DECOMPOSITION AND CHARACTERIZATION OF RISK WITH A CONTINUUM OF RANDOM VARIABLES

BY NABIL IBRAHEEM AL-NAJJAR

The paper studies the representation and characterization of risks generated by a continuum of random variables. The Main Theorem is a characterization of a broad class of continuum processes in terms of the decomposition of risk into aggregate and idiosyncratic components, and in terms of the approximation of the continuum process by finite collections of random variables. This characterization is used to study decision making problems with anonymous and state-independent payoffs. An Extension Theorem shows that if such a payoff function is defined on simple processes, then it has a unique continuous extension to the class of processes characterized in this paper. This extension is formulated without reference to sample realizations and with minimal restrictions on the patterns of correlation between the random variables. As an application, the theory is used to develop a new model of large games which emphasizes the explicit description of the players' randomizations. This model is used to study the class of environments in which Schmeidler's (1973) representation of strategic uncertainty in large games is valid.

KEYWORDS: Idiosyncratic risk, aggregate risk, decision theory, large games.

1. INTRODUCTION

This paper studies the representation and characterization of risks generated by a continuum of random variables \( f = \{ f_t : t \in [0, 1] \} \) (henceforth referred to as a process). Such processes appear in a broad range of economic and game theoretic models. Examples include: (i) models with a continuum of individuals and exogenous uncertainty—e.g., \( f_t \) is an indicator function of an accident or the random endowment of individual \( t \); (ii) models of portfolio selection with a continuum of assets—\( f_t \) is the random return on a set \( t \); (iii) decision making in games with a continuum of players—\( f_t \) is the random strategy chosen by player \( t \); (iv) games with a continuum of privately informed players—\( f_t \) is the randomly selected type of player \( t \); (v) models of matching between the members of a population consisting of a continuum of players—\( f_t \) is the random name of the player matched with player \( t \).

The paper provides a complete characterization of a broad class of continuum processes in terms of three economically relevant properties. Specifically, the Main Theorem asserts that a weak measurability condition on a process \( f \) is equivalent to any one of the following three conditions: (i) Decomposition of Risk: \( f \) can be essentially uniquely decomposed into aggregate and purely idiosyncratic components; (ii) approximation by simple processes: \( f \) is the limit of a sequence of simple processes \( \{ f^m \} \); (iii) Approximation by Large Samples: \( f \)

---

1I am grateful to Harald Uhlig for introducing me to this problem and for our many conversations about the continuum. I also thank the Editor, three referees, Maxwell Stinchcombe, and Harald Uhlig for many thoughtful and detailed comments which helped me improve the paper. Finally, I wish to express my special debt to Marcel K. Richter for his invaluable support and very helpful comments. This research was funded by a grant from the Social Sciences and Humanities Research Council of Canada. Remaining errors and shortcomings are my own.
can be approximated by large finite samples from [0,1] in a consistent manner. The decomposition part reveals that the processes characterized in this paper have a remarkably simple structure, while the approximation parts establish a tight connection between continuum processes and their finite approximations.

This theorem is then used to study the problem of a decision maker who faces the risk represented by a continuum process \( f \) and who has anonymous and state-independent payoffs. Each process \( f \) is assigned a unique probability distribution \( \mu(f) \) which provides an anonymous and state-independent representation of total risk; a convergence concept is also introduced under which any weakly measurable process is the limit of a sequence of simple processes. Two results are established. First, a Continuity Theorem shows that the representation of total risk \( \mu(f) \) in the continuum model is close to the corresponding representations in all nearby simple processes and almost every large sample. Second, an Extension Theorem states that the payoff function of a decision maker who evaluates risk in simple processes in an anonymous and state-independent manner has a unique continuous extension to the set of weakly measurable processes. These results are of basic theoretical interest because there is complete agreement on the representation of risks generated by finite collections of random variables, and because the continuum model is usually thought of as an idealization of large finite models. The statements and proofs of these results are independent of the sample realizations of the continuum process (i.e., the functions \( t \mapsto f_t(\omega) \)), which may be nonmeasurable, and of the possibly complex patterns of correlation which might exist between the random variables.

The paper's abstract framework for the analysis of risk is a tool which can be fruitfully applied to a broad range of problems. As an example, Section 6 examines Schmeidler's (1973) distributional representation of strategic uncertainty in a large game which is formulated in terms of aggregating the individual mixed strategy distributions over pure actions. This can be viewed as a reduced form representation because it does not explicitly model the mechanisms used by the players to select pure actions. Section 6 explicitly models the randomization devices used by the players and shows, essentially, that Schmeidler's representation provides an accurate description of the strategic uncertainty facing individual players if and only if players use uncorrelated randomization devices. This justifies Schmeidler's representation in terms of explicit conditions on the informational constraints faced by the players. On the other hand, these results reveal that strong restrictions on the players' environment must be imposed in order to guarantee the validity of Schmeidler's representation. Some of the economic implications of these restrictions are discussed.

Three aspects of the analysis of risk deserve special emphasis. First, the class of weakly measurable processes characterized in the Theorem is very broad. This is because weak measurability is consistent with rather general patterns of correlation between the random variables \( f_t, t \in [0,1] \). For example, the class of i.i.d. processes, which received a great deal of attention in the literature (e.g., Judd (1985), Feldman and Gilles (1985), and Uhlig (1987)), is included as a
special case. However, the class of processes characterized in this paper is considerably more general. This generality is essential for a unified treatment of a broad range of issues. Modeling random asset returns is an example of a situation where strong distributional assumptions such as independence or identical distributions are inappropriate and should therefore be avoided. Strong distributional assumptions should also be avoided in the study of large games where correlation due to sunspots, correlated types, or correlated randomization devices is quite natural. Finally, a general treatment which imposes minimal restrictions on $f$ is important when $f$ is itself endogenous. This is the case, for example, when $f$ represents an endogenous matching mechanism or the mixed actions chosen by players in a large game.

Second, the paper provides a framework to study the distinction between idiosyncratic and aggregate risks when there is a continuum of random variables. The distinction between these two types of risk in the finite case plays a key role in parts of Game Theory, Finance, and the Economics of Information. In contrast with aggregate risks, idiosyncratic risks are, in principle, insurable in insurance models and diversifiable in portfolio models. Common and individual-specific uncertainty have fundamentally different implications on performance evaluation in agency models, while independent and correlated randomizations lead to very different solution concepts for noncooperative games. Since continuum models often provide a natural framework to examine these issues, defining and characterizing the decomposition of risk in continuum processes is of basic theoretical and applied significance.

Third, the results of the paper provide a clearer link between the continuum model and its finite approximations. This link is important for practical reasons because it provides a simple method for inferring the properties of the limiting continuum process using the information contained in large samples and simple processes. On a more fundamental level, the link with finite approximations is an essential ingredient of an economically motivated interpretation of the assumptions and implications of the continuum model.

The paper is also related to the literature on the law of large numbers for a continuum of i.i.d. random variables, which emphasized what became known as the “measurability problem” of sample realizations (Judd (1985) and Feldman and Gilles (1985)). In this paper, I neither solve this measurability problem nor provide a law of large numbers of continuum models. Instead, my focus will be on the decision problems of individuals facing the risk generated by a continuum of random variables. This approach is conceptually simpler and has the advantage of completely bypassing the issue of the measurability of sample

\footnote{It is worth noting that Proposition 5 contains, as a by-product, a statement that has the flavor of a law of large numbers, namely that the representation of total risk of the continuum process is independent of the idiosyncratic component. It would be wrong, however, to interpret this as a law of large numbers for continuum processes. To begin with, it is not even clear what formal meaning one should give to such law. An important fact to keep in mind is that the present paper views the continuum as a limit of all its approximations by large finite collections of random variables. The elimination of idiosyncratic risk should be interpreted as a law of large numbers for these finite collections rather than for the continuum itself.}
realizations. Essentially, I will argue from first principles that the measurability problem is not important when decision makers regard the continuum as an idealization of large finite models, and when they evaluate risk in an anonymous and state-independent manner.\footnote{Schmeidler (1973, Theorem 2), Mas-Colell (1984), and Green (1984) are examples where anonymity and state-independence are imposed in large games. These assumptions are also standard in many asset pricing models in Finance.}

At a mathematical level, the treatment of this paper was motivated by an observation made by Uhlig (1987) that the i.i.d. process can be viewed as an $L_2$-valued function $t \mapsto f_t$, which can be directly Pettis-integrated. This approach did not yield useful results with clear economic interpretation. One problem was the lack of a good understanding of the relationship between continuum models and their finite approximations. Another problem was the difficulty to justify the claim that Pettis integration somehow delivered a "law of large numbers for the continuum." On the other hand, Uhlig's work pointed to the role of weak measurability in the study of continuum processes involving idiosyncratic risk and suggested that vector-valued integration techniques can be useful in dealing with i.i.d. processes in a way that avoids the difficulty of integrating them "state-by-state." More directly, however, the mathematical treatment of this paper builds on two basic structures. First, the paper extensively uses the geometry of Hilbert spaces (e.g., orthogonal projections and the Radon-Nikodým property). Second, the analysis takes advantage of important recent developments in the theory of integrating nonmeasurable vector-valued functions and its connections with Probability Theory. This was pioneered by Geitz (1981), Talagrand (1984), and Hoffman-Jorgensen (1985), and further developed by Talagrand (1987), and Dobric (1990). It is the interplay between these two structures (rather than each one separately) which makes it possible to carry out the analysis reported here.

2. THE MODEL, MOTIVATION AND EXAMPLES

2.1. Basic Definitions

Let $(\Omega, \Sigma, P)$ be a probability space on which all random variables are defined. The inner product of two random variables $f$ and $g$ in $L_2$ is

$$(f | g) = \int_{\Omega} f(\omega)g(\omega) \, dP,$$

and the $L_2$-norm of $f$ is $\|f\| = (f | f)^{1/2}$. Two random variables $f$ and $g$ are orthogonal if $(f | g) = 0$. From elementary probability, we have $(f | g) = \text{cov}(f, g) + (Ef Eg)$, where $Ef$ denotes the expected value of $f$, and cov denotes the covariance. Let $L_2$ be the space of (equivalent classes of) square-integrable
real-valued random variables on $(\Omega, \Sigma, P)$ with the $L_2$-norm. As usual, two random variables that differ on a set of measure zero have zero $L_2$-distance and will therefore be identified. Abusing notation, I will often use real numbers to denote the corresponding constant random variables. For example “1” may be used to denote the random variable $f(\omega) = 1$, for all $\omega \in \Omega$. The context will always make it clear whether a symbol $\alpha$ is used to refer to the real number $\alpha \in \mathbb{R}$ or to the corresponding constant random variable $\alpha \in L_2$.

Uncertainty is modeled in terms of a collection of random variables indexed by the measure space $(T, \mathcal{F}, \tau)$. Here, $T = [0, 1]$, $\mathcal{F}$ is the set of Lebesgue measurable sets, and $\tau$ is a probability measure which is absolutely continuous with respect to the Lebesgue measure. Formally, this collection of random variables forms a process which can be compactly represented as a function $f : T \rightarrow L_2$.

The process $f$ is bounded if there is $M < \infty$ such that $\|f(t)\| \leq M$ for all $t \in T$. Note that the boundedness of the process $f$ does not require the random variables $f_t$ to have bounded range. The process $f$ is weakly measurable if each member of the family $\{t \mapsto (g|f_t) : g \in L_2\}$ of real-valued functions is measurable. It is easy to see that a process $f$ is weakly measurable if and only if the expected value function $t \mapsto Ef_t$ and each covariance function $t \mapsto \text{cov}(g, f_t)$ is measurable. This observation is convenient and will be repeatedly used in the sequel. In the remainder of the paper, the symbol $\mathcal{F}$ will denote the set of all bounded, weakly measurable processes.

### 2.2. A First Set of Examples

The simplest and best known example of a process $f \in \mathcal{F}$ is the i.i.d. process. In such a process, the random variables $f_t$ are identically distributed and each finite subset $\{f_{t_1}, \ldots, f_{t_n}\}$ is independent. i.i.d. processes are frequently used in the literature to represent exogenous uncertainty. For example, $T$ may represent a set of households and randomness represents independent and identically distributed shocks to their endowments, as in Green (1987). Note that the i.i.d. process is bounded by definition. On the other hand, for a fixed individual $s$, the covariance function $t \mapsto \text{cov}(f_t, f_t)$ represents the way economy-wide shocks are

---

4 The use of $[0, 1]$ as an index set is arbitrary. While the results of the paper can be considerably generalized to other types of index sets, I will use $[0, 1]$ for concreteness and because this is the space found in most current economic applications.

