Classical Subjective Expected Utility

Simone Cerreia-Vioglio [*], **Fabio Maccheroni** [*], **Massimo Marinacci** [*], **and Luigi Montrucchio** [†]

[*]Università Bocconi, Milano, Italy, and [†]Università di Torino, Torino, Italy

Submitted to Proceedings of the National Academy of Sciences of the United States of America

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 (1950). We incorporate this Waldean piece of objective information within an otherwise subjective setting a la Savage (1954) and show that this leads to a two-stage subjective expected utility model that accounts for both state and model uncertainty.

expected utility | model uncertainty | state uncertainty

Abbreviations: DM, decision maker; SEU, subjective expected utility

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 DM 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 DM's decisions. Model uncertainty, in contrast, is not payoff relevant and its role is instrumental relative to state uncertainty. Moreover, models cannot always be observed: while in some cases they have a simple physical description (e.g., urns' compositions), often they do not have it (e.g., fair coins). For these reasons, the purely subjective choice frameworks ą la Savage [1] 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 [2] assumes that the DM knows 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 observations.[1] 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 in some important economic applications it is natural to assume, at least as a working hypothesis, that DMs have this kind of information (see, e.g., Sargent [3]).

Our approach takes the objective information M as a primitive and enriches the standard Savage framework with this datum: DMs 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 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 them through hypothetical (since models are not in general observable) betting behavior. Nevertheless, our basic representation result, Proposition 1, shows that, under Savage's axioms P.1-P.6 and the above consistency condition, acts are ranked according to the criterion

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

where μ is a subjective prior 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 the DM may have, in addition to the objective information M. 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, some important time series models widely used in economic and financial applications satisfy this condition, as discussed later in the paper. For this reason, our Wald-Savage setup provides a proper statistical decision theory framework for empirical works that rely on such time series.

Each prior μ induces a *predictive probability* $\bar{\mu}$ on the sample space S through model averaging:

$$\bar{\mu}(E) = \int_{\Delta} m(E)\, d\mu(m) \qquad \textbf{[2]}$$

In particular, setting $P = \bar{\mu}$,

$$V(f) = \int_S u(f(s))\, dP(s) \qquad \textbf{[3]}$$

is the reduced form of V, its Subjective Expected Utility (SEU) representation ą la Savage. On the other hand, 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

$$V(f) = \int_S u(f(s))\, dm(s) \qquad \textbf{[4]}$$

where subjective probabilities do not play any role.[2] Classical SEU thus encompasses both the Savage and the von Neumann-Morgenstern representations.

Reserved for Publication Footnotes

[1]As Wald [2, p. 1] writes "A characteristic feature of any statistical decision problem is the assumption that the unknown distribution $F(x)$ is merely known to be an element of a given class Ω of distributions functions. The class Ω is to be regarded as a datum of the decision problem."

[2]Lucas [4, p. 15] writes that "Muth [5] ... [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 (4).

In particular, the Savage criterion (3) is what an outside observer, unaware of datum M, would be able to elicit from DM'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 $\mathrm{supp}\mu \subseteq M$, while for the usual Savagean representation (3) is enough a pair (u, P).

Summing up, though the work of Savage [1] was inspired by the seminal decision theoretic approach of Wald [2], his purely subjective setup and the ensuing large literature[3] did not consider the 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 et al. [10] use the Wald-Savage setup of the present paper to study selfconfirming equilibria, while Cerreia-Vioglio et al. [11] use it to provide a behavioral foundation of the robustness approach in Macroeconomics pioneered by Hansen and Sargent [12].

Preliminaries

Subjective expected utility. We consider a standard Savage setting, where (S, Σ) is a measurable state space and X is an outcome space. An *act* is a map $f : S \to X$ that delivers outcome $f(s)$ in state S. Let $\mathcal{F}$ be the set of all simple and measurable acts.[4]

The DM's preferences are represented by a binary relation $\succsim$ over $\mathcal{F}$. We assume that $\succsim$ satisfies the classic Savage's axioms P.1-P.6. By his famous representation theorem, these axioms are equivalent to the existence of a utility function $u : X \to \mathbb{R}$ and a (strongly) nonatomic finitely additive probability P on S such that the SEU evaluation $V(f) = \int_S u(f(s))\, dP(s)$ represents $\succsim$.[5] In this case, u is cardinally unique and P is unique.

Given any $f, g \in \mathcal{F}$ and $E \in \Sigma$, fEg is the act equal to f on E and to g otherwise. The conditional preference $\succsim_E$ is the binary relation on $\mathcal{F}$ defined by $f \succsim_E g$ if and only if $fEh \succsim gEh$ for all $h \in \mathcal{F}$. By P.2, the Sure Thing Principle, $\succsim_E$ is complete. An event $E \in \Sigma$ is said to be *null* if $\succsim_E$ is trivial ([1, p. 24]), in the representation, this amounts to $P(E) = 0$ (E is null if and only if it is P-null).

