Classical subjective expected utility

Simone Cerreia-Vioglio[a], Fabio Maccheroni[a], Massimo Marinacci[a,1], and Luigi Montrucchio[b,c]

[a]Department of Decision Sciences and IGIER, Università Bocconi, 20136 Milan, Italy; [b]Department of Statistics and Applied Mathematics, Università di Torino, 10134 Turin, Italy; and [c]Collegio Carlo Alberto, 10024 Moncalieri, Italy

Edited by Jose A. Scheinkman, Princeton University, Princeton, NJ, and approved February 28, 2013 (received for review May 9, 2012)

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 [Wald A (1950) *Statistical Decision Functions* (Wiley, New York)]. We incorporate this Waldean piece of objective information within an otherwise subjective setting à la Savage [Savage LJ (1954) *The Foundations of Statistics* (Wiley, New York)] and show that this leads to a two-stage subjective expected utility model that accounts for both state and model uncertainty.

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 (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, and 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 the 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: Whereas 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 à 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 à 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. [As Wald (ref. 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."] 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 independent identically distributed (i.i.d.) trials] through maximum-likelihood methods, hypothesis testing, and the like. Although models cannot be observed, in Wald's approach their study is key to better understanding 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. In addition to its theoretical interest, this question is relevant because in some important economic applications it is natural to assume, at least as a working hypothesis, that DMs have this kind of information [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,

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

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 (because 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

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

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:

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

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

$$V(f) = \int_{S} u(f(s)) dP(s) \qquad [3]$$

is the reduced form of V, its subjective expected utility (SEU) representation à 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

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

where subjective probabilities do not play any role. [Lucas (ref. 4, p. 15) writes that "Muth (5) … [identifies] … agents' subjective

Author contributions: S.C.-V., F.M., M.M., and L.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: massimo.marinacci@unibocconi.it.

www.pnas.org/cgi/doi/10.1073/pnas.1207805110

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 modeled by singleton M and results in the expected utility criterion (4).] Classical SEU thus encompasses both the Savage and the von Neumann–Morgenstern representations.

In particular, the Savage criterion [**3**] is what an outside observer, unaware of datum M, would be able to elicit from the 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 supp$\mu \subseteq M$, whereas for the usual Savagean representation [**3**] is enough a pair (u, P).

Summing up, although 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 did not consider the classical datum central in Wald's approach. [See Fishburn (6), Kreps (7), and Gilboa (8). See Jaffray (9) for a different "objective" 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 considering only choice behavior based on observables. Battigalli et al. (10) use the Wald–Savage setup of the present paper to study self-confirming equilibria, whereas we are currently using it to provide a behavioral foundation of the robustness approach in macroeconomics pioneered by Hansen and Sargent (11).

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

The DM's preferences are represented by a binary relation $\succsim$ over $\mathcal{F}$. We assume that $\succsim$ satisfies the classic Savage 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 $\succsim$. [Strong nonatomicity of P means that for each $E \in \Sigma$ and $0 \le c \le P(E)$ there exists $F \in \Sigma$ such that $F \subseteq E$ and $P(F) = c$. See ref. 12, p. 141–143 for the various definitions and properties of nonatomicity of finitely additive probabilities.] 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 (ref. 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 because the primitive preference does [e.g., Kreps (ref. 7, Chap. 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,

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

represents $\succsim_E$, where $P(\cdot|E)$ 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 S. Unless otherwise stated, in the rest of this paper all probability measures are countably additive.

In the sequel, we 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 us to update prior μ through the Bayes rule

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

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

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

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

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 $\{(m_1(E), \ldots, m_n(E)): E \in \Sigma\}$ of a finite collection $\{m_i\}_{i=1}^n$ of nonatomic probability measures is a convex subset of $\mathbb{R}^n$.

ECONOMIC SCIENCES

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,*

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

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,*

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

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. Note that order consistency implies consistency because 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*:

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 supp$\mu \subseteq M$, such that*

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

represents $\succsim$.