5 Most of the analysis in this paper is valid if the boundedness of $f$ is relaxed to the requirement that $\|f_t\|$ is dominated, $\tau$-a.e., by a function $\psi : T \rightarrow \mathbb{R}$ with $\int_T \psi^2 \, d\tau < \infty$.

6 There appears to be some confusion concerning the existence of a probability space on which an i.i.d. process can be defined. A basic result, known as the Kolmogorov Extension Theorem (e.g., Ash (1972, Theorem 4.4.3, p. 191) or any other advanced textbook in Probability), establishes a general condition which ensures the existence of a probability space on which processes with prespecified finite dimensional distributions can be constructed. Judd (1985) showed how this theorem can be used to construct a probability space on which an i.i.d. process can be defined. Such a probability space is not "nice," however: Uhlig (1987, Proposition 1) observed that such a space cannot be a separable metric space with the Borel $\sigma$-algebra.
correlated with the shock received by household \( s \). Independence implies that this function is zero except at \( t = s \), and is therefore measurable.\(^7\)

The source of uncertainty need not be exogenous, as in the last example. To see this, suppose that \( T \) is a set of players; that \( \tau \) is the Lebesgue measure; and that each individual has to choose among two actions \( \{a_0, a_1\} \). In this case, a random variable \( f_t \), taking values in the set \( \{0, 1\} \) can be interpreted as a mixed strategy representing player \( t \)’s randomization over the actions \( a_0 \) and \( a_1 \). Section 7 elaborates on the use of processes to model strategies and information structures in games with a measure space of players.

2.3. Idiosyncratic Risk

The distinction between idiosyncratic and aggregate risks is of fundamental importance in applications which range from contracting in insurance markets, to relative performance evaluation in agency models and asset pricing in financial markets, to name only a few such examples. One goal of this paper is to extend the distinction between idiosyncratic and aggregate risks to environments where risks are generated by a continuum of random variables and to investigate the conditions under which total risk can be decomposed into aggregate and idiosyncratic components. I begin with the following definition:

**Definition:** A bounded process \( h: T \to L_2 \) is purely idiosyncratic if for every \( g \in L_2 \), \((g| h_t) = 0, \tau \text{-a.e.}\)

Note that if we chose \( g = 1 \), then \( E h_t = 0, \tau \text{-a.e.} \). On the other hand, choosing \( g = h_t \) for a fixed \( s \) implies that \( \text{cov}(h_s, h_t) = 0, \tau \text{-a.e.} \). Thus, in an idiosyncratic process, risk is a mean-zero disturbance which is individual-specific in the sense that individual \( t \)’s shock is uncorrelated with the shocks of almost every other individual.

An i.i.d. process with zero mean is an example of an idiosyncratic process. However, the class of idiosyncratic processes is much broader. A particularly revealing example is the following: Let \( f \) be an i.i.d. process, and let \( A \subset T \) be a nonmeasurable subset. Define \( f' \) by \( f'_t = f_t \), \( t \in A \) and \( f'_t = 0, t \notin A \). For any fixed \( s \), the covariance function \( \text{cov}(f'_s, f'_t) \) is 0, except possibly when \( t = s \), and the process is indeed idiosyncratic. Viewed in terms of the \( L_2 \)-norm, \( f' \) is quite wild because \( t \mapsto \text{var}(f'_s) \) is the indicator function of a nonmeasurable subset of \([0, 1]\) and is therefore nonmeasurable. On the other hand, when viewed in terms of the functions \( t \mapsto (g| f'_t) \), \( f' \) is extremely well-behaved because any such function is zero almost everywhere. The observation that a process like \( f' \) may

---

\(^7\)The proof that the i.i.d. process is weakly measurable is not totally trivial and might be instructive: Fix an arbitrary \( g \in L_2 \), and let \( \tilde{L} = \text{span}(f_t: t \in T) \) be the closed linear subspace spanned by the \( f_t \)’s. Take the orthogonal projection \( \tilde{g} \) of \( g \) on \( \tilde{L} \), and note that \( (g| f_t) = (\tilde{g}| f_t) \) for all \( t \). Thus, we only need to show that \( t \mapsto (g| f_t) \) is measurable for every \( g \in L \). This follows from the key fact that for any \( g \in L \) there is a countable set of indices \( T' \subset T \) such that \( g = \sum_{t \in T'} \beta_t f_t \) (see Section A.1), so \( (g| f_t) = 0 \) except when \( t \) belongs to the countable set \( T' \).
be easier to analyze in terms of the family of functions \( \{ t \mapsto (g|f_i): g \in L_2 \} \) is what motivates the approach, followed in this paper, to deal with such problemmatic processes.\(^8\)

### 2.4. Aggregate Risk

In many economic applications, aggregate risk is introduced by assuming that the random variables \( f_i \) share a common random component \( \eta \). Aggregate risks generated by a single factor can be introduced by assuming that \( f \) can be written in the form \( f_i = \beta_i \eta + h_i \), where \( h \) is an idiosyncratic process; \( \eta \in L_2 \) is a random variable representing systematic or aggregate risk; and \( \beta_i \) is a measurable function reflecting the exposure of \( f_i \) to the aggregate risk factor \( \eta \). This formulation of aggregate risk is quite restrictive because there is in principle no reason to think that aggregate risk can be captured by a single factor. For example, \( K \)-factor models are quite common in Finance where aggregate risk is represented using a finite number \( K \geq 1 \) of fundamental factors, each reflecting a different aspect of the aggregate uncertainty in the economy.

To obtain a complete characterization of risk, restricting attention to a finite number of factors will turn out to be a nuisance. For this reason, I allow for the possibility that aggregate risk is countably generated. I will follow the practice, common in factor models, of assuming that the factors are pairwise orthogonal and normalized to have \( L_2 \)-norm of 1 (i.e., they form an orthonormal system). Formally we have the following definition:

**Definition:** A function \( g: T \to L_2 \) is an **aggregate process** if there is a countable orthonormal set of random variables \( \{ \eta_1, \eta_2, \ldots \} \) and measurable functions \( \beta_k: [0, 1] \to \mathbb{R} \) such that

\[
g_t = \sum_{k=1}^{\infty} \beta_{k,t} \eta_k, \text{ \tau-a.e.}
\]

I will refer to the \( \eta_k \)'s as the fundamental or aggregate factors. It is worth noting that the particular set of factors \( \{ \eta_1, \eta_2, \ldots \} \) is an arbitrarily chosen basis for the subspace spanned by the range of \( g \), and is therefore not unique. Proposition B.1 provides a coordinate-free characterization of aggregate processes. It is also important to note that an aggregate process necessarily involves a substantial correlation. Proposition B.1 makes this precise by showing that an aggregate process is characterized by the condition that: Off a set of small measure, the population \( T \) can be divided into a finite number of subpopulations such that the risk of the members in any given subpopulation is highly correlated.

\(^8\)Talagrand (1984) provides an authoritative account of this theory.
An important special case is that of a finitely generated process which occurs when the definition of an aggregate process can be satisfied with a finite set of factors \( \{\eta_1, \ldots, \eta_k\} \). One question of practical interest is whether an arbitrary aggregate process can be approximated by a finitely generated one. The choice of a criterion to evaluate the quality of such approximation involves subtle trade-offs which depend on the nature of the intended economic application. The following result has the advantage of being easy to state and prove.\(^9\)

**Proposition 1:** Given \( \varepsilon > 0 \), any bounded aggregate process \( g \) can be written, \( \tau \)-a.e., as the sum of a finitely generated process \( g^\varepsilon \) and a residual process \( w_t = g_t - g_t^\varepsilon \) satisfying \( \int_T \| w_t \| d\tau < \varepsilon \).

Roughly, the norm of the difference between \( g \) and a finitely generated process can be made uniformly small over a subset \( A \subset T \) of large measure. The approximation offered in Proposition 1 is somewhat crude because \( g^\varepsilon \) may be inefficient in the sense of using an unnecessarily large number of factors. See Al-Najjar (1994) for a more parsimonious approximation.

### 2.5. Decomposition of Risk

**Definition:** A process \( f : T \to L_2 \) is decomposable (with decomposition \( (g, h) \)) if there is an aggregate process \( g \) and an idiosyncratic process \( h \) such that \( f_t = g_t + h_t \), \( \tau \)-a.e.

Intuitively, a process \( f \) is decomposable if, whenever the dependence of a typical pair \( f_t \) and \( f_s \) on the fundamental factors \( \{\eta_1, \eta_2, \ldots\} \) is removed, the residual variations \( h_t \) and \( h_s \) are uncorrelated both with each other and with all of the aggregate factors. This definition of decomposition has a number of attractive features. First, the definition clearly covers the one-factor and \( K \)-factor examples mentioned earlier. Second, the definition also makes sense when there are only finitely many random variables, so finite and continuum models can be directly compared. Third, the definition is tight in the following sense:

**Proposition 2:** If the process \( f \) is decomposable, then the decomposition is essentially unique. That is, if \( (g, h) \) and \( (g', h') \) are two decompositions, then \( g_t = g'_t \) and \( h_t = h'_t \), \( \tau \)-a.e.

### 2.6. Approximation by Simple Processes

The theory of risk presented in this paper needs to be justified on two grounds. First, the theory should ideally provide a useful and tractable framework to study continuum processes. Second, the various definitions and results

\(^9\)All the proofs are found in Appendix B.
DECOMPOSITION AND CHARACTERIZATION OF RISK

derived must be founded on a clear understanding of what the continuum model is a model. To achieve the second goal, I examine two methods of approximating the continuum model by finite collections of random variables. Since we know exactly how to represent risk in finite models (see Section 5.1), the approximation results in Sections 3 and 5 constitute the main economic justification of the theory.

I begin with the notion of approximation by simple processes which plays an important theoretical role in General Equilibrium analysis and Game Theory. Formally, a process \( f \) is simple if there exists a finite partition of \( T \) by measurable sets \( T_1, \ldots, T_M \) such that \( f \) is constant on each of the sets \( T_m \), \( m = 1, \ldots, M \). That is, individuals in each subpopulation have exactly the same risk. Let \( \mathcal{F}_{\text{simple}} \) denote the set of simple processes. Clearly \( \mathcal{F}_{\text{simple}} \subset \mathcal{F} \). Note that a simple process contains no idiosyncratic component since, by definition, each subpopulation receives a perfectly correlated shock.

2.7. Approximation by Large Samples

Another method to model situations with a large number of random variable is in terms of infinite sequences. An early example is Malinvaud's (1972) model of insurance in a large economy. More recent examples can be found in Finance (see Section 2.8 below). The main advantage of the sequence approach is its intuitive appeal and mathematical simplicity. For example, the standard strong law of large numbers of sequences of random variables applies without the measure theoretic complications which arise in continuum models. On the other hand, the continuum model is often more appropriate because it better captures situations in which individuals are exactly negligible and because it allows for a symmetric treatment which does not depend on an artificial arrangement of the random variables.

Sequence models can be given an interesting interpretation within the current model by imagining that an infinite sequence of random variables \( f_1, f_2, \ldots \) is obtained as the result of a random draw of a sample of indices \( (t_1, t_2, \ldots) \) from the underlying population \( T \). Formally, suppose that initially a point \( t_1 \) is randomly selected from \( T \) according to the probability distribution \( \tau \). A second draw is made independently, again using \( \tau \), to produce a sequence \( t_2 = (t_1, t_2) \). More generally, let \( \tau^n \) denote the product measure on \( T^n \), the set of all infinite samples \( t^n \) drawn from \( T \). Identify the set of samples of size \( n \), \( T^n \), with a sub-\( \sigma \)-algebra on \( T^n \) by identifying \( t^n = (t_1, \ldots, t_n) \) with the set of infinite sequences in \( T^n \) which agree with \( t^n \) on the first \( n \) components. With this notation, \( \tau^n(t^n) \) represents the probability of picking the finite sample \( t^n = (t_1, \ldots, t_n) \). Note that the total risk in a sample \( t^n \) is represented by the random variable \( f(t^n) = (1/n) \sum_{i=1}^{n} f_i \).