For each nonnull event E, the conditional preference $\succsim_E$ satisfies P.1-P.6 since the primitive preference does (see, e.g., Kreps [7, 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 nonatomic finitely additive probability P on S such that, for each nonnull event E,

$$V_E(f) = \int_S u(f(s))\, dP(s \mid E) \qquad [5]$$

represents $\succsim_E$ where $P(\cdot \mid E)$ is the conditional of P given E.

Models, priors, and posteriors. As usual, we denote by $\Delta = \Delta(S, \Sigma)$ the collection of all (countably additive) probability measures on Σ. Unless otherwise stated, in the rest of the paper all probability measures are countably additive.

In the sequel, we will consider subsets M of Δ. Each subset M of Δ is endowed with the smallest σ-algebra $\mathcal{M}$ that makes the real valued and bounded functions on M of the form $m \mapsto m(E)$ measurable for all $E \in \Sigma$ and that contains all singletons. In the important special case $M = \Delta$, we write $\mathcal{D}$ instead of $\mathcal{M}$.

Probability measures μ on Δ are interpreted as prior probabilities. The observation of a (non-$\bar{\mu}$-null) event E allows to update prior μ through the Bayes rule

$$\mu(D \mid E) = \frac{\int_D m(E)\, d\mu(m)}{\int_\Delta m(E)\, d\mu(m)}$$

for all $D \in \mathcal{D}$, thus obtaining the posterior of μ given E.

A finite subset $M = \{m_1, ..., m_n\}$ of Δ is *linearly independent* if, given any collection of scalars $\{\alpha_1, ..., \alpha_n\} \subseteq \mathbb{R}$,

$$\sum_{i=1}^n \alpha_i m_i(E) = 0 \quad \forall E \in \Sigma \Longrightarrow \alpha_1 = ... = \alpha_n = 0. \qquad [6]$$

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

If $E \in \Sigma$ and $m(E) = 0$ imply $m'(E) = 0$, m' is *absolutely continuous* with respect to m and we write $m' \ll m$.

Finally, we denote by Δ_{na} the collection of all nonatomic probability measures. By the classical Lyapunov Theorem, the range $\{(m_1(E), ..., m_n(E)) : E \in \Sigma\}$ of a finite collection $\{m_i\}_{i=1}^n$ of nonatomic probability measures is a convex subset of $\mathbb{R}^n$.

Representation

Basic result. The first issue to consider in our normative approach is how DMs' 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 $m(E) = m'(E)$ for all $m, m' \in M$. In other words, all models in M assign the same probability to event E.

Definition 1. *A preference $\succsim$ is consistent with a subset M of Δ if, given $E, F \in \Sigma$, with E unanimous,*

$$m(F) = m(E) \quad \forall m \in M \Longrightarrow xFy \sim xEy \qquad [7]$$

for all outcomes $x \succ y$.

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

Definition 2. *A preference $\succsim$ is order consistent with a subset M of Δ if, given $E, F \in \Sigma$, with E unanimous,*

$$m(F) \geq m(E) \quad \forall m \in M \Longrightarrow xFy \succsim xEy \qquad [8]$$

for all outcomes $x \succ y$.

Both these notions are minimal consistency requirements among information and preference that behaviorally reveal (to an outside observer) that the DM considers M as a datum of the decision problem. Notice that order consistency implies consistency since the premise of (7) implies that also F must be unanimous (this observation also emphasizes how weak an assumption is consistency).

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

Proposition 1. *Let M be a finite subset of Δ_{na}. The following statements are equivalent:*

[3] See Fishburn [6], Kreps [7], and Gilboa [8]. See Jaffray [9] for a different "objective" approach.

[4] Maps $f : S \to X$ such that $f(S)$ is finite and $\{s \in S : f(s) = x\} \in \Sigma$ for all $x \in X$.

[5] Strong nonatomicity of P means that for each $E \in \Sigma$ and $0 \leq c \leq P(E)$ there exists $F \in \Sigma$ such that $F \subseteq E$ and $P(F) = c$. See [13, p. 141-143] for the various definitions and properties of nonatomicity of finitely additive probabilities.

(i) $\succsim$ *is a binary relation on* $\mathcal{F}$ *that satisfies P.1-P.6 and is order consistent with* M*;*

(ii) *there exist a nonconstant utility function* $u : X \to \mathbb{R}$ *and a prior probability* μ *on* Δ *with* $\text{supp}\mu \subseteq M$*, such that*

$$V(f) = \int_{\Delta} \left(\int_S u(f(s))\, dm(s) \right) d\mu(m) \qquad \textbf{[9]}$$

represents $\succsim$.

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

While uniqueness of the utility function u is well known and well discussed in the literature, 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 [14]).

