Long Run Cost Benefit Rules

Antony Millner*[1]

[1]University of California, Santa Barbara & NBER

January 20, 2026

Abstract

I study a preference relation on risky long run public projects induced by a large maturity limit of expected present values. Under common assumptions this relation has a variational representation that is related to a well known model of ambiguity aversion; it is non-probabilistic in general. The formalism generalizes Weitzman's 'lowest possible rate' formula for long run discount rates to a large class of stochastic economies, gives rise to a notion of stochastic dominance adapted to long run valuation, and characterizes features of stochastic processes that cause long run cost benefit rules to be non-probabilistic.

Keywords: Cost benefit analysis, long run social discounting, variational preferences, stochastic dominance.
JEL codes: D61, D81, H43.

Cost benefit analysis prescribes a simple procedure for ranking risky marginal public projects: project A's contribution to social welfare is larger than project B's if the expected present value of A's payoffs is larger than the expected present value of B's payoffs. The justification for the expected present value criterion, with time- and state-contingent payoffs discounted using appropriate maturity- and state-dependent social marginal rates of substitution, appeared in several early contributions (see e.g. Dasgupta et al., 1972; Little & Mirrlees, 1974). The devil is, of course, in the details; a voluminous literature discusses the conceptual and practical issues involved in implementing cost benefit rules, chief amongst them the problem of specifying and parameterizing social marginal rates of substitution, i.e., social discount rates (see Gollier & Hammitt, 2014; Groom et al., 2022; Millner & Heal, 2023, for recent reviews).

This paper focusses not on a specific proposal for social discounting, but on the structure of the cost benefit rule itself, interpreted as a preference ordering on risky projects. I investigate

*Email: amillner@econ.ucsb.edu. I am grateful to Marcus Pivato, Jakub Steiner, Larry Samuelson, Fabio Maccheroni, Luciano Pomatto, Derek Lemoine and the Theory group at UCSB for valuable discussions, and to seminar participants at Berkeley, Oxford, Caltech, Zurich, D-TEA, RUD, and SAET for feedback.

the implications of this rule for rankings of *long run* projects, i.e., risky returns that are realized in the very distant future. Examples of policy problems that are approximated by such long time horizons include climate mitigation, infrastructure projects, and nuclear waste management (Gollier, 2012; Millner & Heal, 2023). I show that for long run projects the cost benefit rule transforms into a preference ordering that looks very different to the familiar expected present value formulation. Under common assumptions, long run cost benefit rules are represented by a variational problem – minimization of a project-specific functional on a set of measures – and not via expectations. The resulting preference ordering provides a dual interpretation of the well known variational model of ambiguity averse preferences (Maccheroni et al., 2006), with quantities that capture ambiguity *attitudes* in that model mapping to quantities that capture *risk* for long run cost benefit rules. The fact that long run cost benefit rules can have qualitatively different structure and properties from their finite maturity counterparts is a consequence of a failure of uniform convergence for expected present value functionals, a phenomenon that arises often when dealing with long run properties of stochastic processes (e.g. Donsker & Varadhan, 1975; Bobenrieth et al., 2002).

With the variational representation of long run cost benefit rules in hand, I proceed to investigate the properties of this preference relation on risky projects. I first define a notion of long run stochastic dominance, an ordering of risks induced by long run cost benefit rules. Stochastic process X is said to long run stochastically dominate another stochastic process Y if the long run value of *any* risky project under X is weakly greater than its value under Y. I characterize this relation in terms of primitives of long run cost benefit rules, and show that it is related to a novel dominance ordering on the persistence of stochastic processes, i.e., an ordering of 'long run risk' (Duffie & Epstein, 1992). Thus, the aspect of project risk captured by long run cost benefit rules is inherently dynamic, although the rules themselves are formally equivalent to a static preference relation.

Next, I turn to investigating an unusual feature of long run cost benefit rules – their generic lack of probabilistic structure. I characterize conditions on stochastic processes that cause families of long run cost benefit rules to be probabilistic. Such families arise from a collection of related stochastic processes, e.g. i.i.d. processes, or ergodic Markov processes with a given stationary distribution. The analysis shows that non-probabilistic long run cost benefit rules occur when the asymptotic dynamics of the stochastic process that generates risk are rich enough that they cannot respect all the symmetries required by probabilistic preferences. A consequence of this analysis is that non-probabilistic long run cost benefit rules are the rule, rather than the exception. I discuss how this feature can lead to unusual behavior, e.g. strict preferences between projects with identical payoffs in two economies that are asymptotically statistically identical with probability 1. This analysis also provides an alternative interpretation of the famous Ellsberg (1961) choice experiments, in which ambiguity averse behavior arises not from discomfort with deep uncertainty, but as a consequence of rich dynamical beliefs.

Background and Related Literature

The duality between long run cost benefit rules and variational preferences that lies at the heart of this paper bridges two disparate parts of the literature; one rooted in social discounting and public project appraisal, and the other in decision theory. To the best of my knowledge this connection and its implications have not appeared in either of these sub-fields, but each of them has extant contributions that contain seeds of the results I present.

The preference relation on long run projects that I study is intimately related to long standing questions about the appropriate long run social discount rate – I make the relationship between these two concepts precise in Section 1 below. The effect of risk on long run discount rates was most famously studied by Weitzman (1998, 2001), with many subsequent contributions on this theme (e.g. Freeman & Groom, 2010; Gollier, 2012; Cropper et al., 2014; Fleurbaey & Zuber, 2015; Millner, 2020). The essence of Weitzman's insight is the following observation: suppose that the future real interest rate is uncertain today, but will be revealed to be some constant r_i in the next time period with probability p_i. From the perspective of today, the present value of a sure payoff of \$1 t years from now is thus

$$\sum_i p_i \exp(-r_i t) \times \$1 \tag{1}$$

The certainty equivalent risk free discount rate $\hat{r}(t)$ on sure payoffs at maturity t is defined via

$$\exp(-t\hat{r}(t)) \equiv \sum_i p_i \exp(-r_i t) \Rightarrow \hat{r}(t) = -\frac{1}{t} \log\left(\sum_i p_i \exp(-r_i t)\right). \tag{2}$$

It is a simple matter to show that

$$\lim_{t\to\infty} \hat{r}(t) = \min_i r_i. \tag{3}$$

This follows since each of the terms in (1) decays exponentially fast as t increases; in the $t \to \infty$ limit, the sum is dominated by the term that decays slowest, and hence the certainty equivalent discount rate $\hat{r}(t)$ tends to the lowest possible rate in the long run. This reasoning is a simple application of a broader mathematical result known as the Laplace Principle, which predates Weitzman's work by more than 200 years (Laplace, 1774). Equation (3) is the basis for the slogan 'the far distant future should be discounted...at the lowest possible rate' (Weitzman, 1998), which has been influential in policy circles (Groom & Hepburn, 2017).

Weitzman's analysis has been criticized from a variety of perspectives (see e.g. Dasgupta, 2008; Gollier & Weitzman, 2010), but two central limitations are most important for this paper. First, the result applies only to risk free discount rates, it does not have much to say about real world projects whose payoffs may be correlated with sources of interest rate (or more generally, valuation) uncertainty. Second, and more importantly, the result relies on a highly stylized representation of uncertainty – all risk is assumed to resolve after a single time period, after which interest rates remain constant for all eternity.[1]

The results of this paper show that both these limitations can be overcome in a general theory of preferences over risky long run projects, which nevertheless maintains a connection to Weitzman's original insight. Using a more powerful version of Laplace's principle that finds its modern expression in large deviation theory, I show that long run cost benefit rules can be represented by a generalization of Weitzman's 'lowest possible rate' formula (3). For a large and empirically relevant class of stochastic economies, long run projects are ranked by the variational formula

$$-\inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}}\lambda\,dq+I(q)\right] \tag{4}$$

where q is a probability measure on a state space $\mathcal{S}$, $\Delta(\mathcal{S})$ is the set of all such probability measures, $\lambda(s)$ is the asymptotic 1-period rate of decay of project value in state $s\in\mathcal{S}$ (i.e., lower values of λ are better), and $I:\Delta(\mathcal{S})\to\mathbb{R}^+$. The infimal q in (4) varies with the project being evaluated (i.e., with λ); this implies that long run cost benefit rules do not, in general, describe uncertainty using a stable probability distribution over a state space – they are not probabilistically sophisticated (Machina & Schmeidler, 1992). A formula that is formally identical to (4) (multiplied by -1) can also be used to analyze long run discount rates in these economies.

The function $I(q)$ in (22) is independent of project returns and discount factors; it depends only on the asymptotic dynamics of the stochastic process that generates risk. As a formal matter, Weitzman's result (3) is the case where

$$I(q)=\begin{cases}0 & q=\delta_s\\ \infty & \text{otherwise}\end{cases}$$

where δ_s is a Dirac measure centered on $s\in\mathcal{S}$, i.e., for $\mathcal{A}\subseteq\mathcal{S}$, $\delta_s(\mathcal{A})$ returns 1 if $s\in\mathcal{A}$, and zero otherwise. This instance of $I(q)$ – which corresponds to Weitzman's 'one-shot' model of risk – is the most patient possible model in the class (22); any other risk model must give rise to weakly larger long run discount rates (holding λ fixed). With richer and more plausible models of risk (e.g. Markov processes) $I(q)$ will turn out to be a convex function on $\Delta(\mathcal{S})$, often with a single zero at a probability measure $p\in\Delta(\mathcal{S})$ that captures the 'typical' long run statistics of the underlying stochastic process. Indeed, any function $I(q)$ satisfying weak

[1] A further important critique is that Weitzman's analysis conflates normative (e.g. how should society discount the wellbeing of future generations?) and empirical (e.g. what will future consumption growth rates be?) questions into a single constant that is difficult to interpret, or justify in the context of more flexible models of social discounting (Gollier, 2002a,b; Dasgupta, 2008). The framework I develop below applies to normative and positive approaches to project evaluation alike; see Millner & Heal (2023) for a detailed discussion of these two methodological approaches to social discounting.

topological properties is compatible with some stochastic model of the economy; Weitzman's 'lowest possible rate' prescription is inaccurate in essentially all such models.

This paper is of course not the first to incorporate project risk and more realistic stochastic dynamics into a model of long run project appraisal. Several authors have studied particular instances of valuation risk and its impacts on long run discounting. Gollier (2002a) studies a model of the impact of i.i.d. consumption growth shocks on the long run risk free discount rate, Gollier (2014) presents a model that incorporates serial correlation in growth rates and project risk in a parametric model, and Fleurbaey & Zuber (2015) investigate a model that incorporates the effects of inequality aversion on long run discount rates in a model with one-shot risk *à la* Weitzman. While these models provide important insights, their treatment of risk and its influence on long run values is limited to simple parametric, or one-shot, models. By contrast I present a fully non-parametric treatment that does not rely on a specific model of social preferences or consumption risk, and encompasses a much larger class of empirically plausible stochastic processes. This generality allows me to investigate the structure of long run cost benefit rules as a general class of preference relations on risky projects.

The finance literature has also studied these questions in a number of important contributions. In a Markovian model Hansen & Scheinkman (2009); Hansen (2012) obtain a representation of long run pricing kernels in terms of a martingale and the Perron-Frobenius (PF) eigenvalue and eigenfunction of a valuation operator.[2] PF eigenvalues of appropriate valuation operators are an instance of long run cost benefit rules; they are long run growth factors that can be used to rank risky projects that mature in the distant future. PF eigenvalues are important examples of these rules, but they are poor guides to their general structure. They are complex objects to compute and perform comparative statics exercises with, and the ergodic Markov property that underlies them corresponds to a particular functional form for $I(q)$ in (4). Importantly, the effect of risk on long run values, while implicit in PF eigenvalues, is highly non-intuitive, and all but impossible to understand directly from the eigenfunction problems that have been the focus of the finance work.

As a simple example of the issues involved, imagine two economies, X and Y, that evolve on a discrete state space $\mathcal{S} = \{1, 2\}$. Both economies are Markovian, with transition probability matrices given by

$$\mathbf{T}_X = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}, \quad \mathbf{T}_Y = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}. \tag{5}$$

Consider a risky project with state dependent 1-period discounted returns $\vec{L} = (L_1 \; L_2)$ where $L_1 \neq L_2$. Assume that the project will mature in the very distant future, and hence its value

[2] Hansen & Scheinkman (2009); Hansen (2012) generalize the discrete time ergodic model of Alvarez & Jermann (2005). Their work has in turn been generalized to semi-martingales by Qin & Linetsky (2017). PF eigenvalues are also important components of the analysis of normative disagreement and long run discounting in Millner (2020); Jaakkola & Millner (2023).

is reflected in the long run growth rate of expected discounted returns (I formalize this notion of long run value below). If investors can invest in a fixed project $\vec{L}$ in either economy, which should they choose? An intuitive response to this question might be to note, by the symmetry of the transition matrices $\mathbf{T}_X$ and $\mathbf{T}_Y$, that both economies have the stationary distribution $(0.5 \quad 0.5)$. The stationary distribution describes the long run statistics of ergodic Markov chains with probability 1, suggesting that investors should be indifferent between X and Y. And yet, a more careful calculation of the long run certainty equivalent rates of return in these two economies – which are increasing transformations of PF eigenvalues of appropriate matrices[3] – shows that this intuitive reasoning would lead investors to make a large mistake. In fact, for *any* risky asset $\vec{L}$,

$$\lim_{t\to\infty} \frac{\text{Expected Present Value of } \vec{L} \text{ at maturity } t|\mathbf{T}_X}{\text{Expected Present Value of } \vec{L} \text{ at maturity } t|\mathbf{T}_Y} \to \infty.$$

How can it be that two economies with identical long run statistics (with probability 1) nevertheless give rise to such different valuations for long run projects? The answer must have something to do with differences in the way these economies produce events that are asymptotically of measure zero. Indeed, the entire contribution of risk to long run values (i.e. PF eigenvalues in this example) comes from such measure zero events, a point emphasized in the qualitative analysis of Martin (2012) in a different context.[4] Any preference relation that ranks long run projects must thus be non-probabilistic in general, as it needs to resolve differences between measure zero events, a task beyond the reach of probabilistically sophisticated preferences. The variational preferences in (22) will turn out to do the job for a large class of stochastic economies, with the function $I(q)$ encoding information about the *rate* at which the probabilities of atypical (i.e., asymptotically measure zero) events decline to zero.

On the other side of the duality between long run cost benefit rules and variational preferences lies the decision theory literature, which has investigated variational preferences of the kind in (4) in a manner largely divorced from their connection to cost benefit analysis.[5]

[3]State-dependent expected discounted returns of project $\vec{L}$ at maturity t are given by the row vector $\vec{E}_t = (\mathbf{TL})^t(1\ 1)^*$, where $\mathbf{L} = \begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}$, $\mathbf{T}$ is the relevant economy's transition matrix, and $*$ is the transpose operator. The long run certainty equivalent rate of return of $\vec{L}$ is $\lim_{t\to\infty} \frac{1}{t} \log \vec{E}_t$. The Perron-Frobenius theorem tells us that this quantity is independent of the initial state of the economy, and given by the log of the PF eigenvalue of $\mathbf{TL}$, which is real and positive.

[4]Martin (2012) assumes a perfect markets framework in which asset prices are martingales. For public cost benefits analysis these assumptions need not hold (perhaps especially for long run projects where markets are highly incomplete, see Millner & Heal (2023)), and it does not hold in the simple example I study here. Moreover, while Martin (2012) identifies the importance of measure zero events for long run values, he does not elucidate the *structure* of preference relations over long run projects, i.e. how risk can be represented for these preferences, associated stochastic orders, when they are probabilistically sophisticated, their relationship to variational ambiguity preferences, etc. The latter questions are the focus of the current paper.

[5]The axiomatic properties of variational preferences were elucidated by Maccheroni et al. (2006). This model is a generalization of the 'minmax expected utility' model of Gilboa & Schmeidler (1989), and the multiplier preferences model popularized by Hansen & Sargent (2001) – see Strzalecki (2011); Cerreia-Vioglio

There is however an important exception where conceptually related ideas can be found. Robson et al. (2023) have demonstrated an equivalence between static rational inattention and 'wishful thinking' models popular in the behavioral literature and i.i.d. stochastic growth processes. Preferences in these models can, but need not, be expressed as variational problems. The results in this paper show that these insights can be considerably generalized once we move away from stochastic models that are dual to probabilistically sophisticated decision theories.[6] Once we do this, a variational representation of preferences becomes necessary, not optional. In addition, the generality of the theory allows for a much richer understanding of the properties of preferences derived from the long run behavior of multiplicative functionals of stochastic processes; these insights would be impossible to obtain in a specific parametric model (e.g. the i.i.d. processes used in Robson et al. (2023)).