The interpretation of sequence models as random samples drawn from an underlying continuum model can be potentially fruitful. For example, such
interpretation makes it possible to think of $T$ as a sample space from which an outside observer (e.g., an econometrician) draws finite samples $t^n = (t_1, \ldots, t_n)$. A natural question is whether this observer can use randomly drawn samples to infer something about the economically relevant properties of the underlying continuum model. The paper’s Main Theorem provides an answer to this question by characterizing the class of processes whose representations of risk can be estimated from the corresponding representations of large finite samples in a statistically consistent manner.

2.8. Examples from Portfolio Theory

As a final set of examples, let $T$ denote the set of assets in a large economy; $\tau$ denotes the supply of assets; and $f_t$ represents the random rate of return on asset $t$. A portfolio is a signed measure $\nu$ on $T$ such that $\nu(T) = 1$. For a (measurable) subset of assets $A \subset T$, the number $\nu(A)$ represents the proportion of the total value of the portfolio invested in the set of assets $A$ (i.e., the portfolio weight of $A$). Allowing $\nu$ to be a signed measure is necessary to cover the case of short selling assets: The set $A$ is short sold if it is a negative set relative to $\nu$.\footnote{A set $A$ is negative if $\nu(A') \leq 0$ for every measurable subset $A' \subset A$. Positive sets are defined similarly.}

It is easy to see that any decomposable process $f$ can be written in the form $f_t = Ef_t + \sum_k \beta_k t \eta_t + h_t$, $\tau$-a.e. In the language of the asset pricing literature, the $\eta_t$’s represent a (finite or infinite) set of aggregate factors; the $\beta_t$’s are the factor loadings of asset $t$; $\sum_k \beta_t \eta_t$ is a random deviation around asset $t$’s expected return representing factor risk; and $h_t$ is an additional asset-specific disturbance. If factor risk is finitely generated, then we obtain a continuum version of the definition of strict factor structure for asset returns familiar from the Finance literature. Even when $f$ is decomposable but its factor risk is not finitely generated, Proposition 1 guarantees that it is always possible to find a nearby process whose factor risk is finitely generated.

A basic feature of many asset pricing models is the asymmetric treatment of factor risk $\sum_k \beta_t \eta_k$ and asset-specific risk $h_t$. A main idea in Arbitrage Pricing Theory (Ross (1976)) is that the idiosyncratic component $h_t$ is diversifiable in large markets, and so will earn no risk premium. Ross’ modeling of a large asset market as an infinite sequence of assets is an example of the sequence approach discussed in Section 2.7. The framework of this paper suggests that one may be able to interpret an infinite sequence of assets as a random sample drawn from a large underlying population of assets. This interpretation emphasizes the role of the statistical properties of the samples (as opposed to the properties of one particular draw) in making inference about the asset pricing relationships in the underlying continuum model. This view is elaborated in Al-Naijjar (1994).
3. THE MAIN THEOREM

I now state the main result of the paper:

**Main Theorem:** For any bounded process \( f \), the following statements are equivalent:

(i) Weak Measurability: The process \( f \) is weakly measurable.

(ii) Decomposition of Risk: The process \( f \) has an essentially unique decomposition into an aggregate process \( g \), and an idiosyncratic process \( h \).

(iii) Approximation by Simple Processes: There is a sequence of simple processes \( \{f^n\} \) such that for any random variable \( g \), \((g|f^n)| \to (g|f), \tau^\times-a.e. \)

(iv) Approximation by Large Samples: There exists a random variable \( f \) such that

\[
\lim_{n \to \infty} \| f(t^n) - f \| \to 0, \quad \tau^\times-a.s.
\]

If (iv) holds, then the random variable \( f \) is equal to the Bochner integral \( \int_\tau^\times g \, d\tau \) of the aggregate component \( g \) of \( f \). In particular, \( f \) is independent of the idiosyncratic component \( h \).

The force of the Theorem lies in establishing the equivalence of four seemingly unrelated aspects of the process \( f \). As indicated earlier, weak measurability is indeed a very weak requirement. The class of weakly measurable processes is broad and includes, for example, all idiosyncratic processes (hence all i.i.d. processes). In a narrow technical sense, it is unlikely that the requirement of weak measurability can be further relaxed. Thus, the most immediately visible implication of the Theorem is the fact that it covers as broad a class of processes as one can reasonably expect.

But the Theorem has another more subtle implication. If one thinks of continuum models as idealizations of large but finite collections of random variables, then parts (iii) and (iv) provide evidence that there is no economically well-motivated reason to look beyond the class of weakly measurable processes. Stated differently, the Theorem suggests that any economically sensible continuum process must be weakly measurable.

Part (iv) and the statement at the end of the Theorem imply that sample averages converge in the \( L_2 \)-norm to a limiting value which is independent of the idiosyncratic component of the process. Thus, idiosyncratic risk is filtered out and the only remaining risk is aggregate risk. It is important to emphasize that the random variable \( f \) whose existence is asserted in part (iv) is the same for all samples, \( \tau^\times \)-almost surely. The Theorem therefore identifies weakly measurable processes as the class for which typical large samples can be used to obtain consistent estimates of total risk of the underlying continuum model.

The decomposition part of the Theorem points to the remarkably simple structure and tractability of weakly measurable processes: Any such process is (essentially uniquely) obtained by combining an idiosyncratic process and an aggregate process. Proposition 5 below will show that aggregate risk in the
continuum process can be computed from finite approximations in a consistent and a straightforward manner. This and the other remarks made earlier suggest that the class of weakly measurable processes is the appropriate class to study continuum models.

Some readers might question the practical usefulness of the decomposition part of the Theorem since the aggregate component $g$ in that decomposition may depend on an infinite number of factors. For example, factor models in Finance are useful only to the extent that aggregate risk can be captured using a finite (and preferably “small”) number of factors. In response, it should be pointed out that the Theorem delivers an exact decomposition and factor structure for $f$ in the sense that: (i) the decomposition holds almost everywhere on $[0,1]$; and (ii) for a typical pair $t$ and $s$, the residual disturbances $h_t$ and $h_s$ are exactly orthogonal. A finite factor representation is possible if these two requirements are slightly relaxed. Specifically, combining the Theorem with Proposition 1 yields the result that any weakly measurable process can be written as the sum of a finitely generated process plus a residual $h^*_t = w_t + h_t$, such that $h$ is idiosyncratic, and $\int_T ||w_t|| d\tau < \epsilon$. While $h^*$ is not exactly idiosyncratic, $h^*_t$ and $h^*_s$ will nevertheless be nearly uncorrelated for most choices of pairs of indices $t$ and $s$.

4. CONVERGENCE CONCEPTS FOR CONTINUUM PROCESSES

This section discusses various convergence concepts of continuum processes which play an important role in interpreting the continuum model as an idealization of finite approximations. Readers not interested in the details can go directly to Section 5, provided they interpret the statement “$f^m \to f$” to mean “$f^m$ approaches $f$ in a reasonable way.”

The convergence concept in part (iii) of the Theorem is due to Geitz (1981) and will be referred to as $G$-convergence (denoted $f^m \overset{G}{\to} f$). Roughly, $f$ and $f'$ are $G$-close if their expected value functions and each of their covariance functions are close. A stronger convergence concept is that of norm convergence: A sequence $\{f^m\}$ converges to $f$ in norm (denoted $f^m \overset{\|\cdot\|}{\to} f$) if $\|f_t - f^m_t\| \to 0$ for $\tau$-a.e. $t$. Clearly, norm convergence implies $G$-convergence. Consequently, the various parts of the Theorem hold a fortiori for any process $f$ which is the norm limit of a sequence of simple processes.

Norm convergence is of limited interest in the study of problems with idiosyncratic risk, however. This was first observed by Uhlig (1987) who noted that the i.i.d. process is not the norm limit of simple processes and that, consequently, this process is not measurable.\textsuperscript{11} It is not too difficult to see that $G$-convergence in the equivalence of parts (i) and (iii) in the Theorem can be replaced by norm convergence if and only if weak measurability is strengthened to measurability. This, in turn, is possible if and only if the idiosyncratic

\textsuperscript{11}Both facts can be viewed as direct consequences of the Pettis Measurability Theorem and that the i.i.d. process is essentially nonseparably valued.
component of the process is trivial. The important work of Geitz (1981) explained the subtle implications of the difference between the two convergence concepts on the measurability of continuum processes.\textsuperscript{12}

While norm convergence is too strong for many applications of economic interest, $G$-convergence, on the other hand, is too weak. Ideally, a convergence concept should satisfy two criteria. First, convergence should be weak enough that the set $\mathcal{F}$ of bounded, weakly measurable processes can be obtained as the closure of the set $\mathcal{F}_{\text{simple}}$ of simple processes. Second, convergence should be strong enough to guarantee that the economically relevant properties of a continuum process change "continuously" as the process itself changes. Norm convergence fails to satisfy the first criterion, while $G$-convergence fails to satisfy the second. To see the latter, consider the following example: Let \( \{\eta_1, \eta_2, \ldots\} \) be a sequence of independent and normally distributed random variables with mean zero and common variance $\sigma^2 > 0$. Define the constant processes $f^m = \eta_m$ and $f_t = 0$ for all $t$. From an economic point of view, the sequence \( \{f^m\} \) is not a good approximation of $f$ because randomness does not disappear along \( \{f^m\} \). Despite this, the sequence \( \{f^m\} \) $G$-converges to $f$.$^\text{13}$

Roughly, the problem is that $G$-convergence does not put uniform bounds on the rate at which the various covariance functions converge to their limits. To circumvent this problem, $G$-convergence must be strengthened in such a way as to guarantee the convergence of the economically relevant aspects of continuum processes. There are many ways to do this. The approach chosen here has the advantages of being direct and simple to state:

**DEFINITION:** Let $f$ and $f^m$, $m = 1, 2, \ldots$, be in $\mathcal{F}$, and let $g$ and $g^m$ be the corresponding aggregate components. Then \( \{f^m\} \) converges to $f \in \mathcal{F}$ (denoted $f^m \rightarrow f$) if $f^m \overset{G}{\rightarrow} f$ and if for every measurable $A \subset T$, $\| \int_A g^m \, d\tau - \int_A g \, d\tau \| \rightarrow 0$.\textsuperscript{14}

Note that this definition makes sense because the Theorem guarantees a unique decomposition for each process in $\mathcal{F}$. Note also that this convergence

---

\textsuperscript{12}The convergence of the sequence of real-valued functions $(g|f^m)$ to $(g|f)$ required in the definition of $f^m \overset{G}{\rightarrow} f$ can fail, for a given $g$, on a $\tau$-null $A_\kappa \subset T$. Geitz (1981, Theorem 7) observed that a bounded process $f$ is Bochner integrable, and hence measurable, if and only if there is a sequence of simple processes $f^m \overset{G}{\rightarrow} f$ such that the exceptional sets $A_\kappa$ can be chosen independently on $g$. Thus, weakly measurable processes which fail to be measurable are precisely those obtained as the $G$-limit of simple processes, but where this limit can be achieved only with mobile exceptional sets $A_\kappa$. This technical observation is crucial for the study of risk because the Theorem and Proposition B.1 imply that the difference between measurability and weak measurability is the existence of nontrivial idiosyncratic risk.

\textsuperscript{13}For any fixed $n$, the function $\text{cov}(\eta_n, f^m_\kappa) = 0$ everywhere for all large $m$. For a general $\eta = \sum_{n=1}^{\infty} \beta_n \eta_\kappa$, the continuity of the inner product implies that $\text{cov}(\eta, f^m) = \sum \beta_n \text{cov}(\eta_\kappa, \eta_\kappa) = \beta_m$, a term which converges to zero.

\textsuperscript{14}Stated differently, the sequence of vector measures $\nu^m$ defined by the indefinite Bochner integral of the aggregate components $\nu^m(A) = \int_A g^m_\kappa \, d\tau$ converges to the vector measure $\nu(A) = \int_A g_\kappa \, d\tau$ setwise in $L_2$-norm. On the other hand, $G$-convergence ensures only that $\nu^m \rightarrow \nu$ setwise weakly (Geitz (1981, Corollary 4, p. 84)).
concept is strong enough to rule out the example presented earlier because
\[
\int_{\mathbb{D}}^{B} g_i \, d\tau = \eta_m \text{ does not converge to } \int_{\mathbb{D}}^{B} g_i \, d\tau = 0 \text{ in norm.}
\]

In some applications (e.g., those in Section 5.4 and 6), a slight strengthening
of the convergence concept is needed. We say that \( \{f^n\} \) converges properly to \( f \)
(written \( f^n \xrightarrow{\text{prop}} f \)) if \( f^n \to f \) and the sequence of random variables \( \{\int_{\mathbb{D}}^{B} g_i^n \, d\tau\} \)
converge to the random variable \( \int_{\mathbb{D}}^{B} g_i \, d\tau \) for P-a.e. \( \omega \) (rather than just in
\( L_2 \)-norm). The next proposition shows that the two convergence concepts
introduced here are not too strong:

**Proposition 3:** For any process \( f \in \mathcal{F} \) there is a sequence \( \{f^n\} \) of simple
processes such that \( f^n \xrightarrow{\text{prop}} f \) (hence \( f^n \to f \)).