Each prior $\mu : \mathcal{D} \to [0,1]$ induces a *predictive probability* $\bar{\mu} : \Sigma \to [0,1]$ on the sample space through the reduction (2). The reduction map $\mu \mapsto \bar{\mu}$ relates subjective probabilities on the space M of models to subjective probabilities on the sample space S, that is, prior and predictive probabilities.[6] Clearly, (9) implies that

$$V(f) = \int_S u(f(s))\, d\bar{\mu}(s) \qquad \textbf{[10]}$$

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 DM's behavior. It is a much weaker representation than the "structural" one (9), which can be equivalently written as

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

since $\text{supp}\mu \subseteq M$ (recall that finite subsets of $\mathcal{D}$ are measurable). 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 $\mu : \mathcal{D} \to [0,1]$ such that $\text{supp}\mu \subseteq M$. The informed observer would be able to focus on the restriction of the reduction map $\mu \mapsto \bar{\mu}$ to $\Delta(M)$. If M is linearly independent, such correspondence is one-to-one and thus allows prior identification from the behaviorally elicited Savagean probability $P = \bar{\mu} \in \Delta$ through inversion.

The structural representation (9) is a version of Savage's representation that may be called *Classical SEU* since it takes into account Waldean information, with its classical flavor.[7] In place of the usual SEU pair (u, P) the representation is now characterized by a triple (u, M, μ), with $\text{supp}\mu \subseteq M$. According to the Bayesian paradigm, the prior μ quantifies probabilistically the DM'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 the Classical SEU criterion (9) reduces to the von Neumann-Morgenstern Expected Utility criterion (4), which is thus the special case of Classical SEU that corresponds to singleton data. In contrast, when M is nonsingleton but the support of a prior μ is a singleton, say $\text{supp}\mu = \{m'\} \subseteq M$, then it is the DM's personal information that prior μ reflects which leads him to a predictive that coincides with m'. In this case,

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

is a Savage's SEU criterion.

Support. In Proposition 1 the support of the prior is included in M, i.e., $\text{supp}\mu \subseteq M$. In fact, because of consistency models are assigned positive probability only if they belong to datum M. But, the DM may well decide to disregard some models in M because of some personal information. This additional information is reflected by his subjective belief μ,[8] with strict inclusion and $\mu(m) = 0$ for some $m \in M$.

Next we behaviorally characterize $\text{supp}\mu$ as the smallest subset of M relative to which $\succsim$ is consistent. These are the models that the DM believes to carry significant probabilistic information for his decision problem. In this perspective it is important to remember that M is an datum of the problem while $\text{supp}\mu$ is a subjective feature of the preferences.

We consider a linearly independent M in view of the identification result of Proposition 1.

Proposition 2. *Let* M *be a finite and linearly independent subset of* Δ_{na} *and* $\succsim$ *be a preference represented as in point (ii) of Proposition 1. A model* $m \in M$ *belongs to* $\text{supp}\mu$ *if and only if* $\succsim$ *is not consistent with* $M \setminus m$.

Therefore not only consistency arguments reveal the acceptance of a datum M, but they also allow to discover what elements of M are subjectively maintained or discarded.

Variations. We close by establishing the conditional and orthogonal versions of Proposition 1. We begin with the conditional version, i.e., with the counterpart of representation (5) under Waldean information.

Proposition 3. *Let* M *be a finite subset of* Δ_{na}*. The following statements are equivalent:*

(i) $\succsim$ *is a binary relation on* $\mathcal{F}$ *that satisfies P.1-P.6 and is order consistent with* M*;*

(ii) *there exist a nonconstant utility function* $u : X \to \mathbb{R}$ *and a prior* μ *on* Δ *with* $\text{supp}\mu \subseteq M$*, such that*

$$V_E(f) = \int_{\Delta} \left(\int_S u(f(s))\, dm(s \mid E) \right) d\mu(m \mid E) \qquad \textbf{[11]}$$

represents $\succsim_E$ *for all non-*$\bar{\mu}$*-null events* $E \in \Sigma$.

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

The representation of the conditional preferences $\succsim_E$ thus depends on the conditional models $m(\cdot \mid E) : \Sigma \to [0,1]$ and on the posterior probability $\mu(\cdot \mid E) : \mathcal{D} \to [0,1]$ that, respectively, update models and prior in the light of E. Criterion (11) shows how the DM currently plans to use the information he may gather through observations to update his inference on the actual data generating process.[9]

The conditional predictive probability is

$$\bar{\mu}(F \mid E) = \int_{\Delta} m(F \mid E)\, d\mu(m \mid E) \qquad \forall F \in \Sigma. \qquad \textbf{[12]}$$

and therefore the reduced form of (11) is

$$V_E(f) = \int_S u(f(s))\, d\bar{\mu}(s \mid E). \qquad \textbf{[13]}$$

---

[6] Notice that probability measures on S can play two conceptually altogether different roles: (subjective) predictive probabilities and (objective) probability models.

[7] Diaconis and Freedman [15] call "classical Bayesianism" the Bayesian approach that considers as a datum of the statistical problem the collection of all possibe data generating mechanisms.