The remainder of the paper is structured as follows. Section 1 defines and characterizes long run cost benefit rules, establishing their variational representation and its connection to long run discount rates. Section 2 defines, characterizes, and interprets a natural notion of stochastic dominance for long run cost benefit rules, illustrating its application to resolving puzzles of the kind in (5), and demonstrating its connection to orderings of the persistence of stochastic processes. Section 3 investigates conditions under which long run cost benefit rules are probabilistic, and the behavioral consequences of violations of those conditions. Section 4 concludes.

1 Variational Representation of Long Run Cost Benefit Rules

Cost benefit analysis is a method for ranking *marginal* public projects, i.e., projects whose payoffs are small relative to a baseline aggregate consumption process.[7] Project-specific information is encoded in a stochastic sequence of *returns* $\mathbf{R} = (R_1, R_2, \ldots)$, where R_t (a random variable) is the return the project yields between maturities $t-1$ and t. Details of the valuation framework are captured by a sequence of 1-period *stochastic discount factors* $\mathbf{M} = (M_0, M_1, \ldots)$, where M_t (also a random variable) converts payoffs realized at maturity $t+1$ into their value in terms of consumption at maturity t. R_t and M_t are positive, bounded, random variables for all $t \in \mathbb{N}^+$; they depend on the (stochastic) state of the economy. Cost benefit rules combine these sequences of random variables into expected present values, which

et al. (2011) for discussions of the relationships between these models of ambiguity aversion.

[6]The paper also partially addresses a conjecture of Robson et al. (2023): 'A natural extension of our setting would allow for serial correlations of the payoff states. We conjecture that the optimal choice arising under serial correlations can be represented by the dynamic extension of the rational inattention problem studied in Steiner, Stewart, and Matejka (2017).' Serially correlated risks and other forms of dynamical structure are an important part of this analysis; I show that the variational ambiguity model is the correct tool for understanding the long run behavior of many stochastic processes that exhibit these features.

[7]We need project payoffs to be small enough that the changes in social welfare they cause are well approximated by a first order Talyor expansion, and general equilibrium effects can be neglected – see Millner & Heal (2023).

are used to rank projects.

1.1 Preliminaries

As an initial example of this framework, suppose that a social planner has preferences over risky consumption streams that can be represented by

$$W = \mathsf{E}_0 \sum_{t=0}^{\infty} e^{-\rho t} U(c_t),$$

where the expectation is over stochastic values of the (per capita) consumption sequence $\{c_t\}_{t=0\ldots\infty}$, and $U(c) = \frac{c^{1-\eta}}{1-\eta}$, where $\eta \geq 0$. Suppose that consumption is a function of a Markov process on a state space $\mathcal{X}$, i.e. $c_t = c(x_t)$ for some policy function $c(x)$, and x_t is a Markov process on $\mathcal{X}$. 1-period consumption growth g_{t+1} between period t and $t+1$ is then given by $g_{t+1} = g(x_{t+1}, x_t) := \log(c(x_{t+1})/c(x_t))$. In this case, the planner's 1-period stochastic discount factor at maturity t is

$$\begin{aligned} M_t = M_t(x_t, x_{t+1}) &:= \frac{\partial W}{\partial c_{t+1}(x_{t+1})} / \frac{\partial W}{\partial c_t(x_t)} = e^{-\rho} U'(c_{t+1}(x_{t+1}))/U'(c_t(x_t)) \\ &= \exp(-[\rho + \eta g(x_{t+1}, x_t)]). \end{aligned} \tag{6}$$

This quantity is the social marginal rate of substitution between consumption at maturity $t+1$ in state x_{t+1}, and consumption at maturity t in state x_t. M_t is itself a random variable, since x_t and x_{t+1} are uncertain today (i.e., at $t = 0$). From the perspective of maturity t, a marginal stochastic return $R_{t+1}(x_{t+1})$ realized at time $t+1$ in state x_{t+1} is valued according to its expected present value, conditional on information available at t:

$$\mathsf{E}_t M_t R_{t+1} = \mathsf{E}_t M_t(x_t, x_{t+1}) R_{t+1}(x_{t+1}) \tag{7}$$

In the Markov case the conditioning information for the expectation E_t would be encapsulated in the value of the state variable x_t at time t, but the left hand side of the formula (7) is completely general; it does not rely on the Markov assumption, or a particular model of planner preferences. Equation (7) says simply that risky marginal payoffs realized at maturity $t+1$ must be valued using a 'price' M_t given by the state-dependent social marginal rate of substitution between consumption at $t+1$ and t. That marginal rate of substitution captures both the relative value of payoffs with maturities $t+1$ and t, and the relative value of payoffs realized in different states of the world, as illustrated in the Markovian example (6). We can iterate the 1-period valuation procedure encapsulated in (7) back to the present to find the value a project with return sequence $\mathbf{R}$ yields at maturity t:

$$v_t(\mathbf{R}; \mathbf{M}) := R_0 \mathsf{E}_0 \Pi_{k=0}^{t-1} M_k R_{k+1}.$$

To this point the formalism has been completely general; I now impose some structure on the stochastic process that generates risk in the economy. Begin by defining a new random variable λ_k through

$$\exp(-\lambda_k) := M_k R_{k+1}. \tag{8}$$

Then we can rewrite

$$v_t(\mathbf{R}; \mathbf{M}) = R_0 \mathsf{E}_0 \exp\left(-\sum_{k=0}^{t-1} \lambda_k\right). \tag{9}$$

My first assumption is that λ_k can be represented as a function of a stochastic process on a state space $\mathcal{S}$:

$$\lambda_k = \lambda(R; M)(s_k),\ \ s_k \in \mathcal{S} \tag{10}$$

for some function $\lambda(R; M)$.[8] This setup is similar to the Markovian example in (6), where the relevant state space was the product space $\mathcal{X} \times \mathcal{X}$. The Markov assumption will turn out to be far more restrictive than is necessary (or desirable) for our purposes, but it is a convenient example that I will often turn to for intuition. The notation in (10) makes the dependence of λ_k on the return process $R_{k+1} = R(s_{k+1})$, and stochastic discount factor $M_k = M(s_k)$, explicit. Note that the sequence of stochastic discount factors is fixed in any cost benefit exercise (although realizations are random); only the return sequences vary across projects.

The properties I impose on the state space $\mathcal{S}$ and the functions $\lambda(R; M)$ are as follows:

Assumption 1. $\mathcal{S}$ is either convex and compact, or finite. The functions $\lambda(R; M) : \mathcal{S} \to \mathbb{R}^+$ are bounded and continuous in the weak topology, with $\inf_{s\in\mathcal{S}} \lambda(R; M)(s) > 0$.

Boundedness and continuity are regularity conditions. Positivity ensures that projects have finite present value for any realization of the stochastic process s_k (this condition will be relaxed later on).

Given this assumption, we can rewrite the argument of the exponent in (9) as follows:

$$\sum_{k=0}^{t-1} \lambda_k = \sum_{k=0}^{t-1} \lambda(R; M)(s_k) = t \int_{\mathcal{S}} \lambda(R; M) dq_t(\kappa), \tag{11}$$

where $q_t(\kappa)$ is the *empirical measure* associated with the sequence $\kappa = \{s_0, s_1, \ldots, s_{t-1}\}$. The empirical measure simply reports the frequency with which κ visits subsets of $\mathcal{S}$. Formally, recall that the Dirac measure is defined by

$$\delta_s(\mathcal{A}) = \begin{cases} 1 & s \in \mathcal{A} \\ 0 & s \notin \mathcal{A} \end{cases},$$

where $s \in \mathcal{S}$, $\mathcal{A} \subseteq \mathcal{S}$. Given a sequence $\kappa = (s_0, s_1, \ldots)$ the empirical measure at maturity t

[8] This assumption may seem to restrict the explicit dependence of M_t on t, but all the results that follow go through if this restriction is only required to hold asymptotically.

is:

$$q_t(\mathcal{A};\kappa) = \frac{1}{t}\sum_{n=0}^{t-1}\delta_{s_n}(\mathcal{A}). \tag{12}$$

The empirical measure captures information about a deterministic realization of the sequence κ. Risk is represented by a measure over sequences, i.e., a measure on empirical measures, due to (11). Let $\mu_t \in \Delta(\Delta(\mathcal{S}))$ be the measure on empirical measures at maturity t generated by the underlying stochastic process on $\mathcal{S}$. I will sometimes abuse notation and write this as $\mu_t(q)$, as a reminder that it is a measure on empirical measures $q \in \Delta(\mathcal{S})$. Thus from (9) and (11) we have

$$v_t(\mathbf{R};\mathbf{M}) = v_t(\lambda(R;M)) := R_0 \mathsf{E}_{\mu_t(q)} \exp\left[-t\int \lambda(R;M)dq\right] \tag{13}$$

where the expectation is taken over all $q \in \Delta(\mathcal{S})$ with the measure $\mu_t(q)$. All the dynamics are now captured by $\mu_t(q)$.

1.2 Large Deviation Principles

My core structural assumption on the stochastic process that generates risk in this model concerns its asymptotic behavior as maturity $t \to \infty$. The important property I will require is that the stochastic process for the empirical measure q_t satisfy a *large deviation principle*. At a high conceptual level, this can be understood as a generalization of the law of large numbers, i.e., the statement that the average of n samples of a random variable with finite mean converges (in probability) to the mean as $n \to \infty$. Just as in that familiar instance, stochastic processes that obey large deviation principles also concentrate all their probability mass on a set of 'typical' events (e.g. sequences consistent with the law of large numbers) as $t \to \infty$. But in addition to this, the probabilities of all 'atypical' events decline to zero at least exponentially fast asymptotically. Large deviation theory allows us to differentiate these asymptotically measure zero atypical events from one another by studying the *rates* at which their probabilities decline to zero. Examples of stochastic processes whose empirical measures obey large deviation principles include i.i.d risks, martingales, ergodic Markov processes, and more general measure preserving dynamical systems (Dembo & Zeitouni, 2009; Budhiraja & Depuis, 2013; Shalizi & Kontorovich, 2007) – almost all of the stochastic processes used in cost benefit analysis fall into one of these categories.[9]

A crucial feature of stochastic processes that satisfy large deviation principles is that all

[9] In applications in finance a no-arbitrage condition is usually imposed. This causes discounted returns to be martingales, and gives rise to a semi-group structure for pricing kernels (Hansen & Scheinkman, 2009; Hansen, 2012). Such processes are known to obey large deviation principles (Kurtz & Feng, 2010). In applications in social discounting and cost benefit analysis more broadly no-arbitrage need not be imposed, but the stochastic processes that generate risk almost always possess other properties – ergodicity, or more generally measure preserving dynamics – that give rise to a large deviation principle (see e.g. Shalizi & Kontorovich, 2007). Gollier (2012); Ljungqvist & Sargent (2004) contain many examples of such models.

long run risk is asymptotically measure zero. We can see this in the familiar case of the law of large numbers – if the large time average of a sequence of i.i.d. random variables tends to the mean with probability 1, the only place left for variability in that average is in the measure zero events. Since expectations are insensitive to measure zero events, it is intuitive that a different formalism is needed if we are to resolve risk in long run outcomes. What will become clear is that resolving these measure zero events is essential if we hope to represent risk in long run cost benefit rules.

As an intuitive initial illustration of a situation where a large deviation principle for the empirical measure holds, suppose that s_t is a sequence of i.i.d. Bernoulli distributed random variables taking values in $\{0, 1\}$ with $\text{Prob}(1) = p \in (0, 1)$. For an arbitrary set $\mathcal{B} \subset [0, 1]$ and positive integer t, define

$$\Gamma_t(\mathcal{B}) = \{n/t | n = 0 \dots t\} \cap \mathcal{B}.$$

The probability that the empirical frequency of ones $q_t = \frac{1}{t}\sum_{n=0}^{t-1} s_n$ in a sequence of length t lies in the set $\mathcal{B}$ is then[10]

$$\mu_t(\mathcal{B}) = \sum_{q \in \Gamma_t(\mathcal{B})} \binom{t}{tq} p^{tq}(1-p)^{t(1-q)}. \tag{14}$$

Using Stirling's formula[11] to approximate the $t \to \infty$ behavior of the binomial coefficient, one can easily show that (ignoring divisibility issues),

$$\lim_{t\to\infty} \frac{1}{t} \log \left[\binom{t}{tq} p^{tq}(1-p)^{t(1-q)} \right] = -D(q||p) \tag{15}$$

where $D(q||p)$ is given by

$$D(q||p) = q \log \frac{p}{q} + (1-q) \log \frac{1-p}{1-q}. \tag{16}$$

$D(q||p)$ is non-negative, convex in q, and equal to zero only if $q = p$. This function is known as the relative entropy of the probability distribution $(1-q, q)$ with respect to $(1-p, p)$. Thus, as t becomes large, the terms in the sum (14) behave like $\exp(-tD(q||p))$. Since (14) is a sum of terms that are declining to zero exponentially as $t \to \infty$, the term with the smallest rate of decay will dominate for large t (see e.g. Hardy et al., 1934), i.e.,

$$\lim_{t\to\infty} \frac{1}{t} \log \mu_t(\mathcal{B}) = -\inf_{q\in\mathcal{B}} D(q||p) \tag{17}$$

for 'nice' sets $\mathcal{B}$ where this limit exists. This is of course another instance of the Laplace principle, this time applied to probabilities. I have played fast and loose with rigor at several

[10] The following formula uses the notation $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ for the binomial coefficient.

[11] Stirling's formula says that $\log n! = n \log n - n + O(\log n)$.

points in this exposition – Dembo & Zeitouni (2009) contains a careful treatment.

The relationship in (17) is an instance of the defining feature of large deviation principles for the empirical measure. The relative entropy $D(q||p)$ is an example of a *rate function*, i.e., a function that controls the asymptotic rate of decay of the probabilities of sets in $\Delta(\mathcal{S})$ under a sequence of measures in $\Delta(\Delta(\mathcal{S}))$. It is apparent that this example can be extended to any i.i.d. risk; moreover, the relative entropy continues to quantify the rate of decay of probabilities in the general i.i.d. case, a result known as Sanov's theorem. The i.i.d. assumption can also be relaxed. There are large classes of stochastic processes that exhibit serial dependence – e.g. ergodic Markov processes and martingales – but satisfy a large deviation principle. The rate function for the empirical measure does *not* correspond to the relative entropy in these more general cases, and can take a wide variety of forms. In fact, given an arbitrary non-negative function $I(q)$ with compact lower contour sets, it is always possible to construct a stochastic process that satisfies a large deviation principle with rate function $I(q)$.[12] This illustrates that requiring a large deviation principle for the empirical measure is much more general than, e.g., restricting the analysis to Markov processes, whose rate functions have a specific functional form (see Section 3).

In the general case, as before let $\Delta(\mathcal{S})$ be the set of probability measures on a set $\mathcal{S}$ that obeys Assumption 1, and let $\{\mu_t\}_{t\in\mathbb{N}^+}$ be a sequence of measures on $\Delta(\mathcal{S})$, i.e., $\mu_t \in \Delta(\Delta(\mathcal{S}))$ for each t. Think of μ_t as describing the distribution of the realized empirical measure q_t at maturity t. The sequence of measures $\{\mu_t\}$ satisfies a large deviation principle if for 'nice' sets $\mathcal{B} \subset \Delta(\mathcal{S})$,

$$\lim_{t\to\infty} \frac{1}{t} \log \mu_t(\mathcal{B}) = -\inf_{q\in\mathcal{B}} I(q) \tag{18}$$

where $I(q) \geq 0$ is a function with compact lower contour sets, i.e., a rate function.[13] In the language of the discussion above sets $\mathcal{B} \in \Delta(S)$ for which $\inf_{q\in\mathcal{B}} I(q) > 0$ are atypical, i.e., their probability declines exponentially to zero as $t \to \infty$ at a rate controlled by $I(q)$. Sets that only contain elements of $I^{-1}(0) \subset \Delta(\mathcal{S})$ are typical – all probability mass concentrates on these sets asymptotically, and the asymptotic rate of decline of their probability is zero. For example, for ergodic Markov processes, the only typical set is the singleton containing

[12] This can be achieved via a change of measure. Consider an i.i.d. stochastic process s_t whose empirical measure q has rate function $D(q||p)$, and denote the product measure on sequences κ at maturity t for this process by K_t. Now consider an alternative process s'_t, and assign probability at maturity t to sequences κ with empirical measure q_t proportional to $\exp(-I(q_t)t)\exp(D(q_t||p)t) \times K_t$. The second factor in this expression cancels out the rate function for i.i.d. processes from Sanov's theorem, and the first factor replaces it with the rate function $I(q)$. If $I(q) \geq 0$ has compact lower contour sets, the resulting process satisfies a large deviation principle with the required rate function.

