. Fishburn (1970), *Utility Theory for Decision Making*, New York: John Wiley and Sons.

[13] I. Gilboa (2009), *Theory of Decision under Uncertainty*, Cambridge: Cambridge University Press.

[14] P. R. Halmos and L. J. Savage (1949), Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *Annals of Mathematical Statistics*, 20: 225-241.

[15] J.-Y. Jaffray (1989), Linear utility theory for belief functions, *Operations Research Letters*, 8: 107-112.

[16] T. H. Johnsen and J. B. Donaldson (1985), The structure of intertemporal preferences under uncertainty and time consistent plans, *Econometrica*, 53: 1451-1458.

[17] Y. M. Kabanov, R. S. Liptser, and A. N. Shiryaev (1977), On the question of absolute continuity and singularity of probability measures, *Math. USSR-Sb.,* 33: 203-221.

[18] S. Kakutani (1948), On equivalence of infinite product measures, *Annals of Mathematics*, 49: 214-224.

[19] J. F. C. Kingman and A. P. Robertson (1968), On a theorem of Lyapunov, *Journal of the London Mathematical Society*, 43: 347-351.

[20] P. Klibanoff, M. Marinacci, and S. Mukerji (2005), A smooth model of decision making under ambiguity, *Econometrica*, 73: 1849-1892.

[21] P. Klibanoff, S. Mukerji, and K. Seo (2010), Relevance and symmetry, mimeo.

[22] D. M. Kreps, *Notes on the Theory of Choice*, Westwiew Press Inc, Colorado, 1988.

[23] D. Lewis (1980), A subjectivist's guide to objective chance, in *Studies in Inductive Logic and Probability* (R. C. Jeffrey, ed.), Berkeley and Los Angeles: University of California Press.

[24] R. E. Lucas Jr. (1977), Understanding business cycles, in *Stabilization of the Domestic and International Economy* (K. Brunner and A. H. Meltzer, eds.), Amsterdam: North-Holland.

[25] M. Marinacci and L. Montrucchio (2003), Subcalculus for set functions and cores of TU games, *Journal of Mathematical Economics,* 39: 1-25.

[26] M. Marinacci and L. Montrucchio (2010), Unique solutions for stochastic recursive utilities, *Journal of Economic Theory,* 145: 1776-1804.

[27] J. Marschak (1946), Neumann's and Morgenstern's new approach to static economics, *Journal of Political Economy*, 54: 97-115.

[28] J. Marschak (1975), Personal probabilities of probabilities, *Theory and Decision*, 6: 121-153.

[29] J. Marschak and R. Radner (1972), *Economic theory of teams*, New Haven: Yale University Press.

[30] J. F. Muth (1961), Rational expectations and the theory of price movements, *Econometrica*, 29: 315-335.

[31] R. Radner (1962), Team decision problems, *Annals of Mathematical Statistics*, 33: 857-881.

[32] T. J. Sargent and N. Williams (2005), Impacts of priors on convergence and escapes from Nash inflation, *Review of Economic Dynamics*, 8: 360-391.

[33] L. J. Savage (1954), *The foundations of statistics.* New York: John Wiley and Sons.

[34] D. W. Stroock (1993), *Probability theory, an analytic view.* Cambridge: Cambridge University Press.

[35] H. Teicher (1963), Identifiability of finite mixtures, *Annals of Mathematical Statistics*, 34: 244-248.

[36] A. Wald (1950), *Statistical decision functions.* New York: John Wiley and Sons.

[37] A. Wald (1952), Basic ideas of a general theory of statistical decision rules, in *Proceedings of the International Congress of Mathematicians*, *1950*, Providence: American Mathematical Society.