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

[9] As Marschak [16, p. 109] remarked "to be an 'economic man' implies being a 'statistical man'." Some works of Jacob Marschak (notably [16], [17], and his classic book [18] with Roy Radner) have been a source of inspiration of our exercise, as we discuss in [19]. Our work addresses, inter alia, the issue that he raised in [17], 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.

The conditional representations (11) and (13) are, respectively, induced by the primitive representations (9) and (10) via conditioning.

Orthogonality is a simple, but important, sufficient condition for linear independence that, as the next section will show, some fundamental classes of time series models satisfy. Because of its importance, the following result shows what form the Classical SEU representation of Proposition 1 takes in this case.

Proposition 4. *Let M be a finite and orthogonal subset of Δ_{na}. The following statements are equivalent:*

(i) $\succsim$ is a binary relation on $\mathcal{F}$ that satisfies P.1-P.6 and is consistent with M;

(ii) there exist a nonconstant utility function $u : X \to \mathbb{R}$ and a prior μ on Δ with $\text{supp}\mu \subseteq M$, such that

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

represents $\succsim$.

Moreover, for each $\succsim$ satisfying (i), u is cardinally unique and μ is unique.

Notice that here consistency suffices and that the prior μ is automatically unique because of the orthogonality of M. In [19] we also show that a representation with an infinite M can be derived in the orthogonal case.

The reduction map $\mu \mapsto \bar{\mu}$ between prior and predictive probabilities is easily seen to be affine. More interestingly, in the orthogonal case it also preserves orthogonality and absolute continuity.

Proposition 5. *Under the assumptions of Proposition 4, two priors μ and ν on Δ with support in M are orthogonal (resp., absolutely continuous) if and only if their predictive probabilities $\bar{\mu}$ and $\bar{\nu}$ on S are orthogonal (resp., absolutely continuous).*

Intertemporal analysis

Setup. Consider a standard intertemporal decision problem where information builds up through observations generated by a sequence $\{Z_t\}$ of random variables taking values on observation spaces $\mathcal{Z}_t$. For ease of exposition, we assume that the observation spaces are finite and identical, each denoted by $\mathcal{Z}$ and endowed with the σ-algebra $\mathcal{B} = 2^{\mathcal{Z}}$.

The relevant state space S for the decision problem is the *sample space* $\mathcal{Z}^\infty = \prod_{t=1}^{\infty} \mathcal{Z}$. Its points are the possible observation paths generated by the process $\{Z_t\}$. W.l.o.g., we identify $\{Z_t\}$ with the coordinate process such that $Z_t(z) = z_t$ for each $z \in \mathcal{Z}^\infty$.

Endow $\mathcal{Z}^\infty$ with the product σ-algebra $\mathcal{B}^\infty$ generated by the elementary cylinder sets $z^t = \{s \in \mathcal{Z}^\infty : s_1 = z_1, ..., s_t = z_t\}$. These sets are the observables in this intertemporal setting. In particular, the filtration $\{\mathcal{B}^t\}$, where $\mathcal{B}^t$ is the algebra generated by the cylinders z^t, records the building up of observations. Clearly, $\mathcal{B}^\infty$ is the σ-algebra generated by the filtration $\{\mathcal{B}^t\}$.

In this intertemporal setting the pair (S, Σ) is thus given by $(\mathcal{Z}^\infty, \mathcal{B}^\infty)$. The space of data generating models Δ consists of all probability measures m on $\mathcal{Z}^\infty$. The outcome space X has also a product structure $X = \mathcal{C}^\infty$, where $\mathcal{C}$ is a common instant outcome space. Acts $f : \mathcal{Z}^\infty \to \mathcal{C}^\infty$ can thus be identified with the processes $\{f_t\}$ of their components. When such processes are adapted, the corresponding acts are called *plans* (here $f_t(s) = f_t(s_1, ..., s_t)$ is the outcome at time t if state s obtains). By Proposition 3, the conditional version of the Classical SEU representation at z^t is:

$$V_{z^t}(f) = \int_\Delta \left(\int_{\mathcal{Z}^\infty} u(f(s))\, dm(s \mid z^t) \right) d\mu(m \mid z^t) \quad \textbf{[14]}$$

where $m(\cdot \mid z^t)$ and $\mu(\cdot \mid z^t)$ are, respectively, the *conditional model* and the *posterior probability* given the observation history z^t. Under standard conditions, the intertemporal utility function $u : \mathcal{C}^\infty \to \mathbb{R}$ in (14) has a classic discounted form $u(c_1, ..., c_t, ...) = \sum_{\tau=1}^{\infty} \beta^{\tau-1} \upsilon(c_\tau)$, with subjective discount factor $\beta \in [0,1]$ and bounded instantaneous utility function $\upsilon : \mathcal{C} \to \mathbb{R}$.

Stationary case. The next known result (e.g., [21, p. 39]) shows that models are orthogonal in the fundamental stationary and ergodic case, which includes the standard i.i.d. setup as a special case.