[13] Compact lower contour sets means that $\{q \in \Delta(\mathcal{S})|I(q) \leq x\}$ is compact for all x. The fact that the rate function has compact lower contour sets ensures that it is lower-semicontinuous, and attains its infimum on closed sets. This property, when combined with the existence of a large deviation principle, implies that there exists a q at which the rate function is zero (see Budhiraja & Depuis, 2013, p.9). The large deviations literature sometimes distinguishes between weak and strong large deviation principles. When the rate function is 'good' (i.e. has compact lower contour sets), these two notions coincide. I follow Budhiraja & Depuis (2013) in simply defining a large deviation principle as a stochastic process with a 'good' rate function.

the unique stationary distribution of the process – the empirical measure approaches the stationary distribution with probability 1. For technical reasons my statement of the definition in (18) is a simplification; requiring the limit in (18) to exist is too strong for some sets (e.g. if $\mathcal{B}$ has measure zero). However, the more complete definition (which amounts to weak convergence on a logarithmic scale) reduces to this one for e.g. convex sets when the function $I(q)$ is continuous on its effective domain $\{q|I(q) < \infty\}$ (La Cour & Schieve, 2015). This continuity property is satisfied in all familiar models (e.g. Markov processes). Since I will not have much need for the finer points of the full definition, I refer the reader to Dembo & Zeitouni (2009) for details.

I summarize my assumptions about the stochastic process that generates risk as follows:

Assumption 2. The sequence of measures on empirical measures $\{\mu_t\}_{t\in\mathbb{N}^+}$ in (13) satisfies a large deviation principle with rate function $I(q)$.

Conditions on the stochastic process q_t that are necessary and sufficient for a large deviation principle to exist are known (Dembo & Zeitouni, 2009), but as I will not use them in what follows, it is more direct to work with this assumption.[14]

1.3 Variational representation

Thus far my analysis has focussed on the value a project realizes at a specific maturity t; cost benefit analysis sums these t-dependent values into total present values, and uses these to rank projects. I say that a project with return sequence $\mathbf{R}$ is *realized* at maturity t if it accrues value from t onwards. The value of such a project is

$$V_t(\mathbf{R};\mathbf{M}) = \sum_{k=0}^{\infty} v_{t+k}(\mathbf{R};\mathbf{M}). \tag{19}$$

where $v_t(\mathbf{R};\mathbf{M})$ is given by the expression in (13). A *cost-benefit rule* for projects that are realized at maturity t is an ordering on return sequences $\mathbf{R}$ generated by $V_t(\mathbf{R};\mathbf{M})$, for some fixed sequence of stochastic discount factors $\mathbf{M}$, and some stochastic process that generates risk. The result I have been working towards is a characterization of 'long run' cost benefit rules, i.e., social preferences over projects generated by (monotonic transformations of) $V_t(\mathbf{R};\mathbf{M})$ in the limit as $t \to \infty$. To state the result I make use of one further definition.

Fixing a stochastic discount factor sequence $\mathbf{M}$, I define a sequence of *certainty equivalent*

[14] Note that this assumption is a constraint on the asymptotic behavior of the stochastic process that drives 1-period *decay rates* of project value λ_k, and not on project value itself (i.e., $v_t(\mathbf{R};\mathbf{M})$), which is unlikely to be asymptotically stationary.

rates of return $\{\phi_t(\mathbf{R}, \mathbf{M})\}_{t=1,\ldots,\infty}$ associated with a project $\mathbf{R}$ as follows:

$$R_0 \exp(t\phi_t(\mathbf{R}; \mathbf{M}))\mathsf{E}_0 \Pi_{k=0}^{t-1} M_k := v_t(\mathbf{R}; \mathbf{M})$$
$$\Rightarrow \phi_t(\mathbf{R}; \mathbf{M}) = \frac{1}{t} \log \left[\frac{\mathsf{E}_0 \Pi_{k=0}^{t-1} M_k R_{k+1}}{\mathsf{E}_0 \Pi_{k=0}^{t-1} M_k} \right]. \tag{20}$$

$\phi_t(\mathbf{R}; \mathbf{M})$ is the (constant) rate of a return a risk-free project would need to earn to yield value equal to $v_t(\mathbf{R}; \mathbf{M})$ (the value of the risky project $\mathbf{R}$) at maturity t. Under Assumption 1 we can re-express $\phi_t(\mathbf{R}; \mathbf{M})$ in terms of expectations over empirical measures:

$$\phi_t(\mathbf{R}; \mathbf{M}) = \phi_t(R, M) := \frac{1}{t} \log \frac{\mathsf{E}_{\mu_t(q)} \exp\left[-t \int \lambda(R; M) dq\right]}{\mathsf{E}_{\mu_t(q)} \exp\left[-t \int \lambda(1; M) dq\right]}. \tag{21}$$

The central result of this section is the following (all proofs are in the appendix):

Theorem 1. *Under Assumptions 1 and 2, the following are equivalent:*

(i)

$$\lim_{t\to\infty} \frac{1}{t} \log V_t(\mathbf{R}; \mathbf{M}) \geq \lim_{t\to\infty} \frac{1}{t} \log V_t(\mathbf{R}'; \mathbf{M})$$

(ii)

$$\lim_{t\to\infty} \phi_t(R, M) \geq \lim_{t\to\infty} \phi_t(R', M)$$

(iii)

$$\inf_{q\in\Delta(\mathcal{S})} \left[\int \lambda(R; M) dq + I(q) \right] \leq \inf_{q\in\Delta(\mathcal{S})} \left[\int \lambda(R'; M) dq + I(q) \right], \tag{22}$$

where $I(q)$ is the rate function associated with the sequence of measures on empirical measures $\{\mu_t\}_{t\in\mathbb{N}^+}$ in Assumption 2. The limits in (i) and (ii) exist.

This result says that long run present values and long run certainty equivalent rates of return are ranked consistently, and that both rankings can be represented by the variational expressions in (22). The rankings in this theorem apply to equivalence classes of return processes, defined by their long run certainty equivalent rates of return, as part (ii) of the result makes plain. This is because, while $\frac{1}{t} \log V_t$ is monotonically related to V_t for any finite t, this is no longer true in the $t \to \infty$ limit, as sub-exponential components of V_t are neglected by $\frac{1}{t} \log V_t$ in this limit. Assumptions 1 and 2 imply that project values decay exponentially asymptotically, making a logarithmic transformation of present values the natural choice for distinguishing between long run projects. Note that the direction of the inequality in part (iii) is opposite to that in parts (i) and (ii), because project value is inversely related to λ, the 1-period decay rate of project value (see (13)).

To get an intuition for part (iii) of the result – the most important and perhaps counter-intuitive part of it – begin by noticing that from (21),

$$\phi_t(R,M) \geq \phi_t(R',M) \iff \frac{1}{t}\log \mathsf{E}_{\mu_t(q)} \exp\left(-t\int_{\mathcal{S}} \lambda(R;M)dq\right) \geq \frac{1}{t}\log \mathsf{E}_{\mu_t(q)} \exp\left(-t\int_{\mathcal{S}} \lambda(R';M)dq\right). \tag{23}$$

Reasoning by analogy to the i.i.d. example above that $d\mu_t(q) \sim_{t\to\infty} \exp(-tI(q))dq$, we see that as $t \to \infty$ the integrand of the expectations in (23) behaves like

$$\exp\left(-t\left[\int_{\mathcal{S}} \lambda(R;M)dq + I(q)\right]\right).$$

Thus the expectations in (23) will be dominated by the terms that decay slowest as $t \to \infty$, i.e.,

$$\lim_{t\to\infty} \frac{1}{t}\log \mathsf{E}_{\mu_t(q)} \exp\left(-t\int_{\mathcal{S}} \lambda(R;M)dq\right) = -\left[\inf_{q\in\Delta(\mathcal{S})} \int_{\mathcal{S}} \lambda(R;M)dq + I(q)\right]. \tag{24}$$

A more rigorous treatment of this finding relies on Varadhan's lemma, a modern incarnation of the Laplace principle that is a cornerstone of large deviation theory. This intuitive discussion illustrates that the formula (24) is the outcome of a horserace between the decreasing *probabilities* of rare events (which scale like $\exp(-I(q)t)$ with t), and their atypical *consequences* (which scale like $\exp(-t\int \lambda dq)$ with t). It is the fact that the measure on empirical measures $\mu_t(q)$ has an exponential dependence on t that causes a failure of uniform convergence (on an exponential scale) in (24). For example, if $I(q)$ has a unique zero at q^* (i.e. all probability mass accumulates at q^* as $t \to \infty$), the measure on empirical measures $\mu_t(q)$ converges pointwise, but not uniformly, to a Dirac measure δ_{q^*} centered on q^*. Thus, in general, the limit of expectations is not equal to the expectation of the limit:

$$\lim_{t\to\infty} \frac{1}{t}\log \mathsf{E}_{\mu_t(q)} \exp\left(-t\int \lambda dq\right) \neq \lim_{t\to\infty} \frac{1}{t}\log \mathsf{E}_{\delta_{q^*}} \exp\left(-t\int \lambda dq\right) = -\int \lambda dq^*.$$

The representation of long run cost benefit rules in (22) will be the focus of the remainder of the paper. I will use the notation $\succeq_{I(q)}$ to denote an instance of the preference ordering over long run projects in (22), when the stochastic process that generates risk has rate function $I(q)$. Readers familiar with the literature on ambiguity aversion will recognize that the preferences $\succeq_{I(q)}$ are reminiscent of Variational Preferences (Maccheroni et al., 2006). The precise relationship between these two models is discussed in Appendix B. The most important technical part of that discussion is that in order for (22) to be an instance of variational preferences the rate function $I(q)$ must be convex. Appendix B explains that convexity of the rate function is guaranteed if we impose a natural differentiability condition on long run cost benefit rules. The large deviations literature has also identified primitive properties of stochastic processes that ensure convexity of the rate function, see Lemma 4.1.21 in Dembo

& Zeitouni (2009) for further details. I will assume this condition is satisfied in what follows:

Assumption 3. The rate function $I(q)$ in (22) is convex.

While long run cost benefit rules and variational preferences are formally very closely related, their interpretations are substantively different. The analogue of the rate function $I(q)$ in the variational ambiguity model is a penalty function $C(q)$, which captures ambiguity *attitudes*. By contrast, the rate function in (22) captures information about *risk*. Thus the roles of tastes and beliefs are reversed in these two models. This has important consequences that I explore in the next section.

Notice that (22) requires us to minimize a functional over all possible probability measures $q \in \Delta(\mathcal{S})$. The infimal q in this problem, call it $q^*(R; M)$, will in general vary with the function $\lambda(R; M)$.[15] Thus we can write the condition in (23) as:

$$\int_{\mathcal{S}} \lambda(R; M) dq^*(R; M) + I(q^*(R; M)) \leq \int_{\mathcal{S}} \lambda(R'; M) dq^*(R'; M) + I(q^*(R'; M)).$$

This comparison involves probability measures, but a *different* measure is needed for each project that is being evaluated – there are no stable, project-independent, probabilistic beliefs that inform preferences in this model. This is a hallmark of a failure of probabilistically sophistication (Machina & Schmeidler, 1992). Indeed long run cost benefit rules are generically non-probabilistic, except when the rate function $I(q)$ has special structure. I investigate the structure on $I(q)$ needed to give rise to probabilistic preferences in Section 3.

1.4 Long run discount rates

Theorem 1 concerns the preference ordering generated by expected present values for long run projects, and will be the focus of most of what follows. However, it is also of interest to study long run discount rates. These quantities are components of cost benefit rules, they do not generate a preference ordering in themselves. Nevertheless, they provide a useful formal connection between long run run cost benefit rules and previous work on long run valuation.

Given a return sequence $\mathbf{R}$ and stochastic discount factor sequence $\mathbf{M}$, define a project specific discount rate $r_t(\mathbf{R}; \mathbf{M})$ at maturity t through:

$$\begin{aligned} v_t(\mathbf{R}; \mathbf{M}) &:= \exp(-t\, r_t(\mathbf{R}; \mathbf{M})) \mathsf{E}_0 \Pi_{k=0}^{t-1} R_k \qquad (25) \\ \Rightarrow r_t(\mathbf{R}; \mathbf{M}) &= -\frac{1}{t} \log \left[\frac{\mathsf{E}_0 \Pi_{k=0}^{t-1} M_k R_{k+1}}{\mathsf{E}_0 \Pi_{k=0}^{t-1} R_{k+1}} \right]. \end{aligned}$$

The *risk-free* discount rate at maturity t is the special case of $r_t(\mathbf{R}; \mathbf{M})$ in which $R_t = 1$ for sure for all t.

[15] I assume that the infimum is attained in this discussion. This can be proved under mild conditions on $I(q)$ (Budhiraja & Depuis, 2013), but the points I make here about probabilistic sophistication do not depend on these assumptions.

Proposition 1. *Under Assumptions 1 and 2,*

$$\lim_{t\to\infty} r_t(\mathbf{R};\mathbf{M}) = \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;M)dq + I(q)\right] - \inf_{q'\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;1)dq' + I(q')\right]. \qquad (26)$$

This proposition demonstrates that, in addition to their role in long run cost benefit rules, variational formulas of the type in (22) are important for understanding long run discount rates.

The comparative exercises that are meaningful for discount rates are however different from those that are meaningful for returns. Since discount rates must be combined with information about undiscounted returns to yield a ranking of projects (this is what cost benefit rules do, see (25)), it is not meaningful to compare project-specific discount rates associated with different return processes, unless we hold a measure of undiscounted returns fixed in this comparison. The long run undiscounted rate of return is represented by the second term in (26). Thus, a comparison of this kind boils down to

$$\inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;1)dq + I(q)\right] = \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R';1)dq + I(q)\right] \Rightarrow$$

$$\lim_{t\to\infty} r_t(\mathbf{R};\mathbf{M}) \geq \lim_{t\to\infty} r_t(\mathbf{R}';\mathbf{M}) \iff \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;M)dq + I(q)\right] \geq \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R';M)dq + I(q)\right].$$

This is of course the mirror image of the ordering generated by long run cost benefit rules. Alternatively, another meaningful comparative exercise for discount rates could be to hold the return process fixed, but vary the stochastic discount factor. In this case we have

$$\lim_{t\to\infty} r_t(\mathbf{R};\mathbf{M}) \geq \lim_{t\to\infty} r_t(\mathbf{R};\mathbf{M}') \iff \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;M)dq + I(q)\right] \geq \inf_{q\in\Delta(\mathcal{S})}\left[\int_{\mathcal{S}} \lambda(R;M')dq + I(q)\right].$$

This is again a mirror image long run cost benefit rule, but this time with the roles of R and M interchanged. The special case $R = 1$ corresponds to a comparison of risk-free long run discount rates for different stochastic discount factors. This shows that comparisons of long run risk free discount rates under different discounting schedules are formally identical to long run cost benefit rules. It is in this sense that this model provides a substantial generalization of the insights of Weitzman (1998, 2001), and the many subsequent commentaries on his approach. Weitzman and his followers focus on long run risk free discount rates but, as noted in the introduction, the models of risk that have been the focus of this literature are static, i.e., risk resolves after a single period, it has no non-trivial dynamics. However, as the formula above makes plain, the asymptotic dynamics of the stochastic process that generates risk, captured by the rate function $I(q)$, will generally strongly affect long run discount rates. The variational formalism developed above provides a tractable and general framework for analyzing long run discounting – as I have noted, a wide variety of empirically relevant stochastic processes satisfy Assumptions 1 and 2. In the remainder of the paper I will focus on long run cost benefit rules themselves, with the understanding that any properties established for these preference relations on returns have direct implications for long run discount rates via Proposition 1.

***

In some respects, long run cost benefit rules are simpler than standard cost benefit rules. Instead of having to rank full dynamic stochastic streams of returns (as required by standard expected present values), all dynamical information is compressed into the rate function in long run cost benefit rules. They can, in effect, be treated as static preferences.[16] However, they are unusual objects, that behave quite differently from more familiar preferences based on expectations.

In the next two sections I study questions that help to elucidate the structure and interpretation of long run cost benefit rules. In Section 2, I define and characterize a notion of long run stochastic dominance, i.e., a ranking of risks induced by long run cost benefit rules. I show that long run stochastic dominance can be conceptualized as a generalized non-parametric ranking of 'long run risk' (Duffie & Epstein, 1992), i.e., the persistence of stochastic processes. In addition, given the unusual variational form of long run cost benefit rules, a natural question is if and when they reduce to something more recognizable. Under what conditions are they probabilistic, and can those conditions be expected to hold in practice? I investigate this question in Section 3, and show its relevance for understanding the distinctions long run cost benefit rules draw between stochastic economies with probabilistically identical long run statistics.