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

Although 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 [e.g., Teicher (13)].

Each prior μ: $\mathcal{D} \to [0, 1]$ induces a predictive probability $\overline{\mu} : \Sigma \to [0, 1]$ on the sample space through the reduction [**2**]. The reduction map $\mu \mapsto \overline{\mu}$ relates subjective probabilities on the space M of models to subjective probabilities on the sample space S, that is, prior and predictive probabilities. [Note that probability measures on S can play two conceptually altogether different roles: (subjective) predictive probabilities and (objective) probability models.] Clearly, [**9**] implies that

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

which is the reduced form of V, its Savage's SEU form. As observed in the introductory section, this is the criterion that an outside observer, unaware of datum M, would be able to elicit from the DM's behavior. It is a much weaker representation than the structural one ([**9**]), which can be equivalently written as

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

because 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. In fact, denote by $\Delta(M)$ the collection of all priors μ: $\mathcal{D} \to [0, 1]$ such that supp$\mu \subseteq M$. The informed observer would be able to focus on the restriction of the reduction map $\mu \mapsto \overline{\mu}$ 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 $P = \overline{\mu} \in \Delta$ through inversion.

The structural representation [**9**] is a version of Savage's representation that may be called classical SEU because it takes into account Waldean information, with its classical flavor. [Diaconis and Freedman (14) call "classical Bayesianism" the Bayesian approach that considers as a datum of the statistical problem the collection of all possible data-generating mechanisms.] In place of the usual SEU pair (u, P) the representation is now characterized by a triple (u, M, μ), with 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 introductory section, 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 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., supp$\mu \subseteq M$. In fact, because of consistency, models are assigned positive probability only if they belong to datum M. However, 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, with strict inclusion and $\mu(m) = 0$ for some $m \in M$. (In fact, the interpretation of μ is purely subjective, not at all logical/objective à la Carnap and Keynes.)

Next we behaviorally characterize suppμ 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 a datum of the problem whereas suppμ 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 suppμ if and only if $\succsim$ is not consistent with $M \backslash m$.*

Therefore, consistency arguments not only reveal the acceptance of a datum M, but also allow us 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 supp$\mu \subseteq M$, such that*

$$V_E(f) = \int_\Delta \left(\int_S u(f(s)) dm(s|E) \right) d\mu(m|E) \qquad [11]$$

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

Moreover, u is cardinally unique for each $\succsim$ satisfying statement i, whereas 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 $m(\cdot|E) : \Sigma \to [0, 1]$ and on the posterior probability $\mu(\cdot|E) : \mathcal{D} \to [0, 1]$ 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. [As Marschak (ref. 15, p. 109) remarked "to be an 'economic man' implies being a 'statistical man'." Some works of Jacob Marschak (notably refs. 15 and 16 and his classic book, ref. 17, with Roy Radner) have been a source of inspiration of our exercise, as we discuss in ref. 18. Our work addresses, inter alia, the issue that he raised in ref. 16, 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 predictive probability is

$$\bar{\mu}(F|E) = \int_{\Delta} m(F|E)d\mu(m|E) \quad \forall F \in \Sigma \qquad [12]$$

and therefore the reduced form of [11] is

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

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 shows, 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* $\operatorname{supp}\mu \subseteq M$, *such that*

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

represents $\succsim$.

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

Note that here consistency suffices and that the prior μ is automatically unique because of the orthogonality of M. In ref. 18 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 $\{Z_t\}$. Without loss of generality, 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 $z^t = \{s \in \mathcal{Z}^\infty : s_1 = z_1, \ldots, s_t = z_t\}$. 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 $f_t(s) = f_t(s_1, \ldots, s_t)$ 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

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

where $m(\cdot|z^t)$ and $\mu(\cdot|z^t)$ 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 $u(c_1, \ldots, c_t, \ldots) = \sum_{\tau=1}^\infty \beta^{\tau-1}\upsilon(c_\tau)$, 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., ref. 19, 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 are a cardinally unique utility function u and a unique prior μ, with $\operatorname{supp}\mu \subseteq M$, such that [14] holds. Its reduced form $V(f) = \int_{\mathcal{Z}^\infty} u(f(s))d\bar{\mu}(s)$ features a predictive probability $\bar{\mu}$ that is stationary (exchangeable in the special i.i.d. case).