Proposition 6. *A finite collection M of models that make the process $\{Z_t\}$ stationary and ergodic is orthogonal.*

By Proposition 4, if $\succsim$ satisfies P.1-P.6 and is consistent with a finite collection M of nonatomic, stationary and ergodic models, then there is a cardinally unique utility function u and a unique prior μ, with $\text{supp}\mu \subseteq M$, such that (14) holds. Its reduced form $V(f) = \int_{\mathcal{Z}^\infty} u(f(s))\, d\bar{\mu}(s)$ features a predictive probability $\bar{\mu}$ which is stationary (exchangeable in the special i.i.d. case).

Since a version of Proposition 6 holds also for collections of homogenous Markov chains, we can conclude that time series models widely used in applications satisfy the orthogonality conditions that ensure the uniqueness of prior μ. The Wald-Savage setup of this paper provides a statistical decision theory framework for empirical works that rely on such time series (as it is often the case in the Finance and Macroeconomics literatures).

Under these orthogonality conditions, there is full learning. Formally, denoting by

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

the continuation value at z^t of any act f and by $m' \in M$ the true model, it can be shown that

$$\left| W_{z^t}(f) - \int_{\mathcal{Z}^\infty} \sum_{\tau=t}^{\infty} \beta^{\tau-t} \upsilon(f_\tau(s))\, dm'(s \mid z^t) \right| \to 0$$

for m' almost every z in $\mathcal{Z}^\infty$. As observations build up, DMs learn and eventually behave as SEU DMs who know the true model that generates observations. Classical SEU thus provides a proper decision theoretic setting where to frame the common justification of rational expectations that "with a long enough historical data record, statistical learning will equate objective and subjective probability distributions." [10] Further intertemporal results are studied in [19] (the working paper version of this paper), which we refer the interested reader to.

Appendix: proofs and related analysis

Let M be a subset of $\Delta(S, \Sigma)$, a probability measure $P \in \Delta(S, \Sigma)$ is said to be a predictive of a prior on M (or to be M-representable) if and only if there exists $\mu \in \Delta(M, \mathcal{M})$ such that $P = \bar{\mu}$. If in addition such μ is unique, then P is said to be M-identifiable (see [14]).

We state the next result for any M since the proof for the finite case is only slightly simpler. We say that a subset M of $\Delta(S, \Sigma)$ is *measure independent* if, given any signed measure $\gamma : \mathcal{M} \to \mathbb{R}$,

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

If M is finite, measure independence reduces to usual notion (6) of linear independence.

Lemma 1. *Let $M \subseteq \Delta(S, \Sigma)$. The following statements are equivalent:*

[10] Sargent and Williams [20, p. 361].

(i) *every predictive of a prior on M is M-identifiable;*
(ii) *the map $\mu \mapsto \bar{\mu}$ from $\Delta(M,\mathcal{M})$ to $\Delta(S,\Sigma)$ is injective;*
(iii) *M is measure independent.*

Proof The equivalence of (i) and (ii) is trivial.

(iii) implies (ii) If $\mu_1,\mu_2 \in \Delta(M,\mathcal{M})$ are such that $\bar{\mu}_1 = \bar{\mu}_2 = P$, then $\mu_1 - \mu_2$ is a signed measure on M and

$$\begin{aligned} & \int_M m(E)\, d(\mu_1 - \mu_2)(m) \\ = & \int_M m(E)\, d\mu_1(m) - \int_M m(E)\, d\mu_2(m) \\ = & P(E) - P(E) = 0 \quad \forall E \in \Sigma. \end{aligned}$$

Since M is measure independent, it follows that $\mu_1 - \mu_2 = 0$, i.e., $\mu_1 = \mu_2$.

(ii) implies (iii) Assume, per contra, that M is not measure independent. Then, there is a signed measure γ on M such that

$$\gamma \neq 0 \text{ and } \int_M m(E)\, d\gamma(m) = 0 \quad \forall E \in \Sigma. \qquad [15]$$

By the Hahn-Jordan Decomposition Theorem, $\gamma = \gamma^+ - \gamma^-$ where γ^+ and γ^- are, respectively, the positive and negative part of γ. By (15),

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

Since $\gamma \neq 0$, this implies that $\gamma^+(M) = \gamma^-(M) = 1/k > 0$. Then $k\gamma^+, k\gamma^- \in \Delta(M,\mathcal{M})$, $k\gamma^+ \neq k\gamma^-$ (else $\gamma = 0$), and, by (15), for each $E \in \Sigma$

$$\begin{aligned} 0 &= k\int_M m(E)\, d\gamma(m) = \int_M m(E)\, d\left(k\gamma^+ - k\gamma^-\right)(m) \\ &= \int_M m(E)\, dk\gamma^+(m) - \int_M m(E)\, dk\gamma^-(m) \\ &= \overline{k\gamma^+}(E) - \overline{k\gamma^-}(E). \end{aligned}$$

Therefore $\overline{k\gamma^+} = \overline{k\gamma^-}$ negating injectivity. ■

Lemma 2. *If $M \subseteq \Delta(S,\Sigma)$ is finite, then*

$$\mathcal{M} = 2^M = \sigma(m \mapsto m(E) : E \in \Sigma).$$

Moreover, the map $\nu \mapsto \bar{\nu}$ from $\Delta(M)$ to $\Delta(S,\Sigma)$ is injective if and only if M is linearly independent.