2 Long Run Stochastic Dominance

Stochastic dominance allows us to rank risks for large classes of preferences. For example, in the case of expected utility theory, risks that are ordered by first order stochastic dominance can be ranked for any increasing utility function, and those that are ordered by second order stochastic dominance are ranked for any increasing and concave utility function.

There is no obvious analogue of the utility function in the variational representation of cost benefit rules (22). The form of this representation is determined by the structure of the cost benefit rule itself, which does not admit a richer set of attitudes to e.g. temporal fluctuations in value, which could be used to introduce nonlinearities into the expectation terms in (22). Preferences enter into this representation indirectly via the 1-period rates of discounted return λ, which incorporate stochastic discount factors, and hence capture social marginal rates of substitution. But viewed as a risk preference over λ, the only way we can define a broad class of such preferences to use as basis for a stochastic dominance relation is in terms of sets of the functions λ themselves. That is the approach I take in this section.

[16] Similarly, variational preferences arose as a model of ambiguity aversion in static choices

2.1 Definition and characterization

To begin to state a result in this direction, I first need to weaken a technical assumption I have been working with until now, i.e., that the functions λ are positive valued. In this section I will allow λ to be *any* bounded and continuous function from $\mathcal{S}$ to $\mathbb{R}$, including negative valued functions. The only place I have used the positivity of λ was in showing that the limiting present values $\lim_{t\to\infty}\frac{1}{t}\log V_t$ and certainty equivalent rates of return $\lim_{t\to\infty}\phi_t$ in Theorem 1 rank returns consistently. If λ may be negative, V_t does not exist in general, so this result does not hold. However, when λ is negative, and Assumption 2 holds, maturity t project value v_t still grows exponentially as $t\to\infty$. While the present values V_t are undefined in this case, certainty equivalent rates of return ϕ_t are perfectly well defined, and continue to provide a meaningful ranking of projects in the $t\to\infty$ limit. Of course, the $t\to\infty$ limit is really the only maturity that matters when project value grows with maturity, so long run cost benefit rules can be thought of as the *only* relevant cost benefit rules in this case. Thus, the introduction of negative λ comes at no cost to the conceptual foundations of the problem I study, and provides a useful generalization to projects whose returns overwhelm the effects of discounting.

Denote the set of bounded and continuous functions from $\mathcal{S}$ to $\mathbb{R}$ by $C_b(\mathcal{S})$. As our discussion of Theorem 1 showed, certainty equivalent rates of return ϕ_t can be expressed as (see (23))

$$\phi_t(\mu_t,\lambda) := \frac{1}{t}\log\frac{\mathsf{E}_{\mu_t(q)}\exp\left[-t\int\lambda dq\right]}{\mathsf{E}_{\mu_t(q)}\exp\left[-t\int\lambda_1 dq\right]}, \tag{27}$$

where $\lambda=\lambda(R;M)\in C_b(\mathcal{S}), \lambda_1=\lambda(1;M)\in C_b(\mathcal{S})$. For reasons that will become clear, this notation emphasizes the dependence of ϕ_t on the measure on empirical measures μ_t and the 1-period rate of discounted return λ, rather than R and M.

Definition 1. Let $\{\mu_t\}$ and $\{\nu_t\}$ $(t\in\mathbb{N}^+)$ be sequences of measures on empirical measures for stochastic processes that satisfy Assumptions 2 and 3, with rate functions $I_\mu(q)$ and $I_\nu(q)$ respectively. I say that $\{\mu_t\}$ *long run stochastically dominates* $\{\nu_t\}$ if

$$\lim_{t\to\infty}\phi_t(\mu_t,\lambda)\geq\lim_{t\to\infty}\phi_t(\nu_t,\lambda) \tag{28}$$

for all $\lambda\in C_b(\mathcal{S})$, where $\phi_t(\cdot,\lambda)$ is defined in (27).

It is instructive to think about what a stochastic order based on such a class of preferences would look like in expected utility theory. The analogous condition for expected utilities – a kind of 'zeroth order' stochastic dominance – would be to require that for probability measures p_μ, p_ν on $\mathcal{S}$, $\mathsf{E}_{p_\mu}U\circ y\geq\mathsf{E}_{p_\nu}U\circ y$ for all bounded, continuous utility functions $U:\mathcal{O}\to\mathbb{R}$ and acts $y:\mathcal{S}\to\mathcal{O}$ (the composition $U\circ y$ is the analogue of λ). This can be seen to require $p_\mu(\mathcal{A})\geq p_\nu(\mathcal{A})$ for all $\mathcal{A}\subset\mathcal{S}$. But since we also have $\int_S dp_\mu=\int_S dp_\nu=1$, this implies

$p_\mu = p_\nu$, and hence this notion makes no meaningful distinction between risks in expected utility theory, i.e., zeroth order stochastic dominance is trivial.

And yet, this order is *not* trivial for long run cost benefit rules. In that case all risk is measure zero; moreover, preferences are sensitive to measure zero events. Thus we can have some hope that ranking a measure of the relative 'size' of measure zero sets, as captured by rate functions, will be a non-trivial exercise. The crucial feature of long run cost benefit rules that allows for this possibility is that we do not need to impose the constraint that probabilities sum to 1. Since the contribution of risk to long run values is entirely through measure zero events, it is possible for rate functions (the long run analogues of probability measures) to be uniformly ordered without falling foul of that constraint.

The next result provides a characterization of long run stochastic dominance:

Proposition 2. $\{\mu_t\}$ *long run stochastically dominates* $\{\nu_t\}$ *if and only if*

$$\forall q \in \Delta(\mathcal{S}), \quad I_\mu(q) \leq I_\nu(q). \tag{29}$$

This result is a consequence of Fenchel-Moreau duality – see Appendix C.

Readers familiar with the ambiguity literature may recognize a similarity between this result and a dual concept in the variational ambiguity model. Recall that the analogue of our rate function $I(q)$ in that model is a penalty function $C(q)$, which captures ambiguity attitudes. Maccheroni et al. (2006) define a dominance relation on ambiguity attitudes,[17] with higher ranked preferences being 'more ambiguity averse'. Their definition is conceptually and technically different from long run stochastic dominance, but is characterized by a dominance relation on penalty functions $C(q)$ that is analogous to (29). Thus we see that variational ambiguity preferences and long run cost benefit rules are again mirror images of one another, with the roles of ambiguity attitudes and risk interchanged. Appendix B explains their result in further detail.

2.2 Markovian example

Long run stochastic dominance is a novel concept, so it is helpful to see how it applies to familiar stochastic processes to build intuition for its meaning. Consider a Markov chain on a discrete state space $\mathcal{S}_N$ with N elements, and denote its transition matrix by $\mathbf{T}$. I assume that the chain is ergodic, i.e., the matrix $\mathbf{T}$ is primitive.[18] Such Markov chains have a unique globally asymptotically stable stationary distribution $\vec{\pi}$ given by the solution of the eigenvector problem

$$\vec{\pi}\mathbf{T} = \vec{\pi}.$$

[17] This is a ranking of attitudes for a large class of risks, contrary to stochastic dominance, which is a ranking of risks for a large class of attitudes.

[18] $\mathbf{T}$ is primitive if there exists an integer k such that $\mathbf{T}^k > 0$.

By the Perron-Frobenius theorem, $\vec{\pi}$ has strictly positive elements, which can be normalized to sum to 1.

I denote the set of primitive transition matrices that have stationary distribution $\vec{\pi}$ by $\Omega(\vec{\pi})$. I show in Appendix D that $\Omega(\vec{\pi})$ is the convex hull of $N^2 - N - 1$ extremal matrices. By definition, all the Markov chains in $\Omega(\vec{\pi})$ have identical long run behavior from the perspective of classical risk measures, as they have identical stationary distributions that are approached with probability 1 – this is a consequence of ergodicity. Nevertheless, we have the following result:

Proposition 3. *Long run stochastic dominance is a complete order on* $\Omega(\vec{\pi})$ *when* $N = 2$. *When* $N > 2$ *there exist convex subsets of* $\Omega(\vec{\pi})$ *on which long run stochastic dominance is a complete order.*

As a simple illustration of this proposition, consider an economy with state space $S_2 = \{1, 2\}$, and state dynamics given by the ergodic Markov transition matrix $\mathbf{T}$. Consider any 2×2 diagonal matrix $\mathbf{\Lambda}$ with diagonal element $\Lambda_{ss} > 0$ representing the value of $\exp(-\lambda(s))$ in (8). Long run certainty equivalent rates of return in this economy are ordinally equivalent to the spectral radius (i.e. largest eigenvalue) of matrices of the form $\mathbf{T\Lambda}$; I denote such spectral radii by $\varrho(\mathbf{T\Lambda})$.[19] Consider the following transition matrices:

$$\mathbf{T}(a) = \begin{pmatrix} a & 1-a \\ 1-a & a \end{pmatrix}, \tag{30}$$

where $a \in (0, 1)$. These matrices have a common stationary distribution – (0.5 0.5) – that describes their long run behavior with probability 1. Nevertheless, Proposition 3 and its proof shows that the stochastic process generated by $\mathbf{T}(a)$ long run stochastically dominates the process generated by $\mathbf{T}(a')$ whenever $a > a'$. This implies that for any $\mathbf{\Lambda}$ and $a > a'$,

$$\varrho(\mathbf{T}(a)\mathbf{\Lambda}) \geq \varrho(\mathbf{T}(a')\mathbf{\Lambda}),$$

with equality only when $\Lambda_{11} = \Lambda_{22}$. Thus, even though the state dynamics associated with $\mathbf{T}(a)$ give rise to identical long run statistics for any a (with probability 1), the long run certainty equivalent rate of return of any risky asset is higher under $\mathbf{T}(a)$ than $\mathbf{T}(a')$ when $a > a'$.

2.3 Long Run Stochastic Dominance and Persistence

Examining the matrices $\mathbf{T}(a)$ in (30), it is clear that the effect of an increase in a is to increase the persistence of the Markov chain. This is suggestive of a more general relationship between long run stochastic dominance and increases in persistence. The notion of persistence

[19] See Footnote 3 for an explanation of this fact.

that is relevant here can be formalized in terms of a class of attitudes to *long run risk*, a concept that has its origins in the work of Duffie & Epstein (1992). A decision maker is long run risk loving if for any random variable that takes the value s_i with probability p_i, they prefer the single lottery over constant sequences $((s_i, s_i, s_i \ldots); p_i)$ to the sequence of lotteries $((s_i; p_i), (s_i; p_i), (s_i; p_i), \ldots)$.

It will be much more instructive for our purposes to express this requirement in terms of the sample paths of stochastic processes. Let $\mu_t^{\text{i.i.d.}}(q)$ be the measure on empirical measures q at maturity t for an arbitrary i.i.d. stochastic process whose realization s_t at time t is in $\mathcal{S}$.[20] Similarly, define

$$\mu_t^{\text{cert}}(q) = \begin{cases} p_i & q = \delta_{s_i} \\ 0 & \text{otherwise} \end{cases}$$

where again δ_{s_i} is the Dirac measure on $\mathcal{S}$ centered on s_i. Clearly then, we have

$$\forall t \in \mathbb{N}^+, \int_{\Delta(\mathcal{S})} q(s_i) d\mu_t^{\text{i.i.d.}}(q) = p_i = \int_{\Delta(\mathcal{S})} q(s_i) d\mu_t^{\text{cert}}(q)$$

and the definition of long run risk loving behavior amounts to

$$\mu_t^{\text{cert}}(q) \succ \mu_t^{\text{i.i.d.}}(q)$$

as $t \to \infty$. This framing of the definition makes the parallels between long run risk and 'conventional' static risk clear: we are comparing two measures over empirical measures that share a stationary 'marginal measure' (s_i, p_i), but where $\mu_t^{\text{cert}}(q)$ is a marginal preserving spread of $\mu_t^{\text{i.i.d.}}(q)$. In fact, $\mu_t^{\text{cert}}(q)$ is a 'maximal' marginal preserving spread of $\mu_t^{\text{i.i.d.}}(q)$, since it only places weight on extremal points in $\Delta(\mathcal{S})$, i.e. the Dirac measures centered on s_i.

By generalizing the kinds of marginal preserving spreads that are considered, we can use this sample-paths perspective on long run risk loving behavior to motivate a dominance ordering of long run risk, or persistence. By way of analogy, recall that in expected utility theory an agent is risk loving if they prefer facing a lottery to receiving its expected value – this is equivalent to the agent having a convex von Neumann-Morgenstern utility function. This definition immediately generates a stochastic dominance ordering: we can say that risk A dominates risk B if they have the same expected value, and A is weakly preferred to B by all risk loving agents with expected utility preferences. Similarly, we can turn the notion of long run risk loving preferences into an ordering of stochastic processes by identifying risks that are ranked consistently by all long run risk-loving preferences within a given class. I focus on the class of long run risk-loving preferences that can be expressed as expected utilities over

[20] This measure can be written down explicitly:

$$\mu_t^{\text{i.i.d.}}(\mathcal{B}) = \int_{\mathcal{B} \subseteq \Delta(S)} \left\{ \begin{array}{cc} \frac{t!}{(tq_1)! \cdots (tq_k)!} \prod_{i=1}^k p_i^{tq_i}, & q_i \in \{\frac{1}{t}, \frac{2}{t}, \ldots, \frac{t}{t}\} \\ 0 & \text{otherwise} \end{array} \right\} dq$$

.

empirical measures – this allows for the development of a theory that closely mirrors static second order stochastic dominance. An important consequence of this restriction is that the preferences in this class are indifferent to the ordering of states in deterministic sequences that share the same empirical measure – they do not exhibit impatience, consumption smoothing etc. This class of preferences thus focuses solely on long run risk; if instead preferences exhibited an intrinsic order dependence, one lottery over sequences could be preferred to another for reasons other than differences in their persistence.

Given this restriction on long run risk-loving preferences, a natural first attempt at a definition of when one stochastic process with sequence of measures on empirical measures $\{\mu_t\}$ 'long run risk dominates' another with the sequence of measures $\{\nu_t\}$ might be:

$\tilde{1}$.
$$\forall t \in \mathbb{N}^+, \forall \mathcal{A} \subseteq \mathcal{S},\ \int_{\Delta(\mathcal{S})} q(\mathcal{A})\, d\mu_t(q) = \int_{\Delta(\mathcal{S})} q(\mathcal{A})\, d\nu_t(q) \tag{31}$$

$\tilde{2}$.
$$\forall t \in \mathbb{N}^+,\ \int_{\Delta(\mathcal{S})} \Phi\left(q\right) d\mu_t(q) \geq \int_{\Delta(\mathcal{S})} \Phi\left(q\right) d\nu_t(q) \tag{32}$$

for all convex functions $\Phi : \Delta(\mathcal{S}) \to \mathbb{R}$.

The first condition is equality of marginal measures (the analogue of equal expected values in second order stochastic dominance). At a fixed t, the second condition requires dominance with respect to all convex expectations over q; this is equivalent to μ_t being a marginal preserving spread of ν_t.[21] The convex expectations in (32) are precisely the long run risk-loving preferences that can be represented as 'expected utilities' over empirical measures.

While this definition is perfectly sensible as a behavioral requirement, it unfortunately has a number of serious problems for the purposes of defining a dominance relation on stochastic processes. First, except in highly specialized cases such as comparing μ_t^{cert} to $\mu_t^{\text{i.i.d.}}$, we cannot expect stochastic processes to have stationary marginal distributions, or to reveal information about their persistence uniformly across time and initial conditions. For example, for Markov processes, the marginal measure on $\mathcal{S}$ at t varies with t, and the measure on empirical measures at a fixed t (and its dynamics over time) depend on the initial conditions of the process. This suggests that requiring the conditions in (31–32) to hold for all $t > 0$ is much too demanding. At least for the purposes of generating a dominance relation on stochastic processes, the definition of long run risk loving behavior must be modified to allow for a richer set of dynamic comparisons. The natural solution to this difficulty is to weaken the requirements (31–32) to hold asymptotically as $t \to \infty$. Any notion of long run risk dominance that results from such an approach will not depend on finite snippets of sequences, which may be non-representative of correlations in full sample paths of the stochastic process, and in addition

[21] A formally similar set of conditions is at the heart of Blackwell's characterization of the informativeness of experiments (Blackwell, 1953).

may be independent of initial conditions if the processes in question are weakly dependent.

Having made that adjustment, we run into a second difficulty: vacuity of the definition. To illustrate, consider again the case of ergodic Markov chains, and suppose that Φ is any bounded and continuous function. Then clearly any Markov chains that share a stationary distribution q^* are not strictly ordered by this definition since we have

$$\lim_{t\to\infty} \int_{\Delta(\mathcal{S})} \Phi(q)\, d\mu_t(q) = \Phi(q^*)$$

for any $\mu_t(q)$ that weakly converges to a Dirac measure centered on q^* as $t \to \infty$. The problem here is that the asymptotic differences between say the Markov chains in (30) are not visible to the convex expectation for fixed Φ. The atypical events that include all the information about differences in persistence between these chains need to be 'amplified' to an exponential scale in t to contribute to the rankings; this is not possible for fixed Φ.