5. ANONYMOUS AND STATE-INDEPENDENT DECISION MAKING

This section uses the Theorem to study the continuity of representations of
risk and its implications for the class of anonymous and state-independent
decision problems.

5.1. Anonymous and State-Independent Representations of Risk

Let \( \mathcal{M} \) denote the set of probability distributions on \( \mathbb{R} \) endowed with the
topology of weak* convergence. To each simple process \( f \), define the random
variable \( f(\omega) = \sum_{m=1}^{M} \tau(A_m) f_m(\omega) \), where \( f_1, \ldots, f_M \) are the constant values
which \( f \) assumes on the measurable sets \( A_1, \ldots, A_M \). Define the anonymous
and state-independent representation of risk in a simple process \( f \) to be the
probability distribution \( \mu(f) \in \mathcal{M} \) of the random variable \( \sum \tau(A_m) f_m \). The
interpretation is straightforward: A state-independent representation of risk
should depend on the distribution over the outcomes in \( \mathbb{R} \), but should be
otherwise independent of the underlying state of nature \( \omega \). An anonymous
representation, on the other hand, depends on the sets \( A_1, \ldots, A_M \) only through
the weights \( \tau(A_m) \) and the corresponding components \( f_m \) but is otherwise
independent of the identity of these sets.

We can define a similar notion for a finite sequence \( t^n = (t_1, \ldots, t_n) \): The
anonymous and state-independent representation of risk for \( t^n \) is defined to be
the probability distribution \( \mu(t^n) \) determined by the random variable
\( (1/n) \sum_{i=1}^{n} t_i f_i \).

It is important to note that the definitions of \( \mu(t^n) \) and \( \mu(f) \) (for a simple
process \( f \)) given above involve only finite collections of random variables. There
is therefore complete agreement about their meaning as representations of risk.

---

15 Note that this is the weighted average of the random variables themselves, rather than the
average of their distributions. This has important implications for the analysis of large games
(Section 6). See also Menahem Yaari's 1987 paper for other subtle decision-theoretic implications of
this distinction.
5.2. Continuity of Representations

The definition of anonymous and state-independent representation given earlier cannot be directly extended to a general process \( f \) because the realizations \( t \mapsto f_t(\omega) \) may be nonmeasurable.\(^\text{16}\) An alternative approach of extending the definition of anonymous and state-independent representations to all of \( \mathcal{F} \) is as follows. If the continuum model is viewed as an idealization of large finite models, then the representation \( \mu(f) \) must be close to the representations of all finite approximations of \( f \). This simple requirement on \( \mu(f) \) yields the following continuity theorem:

**Proposition 4:** For every process \( f \in \mathcal{F} \) with decomposition \( (g, h) \), there exists a probability distribution \( \mu(f) \in \mathcal{M} \) such that

(i) \( \mu(t^n) \to \mu(f) \) as \( n \to \infty \), \( \tau^* \)-a.s.; and

(ii) \( \mu(f^m) \to \mu(f) \) for any sequence of simple processes such that \( f^m \to f \).

The distribution \( \mu(f) \) is unique and coincides with the distribution of the Bochner integral \( \int g \, d\tau \) of the aggregate component \( g \) of \( f \). In particular, \( \mu(f) \) does not depend on the idiosyncratic component \( h \). Furthermore, the mapping \( f \mapsto \mu(f) \) is continuous in the sense that \( f^m \to f \) implies \( \mu(f^m) \to \mu(f) \) for any sequence \( \{f^m\} \) in \( \mathcal{F} \).

Proposition 4 shows that natural continuity and consistency requirements imply that there is a unique way to define \( \mu(f) \) for general processes. It is important to note that \( \mu(f) \) is defined without any reference to the problematic sample realizations \( t \mapsto f_t(\omega) \) which may be nonmeasurable. In addition, \( \mu(f) \) can be consistently calculated from finite approximations of \( f \) because it is independent of the particular approximation chosen. The proposition also justifies describing \( \mu(f) \) as the anonymous and state-independent representation of risk because \( \mu(f) \) is state-independent by construction, and anonymous because it depends on \( f \) only through the integral \( \int g \, d\tau \).

An easy but useful extension of the result is to vector-value processes. These are processes where each \( f_t \) is a random vector taking values in \( \mathbb{R}^N \). The extension is immediate because any such process can be modeled as an \( N \)-vector of ordinary processes \( (f_{1t}, \ldots, f_{Nt}) \), so for each state \( \omega \) the realization of the process at \( t \) is the vector \( (f_{1t}(\omega), \ldots, f_{Nt}(\omega)) \in \mathbb{R}^N \). The concepts of weak measurability, decomposition, and convergence can be defined directly, component by component. For example, \( (f_{1t}, \ldots, f_{Nt}) \) is weakly measurable if each \( f_n \) is, and decomposition is obtained by decomposing each \( f_n \) into \( (g_n, h_n) \). The representation of risk \( \mu(f) \) for a vector-valued process \( f \) is the joint distribution on \( \mathbb{R}^N \) of the random vector \( (\int g_{1t} \, d\tau, \ldots, \int g_{Nt} \, d\tau) \). It is clear that the results of this paper extend, with little or no modifications, to such vector-valued processes.

\(^{16}\) For a simple process \( f \), each realization is a step function (i.e., constant on each set \( A_m \)) since \( f \) is constant on these sets so the weighted sum \( \Sigma \tau(A_m)f_m \) can be directly defined.
5.3. Decision Making and the Extension Theorem

Consider the problem facing a decision maker who must choose an action $a$ in a finite set $A$. This decision maker has a payoff function $U: A \times \mathcal{F}_{\text{simple}} \to \mathbb{R}$ which evaluates risks in a continuous, anonymous, and state-independent manner. By this I mean that there is a continuous function $u: A \times \mathcal{F} \to \mathbb{R}$ such that $U(a, f) = u(a, \mu(f))$ for any $f \in \mathcal{F}_{\text{simple}}$. Since the definition of $U$ is restricted to simple processes, there is no disagreement concerning the meaning of $\mu(f)$ or the interpretation of the statement that “$U$ evaluates risks in an anonymous and state-independent manner.” With these definitions, we have the following Extension Theorem:

**Proposition 5:** There is a unique extension $U^*: A \times \mathcal{F} \to \mathbb{R}$ of $U$ satisfying $U^*(a, f^m) \to U^*(a, f)$ for every $a \in A$ and every sequence $f^m \to f$. Furthermore, this extension satisfies $U^*(a, f) = u(a, \mu(f))$, for all $f \in \mathcal{F}$.

This result, whose proof is quite simple, is nevertheless significant because it enables us to uniquely describe the payoffs of a decision maker over general processes while completely bypassing the difficulties raised by the measurability problem of sample realizations. This construction is possible because a decision maker with anonymous and state-independent payoffs does not care, by definition, about particular sample realizations, but only about aggregate measures of risk. The continuity requirement on the extended payoff function $U^*$ means that the decision maker regards the continuum process as an idealization of simple processes. Proposition 5 says that the combination of the requirements of continuity, anonymity, and state-independence is sufficient to uniquely pin down the decision maker’s payoffs for any process $f \in \mathcal{F}$.

It is important to understand that the results of this section do not constitute a “solution” to the measurability problem. Rather, they identify an important class of environments in which the measurability problem is not a serious obstacle to using models with a continuum of random variables. The measurability problem remains important (but also difficult and, perhaps, intractable)\textsuperscript{17} in models where the decision maker cares about the details of the sample realizations.

5.4. Representation of Risk with Additional Information about the State of Nature

In many applications, a decision maker who faces a process $f$ also has some additional information about the state of nature $\omega$ that might be useful in predicting the distribution $\mu(f)$. For example, in a large economy with aggregate uncertainty and correlated endowment shocks, an individual agent might use the value of his own shock to infer the distribution of shocks in the rest of the economy.

\textsuperscript{17}See, however, Green (1989).
the economy. Similarly, a player in a large game might find it useful to take advantage of his privately observed type or signal to improve his prediction of the aggregate distribution of actions of the other players. In all such examples, the representation \( \mu(f) \) should be modified to reflect the additional information available about \( \omega \).

To introduce this formally, suppose that the additional information about the state of nature is in the form of a sub-\( \sigma \)-algebra \( \mathcal{G} \subset \Sigma \). Using a similar motivation to that in Section 5.1, it is easily seen that the appropriate representation for a simple process \( f \) is the regular conditional distribution of the random variable \( f = \sum \tau(A_m)f_m \) relative to \( \mathcal{G} \).\(^{18}\) This representation will be denoted \( \mu(f, \omega) \) when it is clear from the context, and will be referred to as the anonymous and state-independent representation of risk relative to \( \mathcal{G} \).

This definition can be extended to general processes \( f \in \mathcal{F} \) by letting \( \mu(f, \omega) \) be the regular conditional distribution of the random variable \( \int_B g \, d\tau \). As before, this extension is economically meaningful only to the extent that it can be justified in terms of approximating \( f \) by simple processes.\(^{19}\) That is, we must show that if \( f' \) is a simple process close to \( f \), then the corresponding representations \( \mu(f', \omega) \) and \( \mu(f, \omega) \) are also close. The following proposition makes this statement precise:

**Proposition 6:** For any sequence \( \{f^m\} \) in \( \mathcal{F} \) such that \( f^m \xrightarrow{\text{prop}} f \), the sequence of probability measures \( \{\mu(f^m, \omega)\} \) converges to \( \mu(f, \omega) \) for \( \text{P-a.e.} \) state of nature \( \omega \).

It is clear that the representation \( \mu(f, \omega) \) is essentially unique in the sense that if \( \mu'(\omega) \) is any other collection of measures such that \( \mu(f^m, \omega) \to \mu'(\omega) \) for \( \text{P-a.e.} \) \( \omega \), then \( \mu'(\omega) = \mu(f, \omega) \) for \( \text{P-a.e.} \) \( \omega \). Note also that the stronger requirement of proper convergence is needed because we want the sequences of measures \( \{\mu(f^m, \omega)\} \) to converge for almost every \( \omega \). On the other hand, since any \( f \in \mathcal{F} \) is the proper limit of a sequence of simple processes, Proposition 6 completely and uniquely determines the representations \( \mu(f, \omega) \) for every process \( f \in \mathcal{F} \).

### 6. Applications to Decision Making in Large Games

In 1973, Schmeidler introduced a class of noncooperative games with a continuum of players and established some of their basic properties. Schmeidler's

\(^{18}\)This is a function \( Q \) such that \( Q(\cdot, \omega) \) is a probability measure on \( \mathbb{R} \) for every \( \omega \) and the function \( \omega \to Q(B, \omega) \) is a version of the conditional probability \( P(f(\omega) \in B|\mathcal{G}) \) for every measurable \( B \subset \mathbb{R} \). Such \( Q \) is well-defined only up to sets of measure zero so, strictly speaking, \( Q \) is a "version" of the regular conditional probability. Regular conditional probabilities exist for all real-valued random variables (see Ash (1972, Theorem 6.6.4, p. 265)). Note, finally, that the earlier analysis in this section corresponds to the special case of the trivial \( \sigma \)-algebra which consists of \( \Omega \) and the empty set.

\(^{19}\)Approximation by large samples will not be considered in this subsection.
work generated a great deal of interest because large games provide a natural framework to study situations with anonymous interactions and negligible impact of individual players.

This section provides an alternative model of large games motivated by the work of Aumann (1974, 1987). The model offers a number of advantages, the most important of which is that players' information and randomizations are explicitly described. This section will not attempt to provide a complete analysis of large games, but will focus instead on illustrating how the theory of risk developed in the last few sections can be used to provide a decision-theoretic foundation for their study. Propositions 7 and 8 below describe the class of environments where Schmeidler's representation of the strategic uncertainty facing individual players is valid. The implications of these results on the use and the interpretation of large game models are briefly discussed.