Proof The equality $\mathcal{M} = 2^M$ follows from the fact $\mathcal{M}$ contains all singletons. Next we show that $\sigma(m \mapsto m(E) : E \in \Sigma)$ contains all singletons. Notice that if $p \neq q$ in M, there exists $E_{pq} \in \Sigma$ such that $p(E_{pq}) \neq q(E_{pq})$. Then for each $p \in M$,

$$\{p\} = \{m \in M : m(E_{pq}) = p(E_{pq}) \quad \forall q \in M\}$$

is a finite intersection of $\sigma(m \mapsto m(E) : E \in \Sigma)$-measurable sets and so it is measurable too.

Recall that $\Delta(M) = \{\nu \in \Delta(\Delta(S,\Sigma)) : \nu(M) = 1\}$ while $\Delta(M,\mathcal{M})$ is the set of all probability measures $\mu : 2^M \to [0,1]$.

Let $\nu_1, \nu_2 \in \Delta(M)$. Setting $\nu_i(m) = \nu_i(\{m\})$ for all $m \in M$, it follows $\nu_i = \sum_{m\in M} \nu_i(m)\,\delta_m$ and $\bar{\nu}_i = \sum_{m \in M} \nu_i(m)\, m$. Denote by μ_i the restriction of ν_i to $\mathcal{M} = 2^M$ and notice that $\mu_i = \sum_{m\in M} \nu_i(m)\,\partial_m \in \Delta(M,\mathcal{M})$ where ∂_m is the restriction of δ_m (defined on $\mathcal{D}$) to $\mathcal{M}$ and that $\bar{\mu}_i = \sum_{m\in M} \nu_i(m)\, m = \bar{\nu}_i$. If M is linearly independent, then $\bar{\nu}_1 = \bar{\nu}_2$ implies $\bar{\mu}_1 = \bar{\mu}_2$. By Lemma 1, $\mu_1 = \mu_2$. Thus $\nu_1(m) = \nu_2(m)$ for all $m \in M$ and $\nu_1 = \nu_2$. This proves injectivity.

Conversely, if M is not linearly independent, by Lemma 1 there exist $\eta_1 = \sum_{m\in M} \eta_1(m)\,\partial_m$ and $\eta_2 = \sum_{m\in M}\eta_2(m)\,\partial_m$ in $\Delta(M,\mathcal{M})$ such that $\eta_1 \neq \eta_2$ but $\bar{\eta}_1 = \bar{\eta}_2$. Now, setting $\lambda_i = \sum_{m\in M} \eta_i(m)\,\delta_m \in \Delta(M)$ for $i = 1,2$, it follows that $\lambda_1 \neq \lambda_2$ but $\bar{\lambda}_1 = \bar{\eta}_1 = \bar{\eta}_2 = \bar{\lambda}_2$. This negates injectivity. ■

Proof of Proposition 1 (i) implies (ii) By Savage Representation Theorem, there are a nonconstant function $u : X \to \mathbb{R}$ and a unique (strongly) nonatomic and finitely additive probability P on S such that setting $V(f) = \int_S u(f(s))\, dP(s)$,

$$f \succsim g \iff V(f) \geq V(g).$$

By assumption, each m is nonatomic. By the Lyapunov Theorem, there is a unanimous event $E \in \Sigma$, say with $m(E) = 2^{-1}$ for all $m \in M$. By order consistency, for each $F \in \Sigma$

$$m(F) = m(E) \quad \forall m \in M \Longrightarrow P(F) = P(E) \qquad [16]$$

and

$$m(F) \geq m(E) \quad \forall m \in M \Longrightarrow P(F) \geq P(E). \qquad [17]$$

By [22, Theorem 20], P belongs to the convex cone generated by M, since $P(S) = m(S) = 1$ for all $m \in M$, then $P \in \text{co}M$ and representation (9) holds.

(ii) implies (i) Define $P = \bar{\mu}$. Since each $m \in M$ is a nonatomic probability measure, so is P. By the Savage Representation Theorem, it follows that $\succsim$ satisfies P.1-P.6. Finally, we show that $\succsim$ is order consistent with M. Let $E, F \in \Sigma$ and assume $m(F) \geq m(E)$ for each $m \in \text{supp}\mu \subseteq M$. Then for all outcomes $x \succ y$, normalizing u so that $u(x) = 1 = 1 - u(y)$, $V(xFy) = \bar{\mu}(F) \geq \bar{\mu}(E) = V(xEy)$, and so $xFy \succsim xEy$. A fortiori order consistency is satisfied (both with respect to $\text{supp}\mu$ and M).