This problem suggests a further weakening of the second requirement in our candidate definition. Instead of requiring the ranking to hold for any fixed convex function Φ, I consider a ranking based on sequences of convexity preserving transformations of convex expectations that match the asymptotic behavior of the measures $\mu_t(q), \nu_t(q)$.

Formally, a sequence of measures on empirical measures $\{\mu_t\}$ satisfies a large deviation principle with 'speed' a_t (a sequence in $\mathbb{R}$ with $\lim_{t\to\infty} a_t \to \infty$) if there exists a rate function $I(q)$ such that

$$\lim_{t\to\infty} \frac{1}{a_t} \log \mu_t(\mathcal{B}) = -\inf_{q\in\mathcal{B}} I(q)$$

for 'nice' sets $\mathcal{B} \subset \Delta(\mathcal{S})$ where the limit exists. Once again I am presenting a simplified version of the full definition of a large deviation principle; the goal is merely to observe that we can admit a wide variety of asymptotic behaviors for $\mu_t(q)$; we only require $d\mu_t(q) \sim \exp(-a_t I(q))dq$ as $t \to \infty$ for arbitrary speed a_t, which need not be linear in t.

Definition 2. A sequence of strictly increasing and convex functions $f_t : \mathbb{R} \to \mathbb{R}$ is said to **match** a sequence of measures on empirical measures $\{\mu_t\}$ that satisfies a large deviation principle with speed a_t if

$$\lim_{t\to\infty} \frac{1}{a_t} \log f_t(x) = \bar{f}(x), \quad \text{locally uniformly on } \mathbb{R},$$

where $\bar{f} : \mathbb{R} \to \mathbb{R}$ is a strictly increasing and convex function.

These properties of the limiting function $\bar{f}$ ensure that the following definition captures an asymptotic notion of a marginal preserving spread on $\Delta(\Delta(\mathcal{S}))$:

Definition 3. A stochastic process on $\mathcal{S}$ with a sequence of measures on empirical measures $\{\mu_t\}$ **long run risk dominates** another with the sequence of measures on empirical measures $\{\nu_t\}$ if

1. $\{\mu_t\}$ and $\{\nu_t\}$ satisfy large deviation principles with speeds a_t^μ, a_t^ν respectively, and

$$\lim_{t\to\infty} a_t^\mu / a_t^\nu = 1.$$

2. $$\lim_{t\to\infty} \int_{\Delta(\mathcal{S})} q(\mathcal{A})\, d\mu_t(q) = \lim_{t\to\infty} \int_{\Delta(\mathcal{S})} q(\mathcal{A})\, d\nu_t(q),\ \forall \mathcal{A} \subseteq \mathcal{S} \text{ where the limits exist}$$

3. There exists a sequence of functions $f_t : \mathbb{R} \to \mathbb{R}$ that matches the large deviation principles for μ_t, ν_t such that, for any convex $\Phi : \Delta(\mathcal{S}) \to \mathbb{R}$,

$$\lim_{t\to\infty} f_t^{-1} \left(\int_{\Delta(\mathcal{S})} f_t(\Phi\,(q)) d\mu_t(q) \right) \geq \lim_{t\to\infty} f_t^{-1} \left(\int_{\Delta(\mathcal{S})} f_t(\Phi\,(q)) d\nu_t(q) \right).$$

The first condition requires the two stochastic processes to be asymptotically comparable, i.e. they obey large deviation principles with the same asymptotic speed. The second condition says that the marginal measures on $\mathcal{S}$ associated with μ_t, ν_t must agree asymptotically.[22] The third condition says that μ_t is a 'transformed' marginal preserving spread of ν_t asymptotically when viewed on an appropriate t-dependent scale that makes the convex test functions $\Phi(q)$ of comparable 'size' to $d\mu_t(q)$ as $t \to \infty$. This is clearly a stronger constraint than merely requiring untransformed asymptotic convex expectations to be ranked. Note that $f_t(\Phi(q))$ is always convex in q when f_t is strictly increasing and convex; thus the functions $f_t(\Phi(q))$ (and their logarithmic limit) preserve convexity but introduce a t-dependence that allows the details of the asymptotic behavior of the measures μ_t, ν_t to be resolved.

With this definition in place, the relationship between long run risk dominance and long run stochastic dominance becomes clear:

Proposition 4. *Consider two sequences of measures on empirical measures $\{\mu_t\}, \{\nu_t\}$ ($t \in \mathbb{N}^+$), that obey Assumptions 2 and 3. Denote their rate functions by $I_\mu(q), I_\nu(q)$ respectively. If $I_\mu(q)$ and $I_\nu(q)$ each have a unique zero, $\{\mu_t\}$ long run stochastically dominates $\{\nu_t\} \Rightarrow \{\mu_t\}$ long run risk dominates $\{\nu_t\}$.*

Essentially, for processes satisfying the conditions in this proposition, long run stochastic dominance is the special case of long run risk dominance in which the speeds a_t^μ, a_t^ν are asymptotically linear in t, and the convex transformations $f_t(x) = \exp(tx)$.[23] The dominance requirement on rate functions (29) that characterizes long run stochastic dominance is

[22] If, for example, $I_\mu(q)$ has a unique minimizer q^*, then $\lim_{t\to\infty} \int q(\mathcal{A}) d\mu_t(q) = q^*(\mathcal{A})$ if and only if $q^*(\partial\mathcal{A}) = 0$, where $\partial\mathcal{A}$ is the boundary of $\mathcal{A}$. Hence the restriction to sets where the limit exists; this rules out sets with discontinuities at their boundaries. This qualifier may be dispensed with if $\mathcal{S}$ is discrete, as the limits always exist in that case.

[23] The minus sign in the exponent in (27) is irrelevant, since we require this condition to hold for all $\lambda \in C_b(\mathcal{S})$.

sufficient to rank limits of (transformed) expectations of *all* convex functionals of the empirical measure (i.e., $\Phi(q)$), not just the linear functionals (i.e., $\int \lambda dq$) used to define long run stochastic dominance.

A consequence of this fact is that for stochastic processes with strictly convex rate functions (which ensures the uniqueness of rate function zeros), long run stochastic dominance yields a generalized ranking of the persistence of these processes. Ergodic Markov processes such as those in Section 2.2 are examples of stochastic processes that have strictly convex rate functions.

3 When are long run cost benefit rules probabilistic?

The fact that long run cost benefit rules need not be probabilistic might be superficially surprising, as they are defined as limits of expected present values. Indeed, if the stochastic process s_t that generates risk is an i.i.d. sequence of random variables, it is immediate from (9) that project value at t is given by

$$v_t^{\text{iid}}(\lambda(R;M)) = R_0 \left[\mathsf{E}_s \exp(-\lambda(R;M)(s))\right]^t .$$

and hence from (19) the expected present value V_t of projects realized at t satisfies

$$\lim_{t\to\infty} \frac{1}{t} \log V_t(\lambda(R;M)) = \log(\mathsf{E}_s \exp(-\lambda(R;M)(s))).$$

In this case the long run cost benefit rule is clearly ordinally equivalent to an expected utility preference over the random variable s. We have also seen that in the i.i.d. case Sanov's theorem tells us that the large deviations rate function for the empirical measure q is given by the relative entropy. In its general form, given $p, q \in \Delta(\mathcal{S})$, the relative entropy of q with respect to p is

$$D(q||p) = \begin{cases} \int_{\mathcal{S}} \log\left(\frac{dq}{dp}\right) dq & q \ll p \\ \infty & \text{otherwise} \end{cases} \tag{33}$$

where $q \ll p$ means that q is absolutely continuous relative to p. This is a generalization of the binary expression in (16). From the proof of Theorem 1, it is immediate that

$$\log \mathsf{E}_s \exp(-\lambda(s)) = -\inf_{q\in\Delta(S)} \left[\int_S \lambda dq + D(q||p)\right]. \tag{34}$$

This expression is known as the Donsker-Varadhan variational formula for the cumulant; it forms the basis for the analysis in Robson et al. (2023). This begs two questions: are there other rate functions for which long run cost benefit rules reduce to expected utility theory, or more generally any probabilistic preference? In addition, what causes long run cost benefit rules to fail to be probabilistic?

3.1 Rearrangement invariance

To begin to address these questions, notice that the requirement that preferences be 'probabilistic' is really a symmetry property – it says that if two acts yield the same distribution of consequences according to some probability measure p we should be indifferent between them. Formally, a preference over generic acts $f, g : \mathcal{S} \to \mathcal{O}$ with symmetric part $\sim$ is *probabilistically sophisticated* if there exists a probability measure $p \in \Delta(\mathcal{S})$ such that

$$[\forall y \in \mathcal{O},\ p(\{s \in \mathcal{S} | f(s) = y\}) = p(\{s \in \mathcal{S} | g(s) = y\})] \Rightarrow f \sim g. \tag{35}$$

Probabilistic sophistication, initially defined in Machina & Schmeidler (1992), allows us to describe beliefs using a probability distribution on $\mathcal{S}$, but is much more general than expected utility theory (which requires preferences to be linear in probabilities).

To begin to investigate the implications of this symmetry requirement for the rate function, I start with a definition:

Definition 4. Fix probability measures $q, q', p \in \Delta(\mathcal{S})$. We say that q' is a *rearrangement* of q relative to p if there is a measurable bijection $\pi : \mathcal{S} \to \mathcal{S}$ such that $\forall \mathcal{A} \subseteq \mathcal{S}$:

1. $p(\pi(\mathcal{A})) = p(\mathcal{A})$ (i.e., π preserves p)
2. $q'(\mathcal{A}) = q(\pi(\mathcal{A}))$

The bijections π in this definition are just permutations of the state space that don't disturb the probability masses of events under p. If $\mathcal{S}$ is finite, π just permutes any subset of states S for which $\forall i, j \in S, p_i = p_j$. A rate function $I(q||p)$ is *rearrangement invariant relative to* p if for any two measure q, q' that are rearrangements relative to p, $I(q||p) = I(q'||p)$. As long as the set of permutations that preserve p is rich enough, there is a tight connection between rearrangement invariance and probabilistic sophistication

Proposition 5 (Maccheroni et al. (2006))**.** *For non-atomic p, or p uniform on finite $\mathcal{S}$, $I(q||p)$ rearrangement invariant relative to p $\iff$ $\succeq_{I(q||p)}$ is probabilistically sophisticated.*

In the case of a finite state space $\mathcal{S}_N$, the set of permutations that preserve a probability vector $\vec{p} \in \Delta(\mathcal{S}_N)$ is usually trivial, since p_i is generically unique for each state $i \in \mathcal{S}_N$. The only permutation that preserves $\vec{p}$ is the identity in this case, and so rearrangement invariance places no constraints on preferences.[24] This limitation arises because rearrangement invariance is a local property: it constraints the behavior of the rate function at a *fixed* vector $\vec{p}$. This feature is an advantage in applications of variational preferences to ambiguity aversion, where $\vec{p}$ can be considered a fixed part of an agent's behavioral endowment. For long run cost benefit rules however, $\vec{p}$ arises as part of the long run behavior of a stochastic processes, and

[24] This is one reason why the definition of probabilistic sophistication in Machina & Schmeidler (1992) assumes a non-atomic state space.

we can always consider families of such processes that give rise to different $\vec{p}$. In this context it is sensible to investigate global symmetries that ensure consistency within a family of processes indexed by different, but related, values of $\vec{p}$. Such global symmetries are meaningful constraints on families of stochastic processes, whether the state space is finite or not. For the sake of notational simplicity, I discuss the case of finite $\mathcal{S}$ below; the extension to non-atomic measures is straightforward.

3.2 State-label invariance

For any vector $\vec{p} \in \Delta(\mathcal{S}_N)$ and permutation σ of $\mathcal{S}_N$, define $\vec{p} \circ \sigma := (p_{\sigma(1)}, \ldots, p_{\sigma(N)})$. A condition that captures when a *family* of long run cost benefit rules indexed by $\theta \in \Theta$ is probabilistic is the following:

Definition 5. A family of long run cost benefit rules $\{\succeq_{I_\theta(\vec{q}||\vec{p}(\theta))}\}_{\theta\in\Theta}$ on a finite state space $\mathcal{S}_N$ is **state-label invariant** if

1. $\exists\, \theta \in \Theta$ with $\vec{p}(\theta) = \vec{r} \Rightarrow$ for any permutation σ of $\mathcal{S}_N$, $\exists\, \theta' \in \Theta$ with $\vec{p}(\theta') = \vec{r} \circ \sigma$.
2. For any $\theta, \theta' \in \Theta$ such that
$$\vec{p}(\theta') = \vec{p}(\theta) \circ \sigma$$
for some permutation σ of $\mathcal{S}_N$,
$$f \succeq_{I_\theta(\vec{q}||\vec{p}(\theta))} g \iff f \circ \sigma \succeq_{I_{\theta'}(\vec{q}||\vec{p}(\theta'))} g \circ \sigma$$
for any acts f, g.

State-label invariant families include not just the rule $\succeq_{I(\vec{q}||\vec{p})}$ for fixed $\vec{p}$, but rules indexed by all permutations of $\vec{p}$. In addition, rules related by a permutation σ of the state space are required to rank σ-permuted acts consistently. Like probabilistic sophistication, this precludes the labels of states from mattering, only the *distribution* of payoffs matters for preferences. Unlike the definition of probabilistic sophistication in (35) however, this requirement places meaningful constraints on preferences even if the state space is discrete and $\vec{p}$ is non-uniform. If the index set Θ maps only to a single probability vector $\vec{p}$ and its permutations, and the associated preference family is state-label invariant, this family essentially encodes a single probabilistic preference, up to permutations. However, in general the mapping from Θ to a probability vector $\vec{p}$ may be many-to-one, and state-label invariance is required to hold over the entirety of Θ.

The following result characterizes state-label invariant families of long run cost benefit rules:

Proposition 6. *A family of long run cost benefit rules* $\{\succeq_{I_\theta(\vec{q}||\vec{p}(\theta))}\}_{\theta\in\Theta}$ *is state-label invariant if and only if for any* $\theta, \theta' \in \Theta$ *such that* $\vec{p}(\theta') = \vec{p}(\theta) \circ \sigma$ *for some permutation* σ *of* $\mathcal{S}^N$*,*

$$I_{\theta'}(\vec{q} \circ \sigma||\vec{p}(\theta')) = I_\theta(\vec{q}||\vec{p}(\theta)). \tag{36}$$

This result shows that there are two ways in which preference families can fail to be state-label invariant. First, the rate function may fail to be invariant under joint permutations of $\vec{q}$ and $\vec{p}$. Second, the mapping from the index set Θ to probability vectors in $\Delta(\mathcal{S}_N)$ may be many-to-one, and the rate function may vary on sets $\{\theta \in \Theta|\vec{p}(\theta) = \vec{r}\}$.

A canonical example of preferences that are state-label invariant is the family generated by f-divergences:

$$I_f(\vec{q}||\vec{p}) = \begin{cases} \sum_{i\in\mathcal{S}_N} q_i f\left(\frac{p_i}{q_i}\right) & \vec{q} \ll \vec{p} \\ \infty & \text{otherwise} \end{cases} \tag{37}$$

where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a convex function with $f(1) = 0$ (relative entropy is the case $f(t) = t\log t$). The preference family corresponding to the set

$$H(\vec{p}) = \{I_f(\vec{q}||\vec{p} \circ \sigma)|\sigma \text{ permutes } \mathcal{S}_N\}$$

is clearly state-label invariant, as indeed is the family corresponding to $\bigcup_{\vec{p}>0} H(\vec{p})$. This tells us that there is an entire family of long run cost benefit rules associated with f-divergences, indexed by arbitrary probability vectors $\vec{p}$, that is probabilistic.