6.1. Information and Strategies

The set of layers is modeled as the measure space \((T, \mathcal{F})\) with the Lebesgue measure \(\tau\). There is a finite set of possible actions \(A = \{a_1, \ldots, a_N\}\), which will be identified with the vertices of the unit simplex \(\Delta\) in \(\mathbb{R}^N\). This is done by identifying action \(a_1\), viewed as a degenerate probability distribution, with the vertex \((1,0,\ldots,0)\) of \(\Delta\); action \(a_2\) with \((0,1,0,\ldots,0)\), and so on. More generally, the simplex \(\Delta\) represents the set of all probability measures on \(A\).

There is a fixed probability space \((\Omega, \Sigma, P)\) on which all uncertainty is defined. Player \(t\) has a sub-\(\sigma\)-algebra \(\mathcal{F}_t \subset \Sigma\) representing his limited knowledge about the true state \(\omega\). I will refer to the continuum of \(\sigma\)-algebras \(G = \{\mathcal{F}_t:\ t \in T\}\) as the information structure of the game. To make the earlier analysis of risk applicable, assume that \(G\) is generated by a fixed process \(f \in \mathcal{F}\) in the sense that \(\mathcal{F}_t\) is the \(\sigma\)-algebra determined by observing the random variable \(f_t\). Without loss of generality, assume that \(Ef_t = 0\) for all \(t\). To eliminate trivial cases assume also that \(\text{var}(f_t) > 0\) for all \(t\). I will say that the information structure \(G\) is independent if any finite collection of \(\mathcal{F}_t\)'s is independent, and idiosyncratic if \(f\) is idiosyncratic.

A strategy for player \(t\) is a choice of an action \(a \in A\) for every state \(\omega\) which respects this player's informational constraint \(\mathcal{F}_t\). Formally, a strategy is an \(\mathbb{R}^N\)-valued random vector \(s_t : \Omega \rightarrow A\) on the probability space \((\Omega, \mathcal{F}_t, \mathbb{P})\). Note that \(s_t\) takes values in the set of actions \(A\), so that every aspect of the behavior

---

20 This assumption is general enough to allow for many situations encountered in applications such as the case when players randomize independently, or where a subset of players make a noisy observation of a sunspot. More generally, any information \(\mathcal{F}_t\) obtained by observing a random signal \(f_t\) with values in a complete separable metric space can be equivalently generated by observing a random variable \(f_t\). This follows from the fact that any complete separable metric space is Borel equivalent to a subset of \(\mathbb{R}\).
of player $t$, including any randomization device he might be using, is explicitly modeled through the strategy $s_t$ and the probability space $(\Omega, \mathcal{F}_t, \mathcal{P})$.

A strategy profile is a weakly measurable process $s$ such that $s_t$ is $\mathcal{F}_t$-measurable for every player $t$. The Main Theorem implies that $s$ can be decomposed into aggregate and idiosyncratic components; Proposition 4 derives the (unique) anonymous and state-independent representation of risk and shows that it depends on the aggregate component of $s$ only. Since $s$ is a vector-valued process (see Section 5.2), this representation is a probability distribution $\mu(s)$ on $\Delta$.

In interpreting $\mu(s)$, one should think of a vector $d = (d_1, \ldots, d_N) \in \Delta$ as representing an aggregate action distribution in the sense that $d_1$ is the fraction of the population taking action $a_1$, $d_2$ is the fraction taking action $a_2$, and so on. Thus, the definition of $\mu(s)$ involves two layers of distributions: First, each point $d \in \Delta$ is a measure on the set of actions $A$ representing the fractions of the population which took various actions; second, $\mu(s)$ is a probability measure on $\Delta$ which represents the way these fractions are distributed. To take a simple example, suppose that all players are able to coordinate their actions on a perfectly observed, payoff-irrelevant event (a sunspot) which occurs with probability $\frac{1}{2}$. Consider the profile $s^a$ in which each player takes action $a_1$ if this payoff-irrelevant event occurs and action $a_2$ otherwise. Then $\mu(s^a)$ is a probability measure on $\Delta$ which assigns probability $\frac{1}{2}$ to the event $d = (1, 0, \ldots, 0)$ that all players take action $a_1$, and probability $\frac{1}{2}$ to the event $d = (0, 1, 0, \ldots, 0)$ that all players take the action $a_2$. This is very different from a profile $s^b$ in which players randomize independently and choose the actions $a_1$ and $a_2$ with equal probability regardless of the state of nature. The two profiles generate different risks because $\mu(s^a)$ involves aggregate uncertainty, while $\mu(s^b)$ puts unit mass on the event $d = (\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$, thus yielding a perfectly deterministic outcome.

6.2. Payoffs

As in Schmeidler (1973), a game is an assignment $t \rightarrow u_t$ of a continuous payoff function $u_t : A \times \Delta \rightarrow \mathbb{R}$ to each player $t$.\textsuperscript{22} Here, $u_t(a, d)$ is the payoff of

\textsuperscript{21}The decision problem of player $t$ can be interpreted in two different ways. In the first interpretation, $(\Omega, \mathcal{F}_t, P)$ is an objective randomization device used by player $t$ to make up his mind about which action to take. This interpretation is in the spirit of Aumann's (1974) definitions of correlated strategies and correlated equilibria. Note that the independent randomization usually assumed in the study of Nash equilibria can be viewed as a special case. A second more subtle interpretation was proposed in Aumann's 1987 paper, which the reader should consult for a fuller and deeper motivation. Here, $(\Omega, \mathcal{F}_t, P)$ represents the subjective beliefs of player $t$ about all the events in the game, including his strategic uncertainty about the behavior of other players. Under this interpretation, each player chooses a definite action so no one actually randomizes. Rather, the strategy $s_t$ represents the uncertainty of other players about the action taken by player $t$. The formal analysis presented here is consistent with either interpretation of $(\Omega, \mathcal{F}_t, P)$.

\textsuperscript{22}For proofs of existence of equilibrium, measurability conditions on the assignment of payoffs $t \rightarrow u_t$ and information $t \rightarrow \mathcal{F}_t$ will be required. This is not needed here, however, because I am only concerned with the representation of uncertainty and the definition of strategies.
player $t$ given the action $a$ and the aggregate action distribution $d = (d_1, \ldots, d_N) \in \Delta$. Anonymity is built into the description of the game because each payoff function $u_t$ depends on the actions of other players only through the aggregate action distribution.

Player $t$ might regard $d$ to be random since he has only limited knowledge about the true state $\omega$, in which case this player will use the available information $\mathcal{F}_t$ to form his belief about the likely values of $d$. Formally, this belief is represented by the regular conditional distribution $\mu_t(s, \omega)$ relative to $\mathcal{F}_t$ defined in Section 5.4. Intuitively, for each state $\omega$, $\mu_t(s, \omega)$ is a probability distribution on $\Delta$ which represents player $t$'s assessment of the way $d$ is distributed given his information $\mathcal{F}_t$. If player $t$ takes action $a$ and faces the profile $s$, his expected payoff is

$$u_t(a, \mu_t(s, \omega)) = \int_{\Delta} u_t(a, d) \, d\mu_t(s, \omega),$$

where the integral sign denotes the Lebesgue integral. Call a strategy $s$, a best response to the profile $s$, written $s_t \in \text{Br}_t(s)$, if $s_t(\omega)$ maximizes $u_t(a, \mu_t(s, \omega))$ for $P$-a.e. state $\omega$.

The model assumes that uncertainty is purely extrinsic in the sense that the state $\omega$ has no direct effect on the “fundamentals” such as payoff functions and action sets. Knowledge about $\omega$ enters in a player’s decision problem by affecting his beliefs about what other players will do. The restriction to extrinsic uncertainty is made only to simplify the exposition and to allow for a direct comparison with Schmeidler’s model. Uncertainty about the fundamentals can be introduced in a variety of ways with minor changes and with little effect on the results. One simple way of introducing such uncertainty is to assume that there is a finite set $W$ of possible types; define the type function for player $t$ to be a $\mathcal{F}_t$-measurable function $w_t: \Omega \to W$; and write the payoff of player $t$ in the form $u_t(a, d, w_t)$.

For example, $w_t(\omega)$ may be the random endowment of player $t$ or his privately known type in a game with incomplete information. An important special case is when $u_t$ is assumed to be independent of $d$, so there is no strategic interaction among the players and uncertainty is purely exogenous.

6.3. Schmeidler’s Model and the Distributional Representation of Strategic Uncertainty

Each strategy $s_t$ uniquely determines a mixed strategy distribution $\bar{s}_t \in \Delta$. Two strategies, $s_t$ and $s_t'$, are distribution-equivalent if they induce the same distribution on actions (i.e., if $\bar{s}_t = \bar{s}_t'$). The relationship between strategies and their distributions is sometimes blurred in the literature because strategies are often

---

23The measurability with respect to $\mathcal{F}_t$ ensures that $w_t$ does not convey more information to player $t$ than what he knew already through $\mathcal{F}_t$.

24Using the current notation, this is the probability distribution $\bar{s}_t = (\bar{s}_t^1, \ldots, \bar{s}_t^N) \in \Delta$, where $\bar{s}_t^a = P(\omega: s_t(\omega) = a)$ is the probability that action $a$ is chosen.
DECOMPOSITION AND CHARACTERIZATION OF RISK

introduced directly as distributions. Here the distinction between the two is crucial: A strategy \( s_t \) is a measurable function while a distribution \( \bar{s}_t \) is a point in \( \Delta \). The strategy \( s_t \) determines not only the distribution \( \bar{s}_t \) on actions, but also describes when (i.e., in which states of nature) a particular action is taken.

In Schmeidler's model, each player \( t \) chooses a mixed strategy distribution \( \bar{s}_t \), and a strategy profile is defined to be a \( \Delta \)-valued measurable function \( t \mapsto \bar{s}_t \). In his main theorem, Schmeidler assumes that the payoff of each player \( t \) is given by \( u_t(a, \bar{s}) \) where \( \bar{s} = \int_T \bar{s}_t \, dt \in \Delta \) is the population average of the players' mixed strategy distributions. Since this formulation is entirely in terms of distributions and makes no reference to the underlying strategies, I will refer to it as the distributional representation of strategic uncertainty facing players in a large game.\(^{25}\)

Interpreting \( \bar{s} \) within Schmeidler's model is difficult because players' strategies are not formally introduced. In our framework, \( \bar{s} \) is obtained by: (i) taking distributions relative to \( \Omega \) to obtain the \( \bar{s}_t \)'s; and (ii) aggregating the \( \bar{s}_t \)'s over the population of players \( T \) to obtain \( \bar{s} \). The theory of Section 5 suggests that the generally correct representation of risk is obtained by reversing the order in which these two operations are performed. That is, one should first aggregate the strategies \( s_t \) over the population, then take the distribution of the aggregate component. If \( \mu_s(s, \omega) \) assigns unit mass to the point \( \bar{s} \) for almost every \( \omega \), then the order of taking distributions and aggregating over the population has no effect on the representation of the strategic uncertainty facing player \( t \). On the other hand, if \( \bar{s} \) and \( \mu_s(s, \omega) \) do not coincide, then \( \bar{s} \) is no longer a correct representation of the strategic uncertainty facing player \( t \). This motivates the following definition:

**Definition:** The distributional representation is valid, for each player \( t \), \( u_t(a, \bar{s}) = u_t(a, \mu_s(s, \omega)) \) for every action \( a \), aggregate profile \( s \), continuous payoff function \( u_t \), and \( P \)-a.e. state of nature \( \omega \).