Moreover, for each $\succsim$ satisfying (i), the cardinal uniqueness of u and the uniqueness of $\bar{\mu}$ follow from Savage Representation Theorem. If M is linearly independent, for each $\succsim$ satisfying (i), $\bar{\mu}$ is unique and Lemma 2 delivers the uniqueness of μ. Conversely, if M is not linearly independent, by Lemma 2 there exist two different $\mu_1, \mu_2 \in \Delta(M)$ such that $\bar{\mu}_1 = \bar{\mu}_2$; arbitrarily choose a nonconstant $u : X \to \mathbb{R}$ to obtain a binary relation $\succsim$ satisfying (i) which is represented both by μ_1 and by μ_2 (together with u) in the sense of (ii). ■

Proof of Proposition 2 Let $m \in M$. Replicating the last part of the previous proof, if m does not belong to $\text{supp}\mu$ then $\succsim$ is consistent with $M \setminus m$. Now assume that $\succsim$ is consistent with $M \setminus m$. Take $E \in \Sigma$ such that $m'(E) = 2^{-1}$ for all $m' \in M \setminus m$, by consistency, if $F \in \Sigma$, then

$$m'(F) = m'(E) \quad \forall m' \in M \setminus m \Longrightarrow \bar{\mu}(F) = \bar{\mu}(E).$$

If m belongs to $\text{supp}\mu$, then

$$\begin{aligned} m(F) &= \frac{1}{\mu(m)}\left(\bar{\mu}(F) - \sum_{m' \in M\setminus m} \mu(m')\, m'(F)\right) \\ &= \frac{1}{\mu(m)}\left(\bar{\mu}(E) - \sum_{m' \in M\setminus m} \mu(m')\, m'(E)\right) \\ &= m(E) \end{aligned}$$

Since each element of M is nonatomic, by [22, Theorem 20] $m \in \text{span}(M \setminus m)$, which contradicts the linear independence of M. ■

Proof of Proposition 3 Clearly (ii) of this proposition implies point (ii) of Proposition 1 which in turn implies (i).

Conversely, (i) of this proposition implies point (ii) of Proposition 1 which together with (5) implies that $V_E(f) = \int_S u(f(s))\, d\bar{\mu}(s \mid E)$ represents $\succsim_E$ for all nonnull $E \in \Sigma$. But $\text{supp}\mu(\cdot \mid E) = \{m \in \text{supp}\mu : m(E) > 0\}$ and hence

$$\begin{aligned} V_E(f) &= \frac{1}{\bar{\mu}(E)} \int_E u(f)\, d\bar{\mu} \\ &= \frac{1}{\bar{\mu}(E)} \sum_{m \in \text{supp}\mu} \mu(m) \int_E u(f)\, dm \\ &= \frac{1}{\bar{\mu}(E)} \sum_{m \in \text{supp}\mu : m(E)>0} \mu(m) \frac{m(E)}{m(E)} \int_E u(f)\, dm \\ &= \sum_{m \in \text{supp}\mu(\cdot \mid E)} \left(\frac{\mu(m)\, m(E)}{\bar{\mu}(E)} \right) \int_S u(f)\, dm(\cdot \mid E) \\ &= \int_\Delta \left(\int_S u(f(s))\, dm(s \mid E) \right) d\mu(m \mid E) \end{aligned}$$

so that (ii) holds.

The rest follows immediately from Proposition 1. ■

Proof of Proposition 4 The proof of (i) implies (ii) of Proposition 1 has to be modified since consistency only yields (16). Then [22, Theorem 20] only yields that P belongs to the vector subspace generated by M. In any case, there exists a collection $\{\mu(m)\}_{m \in M}$ of scalars such that $P(E) = \sum_{m \in M} \mu(m)\, m(E)$ for all $E \in \Sigma$. From $P(S) = m(S) = 1$ for all $m \in M$, it follows that $\sum_{m \in M} \mu(m) = 1$. Moreover, by orthogonality, there exists a partition $\{E_m\}_{m \in M}$ of S in Σ such that $m(E_m) = 1$ and $m'(E_m) = 0$ for all distinct $m, m' \in M$ (see the beginning of the next proof). Hence, for each m it holds $P(E_m) = \mu(m)$, and so $\mu(m) \geq 0$. We conclude that $P \in \text{co}M$ again. The rest of the proof is very similar to that of Proposition 1. ■