On the other hand, an important example of a preference family that is *not* state-label invariant is the collection generated by the rate functions of ergodic Markov chains that share a stationary distribution $\vec{p}$. As in Section 2.2, let $\Omega(\vec{q})$ be the set of primitive transition matrices that has the stationary distribution $\vec{q} = (q_1, \ldots, q_N) \in \Delta(\mathcal{S}_N)$. The transition matrix for a generic chain in $\Omega(\vec{q})$ is denoted by $\mathbf{Q}$. Using the 'chain rule' for relative entropy, it can be shown (see Budhiraja & Depuis, 2013, Ch. 6.4) that the (convex) rate function for the empirical measure $\vec{q}$ of an ergodic Markov chain with transition matrix $\mathbf{P} \in \Omega(\vec{p})$ is given by

$$I_{\mathbf{P}}(\vec{q}||\vec{p}) = \begin{cases} \inf_{\mathbf{Q}\in\Omega(\vec{q})} \sum_i D(\mathbf{Q}_{i,\cdot}||\mathbf{P}_{i,\cdot})q_i & \vec{q} \ll \vec{p} \\ \infty & \text{otherwise} \end{cases}. \tag{38}$$

Consider the family of preferences induced by

$$J(\mathbf{P}) := \left\{I_{\mathbf{\Sigma P \Sigma}^{-1}}(\vec{q}||\,\vec{p}\,\mathbf{\Sigma}^{-1})|\mathbf{\Sigma} \text{ any } N \times N \text{ permutation matrix}\right\}.$$

This family is again state-label invariant – the Markov chain's dynamics are invariant to relabelings of the state space. However, now consider the collection induced by

$$\mathcal{J}(\vec{p}) = \bigcup_{\mathbf{P} \in \Omega(\vec{p})} J(\mathbf{P}).$$

Each of the preferences in this collection has stationary distribution equal to a permutation of $\vec{p}$, and the empirical measure of any chain in $\mathcal{J}(\vec{p})$ converges to a permutation of $\vec{p}$ with probability 1. But $\mathcal{J}(\vec{p})$ is not state-label invariant. This occurs since the rate function (38) varies on the set $\Omega(\vec{p})$. One way to see this is that transition matrices in $\Omega(\vec{p})$ have $(N-1)^2$ free parameters,[25] and the relative entropy factors in (38) vary continuously over $\Omega(\vec{p})$. Thus when $N \geq 2$, we can find an infinite number of matrices that share the stationary distribution $\vec{p}$, but yield different rate functions; this violates the state-label invariance condition (36).

3.3 Behavioral implications

To see the implications of failures of state-label invariance, consider again the example of two state ergodic Markov chains on a state space $\mathcal{S} = \{\text{Red}, \text{Black}\}$ with the transition matrices $\mathbf{T}(a)$ in (30). All such matrices have the stationary distribution $\vec{u} = (0.5\ 0.5)$. In addition, because each chain is state-label invariant for any *fixed* value of a, the long run cost benefit rule $\succeq_{I_{\mathbf{P}(a)}(\vec{q}||\vec{u})}$ is indifferent between projects whose 1-period rates of discounted return λ are permutations of one another. Thus, if we define two projects $\lambda_a^{\text{Red}}, \lambda_a^{\text{Black}}$ with

$$\lambda_a^{\text{Red/Black}}(s) = \begin{cases} 1 & s = \text{Red/Black in chain } \mathbf{T}(a) \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

we have

$$\lambda_a^{\text{Red}} \sim_{I_{\mathbf{T}(a)}(\vec{q}||\vec{u})} \lambda_a^{\text{Black}} \tag{40}$$

for any $a \in (0,1)$. Now consider an agent who can choose amongst projects in *different* economies, indexed by a; denote the asymmetric and symmetric parts of this agent's meta-preference relation by $\succeq, \sim$ respectively. If the family of long run cost benefit rules induced by $\{I_{\mathbf{T}(a)}(\vec{q}||\vec{u})\}$ were state-label invariant relative to the probability vector $\vec{u}$, the set of relations (40) would further imply, for example,

$$\lambda_a^{\text{Red}} \sim \lambda_{a'}^{\text{Red}}$$

for any $a, a' \in (0,1)$. This follows since each value of a maps to the same probability vector $\vec{u}$; state-label invariance thus requires consistency across all values of a. However, by long

[25]The set of $N \times N$ matrices has dimension N^2, but row sums must equal 1 (N constraints), and we have $N-1$ further constraints from the stationarity condition $\vec{p}\mathbf{P} = \vec{p}$ (N equations, but one is redundant since $\sum_i p_i = 1$). $N^2 - N - (N-1) = (N-1)^2$.

run stochastic dominance (see Proposition 3), we actually have

$$a > a' \Rightarrow \lambda_a^{\text{Red}} \succ \lambda_{a'}^{\text{Red}}. \tag{41}$$

Thus, even though the asymptotic empirical frequency of Red is 0.5 with probability 1 for any a, we strictly prefer long run bets in economies with larger values of a to equivalent bets in economies with smaller values of a.

Readers familiar with the ambiguity literature may realize that this example is formally related to a sequential version of Ellsberg's 'Two Urn' decision problem (Ellsberg, 1961). In that problem Urn I contains 50 red and 50 black balls, and Urn II contains 100 balls, each or which is red or black, in unknown proportions. A 'Red' bet pays out \$1 if a red ball is drawn and zero otherwise, and a 'Black' bet pays out \$1 if a black ball is drawn and zero otherwise. Assuming that balls will be drawn uniformly at random from each urn, the modal preferences in Ellsberg's experiment are:

$$\begin{aligned}&\text{Red on Urn I} \sim \text{Black on Urn I}\\&\text{Red on Urn II} \sim \text{Black on Urn II}\\&\text{Red on Urn I} \succ \text{Red on Urn II}\end{aligned}$$

These preferences cannot be probabilistically sophisticated. If they were, indifference in the first two choices would reveal a 50% subjective probability that the drawn ball will be Red in both Urn I and Urn II. Probabilistic sophistication then requires indifference on the third choice as well, contrary to the stated preference. The standard interpretation of this finding is that people are ambiguity averse – they are sensitive to the subjective uncertainty over the composition of Urn II, and do not reduce the compound lottery over subjective uncertainty (share of Red balls in Urn II) and objective risk (chance of Red conditional on Urn II's composition) to a single lottery.

Suppose that we replace this one-shot decision problem with one in which sequences of balls are drawn from two sequences of urns. The first sequence (I) contains infinitely many copies of Ellberg's Urn I, and the second sequence (II) contains infinitely many urns, each of whose composition is unknown (and not necessarily constant across urns). Assume that payoffs accumulate multiplicatively across draws, so that Ellsberg-style choices become choices between 'projects' of the kind in (39), in the two sequences of urns. Because this setting preserves the symmetry of payoffs and information states, the choice between e.g. λ_I^{Red} and $\lambda_{II}^{\text{Red}}$ is isomorphic to Ellsberg's original choice problem.

The fact that long run cost benefit rules may be non-probabilistic offers an alternative explanation for the preference $\lambda_I^{\text{Red}} \succ \lambda_{II}^{\text{Red}}$ in this sequential Ellsberg problem: non-trivial dynamical beliefs. Sequential draws from Urn I are verifiably i.i.d. (one simply looks inside the urns before a sequence of balls is drawn), and so there is no rational basis for believing that outcomes are correlated across urns. Restricting attention to beliefs consistent with $\mathbf{T}(a)$ for some a, this implies that we can model beliefs about the Urn I sequences as $a = 0.5$, the i.i.d. case. In contrast, because it is impossible to verify the composition of urns in the second sequence ex-ante, there is nothing to restrict the agents' *dynamical* beliefs in this case. The agent could fear that the experimenter will 'cheat' by conditioning the composition of Urn II at time t on information at time $t-1$, e.g. the colors of previous balls drawn. Because of the impossibility of verifying that Urn II draws are i.i.d. (revealing the composition of an urn destroys the state of uncertainty that the experiment is designed to achieve), there is no way for the experimenter to allay these fears. For example, there is nothing to prevent beliefs being described by $\mathbf{T}(a)$ for $a < 0.5$, i.e. negatively correlated draws. As we have seen (41), if decision-makers care about long run returns, this would cause them to prefer a bet on the Urn I sequence to a bet on the Urn II sequence.

This argument is so far particular to the sequence version of the Ellsberg experiment, but, by modifying the arguments in Robson et al. (2023), it can be turned into an evolutionary explanation of ambiguity averse behavior in the original, static, Ellsberg decision problem. In a dynamic choice environment evolutionary selection generates a connection between long run cost benefit rules and preferences in static environments. The details of this argument are somewhat tangential to the thrust of this paper, so I have relegated them to Appendix G.

This line of argument further solidifies the conceptual connection between long run cost benefit rules and seemingly ambiguity averse behavior. Non-trivial dynamical beliefs, i.e., beliefs described by families of stochastic processes that violated state-label invariance, can yield violations of probabilistic behavior in a manner closely related to ambiguity aversion. A connection between the properties of dynamical beliefs (e.g. long run risk/persistence) and ambiguity has appeared elsewhere in the literature. Halevy & Feltkamp (2005) showed that risk averse Bayesian agents who must choose between *sequences* of bets on Ellsberg urns – one with two urns containing 50 red and 50 black balls, one with two urns of unknown but identical composition – will prefer to bet on the 'known' urns. All that is required for their result is risk aversion, and correlation between the compositions of the unknown urns; ambiguity aversion does not enter into the analysis. Strzalecki (2013) also demonstrates a relationship between ambiguity aversion and aversion to long run risk in a class of recursive models of dynamic choice. On this dimension, the discussion above can be considered as illustrating the relevance of related insights in a frequentist paradigm to long run cost benefit analysis. Variational preferences and their properties are the appropriate tool for capturing this connection in this context.

4 Conclusion

This paper was motivated by the observation that, under common assumptions on the stochastic process that generates risk in returns, long run cost benefit rules (i.e., rankings of risky public projects that mature in the distant future) are variational in nature; they cannot be represented by probabilities and expectations. The variational representation of long run cost benefit rules provides a flexible microeconomic foundation for studying existing problems, including the influence of risk on long run social discount rates, and related long run financial valuation problems. By studying long run cost benefit rules as a general preference relation, rather than under detailed structural or parametric assumptions on a stochastic process, I have been able to leverage the tools of decision theory to provide deeper non-parametric insights into how risk can be represented, compared, and interpreted for long run projects. I've shown that long run cost benefit rules induce a natural notion of long run stochastic dominance, which can rank stochastic processes even when they are asymptotically identical from a probabilistic perspective. This dominance relation can be interpreted as a novel ranking of the persistence of stochastic processes. I've also shown how and when long-run cost benefit rules fail to be probabilistic, and illustrated how such failures are related to a dual interpretation of variational preferences as a model of ambiguity aversion.

An upshot of this analysis is that a meaningful and decisive ranking of long run projects can be achieved with a preference relation that is 'almost' static, and thus rather simpler than the standard expected present value rule. All that is required is a change of perspective from a probabilistic conception of risk to a conception that elevates and resolves the role of asymptotically vanishingly likely events. Variational preferences – interpreted not as a model of static ambiguity aversion, but as the natural large maturity limits of expected present value functionals – achieve this for a large class of stochastic processes that encompasses many of the models commonly studied in cost benefit analysis and social discounting.

References

F. Alvarez & U. J. Jermann (2005). 'Using asset prices to measure the persistence of the marginal utility of wealth'. *Econometrica* **73**(6):1977–2016.

D. Blackwell (1953). 'Equivalent Comparisons of Experiments'. *The Annals of Mathematical Statistics* **24**(2):265–272.

E. S. Bobenrieth, et al. (2002). 'A commodity price process with a unique continuous invariant distribution having infinite mean'. *Econometrica* **70**(3):1213–1219.

A. Budhiraja & P. Depuis (2013). *Analysis and Approximation of Rare Events*. Springer.

S. Cerreia-Vioglio, et al. (2011). 'Uncertainty averse preferences'. *Journal of Economic Theory* **146**(4):1275–1330.

M. L. Cropper, et al. (2014). 'Declining Discount Rates'. *American Economic Review: Papers & Proceedings* **104**(5):538–543.

P. Dasgupta (2008). 'Discounting climate change'. *Journal of Risk and Uncertainty* **37**(2):141–169.

P. Dasgupta, et al. (1972). *Guidelines for project evaluation.* United Nations, New York.

A. Dembo & O. Zeitouni (2009). *Large Deviations Techniques and Applications.* Springer-Verlag, 2nd edn.

M. Donsker & S. Varadhan (1975). 'Asymptotic evaluation of certain markov process expectations for large time, I'. *Communications on Pure and Applied Mathematics* **28**:1–47.

D. Duffie & L. G. Epstein (1992). 'Stochastic Differential Utility'. *Econometrica* **60**(2):353–394.

D. Ellsberg (1961). 'Risk, Ambiguity, and the Savage Axioms'. *The Quarterly Journal of Economics* **75**(4):643–669.

M. Fleurbaey & S. Zuber (2015). 'Discounting, risk and inequality: A general approach'. *Journal of Public Economics* **128**:34–49.

M. Freeman & B. Groom (2010). 'Gamma Discounting and the Combination of Forecasts'. *SSRN working paper 1676793* .

I. Gilboa & D. Schmeidler (1989). 'Maxmin expected utility with non-unique prior'. *Journal of Mathematical Economics* **18**(2):141–153.

C. Gollier (2002a). 'Discounting an uncertain future'. *Journal of Public Economics* **85**(2):149–166.

C. Gollier (2002b). 'Time Horizon and the Discount Rate'. *Journal of Economic Theory* **107**(2):463–473.

C. Gollier (2012). *Pricing the Planet's Future: The Economics of Discounting in an Uncertain World.* Princeton University Press.

C. Gollier (2014). 'Gamma discounters are short-termist'. *IDEI Working Paper* .

C. Gollier & J. K. Hammitt (2014). 'The Long-Run Discount Rate Controversy'. *Annual Review of Resource Economics* **6**(1):273–295.

C. Gollier & M. L. Weitzman (2010). 'How should the distant future be discounted when discount rates are uncertain?'. *Economics Letters* **107**:350–353.

B. Groom, et al. (2022). 'The Future, Now: A Review of Social Discounting'. *Annual Review of Resource Economics* **14**:467–491.

B. Groom & C. Hepburn (2017). 'Reflections—Looking Back at Social Discounting Policy: The Influence of Papers, Presentations, Political Preconditions, and Personalities'. *Review of Environmental Economics and Policy* **11**(2):336–356.

Y. Halevy & V. Feltkamp (2005). 'A Bayesian Approach to Uncertainty Aversion'. *Review of Economic Studies* **74**:449–466.

L. P. Hansen (2012). 'Dynamic Valuation Decomposition Within Stochastic Economies'. *Econometrica* **80**(3):911–967.

L. P. Hansen & T. J. Sargent (2001). 'Robust Control and Model Uncertainty'. *The American Economic Review* **91**(2):60–66.

L. P. Hansen & J. A. Scheinkman (2009). 'Long-term Risk: An Operator Approach'. *Econometrica* **77**(1):177–234.

G. Hardy, et al. (1934). *Inequalities*. Cambridge University Press.

N. Jaakkola & A. Millner (2023). 'Non-dogmatic climate policy'. *Journal of the Association of Environmental and Resource Economists* **9**(4):807–841.

T. Kurtz & J. Feng (2010). 'Weak convergence and large deviation theory (lecture notes)'. *https://people.math.wisc.edu/~tgkurtz/Lectures/bathldp.pdf* .

B. R. La Cour & W. C. Schieve (2015). 'A general conditional large deviation principle'. *Journal of Statistical Physics* **161**.

P.-S. Laplace (1774). *Mémoires de Mathématique et de Physique, Tome Sixième.*

I. M. D. Little & J. A. Mirrlees (1974). *Project appraisal and planning for developing countries.* Heinemann.

L. Ljungqvist & T. J. Sargent (2004). *Recursive Macroeconomic Theory.* MIT Press, Cambridge, Mass, 2nd edn.

F. Maccheroni, et al. (2006). 'Ambiguity Aversion, Robustness, and the Variational Representation of Preferences'. *Econometrica* **74**(6):1447–1498.

M. J. Machina & D. Schmeidler (1992). 'A More Robust Definition of Subjective Probability'. *Econometrica* **60**(4):745–780.

I. Martin (2012). 'On the valuation of long-dated assets'. *Journal of Political Economy* **120**(2):346–358.

A. Millner (2020). 'Nondogmatic Social Discounting'. *American Economic Review* **110**(3):760–775.

A. Millner & G. Heal (2023). 'Choosing the future: Markets, Ethics, and Rapprochement in Social Discounting'. *Journal of Economic Literature* **61**(3):1037–1087.

L. Qin & V. Linetsky (2017). 'Long-Term Risk: A Martingale Approach'. *Econometrica* **85**(1):299–312.

A. Robson, et al. (2023). 'Decision Theory and Stochastic Growth'. *American Economic Review: Insights* **5**(3):357–376.

R. T. Rockafellar (1970). *Convex Analysis*. Princeton University Press.

C. Shalizi & A. Kontorovich (2007). *Almost None of the Theory of Stochastic Processes*. `https://www.stat.cmu.edu/~cshalizi/almost-none/`.