By focusing on the mixed strategy distributions \( \bar{s}_t \), the distributional representation suppresses the additional information contained in the strategies \( s_t \) concerning when players choose particular actions. For example, the distributional representation does not distinguish between the profiles \( s^a \) and \( s^b \) defined earlier because \( \tilde{s}^a = \tilde{s}^b = (\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0) \), and this despite the fact that individual players face very different risks under the two profiles. The distributional representation may be thought of as a "reduced-form approach" which

\(^{25}\)This should not be confused with the distributional approach used to refer to Mas-Colell's (1984) model (which is equivalent to Schmeidler's under the current assumptions; see Khan and Sun (1995)); or with the distributional approach of Milgrom and Weber (1985) for games of incomplete information with a continuum of types and finitely many players. The model of this section points out that this last class of games differs from large games in a fundamental way. The reason is that if \( T \) were to represent the set of player types, then \( \tau \) would be a lottery with mutually exclusive outcomes since only one player type will be actually selected. This should be contrasted with the common practice of treating strategies in large games and in games with a continuum of types in an identical manner (e.g., Mas-Colell (1984, Remark 4, p. 204)).
trades off a detailed description of the behavior of the players for greater tractability and simpler description of strategy profiles. The question is then whether the suppressed details are relevant to understanding the way a large game is played. Stated differently: What do we lose by assuming that each player \( t \) ignores the details of the behavior of other players and uses \( \hat{s} \) instead of \( \mu_t(s, \omega) \) to evaluate the strategic uncertainty facing him? The following proposition gives an answer:

**Proposition 7:** If \( G \) is independent, then the distributional representation is valid.

To appreciate the result, it is worth noting that the measurability problem identified in the literature on exogenous i.i.d. processes also appears in the context of large games. This means that a decision-theoretic foundation based on the explicit description of the players' randomizations is, in principle, problematic.\(^{26}\) Proposition 7 is significant because it identifies a set of conditions under which such foundation can be secured. The most important and substantive of these conditions are the ones imposed on the players' payoffs, namely anonymity and state-independence. One can easily imagine large games where some players care about the action of every other player (as in Schmeidler's Theorem 1, for example), in which case Proposition 7 is no longer applicable.

The key advantage of the distributional representation is that it allows each player to think of mixed strategy distributions \( \hat{s} \), as providing sufficient summary information about all the strategically relevant aspects of the other players' behavior. This makes it unnecessary to examine the game at the greater level of detail required by the strategies \( s_t \). To see this, suppose that \( G \) is independent, and that \( s \) and \( s' \) are two profiles such that \( s_t = \hat{s}_t \) for each \( t \). Under independence, the proof of Proposition 7 implies that for every player \( t \), \( \mu_t(s, \omega) = \mu_t(s', \omega) = \hat{s} \) for almost every \( \omega \). In particular, \( s_t \in \text{Br}_t(s) \) if and only if \( s'_t \in \text{Br}_t(s') \). This means that replacing \( s \) by \( s' \) has no impact on the risk facing the players, on the optimality of their strategies, or on whether the game is in equilibrium. This is quite remarkable because the actual behavior of individual players can be very different under \( s \) and \( s' \). The point is that when the distributional representation is valid, details of the individual players' behavior are irrelevant because the impact of these details disappears in the aggregate.

Proposition 7 also explains that the simplex \( \Delta \) in Schmeidler's model has two very different meanings. First, a point \( \hat{s}_t \in \Delta \) is used to represent player \( t \)'s mixed strategy distribution, which is obtained by integrating over the set of states of nature \( \Omega \) for a given player. On the other hand, a point \( \bar{s} \in \Delta \) represents the aggregate action distribution associated with a given strategy profile, and this involves aggregation over the set of players \( T \). Since it suppresses the details of the players' choice of actions, the distributional represen-

\(^{26}\) See, Pascoa (1993, Section 4.2) for a detailed discussion.
tation cannot provide a formal decision-theoretic interpretation of \( \tilde{s} \) or explain how it differs from \( \bar{s} \). The significance of Proposition 7 is to identify conditions under which one can derive \( \tilde{s} \) as the correct description of the strategic uncertainty facing players, hence justifying its interpretation as an aggregate action distribution. The proposition also reveals that the distributional representation is fundamentally based on an implicit "law of large numbers" argument.

6.4. Correlated Randomizations

The assumption in Proposition 7 that all players have independent information is obviously quite strong. It is therefore important to know the extent to which one can relax this assumption while maintaining the validity of the distributional representation. To give an answer to this question, we will need a mild measurability condition that for every \( A \subset \Omega \) with \( P(A) > 0 \), the probability \( P(f_i \leq 0 | A) \) is measurable in \( t \). Our final result gives an idea of how stringent the conditions needed to ensure that the distributional representation is valid:

PROPOSITION 8: If the distributional representation is valid then \( G \) must be idiosyncratic.

Correlation in the underlying information structure is important because it can cause correlation in the players' strategies. In this case two things might happen. First, a player \( t \) might think that other players coordinate their actions on some private signal which these players share, but which is unknown to player \( t \). This introduces strategic uncertainty in the form of a nondegenerate probability measure \( \mu_\omega(s, \omega) \). Second, player \( t \) might believe that his private information \( \tilde{\mathcal{F}}_t \) is correlated with the aggregate action distribution, and that this correlation is useful in improving his prediction of the behavior of other players. This will be reflected in a belief \( \mu_\omega(s, \omega) \) which varies with \( \omega \). Roughly, for the distributional representation to be valid, both possibilities must be ruled out. The force of Proposition 8 is in showing that these possibilities cannot be ruled out if there is even the slightest correlation in \( G \).

Recall that the validity of the distributional representation is defined in terms of a requirement which must hold for all \( t, a, s, u, \omega \), and almost every \( \omega \). This can appear too strong at first, so some readers might be under the impression that Proposition 8 is an artificial consequence of too strong a definition. This is not the case, however. Strategy profiles \( s \) in a large game are determined by the players, so it would be inappropriate to rule out particular strategy profiles from the outset (except for technical conditions like weak measurability). On the other hand, payoff functions in models of large games are usually unrestricted.

---

27 This is a mild condition because it imposes no measurability restrictions on the actual signals received by the players, but only on the probabilities of receiving these signals. This condition is satisfied, for example, whenever \( f \) has an i.i.d. idiosyncratic component \( h \) and such that each \( h \) is independent from the factors in the aggregate component. Much more general conditions are also available.
except by technical conditions such as continuity. Thus, to be sure that the distributional representation is valid in general, all admissible payoff functions must be covered. That the definition must cover all $t, a$ and almost every $\omega$ is obvious.\footnote{Schmeidler has shown that equilibria in pure strategies always exist. Thus, some readers might be inclined to think that Proposition 8 is of minor relevance if, for whatever reason, one focuses only on pure strategy equilibria in large games. In response, note that any notion of equilibrium must be built on the players' decision problems and their best responses. By focusing on the decision-theoretic aspects of the large game, this section points to the possibility that once the players' informational constraints are explicitly modeled, some of the equilibria in Schmeidler's model are eliminated while new ones are created.}

An important feature of the present model is that it is consistent with strategic uncertainty about the aggregate action distribution which cannot be eliminated by the presence of a continuum of players. To an outside observer who does not have access to the information $\mathcal{F}_t$, the uncertainty about the value of the aggregate action distribution takes the form of a nondegenerate probability measure $\mu(s)$. Some might dismiss such uncertainty as artificial on the grounds that the "fundamentals" of the game, such as payoff functions and action sets, are not affected by $\omega$. This is misleading, however. Correlation in the information structure $G$ plays an important role because it can affect the players expectations about the actions of their opponents. For example, a player might try to use this correlation and his knowledge of the profile $s$ to "guess" the actions of other players. This can have an impact on the players' optimal actions, on their payoffs, and on the aggregate action distribution. The point is that one should not dismiss an uncertainty as artificial just because its effect is not directly on the players' payoff functions.

The results of this section have potentially interesting consequences on the study of sunspots in large games. On the one hand, the most natural assumption needed to ensure the validity of the distributional representation, namely independent randomizations, necessarily leads to a deterministic aggregate action distribution and to the absence of aggregate strategic uncertainty. On the other hand, Proposition 8 points out that even a slight correlation in $G$ might cause aggregate uncertainty. The information structure $G$ will not be idiosyncratic if players in a set of positive measure happen to make even a noisy observation of a payoff-irrelevant event (or sunspot). In the real-world environments which large games attempt to model, it seems rather strong to suppose that players never make such observations.

Département des sciences économique, Université du Québec à Montréal, Montréal, Québec, H3C 3P8, Canada.

and

Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208, U.S.A.

Manuscript received December, 1993; final revision received October, 1994.
DECOMPOSITION AND CHARACTERIZATION OF RISK

APPENDIX A

TECHNICAL BACKGROUND

For the convenience of the reader, this Appendix collects some of the mathematical facts used in the paper. Most of this material can be found in Dunford and Schwartz (1958) and Diestel and Uhl (1977).

A.1. Hilbert Spaces

For a detailed discussion of Hilbert spaces and their geometry, see Dunford and Schwartz, Section IV.4. A set of vectors is an orthonormal system if for every distinct pair of vectors and , , and . For an arbitrary set of vectors , span( ) will denote the linear subspace spanned by . That is, span( ) consists of all finite linear combinations of elements of . The closure of span( ) in the -norm, denoted span( ), is the smallest closed linear subspace containing span( ). An orthonormal system satisfies is called an orthonormal basis for if span( ) = . A consequence of Zorn’s Lemma is that every closed linear subspace has an orthonormal basis. A linear subspace of is separable if it has a countable orthonormal basis. If is an orthonormal system, then for any and , , for at most countably many ’s in , and . If is a closed linear subspace of , the orthogonal complement of , denoted , is the set of all vectors orthogonal to every vector in . If is closed, then so is .

A.2. Measurability

A set of vectors is an essential range for if there is a set such that and whenever . A function is measurable if it is the a.e. norm limit of a sequence of simple functions. A weakly measurable function from to has an essentially separable range if it has an essential range which is contained in a separable subspace of . The following is a version of a fundamental result which clarifies the connection between measurability and weak measurability (Diestel and Uhl (1977, II.1.2) or Dunford and Schwartz (1958, III.6.11)).

The Pettis Measurability Theorem: A function is measurable if and only if (i) is weakly measurable; and (ii) has an essentially separable range.

This theorem points to the main limitation of the concept of measurability in dealing with problems involving idiosyncratic risk: A nontrivial idiosyncratic process cannot have an essentially separable range, so such a process cannot be measurable.

A.3. Integration

For a measurable function , the natural notion of integral is the Bochner integral. This integral is a natural (countably infinite dimensional analogue of the Lebesgue integral and has most of the properties familiar from the real-valued case. For example, a bounded function is Bochner integrable if and only if it is measurable, and appropriate versions of the classic convergence theorems hold.

Any aggregate process has, off a set of measure zero, a countably generated range, and must therefore be measurable. Consequently, every bounded aggregate process is Bochner integrable. Let denote the corresponding indefinite Bochner integral. For valued processes, the Bochner integral has a particularly simple form; Fix the representation , and define the sequence of finitely generated processes , . The Bochner integral of over a subset is the random variable (the integral within square brackets is the usual Lebesgue integral). Intuitively, we first think of the factors as an orthonormal basis for the -dimensional subspace they span, then Lebesgue-integrate the coordinate functions “dimension-by-dimension” (very much like Lebesgue integration for -valued
functions). Since the sequence \( (g^K) \) is uniformly norm bounded and norm-converges to \( g \), \( \tau \)-a.e., the Dominated Convergence Theorem for vector-valued functions (Dunford and Schwartz (1958, III.6.16)) implies that

\[
\int_A^B g_t \, d\tau = \lim_{K \to \infty} \int_A^B g^K_t \, d\tau = \sum_{k=1}^\infty \left( \int_A^B \beta_k \, d\tau \right) \eta_k.
\]

As noted earlier, any process \( f \) involving nontrivial idiosyncratic risk cannot be measurable, so such \( f \) cannot be Bochner integrated. A weaker notion of vector-valued integration is the Pettis integral, which will be used in some of the proofs. A function \( f: T \to L_2 \) is Pettis integrable if for every measurable \( A \subset T \), there is a random variable \( \int_A^B f_t \, d\tau \in L_2 \) such that \( (g||f||^2 \, d\tau) = \int_A^B (g|f|) \, d\tau \). The symbol \( \int_A^B f_t \, d\tau \) will denote the indefinite Pettis integral (see Diestel and Uhl (1977, II.3) for more details).