Because 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 is often the case in the finance and macroeconomics literatures).

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

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

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

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

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. The above convergence result shows how the Classical SEU framework of this paper allows to formalize, in terms of learning, the common justification of rational expectations according to which "with a long enough historical data record, statistical learning will equate objective and subjective probability distributions" [Sargent and Williams (ref. 20, p. 361)]. Further intertemporal results are studied in ref. 18 (the working paper version of this paper), which we refer the interested reader to.

Appendix: Proofs and Related Analysis

Letting 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 there exists $\mu\in\Delta(M,\mathcal{M})$ such that $P=\overline{\mu}$. If in addition such μ is unique, then P is said to be M-identifiable (13).

We state the next result for any M because 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 $\gamma:\mathcal{M}\to\mathbb{R}$,

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

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

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

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

Proof: The equivalence of statements *i* and *ii* is trivial.

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

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

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

Statement *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 parts of γ. By [15],

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

Because $\gamma\neq 0$, 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 $E\in\Sigma$

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

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 $\nu\mapsto\overline{\nu}$ 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 that $\mathcal{M}$ contains all singletons. Next we show that $\sigma(m\mapsto m(E):E\in\Sigma)$ contains all singletons. Note 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 $\Delta(M)=\{\nu\in\Delta(\Delta(S,\Sigma)):\nu(M)=1\}$ whereas $\Delta(M,\mathcal{M})$ 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 that $\nu_i=\sum_{m\in M}\nu_i(m)\delta_m$ and $\overline{\nu}_i=\sum_{m\in M}\nu_i(m)m$. Denote by μ_i the restriction of ν_i to $\mathcal{M}=2^M$ and note 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 $\overline{\mu}_i=\sum_{m\in M}\nu_i(m)m=\overline{\nu}_i$. If M is linearly independent, then $\overline{\nu}_1=\overline{\nu}_2$ implies $\overline{\mu}_1=\overline{\mu}_2$. 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 $\overline{\eta}_1=\overline{\eta}_2$. 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 $\overline{\lambda}_1=\overline{\eta}_1=\overline{\eta}_2=\overline{\lambda}_2$. This negates injectivity. ∎

Proof of Proposition 1: Statement *i* implies *ii*. By the 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 $V(f)=\int_S u(f(s))dP(s)$,

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

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 $F\in\Sigma$

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

and

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

By ref. 21, Theorem 20, P belongs to the convex cone generated by M, because $P(S)=m(S)=1$ for all $m\in M$, and then $P\in\mathrm{co}M$ and representation [9] holds.

Statement *ii* implies *i*. Define $P=\overline{\mu}$. Because 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 $m\in\mathrm{supp}\mu\subseteq M$. Then for all outcomes $x\succ y$, normalizing u so that $u(x)=1=1-u(y)$, $V(xFy)=\overline{\mu}(F)\geq\overline{\mu}(E)=V(xEy)$, and so $xFy\succsim xEy$. *A fortiori* order consistency is satisfied (with respect to both $\mathrm{supp}\mu$ and M).

Moreover, for each $\succsim$ satisfying statement *i*, the cardinal uniqueness of u and the uniqueness of $\overline{\mu}$ follow from the Savage representation theorem. If M is linearly independent, for each $\succsim$ satisfying *i*, $\overline{\mu}$ 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 $\overline{\mu}_1=\overline{\mu}_2$; arbitrarily choose a nonconstant $u:X\to\mathbb{R}$ to obtain a binary relation $\succsim$ satisfying *i* that 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 suppμ, then $\succsim$ is consistent with $M\backslash m$. Now assume that $\succsim$ is consistent with $M\backslash m$. Take $E \in \Sigma$ such that $m'(E) = 2^{-1}$ for all $m' \in M\backslash m$; by consistency, if $F \in \Sigma$, then

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

If m belongs to suppμ, then

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

Because each element of M is nonatomic, by ref. 21, Theorem 20, $m \in \mathrm{span}(M\backslash m)$, which contradicts the linear independence of M. ■

Proof of Proposition 3: Clearly statement *ii* of this proposition implies point *ii* of Proposition 1, which in turn implies *i*.