T. Strzalecki (2011). 'Axiomatic Foundations of Multiplier Preferences'. *Econometrica* **79**(1):47–73.

T. Strzalecki (2013). 'Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion'. *Econometrica* **81**(3):1039–1074.

M. L. Weitzman (1998). 'Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate,'. *Journal of Environmental Economics and Management* **36**(3):201 – 208.

M. L. Weitzman (2001). 'Gamma Discounting'. *The American Economic Review* **91**(1):260–271.

A Proof of Theorem 1

(ii) $\iff$ (iii):

We have

$$\begin{aligned}
&\lim_{t\to\infty} \phi_t(R, M) \geq \lim_{t\to\infty} \phi_t(R', M) \\
\iff &\lim_{t\to\infty} \frac{1}{t} \log \mathsf{E}_{\mu_t(q)} \exp\left[-t \int \lambda(R; M) dq\right] \geq \lim_{t\to\infty} \frac{1}{t} \log \mathsf{E}_{\mu_t(q)} \exp\left[-t \int \lambda(R'; M) dq\right] \\
\iff & -\inf_{q\in\Delta(S)} \left[\int_S \lambda(R; M) dq + I(q)\right] \geq -\inf_{q\in\Delta(S)} \left[\int_S \lambda(R'; M) dq + I(q)\right]
\end{aligned}$$

where the last line follows from the assumption that μ_t satisfies a large deviation principle with rate function $I(q)$, and an application of Varadhan's lemma (see Budhiraja & Depuis, 2013, p.7).

(i) $\iff$ (ii):

Let $(s_k)_{k\in\mathbb{N}^+}$ be a stochastic process on a set $\mathcal{S}$ that satisfies Assumption 1, and let $\lambda : \mathcal{S} \to \mathbb{R}^+$ be bounded and continuous, with

$$\lambda_{\min} := \inf_{s\in S} \lambda(s) > 0, \qquad \lambda_{\max} := \sup_{s\in S} \lambda(s) < \infty.$$

We have

$$\frac{1}{t} \log V_t(\mathbf{R}; \mathbf{M}) := \frac{1}{t} \log \mathsf{E}\left[\sum_{n=t}^{\infty} \prod_{k=0}^{n-1} e^{-\lambda(s_k)}\right] = \frac{1}{t} \log \mathsf{E}\left[\sum_{n=t}^{\infty} \exp\Big(-\sum_{k=0}^{n-1} \lambda(s_k)\Big)\right].$$

For any fixed deterministic sequence $\kappa = (s_0, s_1, \ldots)$ and $n \geq t$, we can split

$$\sum_{k=0}^{n-1} \lambda(s_k) = t \int \lambda dq_t(\kappa) + \sum_{j=t}^{n-1} \lambda(s_j).$$

where $q_t(\kappa)$ is the empirical measure associate with the first t elements of κ. Thus

$$\sum_{n=t}^{\infty} \exp\Big(-\sum_{k=0}^{n-1} \lambda(s_k)\Big) = e^{-t \int \lambda dq_t(\kappa)}\, H_t,$$

where

$$H_t := \sum_{m=0}^{\infty} \exp\Big(-\sum_{j=t}^{t+m-1} \lambda(s_j)\Big).$$

Since $\lambda_{\min} \le \lambda(s) \le \lambda_{\max}$,

$$m\,\lambda_{\min} \le \sum_{j=t}^{t+m-1} \lambda(s_j) \le m\,\lambda_{\max}, \qquad (m \ge 1).$$

Hence

$$e^{-m\lambda_{\max}} \le \exp\Big(-\sum_{j=t}^{t+m-1} \lambda(s_j)\Big) \le e^{-m\lambda_{\min}}.$$

Summing this inequality from $m = 0$ to ∞ gives the deterministic bounds

$$\frac{1}{1-e^{-\lambda_{\max}}} \le H_t \le \frac{1}{1-e^{-\lambda_{\min}}}.$$

Thus we have

$$\frac{1}{1-e^{-\lambda_{\max}}}\,\mathsf{E}_{\mu_t(q)}[e^{-t\int \lambda dq}] \;\le\; \mathsf{E}\left[\sum_{n=t}^{\infty} \exp\Big(-\sum_{k=0}^{n-1} \lambda(s_k)\Big)\right] \;\le\; \frac{1}{1-e^{-\lambda_{\min}}}\,\mathsf{E}_{\mu_t(q)}[e^{-t\int \lambda dq}]$$

where $\mu_t(q)$ is the measure on empirical measures q at maturity t. Taking $(1/t)\log$ of this inequality, and sending $t \to \infty$, we find

$$\lim_{t\to\infty} \frac{1}{t} V_t(\mathbf{R};\mathbf{M}) = \lim_{t\to\infty} \frac{1}{t} \log \mathsf{E}_{\mu_t(q)} \exp\left(-t\int \lambda dq\right) \propto \lim_{t\to\infty} \phi_t(\mathbf{R};\mathbf{M})$$

where the proportionality constant is positive and independent of $\mathbf{R}$, and hence drops out of rankings of returns.

B Relationship between Variational Ambiguity Preferences and Long Run Cost Benefit Rules

In the variational ambiguity model preferences $\succeq^{(2)}$ are defined over static measurable acts $f \in \mathcal{F}$, which map a state space $\mathcal{S}$ to an outcome space $\mathcal{O}$. I assume that $\mathcal{S}$ satisfies one of the conditions stated in Assumption 1. The set $\mathcal{O}$ is assumed to be a convex subset of a vector space, e.g. the set of lotteries over some payoffs. This implies that if $f, g \in \mathcal{F}$, then $\alpha f + (1-\alpha) g \in \mathcal{F}$ for any $\alpha \in (0,1)$. The representation of preferences that is axiomatized by Maccheroni et al. (2006) is:[26]

$$f \succeq^{(2)} g \iff \inf_{q \in \Delta(S)} \left[\int_{\mathcal{S}} f dq + C(q) \right] \geq \inf_{q \in \Delta(S)} \left[\int_{\mathcal{S}} g dq + C(q) \right]. \tag{42}$$

Here $\Delta(S)$ is again the set of probability measures over $\mathcal{S}$, and $C(q)$ is a convex, grounded,[27] lower semi-continuous function.[28]

Think of this setup in the Anscombe-Aumann paradigm, i.e. acts f map subjective states $s \in \mathcal{S}$ to objective lotteries $f(s) \in \mathcal{O}$. We say that a preference $\succeq$ is ambiguity averse if for all constant acts x, $f \succeq x \Rightarrow f \succeq^{SEU} x$ for some Subjective Expected Utility preference $\succeq^{SEU}$. The crucial feature of variational ambiguity preferences $\succeq^{(1)}$ is that they generically exhibit ambiguity aversion. Naturally, this implies that they do not reduce to subjective expected utility theory, or indeed any probabilistically sophisticated model of choice, in general.

What is the precise relationship between the variational ambiguity model and long run cost benefit rules? There are three differences between the long run cost benefit rules in (22) and the variational ambiguity preferences in (42):

1. The functions λ in (22), which are the objects of choice in long run cost benefit rules (i.e., rates of discounted return), are required to be bounded and continuous on $\mathcal{S}$, whereas acts in the variational ambiguity model are only required to measurable.
2. The rate function $I(q)$ in (22) was assumed to have compact sub-level sets, whereas the penalty function $C(q)$ in (42) is only lower-semicontinuous.
3. The penalty function $C(q)$ in (42) must be convex, but this is not an absolute requirement for the rate function $I(q)$ in (22).

All these differences can be attributed to requirements placed on the existence and properties of large maturity limits of cost benefit functionals. Consider limits of the form

$$Z(\lambda) := \lim_{t \to \infty} \frac{1}{t} \log \mathsf{E}_{\mu_t(q)} \exp \left[-t \int_{\mathcal{S}} \lambda dq \right]$$

where $\lambda \in C_b(\mathcal{S})$.

[26] The representation in Maccheroni et al. (2006) is slightly different to that presented here. They show that their axioms are represented by the function in (42) where the first term is the expectation of an affine transformation $u(f)$, rather than of f. They go on to show that preferences are invariant under common (positive) rescalings of u and C. Our presentation thus corresponds to a particular normalization in which u is the identity function, which can be implemented without loss of generality.

[27] A positive function is grounded if its infimum value is zero.

[28] The axioms that deliver this representation are as follows. First, there is a group of technical axioms that require that preferences over acts should be complete, transitive, monotone, continuous, and non-trivial (i.e., there exist at least two acts that are strictly ranked). A weak independence axiom requires that $\alpha f + (1-\alpha)x \succeq^{(1)} \alpha g + (1-\alpha)x \Rightarrow \alpha f + (1-\alpha)y \succeq^{(1)} \alpha g + (1-\alpha)y$ for arbitrary acts f, g, constant acts x, y and $\alpha \in (0,1)$. Informally, mixing acts with constant acts doesn't change their ranking. Finally, an uncertainty aversion axiom requires $f \sim g \Rightarrow \alpha f + (1-\alpha)g \succeq^{(1)} g$ for all $\alpha \in (0,1)$. Informally, the decision-maker has a weak preference for 'hedging' between two equally preferred acts.

Differences 1. and 2. above are attributable to the assumption that $Z(\lambda)$ exists for any $\lambda \in C_b(\mathcal{S})$. Bryc's theorem (see e.g. Budhiraja & Depuis, 2013) shows that this requirement is in fact equivalent to the existence of a large deviation principle for $\mu_t(q)$ (i.e., Assumption 2), where the rate function $I(q)$ has compact sub level sets, and is not just lower semicontinuous. Note that even with this assumption $Z(\lambda)$ may fail to exist if λ is only measurable, but not necessarily bounded and continuous. Thus *both* the restriction to bounded and continuous λ, and the assumption that $I(q)$ has compact sub-level sets, are required for $Z(\lambda)$ to exist for all $\lambda \in C_b(\mathcal{S})$.

Difference 3., i.e. convexity of the penalty function $C(q)$, is required in the variational ambiguity model, to ensure that it is ambiguity averse, and that an 'inverse' variational formula for the penalty function exists. The convexity of rate functions (the relevant quantities for long run cost benefit rules) has been studied extensively in the large deviations literature.[29] The most easily understood condition that ensures convexity is the following: if $Z(\lambda)$ is Gateaux differentiable as a function of λ, then the rate function $I(q)$ must be convex.[30] Convexity of the rate function thus follows when we impose smoothness restrictions on long run values.

B.1 'More ambiguity averse' and Long Run Stochastic Dominance

As mentioned in the body of the manuscript, there is a duality between my notion of long run stochastic dominance and the notion of 'more ambiguity averse' preferencecs in Maccheroni et al. (2006).

Denoting a variational ambiguity preference with penalty function $C(q)$ by $\succeq_C^{(2)}$, Maccheroni et al. (2006) say that $\succeq_{C_A}^{(2)}$ is *more ambiguity averse* than $\succeq_{C_B}^{(2)}$ if for all acts $f : \mathcal{S} \to \mathcal{X}$ and constant acts x,

$$f \succeq_{C_A}^{(2)} x \Rightarrow f \succeq_{C_B}^{(1)} x$$

To understand this definition, recall that in the Anscombe-Aumann interpretation of the variational ambiguity model acts map subjectively uncertain states $s \in \mathcal{S}$ to objective lotteries $f(s)$ – a constant act is therefore an objective lottery. These unambiguous acts are used to benchmark ambiguity attitudes in the same way that deterministic outcomes benchmark risk attitudes in expected utility theory.

Maccheroni et al. (2006) characterized the 'more ambiguity averse' relation thus:

[29]See Section 4 in Dembo & Zeitouni (2009), and in particular Lemma 4.1.21 and Corollary 4.6.14, for a detailed treatment.

[30]$Z(\lambda)$ is Gateaux differentiable if for all functions $z : \mathcal{S} \to \mathbb{R}$, $\lim_{h\to 0} \frac{Z(\lambda+hz)-Z(\lambda)}{h}$ exists. The convexity result follows from Corollary 4.6.14 in Dembo & Zeitouni (2009). The two conditions that are required in their statement of this result – exponential tightness of $\mu_t(q)$ and existence of the limiting (rescaled) cumulant generating function (i.e. $Z(\lambda)$ in our notation) – are implied by the assumption that $\mu_t(q)$ satisfies a large deviation principle. Theorem 18 in Maccheroni et al. (2006) states a version of this result for the variational ambiguity model.

Proposition 7 (Maccheroni et al. (2006)). $\succeq_A^{(1)}$ *is more ambiguity averse than* $\succeq_B^{(1)}$ *if and only if*

$$\forall q \in \Delta(S),\ C_A(q) \leq C_B(q).$$

The parallels between this finding and long run stochastic dominance are immediate.

C Proof of Proposition 2

Given a measure on empirical measures $\mu_t(q)$ satisfying Assumptions 1 and 2, with rate function $I_\mu(q)$, define

$$Z_{I_\mu}(\lambda) := \lim_{t\to\infty} \frac{1}{t} \log \mathsf{E}_{\mu_t(q)} \exp\left[-t \int_{\mathcal{S}} \lambda dq\right] = - \inf_{q\in\Delta(\mathcal{S})} \left[\int_{\mathcal{S}} \lambda dq + I_\mu(q)\right], \tag{43}$$

where the latter equality holds for any $\lambda \in C_b(\mathcal{S})$. Clearly,

$$\lim_{t\to\infty} \phi_t(\mu_t, \lambda) \geq \lim_{t\to\infty} \phi_t(\nu_t, \lambda) \iff Z_{I_\mu}(\lambda) \geq Z_{I_\nu}(\lambda) \tag{44}$$

as the denominator in the definition of $\phi_t(\mu_t, \lambda)$ in (27) is common to all return functions R.

For an arbitrary function $I(q)$, define its convex conjugate

$$I^*(\lambda) = \sup_{q\in\Delta(\mathcal{S})} \left[\int_{\mathcal{S}} \lambda dq - I(q)\right]. \tag{45}$$

For an arbitrary function $I^*(\lambda)$, define its convex conjugate

$$I^{**}(q) = \sup_{\lambda\in C_b(\mathcal{S})} \left[\int_{\mathcal{S}} \lambda dq - I^*(\lambda)\right].$$

If $I(q)$ is a proper convex function (i.e., Assumption 3 holds – properness is guaranteed by the assumption that $I(q)$ has compact sub-level sets), Fenchel-Moreau duality (Rockafellar, 1970) implies that:

$$I^{**} = I.$$

We have $I^*(-\lambda) = Z_I(\lambda)$, and requiring (44) to hold for all λ is equivalent to requiring

$$I_\mu^*(-\lambda) \geq I_\nu^*(-\lambda)$$

for all λ. Since λ is an arbitrary function in $C_b(\mathcal{S})$ we can rewrite this condition as $I_\mu^*(\lambda) \geq I_\nu^*(\lambda)$ for all $\lambda \in C_b(\mathcal{S})$. This fact, combined with the Fenchel-Moreau duality condition, then tells us that

$$I_\mu(q) = I_\mu^{**}(q) = \sup_{\lambda\in C_b(\mathcal{S})} \left[\int_{\mathcal{S}} \lambda dq - I_\mu^*(\lambda)\right] \leq \sup_{\lambda\in C_b(\mathcal{S})} \left[\int_{\mathcal{S}} \lambda dq - I_\nu^*(\lambda)\right] = I_\nu^{**}(q) = I_\nu(q)$$

for all $q \in \Delta(\mathcal{S})$.

D Proof of Proposition 3

Consider an ergodic Markov chain on a discrete state space S_N with N elements, and generic element s. Such Markov chains have a unique globally asymptotically stable stationary distribution given by a left (row) eigenvector of the transition matrix $\mathbf{P}$:

$$\vec{\pi}\mathbf{P} = \vec{\pi}.$$

Moreover, $\vec{\pi}$ has strictly positive entries, by the Perron-Frobenius theorem.

We begin by characterizing the set of matrices $\Omega(\vec{\pi})$ that are consistent with a given positive stationary distribution $\vec{\pi}$. First note that if $\mathbf{P}$ and $\mathbf{P}'$ both have stationary distribution $\vec{\pi}$, then clearly for any $\alpha \in (0,1)$

$$\vec{\pi}(\alpha\mathbf{P} + (1-\alpha)\mathbf{P}') = \vec{\pi}$$

Thus $\Omega(\vec{\pi})$ is convex. We can characterize this set further by writing out the defining equation for the stationary distribution explicitly:

$$\begin{aligned}
\pi_1 p_{11} + \pi_2 p_{21} + \dots \pi_N p_{N1} &= \pi_1 \\
\pi_1 p_{12} + \pi_2 p_{22} + \dots \pi_N p_{N2} &= \pi_2 \\
&\vdots \\
\pi_1 p_{1N} + \pi_2 p_{2N} + \dots \pi_N p_{NN} &= \pi_N
\end{aligned}$$

We define a set of *extremal* matrices consistent with $\vec{\pi}$ by choosing exactly one non-zero value of p_{ij} for each row i. If j^* is the index of the non-zero column element in row i, we set $p_{ij^*} = \pi_{j^*}/\pi_i$, and $p_{ij} = 0$ for $j \neq j^*$. There are $N^2 - N - 1$ such extremal matrices (N^2 square matrices, N constraints of the form $\sum_j p_{ij} = 1$, and one further constraint $\sum_i \pi_i = 1$). The set $\Omega(\vec{\pi})$ is the convex hull of these extremal matrices.