### A.4. Vector Measures and Spaces with the Radon-Nikodým Property

Let \( (L, \mathcal{L}, \lambda) \) be an arbitrary measure space. A vector measure on \( L \) with values in a Banach space \( (B, ||\cdot||) \) is a countably additive set function \( \eta: \mathcal{L} \to B \). A vector measure has bounded variation if \( \sum_{s \in A} \|\eta(A_s)\| < \infty \), where the supremum is taken over all finite partitions of \( L \) by sets in \( \mathcal{F} \). The vector measure \( \nu \) is absolutely continuous with respect to \( \lambda \) if \( \lambda(A) = 0 \) implies \( \nu(A) = 0 \) for every measurable \( A \subset T \). A Banach space \( B \) has the Radon-Nikodým property if for every finite measure space \( (L, \mathcal{F}, \lambda) \), any vector measure \( \eta: \mathcal{F} \to B \) which is of bounded variation and is absolutely continuous with respect to \( \lambda \) has a Bochner integrable density. An important result in vector measure theory is that all reflexive spaces (hence Hilbert spaces and, of course, \( L_2 \)) have the Radon-Nikodým property (Diestel and Uhl (1977, III.2.13, or IV.1.4)).

### APPENDIX B

#### PROOFS

I begin with a characterization of aggregate processes:

**Proposition B.1:** For any bounded, weakly measurable process \( f \), the following statements are equivalent:

(i) \( f \) is an aggregate process.

(ii) \( f \) has an essentially separable range.

(iii) \( f \) is measurable (hence \( f \) is the \( \tau \)-a.e. norm limit of a sequence of simple processes).

(iv) For every \( \epsilon > 0 \) there is a partition of \( T \) into \( N + 1 \) measurable subsets \( \{T_0, \ldots, T_N, T_{N+1}\} \) such that \( \tau(T_{N+1}) < \epsilon \) and for every \( 1 \leq n \leq N \), we have \( \|f_s - f_{t'}\| < \epsilon \) for every \( s, t' \in T_n \).

**Proof:** (i) \( \iff \) (ii): If \( f \) is an aggregate process then, off a set of measure zero, the range of \( f \) is spanned by a countable set \( \{\eta_1, \eta_2, \ldots\} \), and is therefore separable. By the Pettis Measurability Theorem, \( f \) is measurable. Conversely, if \( f \) is measurable, then its range is essentially separable by the Pettis Measurability Theorem. If \( \mathcal{F} \) is a separable essential range, then \( \text{span}(\mathcal{F}) \) is a closed separable linear subspace of a Hilbert space and it must therefore have a countable orthonormal basis \( \{\eta_1, \eta_2, \ldots\} \). This means that \( f \) can be written in the form \( f = \sum \beta_i \eta_i \), \( \tau \)-a.e. Since \( f \) is weakly measurable, and \( ||\eta|| = 1 \) by construction, \( \beta = (\eta_i | f) \) is a measurable function.

(ii) \( \Rightarrow \) (iv): If \( f \) is measurable, then there must be a sequence of simple processes \( \{f^n\} \) which converges to \( f \) \( \tau \)-a.e. By the Egoroff Theorem (Dunford and Schwartz (1958, III.6.12)), for any \( \epsilon > 0 \), there is \( T' \) with \( \tau(T') < \epsilon \) and \( m \) large enough such that \( ||f^n - f|| < \epsilon/2 \) for all \( t \in T' \). Fix such \( m \), and let \( E_0^n, \ldots, E_N^n \) be the component sets of the simple function \( f^n \). Set \( T_{n+1} = T' \), and \( T_n = E_0^n \cap (T - T_{n+1}) \). Since for every \( t \in T_n \), \( ||f^n - f|| < \epsilon/2 \), and since \( f^n \) is constant on \( T_n \), we must have \( ||f - f_n|| < \epsilon \) for all \( s, t \in T_n \).

(iv) \( \Rightarrow \) (i): Conversely, for each \( m = 1, 2, \ldots \), let \( T_1^m, \ldots, T_{N_m+1}^m \) denote the partition of \( T \) given by the definition and corresponding to the value \( \epsilon = 1/m \). Choose \( t_n^m \in T_n^m \) for all \( n \) and \( m \). Construct a simple function \( f^m \) by letting \( f^m_t = f_t \) for \( t \in T_{n+1}^m \). Then for all \( n, m \), \( t \in T_n^m \), we must have \( ||f^m - f|| < \epsilon \). Consequently, for any \( m \), \( \tau(||f^m - f|| < 1/m) > \tau(T_{N_m+1}^m) > 1/m \). This implies \( f^m \to f \) in measure. It is therefore possible to extract a subsequence of simple functions which converges to \( f \), \( \tau \)-a.e. (Dunford and Schwartz (1958, III.6.13)). This means that \( f \) is measurable.

Q.E.D.
DECOMPOSITION AND CHARACTERIZATION OF RISK

PROOF OF PROPOSITION 1: Fix $g_i = \sum_{k=1}^{\infty} \beta_k n_k$ and define $g_k = \sum_{k=1}^{K} \beta_k n_k$. Clearly, $\|g_k - g^K\| \to 0$ for $\tau$-a.e. $t$, and $t \mapsto \|g_k - g^k\|$, $K = 1, 2, \ldots$, is a uniformly bounded sequence of measurable real-valued functions. This implies that $\int_{\Omega} g_k - g^k \, d\tau \to 0$. This result now follows by defining $w = g - g^k$, and $g^k = g^k$ for a large enough $K$.

Q.E.D.

PROOF OF PROPOSITION 2: Suppose that $(g, h)$ and $(g', h')$ are two decompositions of $f$. Since $h$ and $h'$ are purely idiosyncratic, then for any $g \in L_2$, we must have $(g|h) = (g|h') = 0$, $\tau$-a.e. Thus, for any $g$, $(g|g) = (g|f) = (g|g)$, $\tau$-a.e. By Corollary II.2.7 of Diestel and Uhl, we must have $g = h$, $\tau$-a.e. This in turn implies that $h = h'$, $\tau$-a.e., as required.

Q.E.D.

PROOF OF THE MAIN THEOREM: (i) $\Rightarrow$ (ii): Since $f$ is bounded and weakly measurable, then for any $g \in L_2$ the real-valued function $(g|f)$ is Lebesgue integrable. By Lemma II.3.1 of Diestel and Uhl, $f$ is Dunford integrable. Since the Dunford and the Pettis integrals coincide in reflexive spaces, and since $L_2$ is self-dual (hence reflexive), $f$ is Pettis integrable. Use the indefinite Pettis integral $\int f \, d\tau$ to define a vector measure $\nu$ on $[0, 1]$ by the formula $\nu(A) = \int f \, d\tau$. By Theorem II.3.5 of Diestel and Uhl (1977, p. 53), $\nu$ is $L_2$-valued, countably additive vector measure which is absolutely continuous with respect to $\tau$. Since there is $M$ such that $\|f\| \leq M < \infty$ for all $t$, we must have $\|\nu(A)\| \leq M\rho(A)$, $\tau$-a.e. Thus, for any measurable partition $A_1, \ldots, A_n$, $\|\nu(A)\| \leq M$, implying that $\nu$ has bounded variation. Since $L_2$ has the Radon-Nikodym property, $\nu$ has a bounded Pettis integrable density $g$. That is, there is a Pettis integrable process $g$ such that $\nu(A) = \int_A g \, d\tau$, for every $A \in \mathcal{A}$. By Proposition B.1, $g$ is an aggregate process. Define $h = f - g$, and note that for any $g \in L_2$ and $A \in \mathcal{A}$, we have $\int_A (g|h) \, d\tau = (g|f - g) \, d\tau = (g|\int f \, d\tau - \int g \, d\tau) = 0$. This implies that $(g|h) = 0$, $\tau$-a.e. so $h$ is idiosyncratic. Thus, $(g, h)$ is a decomposition of $f$, which is essentially unique by Proposition 2.

(ii) $\Rightarrow$ (iii): If $f$ is decomposable, then the proof of Proposition 3 below can be used to show that there is a sequence of simple processes such that $f^m \to f$, hence $f^m \leq f$.

(iii) $\Rightarrow$ (i): Suppose that there is a sequence of simple processes such that $f^m \leq f$. Since $f^m$, $T \to L_2$, is simple, the functions $t \mapsto (g|f^m)$, which map $T$ into $\mathbb{R}$, are simple for each $m$. Thus, for any $g \in L_2$, the function $t \mapsto (g|f^m)$ is the $\tau$-a.e. limit of a sequence of simple functions $t \mapsto (g|f^m)$, and is therefore measurable. This shows that $f$ is weakly measurable.

(iii) $\Rightarrow$ (iv): Say that a process satisfies the law of large numbers (LLN) if the condition in (iv) is satisfied for that process. The set of processes satisfying LLN is a linear space. Since every Pettis integrable process satisfies LLN (e.g., Hoffman-Jorgensen (1987, Theorem 2.4, p. 310)), we only have to show that $h$ satisfies LLN. Since the vectors $h_i$ are all orthogonal off a set $A \subset T$ of $\tau$-measure zero, an infinite sequence $(t_1, t_2, \ldots)$ will not intersect $A$ almost surely. This means that, with $\tau$-probability 1, any sequence will consist of orthogonal random variables. The result now follows from the fact that the range of $h$ is bounded.

(iv) $\Rightarrow$ (i): By Theorem 8 of Talagrand (1987), $f$ is Pettis integrable and hence weakly measurable.

Finally, the statement at the end of the Theorem follows from Theorem 8 of Talagrand (1987).

Q.E.D.

I now provide some comments on the proof of the Theorem. The idea of constructing the measure $\nu$, then taking its Radon-Nikodym derivative, in the proof that every weakly measurable process is decomposable is due (to my knowledge) to Dobric (1990). The result obtained above is distinct from the one found in Dobric because the special structure of the problem (e.g., special features of $L_2$) is used to derive different results from different assumptions. Note that the basic argument works for all spaces with the Radon-Nikodym property. An alternative proof of decomposition using an orthogonal projection argument can be found in Al-Najjar (1994). The characterization by simple processes is motivated by Geitz (1981, Theorem 7) who provided a general version of this result for functions taking values in arbitrary Banach spaces. The simple geometry of $L_2$ and the decomposition part of the Theorem are used to provide a simpler and more direct proof. Finally, a direct proof that (iv) implies weak measurability can be given using the argument found in the last paragraph on page 861 of Talagrand (1987). However, this argument is quite delicate and involves new definitions, so the relevant result is quoted directly from Talagrand.

PROOF OF PROPOSITION 3: Fix the decomposition $(g, h)$ of $f$ and let $\mathcal{A}$ be a separable essential range of $g$. I first show that there is a sequence $(f^m)$ of simple processes such that $f^m \to f$. Note that it is always possible to find a sequence of simple processes $f^m$, $T \to \text{span}(\mathcal{A})$ such that $f^m \to f$. Since $g$ is bounded, the sequence $(f^m)$ may be chosen to be uniformly bounded (i.e., there
is $r$ such that $\|f^m\| \leq r$ for all $t$ and $m$). By the vector-valued version of the Dominated Convergence Theorem for vector-valued functions (Dunford and Schwartz (1958, III.6.16)), the sequence $\int f^m dt$ converges to $\int f dt$ in the $L_2$-norm for every measurable $A \subseteq T$. To complete the proof that $f^m \rightarrow f$, it only remains to show that $f^m \rightarrow f$ $\tau$-a.e. That is, for any $g \in L_2$, $(\langle g, f^m \rangle \rightarrow \langle g, f \rangle)$, $\tau$-a.e. To do this, let $\mathcal{A}$ denote the orthogonal complement of $\text{span}(\mathcal{A})$ in $L_2$. Any vector $g \in L_2$ can be uniquely written as the sum $g = g_1 + g_2$ where $g_1 \in \text{span}(\mathcal{A})$ and $g_2 \in \mathcal{A}$. It is clearly sufficient to show that $\|g - f^m\| \rightarrow 0$, $\tau$-a.e. for $i = 1, 2$. First, the fact that $f^m \rightarrow f$ implies that $(g_1, f^m) \rightarrow (g_1, f)$, $\tau$-a.e. Next, for every $\epsilon$, we have $\|g_2\| = 0$, while $(g_2, f) = (g_2, g_2) + (g_2, h) = (g_2, h) = 0$, $\tau$-a.e. by the definition of an idiosyncratic process. This means that $(g_2, f^m) \rightarrow (g_2, f)$, $\tau$-a.e.

Since we have $\int f^m dt \rightarrow \int f dt$ in the $L_2$-norm, hence in measure, there must be a subsequence $(f^{m_i})$ such that $\int f^{m_i} dt \rightarrow \int f dt$, $\tau$-a.e. This means that there is a sequence of simple processes $(f^{m_i})$ such that both $f^{m_i} \rightarrow f$ and $f^{m_i} \rightarrow f$.