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

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

so that *ii* holds.

The rest follows immediately from Proposition 3. ■

Proof of Proposition 4: The proof of statement *i* implies *ii* of Proposition 3 has to be modified because consistency yields only [**16**]. Then ref. 21, Theorem 20, yields only 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 that $P(E_m) = \mu(m)$, and so $\mu(m) \geq 0$. We conclude that $P \in \mathrm{co}M$ again. The rest of the proof is very similar to that of Proposition 3. ■

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 $m, m' \in M$. [Note 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 $m \in M$.] Let $M = \{m_1, \ldots, m_n\}$. 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 $i = 2, \ldots, n$. However, $m_1 \perp m_i$ for each $i \neq 1$, and hence there is $E_{1i} \in \Sigma$ such that $m_1(E_{1i}) = 1 = m_i(E_{1i}^c)$. By setting $F_1 = \cap_{i \neq 1} E_{1i}$ 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 $E_1 = S \backslash \cup_{i \neq 1} E_i$.

Set $E = \cup\{E_m : m \in A\}$. 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

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

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 $A \in \mathcal{D}$ because A is finite. It holds that

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

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

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

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

ACKNOWLEDGMENTS. The financial support of ERC (advanced Grant BRSCDP-TEA) is gratefully acknowledged.

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. *Am Econ Rev* 98:3–37.
4. Lucas RE, Jr (1977) Understanding business cycles. *Stabilization of the Domestic and International Economy*, eds Brunner K, Meltzer AH (North-Holland, Amsterdam), pp 7–29.
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* (Westview Press, Boulder, CO).
8. Gilboa I (2009) *Theory of Decision under Uncertainty* (Cambridge Univ Press, Cambridge, UK).
9. Jaffray J-Y (1989) Linear utility theory for belief functions. *Oper Res Lett* 8:107–112.
10. Battigalli P, Cerreia-Vioglio S, Maccheroni F, Marinacci M (2011) Selfconfirming equilibrium and model uncertainty. Working paper 428 (Innocenzo Gasparini Institute for Economic Research, Milan).
11. Hansen LP, Sargent TJ (2007) *Robustness* (Princeton Univ Press, Princeton).
12. Bhaskara Rao KPS, Bhaskara Rao M (1983) *Theory of Charges* (Academic, New York).
13. Teicher H (1963) Identifiability of finite mixtures. *Ann Math Stat* 34:244–248.
14. Diaconis P, Freedman D (1986) On the consistency of Bayes estimates. *Ann Stat* 14: 1–26.
15. Marschak J (1946) Neumann's and Morgenstern's new approach to static economics. *J Polit Econ* 54:97–115.
16. Marschak J (1975) Personal probabilities of probabilities. *Theory Decis* 6:121–153.
17. Marschak J, Radner R (1972) *Economic Theory of Teams* (Yale Univ Press, New Haven).
18. Cerreia-Vioglio S, Maccheroni F, Marinacci M, Montrucchio L (2011) Classical subjective expected utility. Working paper 400 (Innocenzo Gasparini Institute for Economic Research, Milan).
19. Billingsley P (1965) *Ergodic Theory and Information* (Wiley, New York).
20. Sargent TJ, Williams N (2005) Impacts of priors on convergence and escapes from Nash inflation. *Rev Econ Dyn* 8:360–391.
21. Marinacci M, Montrucchio L (2003) Subcalculus for set functions and cores of TU games. *J Math Econ* 39:1–25.