Now for an arbitrary $1 \times N$ vector $\vec{q}$ define

$$J(\vec{q};\mathbf{P}) = \begin{cases} \min_{\mathbf{P}' \in \Omega(\vec{q})} \sum_{\sigma \in \Sigma_N} D(P'(\cdot|s)||P(\cdot|s))\vec{q}(s) & \vec{q} \ll \vec{\pi} \\ \infty & \text{otherwise} \end{cases} \tag{46}$$

where the minimisation is over all matrices $\mathbf{P}' \in \Omega(\vec{q})$. From the chain rule for relative entropy, $J(\vec{q};\mathbf{P})$ is the rate function for the empirical distribution of the Markov chain with transition matrix $\mathbf{P}$ (Dembo & Zeitouni, 2009; Budhiraja & Depuis, 2013).

Next we show that $J(\vec{q};\mathbf{P})$ is convex in $\mathbf{P}$. This follows from the fact that the relative entropy is convex in both its arguments. For $\alpha \in (0,1)$, and $\mathbf{P}, \mathbf{Q} \in \Omega(\vec{\pi})$,

$$\begin{aligned}
J(\vec{q}; \alpha\mathbf{P} + (1-\alpha)\mathbf{Q}) &= \min_{\mathbf{P}' \in \Omega(\vec{q})} \sum_s [D(P'(\cdot|s) || \alpha P(\cdot|s) + (1-\alpha) Q(\cdot|s))] \vec{q}(s) \\
&\leq \min_{\mathbf{P}' \in \Omega(\vec{q})} \left[\alpha \sum_s D(P'(\cdot|s)||P(\cdot|s))\vec{q}(s) + (1-\alpha) \sum_s D(P'(\cdot|s)||Q(\cdot|s))\vec{q}(s) \right] \\
&\leq \alpha \min_{\mathbf{P}' \in \Omega(\vec{q})} \sum_s D(P'(\cdot|s)||P(\cdot|s))\vec{q}(s) + (1-\alpha) \min_{\mathbf{P}' \in \Omega(\vec{q})} \sum_s D(P'(\cdot|s)||Q(\cdot|s))\vec{q}(s) \\
&= \alpha J(\vec{q}; \mathbf{P}) + (1-\alpha) J(\vec{q}; \mathbf{Q}).
\end{aligned}$$

The first inequality follows from the convexity of relative entropy, and the second inequality from the convexity of the 'min' function. Thus the rate function $J(\vec{q}; \mathbf{P})$ is a convex function on the convex set $\Omega(\vec{\pi})$.

We now prove that long run stochastic dominance is a complete order on $\Omega(\vec{\pi})$ when $N = 2$:

Proof. For $N = 2$ the set $\Omega(\vec{q})$ is the convex hull of $2^2 - 2 - 1 = 2$ extremal matrices. Writing $\vec{q} = (q \ 1-q)$ it is straightforward to show that the extremal matrices of $\Omega(\vec{q})$ are

$$1_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ E_q = \begin{pmatrix} 0 & 1 \\ \frac{q}{1-q} & \frac{1-2q}{1-q} \end{pmatrix}.$$

Thus,

$$\Omega(\vec{q}) = \{(1-\alpha)1_2 + \alpha E_q | \alpha \in (0,1)\}. \tag{47}$$

The crucial feature of $\Omega(\vec{q})$ is that it is a one dimensional convex set. Notice also that

$$\forall \vec{q}, \lim_{\alpha \to 0^+} J(\vec{q}; (1-\alpha)1_2 + \alpha E_q) = 0.$$

From the representation (47), we can index $\Omega(\vec{\pi})$ by a parameter $\alpha \in (0,1)$. Slightly abusing notation, we thus write the rate function $J(\vec{q}; \mathbf{Q})$ as $J(\vec{q}; \alpha)$. The function $J(\vec{q}; \alpha)$ is convex in α, with $J(\vec{q}; 0) = 0$. Hence for any $\alpha' < \alpha$, there exists $\beta \in (0,1)$ such that $J(\vec{q}; \alpha') = J(\vec{q}; \beta(0) + (1-\beta)\alpha) \leq \beta J(\vec{q}; 0) + (1-\beta) J(\vec{q}; \alpha) = (1-\beta) J(\vec{q}; \alpha)$. Sending $\beta \to 0$, and using the fact that convex functions are continuous, we obtain $J(\vec{q}; \alpha') \leq J(\vec{q}; \alpha)$ for all $\vec{q}$. Since α, α' were arbitrary, we conclude that long run stochastic dominance is a complete order on $\Omega(\vec{\pi})$. $\square$

The proof immediately suggests a partial generalization to the case $N > 2$. The proof works unmodified in higher dimensions if we replace the matrix 1_2 with the N-dimensional identity matrix, and the matrix E_q with any other extremal matrix of $\Omega(\vec{\pi})$. In this case the long run stochastic order is complete on convex subsets of $\Omega(\vec{\pi})$ of the form $(1-\alpha)1_2 + \alpha E_q$

for $\alpha \in (0,1)$. The difference is that for $N > 2$ these subsets no longer coincide with the entirety of $\Omega(\vec{q})$.

E Proof of Proposition 4

Let $\{\mu_t\}_{t\in\mathbb{N}^+}$ and $\{\nu_t\}_{t\in\mathbb{N}^+}$ be sequences of probability measures on $\Delta(\mathcal{S})$, each satisfying a large deviation principle with rate functions I_μ and I_ν respectively, and speeds $a_t^\mu, a_t^\nu \to \infty$ such that $\lim_{t\to\infty} a_t^\mu / a_t^\nu \to 1$.

Let $f_t : \mathbb{R} \to (0,\infty)$ be strictly increasing and convex, and assume that

$$\lim_{t\to\infty} \frac{1}{a_t} \log f_t(x) \to g(x)$$

where the convergence is locally uniform, and $g : \mathbb{R} \to \mathbb{R}$ is strictly increasing and convex.

For any convex and bounded below function $\Phi : \Delta(\mathcal{S}) \to \mathbb{R}$ define the asymptotic functional

$$L_\sigma(\Phi) := \lim_{t\to\infty} f_t^{-1}\Big(\int f_t(\Phi(q))\, d\sigma(q) \Big), \qquad \sigma \in \{\mu, \nu\}.$$

Varadhan's lemma tells us that these limits exist and satisfy

$$L_\sigma(\Phi) = g^{-1}\Big(\sup_{q\in\Delta(\mathcal{S})} \{\, g(\Phi(q)) - I_\sigma(q) \,\} \Big). \tag{48}$$

I claim the following are equivalent:

(A) For every convex Φ,

$$L_\mu(\Phi) \geq L_\nu(\Phi).$$

(B) The rate functions are ordered:

$$I_\mu(q) \leq I_\nu(q) \qquad \forall\, q \in \Delta(\mathcal{S}).$$

Proof. B $\Rightarrow$ A: Immediate from (48).

A $\Rightarrow$ B: Define the g-conjugate of a proper convex function I by

$$I^g(\Phi) := \sup_{q\in\Delta(\mathcal{S})} \{ g(\Phi(q)) - I(q) \}.$$

Since g is convex and strictly monotone, a generalized Fenchel–Moreau identity holds (Rockafellar, 1970):

$$I(q) = \sup_{\Phi \text{ convex}} \big\{ \Phi(q) - g^{-1}(I^g(\Phi)) \big\}.$$

Condition (A) means $I^g_\mu(\Phi) \geq I^g_\nu(\Phi)$ for all convex Φ, hence

$$I_\mu(q) = \sup_{\Phi \text{ convex}} \{\Phi(q) - g^{-1}(I^g_\mu(\Phi))\} \leq \sup_{\Phi \text{ convex}} \{\Phi(q) - g^{-1}(I^g_\nu(\Phi))\} = I_\nu(q),$$

which yields $I_\mu \leq I_\nu$.

The final step of the proof is to note first that asymptotic equality of marginals of μ_t and ν_t is guaranteed if $\text{argmin}_q I_\nu(q) = \text{argmin}_q I_\mu(q) = q^*$ for some unique $q^* \in \Delta(\mathcal{S})$, and second that the case of long run stochastic dominance corresponds to $f_t(x) = \exp(tx)$, and speeds $a^\mu_t = a^\nu_t = t$. □

F Proof of Proposition 6

As in (43) define

$$Z_I(\vec{f}) := -\inf_{\vec{q} \in \Delta(\mathcal{S}_N)} \left[\vec{f} \cdot \vec{q} + I(\vec{q})\right], \tag{49}$$

where $I(\vec{q})$ is a rate function with compact sublevel sets on $\Delta(\mathcal{S}_N)$ satisfying Assumption 3.

We begin with a lemma:

Lemma 1. Given two rate functions $I(\vec{q}), J(\vec{q})$, if for any $\vec{f}, \vec{g} \in \mathbb{R}^N$

$$Z_I(\vec{f}) \geq Z_I(\vec{g}) \iff Z_J(\vec{f}) \geq Z_J(\vec{g})$$

then $I = J$ pointwise on $\Delta(\mathcal{S}_N)$.

Proof. The fact that Z_I and Z_J yield numerically representable rankings of acts $\vec{f}$ that are ordinally equivalent implies that there exists an increasing function $U : \mathbb{R} \to \mathbb{R}$ such that

$$Z_J(\vec{f}) = U(Z_I(\vec{f}))$$

for all $\vec{f}$. Now notice that for any rate function I, and any constant act $a\vec{1}, a \in \mathbb{R}$, we have

$$Z_I(\vec{f} + a\vec{1}) = Z_I(\vec{f}) - a$$

for any $\vec{f}$. Thus for any $a \in \mathbb{R}$, and any $\vec{f}$

$$Z_J(\vec{f} + a\vec{1}) = U(Z_I(\vec{f} + a\vec{1})) = U(Z_I(\vec{f}) - a) = Z_J(\vec{f}) - a = U(Z_I(\vec{f})) - a.$$

Since $Z_I(\vec{f})$ can take values in any bounded interval in $\mathbb{R}$, we have

$$U(\epsilon - a) = U(\epsilon) - a$$

for any ϵ in any bounded interval of $\mathbb{R}$, and any $a \in \mathbb{R}$. Defining $\tilde{U}(x) = U(x) - x$, we have

$$\tilde{U}(x-a) = \tilde{U}(x-a) - (x-a) = \tilde{U}(x) - x = \tilde{U}(x)$$

for any $a \in \mathbb{R}$, and any x in a bounded interval of $\mathbb{R}$. This implies that any real value a is a period of $\tilde{U}(x)$; the only function that has any real number as a period is a constant function. Thus

$$U(x) = x + c$$

for some $c \in \mathbb{R}$. However, since $Z_I(\vec{0}) = -\inf_{\vec{q}} I(\vec{q}) = 0$ for any rate function $I(\vec{q})$, we must have $U(0) = 0 \Rightarrow c = 0$. Thus, $Z_I(\vec{f}) = Z_J(f)$ for any $\vec{f}$.

Now,

$$Z_I(\vec{f}) = I^*(-\vec{f})$$

where I^* is the convex conjugate of I (see (45)) and hence from Assumption 3 and Fenchel-Moreau duality,

$$I(q) = I^{**}(q) = \sup_{\vec{f}} \vec{f} \cdot \vec{q} - I^*(\vec{f}) = \sup_{\vec{f}} \vec{f} \cdot \vec{q} - Z_I(-\vec{f}).$$

Thus $Z_I(\vec{f}) = Z_J(\vec{f})$ for all $\vec{f}$ implies $I = J$ pointwise. □

For any vector $\vec{f} \in \mathbf{R}^N$ and any permutation σ of $\mathcal{S}_N$, define $\vec{f} \circ \sigma = (f_{\sigma(1)}, \ldots, f_{\sigma(N)})$. We wish to prove that, if for any $\vec{f}, \vec{g}$ and any permutation σ,

$$\vec{f} \succeq_{I(\vec{q}||\vec{p})} \vec{g} \iff \vec{f} \circ \sigma \succeq_{I(\vec{q}||\vec{p}\circ\sigma)} \vec{g} \circ \sigma \tag{50}$$

then

$$I(\vec{q} \circ \sigma || \vec{p} \circ \sigma) = I(\vec{q}||\vec{p})$$

for any $\vec{q}$ and any permutation σ. Under $\succeq_{I(\vec{q}||\vec{p}\circ\sigma)}$ the value of an act $\vec{f} \circ \sigma$ is represented by

$$-\inf_{\vec{q}} \vec{f} \circ \sigma \cdot \vec{q} + I(\vec{q}||\vec{p} \circ \sigma) = -\inf_{\vec{q}} \sum_i f_{\sigma(i)} q_i + I(\vec{q}||\vec{p} \circ \sigma) = -\inf_{\vec{q}} \sum_i f_i q_{\sigma^{-1}(i)} + I(\vec{q}||\vec{p} \circ \sigma).$$

Defining $\vec{q}\,' = \vec{q} \circ \sigma^{-1}$, this last expression is equal to

$$-\inf_{\vec{q}\,'} \sum_i f_i q_i' + +I(\vec{q}\,' \circ \sigma||\vec{p} \circ \sigma).$$

Thus, by Lemma 1, (50) is satisfied if and only if

$$I(\vec{q}||\vec{p}) = I(\vec{q} \circ \sigma||\vec{p} \circ \sigma)$$

for all $\vec{q} \in \Delta(\mathcal{S}_N)$. As σ was arbitrary, this must hold for all permutations of $\mathcal{S}_N$.

G Evolutionary foundation for variational ambiguity preferences

The basic setup follows that in Robson et al. (2023). We imagine a continuum of agents, indexed by $i \in \mathcal{I}$, each of whom faces an idiosyncratic risk $s_t^i \in S$ at time t. Individuals are characterized by a state variable w_t^i. If we interpret the model in economic terms, w_t^i can be viewed as individual i's wealth; if we interpret it biologically w_t^i is a measure of i's reproductive fitness. For simplicity we will refer to w_t^i as wealth, but the biological interpretation should be born in mind.

If individual i invests in an asset with rate of return $-r(s) \in \mathbb{R}$, w^i evolves according to

$$w_t^i = e^{-r(s_t^i)} w_{t-1}^i.$$

We assume that for each i, s_t^i is stochastic process whose empirical measures obeys a large deviation principle with rate function $I(q)$, and that s_t^i is independent of s_t^j for $i \neq j$, i.e., risk is idiosyncratic. Unlike Robson et al. (2023), we allow for the possibility of dynamical structure in the wealth shocks, i.e., they need not be i.i.d. over time.

As was demonstrated in the discussion of Theorem 1, the proportion of the population that experiences a realization of the sequence $\kappa^i = (s_0^i, s_1^i, \ldots)$ with empirical measure $q_t(\cdot;\kappa^i)$ declines like $\exp(-tI(q_t))$ asymptotically. The wealth of an individual who invests in asset r and experiences the empirical measure q grows like $\exp(-t\int_S r dq)$ (recall that $r(s)$ may be negative). As $t \to \infty$, the growth of aggregate wealth (or fitness) in a population of like individuals will be determined by the vanishingly small set of agents who experience the goldilocks empirical measure that maximizes the asymptotic growth factor

$$\exp\left[-t\left(\int_S r dq + I(q)\right)\right].$$

These agents' contribution to the aggregate wealth of the population will be infinitely larger than that of all other agents, asymptotically. Following Robson et al. (2023), I assume that evolution will select for individuals who rank assets according to the long run growth rate of aggregate wealth in a population of like agents. Individuals that do not do this will suffer an infinite penalty at the population level asymptotically, and will thus be driven out by evolution. As such, asset r will be 'evolutionarily preferred' to asset r' if and only if

$$-\inf_{q\in\Delta(S)}\left[\int_S r dq + I(q)\right] \geq -\inf_{q\in\Delta(S)}\left[\int_S r' dq + I(q)\right] \iff r \preceq_{I(q)} r'.$$

As I discussed in footnote 12, we can always construct a stochastic process that gives rise to a pre-specified rate function $I(q)$, and so we have found an evolutionary foundation for arbitrary variational ambiguity preferences. This analysis shows that preferences that violate probabilistic sophistication can emerge endogenously from a classical (i.e., objective) risk environment.