**Proof of Proposition 4:** Define $\mu(f)$ to be the probability distribution of the Bochner integral $\int f d\tau$. Part (iv) of the Theorem implies that $f(t^n) \rightarrow f(t)$ in the $L_2$-norm, $\tau$-a.e. Since convergence in $L_2$ implies weak convergence of probability measures, we have $\mu(t^n) \rightarrow \mu(f)$, $\tau$-a.e. as required. On the other hand, if $(f^m)$ is any sequence of $\mathcal{F}$ such that $f^m \rightarrow f$, then we must have $\int f^m dt \rightarrow \int f dt$ in $L_2$ by definition. This, in turn, implies that $\mu(f^m) \rightarrow \mu(f)$. 

**Proof of Proposition 5:** Fix $\alpha$ (for notational simplicity, I will drop any references to $\alpha$), and define $U^*(f)$ by $U^*(\mu(f))$. Clearly, $U^*$ is an extension of $U$ which is defined on all of $\mathcal{F}$, and which automatically satisfies the second part of the statement of the Proposition.

For any sequence $(f^m)$ in $\mathcal{F}$ (not necessarily simple) converging to $f$, we have $\mu(f^m) \rightarrow \mu(f)$ by Proposition (ii). Since $\mu$ is $\tau$-a.s. continuous, $U^*(\mu(f^m)) \rightarrow U^*(\mu(f))$.

**Proof of Proposition 6:** Suppose that $(f^m)$ is any sequence in $\mathcal{F}$ such that $f^m \rightarrow f$ in $\mathcal{F}$, and let $f^m$ and $f$ denote the corresponding Bochner integrals $\int f^m dt$ and $\int f dt$. By definition, $f^m \rightarrow f$, $\tau$-a.e.

For the purposes of this proof, it will be more convenient to work with distribution functions instead of probability measures. To fix notation, for any $x \in \mathbb{R}$ define the random variable $I_x$ by $I_x(\omega) = 1$ if $f(\omega) \leq x$, and zero otherwise. Let $F$ denote the (unconditional) distribution function of $f$ on $\mathbb{R}$ defined by $F = \int f \, d\mathbb{P} = \mathbb{P}(f(x) \leq x)$. Also, let $F^m$ be a version of the regular conditional distribution function of $f$ relative to $\mathcal{F}$. That is: (i) for each $x$, $F^m$ is a distribution function on $\mathbb{R}$; and (ii) for every $x \in \mathbb{R}$, $F^m(x)$ is a version of the conditional probability $\mathbb{P}(f \leq x | \mathcal{F}(\omega))$ (Ash (1972, Definition 6.6.1 and Theorem 6.6.2, p. 273)). In fact, for each $x$, we may take $F^m(x)$ to be a version of $E[I_{x, \mathcal{F}(\omega)}(f)]$ (Ash (1972, Theorem 6.4.6, p. 253)). For each $m$, use $f^m$ to define $I^m_x(\omega)$, $F^m_x$ similarly.

Suppose that $x$ is a point of continuity of $F$. Let $B_x = \{\omega: f(\omega) = x\}$ and $C = \{\omega: f^m(\omega) \rightarrow f(\omega)\}$. Note that $P(B_x) = 0$ because $x$ is a point of continuity of $f$, while $P(C) = 0$ since $f^m \rightarrow f$ for $\tau$-a.e. $\omega$. Since for any $\omega \in B_x \cup C$, $f^m(\omega) \rightarrow f(\omega)$ and either $f(\omega) > x$ or $f(\omega) < x$, we must have $I^m_x(\omega) = I_x(\omega)$ for all large enough $m$. We therefore conclude that $I^m_x \rightarrow I_x$ for all $\omega$ outside $B_x \cup C$. This implies $E[I^m_x(\mathcal{F}) \rightarrow E[I_x(\mathcal{F})], P$-a.e. (Ash (1972, Theorem 6.5.5, p. 257)). Thus, for every point of continuity $x$, we have $F^m_x \rightarrow I_x$, $P$-a.e.

Let $\{r_n\}_{n=1}^{\infty}$ be a countable dense subset of $\mathbb{R}$ which excludes all points of discontinuity of $F$. To simplify notation, I replace $r_i$ by $i$ in subscripts (e.g., $F^m_i$ will replace $F^m_{r_i}$). From the last paragraph, for every $i = 1, 2, \ldots$, the set $A_i = \{\omega: F^m_1(\omega) \rightarrow F^m_{r_i}(\omega)\}$ is such that $P(A_i) = 0$. Define $A = \bigcup A_i$, and note that $P(A) = 0$ since $P(A_i) = 0$ for each $i$. Thus, for $\omega \in A$ (an event which has probability 1), the sequence of distribution functions $F^m(\omega)$ converges to $F(\omega)$ on a countable dense subset $\{r_i\}$ of $\mathbb{R}$.

We now use a key fact about weak convergence: A sequence of probability measures $\mu^m$ converges weakly to a probability measure $\mu$ if and only if the corresponding sequence of distribution functions converges at every point of continuity of the distribution function of $\mu$ (Ash 1972, Theorem 6.6.1, p. 273). Since $F^m \rightarrow F$ for $\tau$-a.e. $\omega$, we conclude that $\mu^m \rightarrow \mu$ for $\tau$-a.e. $\omega$. Thus, $\mu^m \rightarrow \mu$ for $\tau$-a.e. $\omega$. This completes the proof of Theorem 6.
PROOF OF PROPOSITION 7: Fix the profile \( \mathbf{s} = (s_1, \ldots, s_N) \) (subscripts will refer to the components of this \( N \)-vector of random processes). Since any profile is weakly measurable by assumption, we may define \( \hat{s}_{nt} = \int_0^1 \hat{s}_{nt} \, dP = P(s_{nt} = 1) \). Furthermore, the independence of \( G \) implies that the random variables \( \{s_{nt} : t \in [0,1]\} \) are also independent. This means that each process \( t \mapsto [s_{nt} - \hat{s}_{nt}] \) is idiosyncratic, so the pair of processes \( t \mapsto [s_{nt}] \) and \( t \mapsto [s_{nt} - \hat{s}_{nt}] \) constitutes a decomposition of \( s_{nt} \).

The term \( \hat{s}_{nt} \) here refers to the degenerate random variable which assumes the constant value \( \hat{s}_{nt} \) in all states of nature. The result now follows by noting that, by definition, \( \mu_\iota(s, \omega) \) is the distribution of the random vector \( \{f_1^0 s_1, \ldots, f_N^0 s_N, d\epsilon\} \), and that this random vector is one which takes the value \( \hat{s} \) with probability 1.

PROOF OF PROPOSITION 8: I assume that \( f \) is not idiosyncratic, and construct a profile \( \mathbf{s} \) at which the distributional representation is not valid. Let \((g, h)\) be the (essentially unique) decomposition of \( f \) and note that since \( EF_t = 0 \) for all \( t \), we must have \( EG_t = 0 \), \( \tau \)-a.e. The assumptions that \( f \) is not idiosyncratic and \( \text{var}(f_t) > 0 \) for all \( t \) imply that \( \text{var}(g_t) > 0 \) for a set of players of positive measure.

By Proposition B.1, for some \( \epsilon > 0 \), it is possible to find a set \( A \) with \( \tau(A) > 0 \) such that \( \text{cov}(g_t, g_r) > \epsilon > 0 \) for all \( t, r \in A \).

Each player in the profile \( \mathbf{s} \) takes either \( a_1 \) or \( a_2 \). For \( t \in A \), let \( s_{tj}(\omega) = 0 \) for all \( \omega \) (so these players always play \( a_2 \)); for \( t \in A \), let \( s_{tj}(\omega) = 1 \) if \( f_t(\omega) > 0 \) and \( s_{tj}(\omega) = 0 \) if \( f_t(\omega) < 0 \). For \( C \subset \Omega \), let \( I_C \) denote the indicator function of \( C \). Our measurability assumption ensures that for every \( C \subset \Omega \) with \( P(C) > 0 \), the function \( t \mapsto (I_C | s_{1t}) \) is measurable; hence \( t \mapsto (g | s_{1t}) \) is measurable for every simple random variable \( g \). Since every \( g \in L_2 \) is the limit of a sequence of simple random variables (Ash 1972, Theorem 2.4.13, p. 88), we conclude that the process \( s_{1t} \) is weakly measurable.

Since \( EF_t = 0 \) and \( \text{var}(f_t) > 0 \) on \( A \), the probability \( \beta_t = ES_{1t} \) that player \( t \) takes action \( a_1 \) satisfies \( 0 < \beta_t < 1 \) on \( A \). Also, we have \( \{\omega : s_{1t} = 1\} = \{g_t : 0\} \) (recall that \( s_{1t} \) assumes only two possible values).

Since the \( f_t \)'s are positively correlated over \( A \), the \( s_{1t} \)'s are also positively correlated over \( A \). Let \((\hat{g}_{1t}, \hat{h}_t)\) be the (essentially) unique decomposition of the process \( s_{1t} - \beta_t \), and note that \( \hat{h}_t = 0 \) outside \( A \). Since the \( \hat{h}_t \)'s are by definition uncorrelated with each other and with the aggregate component \( \hat{g}_t \), we must have \( \text{cov}(\hat{g}_{1t}, \hat{g}_{1t}) > 0 \), hence \( \text{var}(\hat{g}_{1t}) > 0 \), over \( A \). These facts are seen to imply that \( f_1^0 \hat{g}_{1t}, d\epsilon \) is a random variable with strictly positive variance. By definition, \( f_1^0 \hat{g}_{1t}, d\epsilon + f_1^0 \beta_t, d\epsilon = f_1 \) is the Bochner integral of the aggregate component of \( s_{1t} \).

Fix a player \( t \) and recall that \( \mu(s, \omega) \) is a fixed version of the regular conditional distribution of \( f \). If \( \hat{f} \) is independent of \( f \), then \( \mu(s, \omega) \) equals the unconditional distribution \( \mu(s) \) for \( P\)-a.e. \( \omega \). Since \( \hat{f} \) has positive variance, \( \mu(s) \) is a nondegenerate probability distribution. In this case, the distributional representation can be shown to be invalid by choosing a continuous payoff function \( u(a, \cdot) \) which, for some fixed action \( a \), attains its unique maximum at \( \hat{s}_t \).

If, on the other hand, \( \hat{f} \) and \( f \) are not independent, then the measure-valued mapping \( \omega \mapsto \mu(s, \omega) \) is not essentially constant (i.e., there does not exist a measure \( \mu \) such that \( \mu(s, \omega) = \mu \) for \( P\)-a.e. \( \omega \)). I prove that the distributional representation is not valid by showing that there is a continuous payoff function \( u \), and an action \( a \) such that \( \omega \mapsto u(a, \mu(s, \omega)) \) is not essentially constant (i.e., there does not exist a constant \( c \) such that \( u(a, \mu(s, \omega)) = c \) \( P\)-a.e.) I begin by simplifying notation: Restrict attention to payoff functions \( u(a, d) \) which depend on \( \omega \) only through \( d_1 \). Also, since \( a \) can be chosen arbitrarily, all references to it will be dropped. Finally, by construction, the measures \( \mu(s, \omega) \) are viewed as measures on \( d_1 \) only. To summarize it will be enough to show that there is a continuous function \( s : [0,1] \rightarrow \mathbb{R} \) such that \( \omega \mapsto \int_0^1 u(r) \, d\mu(s, \omega) \) is not essentially constant. To prove this, let \( \{a_n\} \) be a countable dense subset of \( C[0,1] \) and assume, by way of contradiction, that \( [f(x) \, d\mu(s, \omega) = c] \) except for \( \omega \in A \), where \( P(A) > 0 \). Define \( A = \)
and note that \( P(A) = 0 \). Since \( \omega \rightarrow \mu_\omega(s, \omega) \) is not essentially constant, there is at least one pair of points \( \omega, \omega' \in A \) such that \( \mu_\omega(s, \omega) \neq \mu_\omega(s, \omega') \). Since \( C[0, 1] \), viewed as a collection of linear functionals on the space of measures on \([0, 1]\), separates points, and since \((u_i)\) is dense in \( C[0, 1] \) by assumption, there must be some \( u_i \) such that \( \int u_i(r) d\mu_\omega(s, \omega) \neq \int u_i(r) d\mu_\omega(s, \omega') \). This contradicts the definition of \( A \).

\[ Q.E.D. \]

REFERENCES