Proof of Proposition 5 We consider orthogonality and leave absolute continuity to the reader. Suppose $\mu \perp \nu$, i.e., there is $A \in \mathcal{D}$ such that $\mu(A) = 1 = \nu(A^c)$. Next we show that there exists a partition $\{E_m\}_{m \in M}$ of S in Σ such that $m(E_m) = 1$ and $m'(E_m) = 0$ for all distinct $m, m' \in M$.[11] Let $M = \{m_1, ..., m_n\}$. For $n = 2$, the result is true by definition of orthogonality. Assume $n \geq 3$ and the result holds for $n-1$. Then there exists a partition $\{F_i\}_{i=2}^n$ of S in Σ such that $m_i(F_i) = 1$ for all $i, j = 2, ..., n$. But $m_1 \perp m_i$ for each $i \neq 1$, hence there is $E_{1i} \in \Sigma$ such that $m_1(E_{1i}) = 1 = m_i(E_{1i}^c)$. By setting $F_1 = \bigcap_{i \neq 1} E_{1i}$ and $E_i = E_{1i}^c \cap F_i$ we then have $m_1(F_1) = 1$ and $m_i(E_i) = 1$ for each $i \neq 1$. The desired partition is obtained by setting $E_1 = S \setminus \bigcup_{i \neq 1} E_i$.

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,

$$\begin{aligned} \bar{\mu}(E) &= \sum_{m \in M} m(E)\, \mu(m) = \sum_{m \in A} m(E)\, \mu(m) \\ &= \sum_{m \in A} \mu(m) = \mu(A) = 1 \end{aligned}$$

and

$$\bar{\nu}(E) = \sum_{m' \in M} m'(E)\, \nu(m') = \sum_{m' \in A^c} m'(E)\, \nu(m') = 0 \quad [18]$$

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

$$\begin{aligned} 1 &= \bar{\mu}(E) = \sum_{m \in M} m(E)\, \mu(m) = \sum_{m \in A} m(E)\, \mu(m) \\ &\leq \sum_{m \in A} \mu(m) = \mu(A) \leq 1 \end{aligned}$$

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

$$0 = \bar{\nu}(E) = \sum_{m \in M} m(E)\, \nu(m) = \sum_{m \in A} m(E)\, \nu(m) \quad [19]$$

whence $\nu(m) = 0$ for all $m \in A$ because $m(E) > 0$. We conclude that $\nu(A) = 0$ and $\mu \perp \nu$. ■

1. Savage LJ (1954) *The foundations of statistics*. Wiley, New York.
2. Wald A (1950) *Statistical decision functions*. Wiley, New York.
3. Sargent TJ (2008) Evolution and intelligent design. *American Economic Review* 98: 3-37.
4. Lucas Jr. RE (1977) Understanding business cycles, in *Stabilization of the Domestic and International Economy*, eds Brunner K, Meltzer AH (North-Holland, Amsterdam).
5. Muth JF (1961) Rational expectations and the theory of price movements. *Econometrica* 29: 315-335.
6. Fishburn PC (1970) *Utility Theory for Decision Making*. Wiley, New York.
7. Kreps DM (1988) *Notes on the Theory of Choice*. Westwiew Press Inc, Colorado.
8. Gilboa I (2009) *Theory of Decision under Uncertainty*. Cambridge University Press, Cambridge.
9. Jaffray J-Y (1989) Linear utility theory for belief functions. *Operations Research Letters* 8: 107-112.
10. Battigalli P, Cerreia-Vioglio S, Maccheroni F, Marinacci M (2011) Selfconfirming equilibrium and model uncertainty. Igier WP 428.
11. Cerreia-Vioglio S, Maccheroni F, Marinacci M (2012) Robust statistical decision theory. mimeo.
12. Hansen LP, Sargent TJ (2007) *Robustness*. Princeton University Press, Princeton.
13. Bhaskara Rao KPS, Bhaskara Rao M (1983) *Theory of Charges*. Academic Press, New York.
14. Teicher H (1963) Identifiability of finite mixtures. *Annals of Mathematical Statistics* 34: 244-248.
15. Diaconis P, Freedman D (1986) On the consistency of Bayes estimates. *Annals of Statistics* 14: 1-26.
16. Marschak J (1946) Neumann's and Morgenstern's new approach to static economics. *Journal of Political Economy* 54: 97-115.
17. Marschak J (1975) Personal probabilities of probabilities. *Theory and Decision* 6: 121-153.
18. Marschak J, Radner R (1972) *Economic theory of teams*. Yale University Press, New Haven.
19. Cerreia-Vioglio S, Maccheroni F, Marinacci M, Montrucchio L (2011) Classical subjective expected utility, IGIER WP 400.
20. Sargent TJ, Williams N (2005) Impacts of priors on convergence and escapes from Nash inflation. *Review of Economic Dynamics* 8: 360-391.
21. Billingsley, P. (1965), *Ergodic theory and information*, Wiley, New York.
22. Marinacci M, Montrucchio L (2003) Subcalculus for set functions and cores of TU games. *Journal of Mathematical Economics* 39: 1-25.

[11] Notice that $m'(E_m) = 0$ for all $m, m' \in M$ such that $m \neq m'$ actually follows from the fact that $\{E_m\}$ is a partition and $m(E_m) = 1$ for all $m \in M$.

Reserved for Publication Footnotes