Decision Theory and Optimal Transport

18th e-Summer School on Risk Finance and Stochastics (ε-RFS$^{\iota\eta'}$), Athens

A. N. Yannacopoulos

AUEB

September 6, 2021

Joint work with:

- G.I. Papayiannis (HNA)
  Statistical Learning, Model averaging, Risk management
- E. V. Petracou (U. of Aegean)
  Group decision making
- A. Xepapadeas (AUEB and Bologna)
  Environmental decision making

Table of contents

Motivation I : Decision under uncertainty

- Consider an agent facing a random consequence $X \sim P$.
- If P is known for certain then X can be evaluated using VnM utilities

$$U(X) = \mathbb{E}_P[u(X)]$$

- What if a single probability law (model) is not available for the description of X?

$$\mathbb{M} = \{P_1, \cdots, P_N\} \text{ set of plausible models for } X$$

- Various classical choices (incomplete list)
  - Minimax preferences (Gilboa - Schmeidler)

$$U(X) = \min_{P \in \mathbb{M}} \mathbb{E}[u(X)]$$

To paraphrase Leibnitz "the worst of all possible worlds" - an over cautious pessimistic view of the risk, a worst case scenario.

  - Multiplier preferences (Hansen-Sargent)

Pick the most plausible model P_0 (reference model) and then consider all possible models $P \in \mathbb{P}$ as models for X "penalizing" them using the KL entropy $H(P \mid P_0)$:

$$U(X) = \min_{P \in \mathbb{P}} (\mathbb{E}_P(u(X)) + \frac{1}{\theta} H(P \mid P_0)), \;\; \theta > 0$$

  - Variational preferences (Maccheroni, Marinacci, Rustichini)

Consider a convex lower semicontinuous penalty function φ in the space of all possible models for X, $P \in \mathbb{P}$, and solve the variational problem

$$U(X) = \min_{P \in \mathbb{P}} (\mathbb{E}_P[u(X)] + \varphi(P)).$$

Variational preferences include minimax preferences and multiplier preferences as special cases.

Moreover they follow an interesting axiomatic framework developed by Maccheroni, Marinacci, Rustichini (MMR2006) and importantly satisfy the uncertainty aversion axiom

$$f_1 \sim f_2 \Longrightarrow \lambda f_1 + (1-\lambda) f_2 \succeq f_1, \text{ for any acts } f_1, f_2, \;\; \lambda \in (0,1).$$

- What if there is some belief that all models in $\mathbb{M}$ share a fraction of the truth or capture different aspects of the phenomenon under study?
- Then we may treat each of the models in $\mathbb{M}$ as "random" observations of the true (unknown) model and try to do some type of least squares approximation in model space i.e. find a mean model consisting of the above which will be used to describe X.
- Motivated by the concept of the Fréchet mean such a model will be

$$P_B = \arg\min_{P \in \mathbb{P}} (w_1 d^2(P, P_1) + \cdots + w_N d^2(P, P_N))$$

where

  - $w_1, \cdots, w_N$ are weights on the various models
  - $d(P, P_i)$ is some distance in the space of models (probability measures)
- P_B is called the barycenter of $\mathbb{M}$ and is the model in $\mathbb{P}$ of least distance from all models in $\mathbb{M}$.
- In Euclidean space the barycenter corresponds to the standard notion of the mean - minimizer of variance.

- Having obtained P_B one could consider

$$U(X) = \mathbb{E}_{P_B}[u(X)] \tag{1}$$

  as a decision making tool.
- However, as we shall see, building on the theory of variational preferences, we can go further than our proposal in (1).
- At any rate, P_B sound like a good way to combine all models in a single one - weighted by an appropriate weight of belief w_i - to be used subsequently for evaluating X.

Motivation II: Group decision making

- Consider a group of N agents wishing to agree on the valuation of a common resource (asset or risk) X
- Each agent i has a different model P_i for X
- How can we get a valuation for X such that it is most likely that all agents will agree on and thus validate the common decision?
- Examples: Climate change negotiations, CAT bonds valuation etc

- If the probability of agent i accepting a proposal P concerning X depends on how distant P is from her anchor P_i then the proposal which is most likely to be accepted by the group will be the one that satisfies

$$d(P, P_1) \leq \epsilon_1, \\ \dots \\ d(P, P_N) \leq \epsilon_N,$$

where ϵ_i represent the disposition of agent i to deviate from her anchor position.

- By a scalarization argument one could argue (informally) that a model P with sufficiently high probability of being accepted by all agents will be

$$P_B = \arg\min_{P \in \mathbb{P}} (w_1 d^2(P, P_1) + \cdots + w_N d^2(P, P_N)),$$

with w_i related to ϵ_i.

- This is a model suitable for valuation of X, in terms of

$$U(X) = \mathbb{E}_{P_B}[u(X)], \tag{2}$$

or some refinements upon it.

Frechét mean preferences

- Consider the space of models (probability measures) $\mathbb{P}$ endowed with a notion of distance (metric) d.
- Consider the set of plausible models

$$\mathbb{M} = \{P_1, \cdots, P_N\},$$

and define the Fréchet function

$$F_{\mathbb{M}}(P) = \sum_{i=1}^{N} w_i d^2(P, P_i).$$

- The baryenter of $\mathbb{M}$ is defined as

$$P_B = \arg\min_{P \in \mathbb{P}} F_{\mathbb{M}}.$$

Definition 1

For any $\phi : \mathbb{R} \to \mathbb{R}_{++}$ convex and increasing we define the Fréchet mean utility functional

$$U(X) = \min_{P \in \mathbb{P}} (\mathbb{E}_P[u(X)] + \phi(F_{\mathbb{M}}(P))$$

- Fréchet mean utilities are a special case of the class of variational utilities (Maccheroni, Marinacci, Rustichini MMR2006) and satisfy their axiomatic framework.
- Importantly on account of that they display ambiguity aversion effects
- They penalize large variability in model space, if $\phi(F_{\mathbb{M}}(P_B)) = 0$ (normalization) then

$$U(X) \leq \mathbb{E}_{P_B}[u(X)].$$

- $P \mapsto \phi(F_{\mathbb{M}}(P))$ plays the role of a penalty function penalizing the less plausible models in $\mathbb{M}$ (e.g. outliers of some sort).
- The relative magnitude of the penalty term with respect to the utility term plays an important role
  - If $\phi >>$ then $P^* \simeq P_B$ and $U(X) \simeq \mathbb{E}_{P_B}[u(X)]$.
  - If $\phi <<$ then P^* may deviate considerably from P_B and $U(X)$ can be significantly smaller than $\mathbb{E}_{P_B}[u(X)]$.

Fréchet multiplier preferences

- The special case where

$$\phi(F_{\mathbb{M}}(P) = \frac{\theta}{2}(F_{\mathbb{M}}(P) - F_{\mathbb{M}}(P_B), \;\; \theta > 0,$$

corresponds to Fréchet multiplier preferences,

$$U_\theta(X) = \min_{P \in \mathbb{P}}[\mathbb{E}_P[u(X)] + \frac{\theta}{2}(F_{\mathbb{M}}(P) - F_{\mathbb{M}}(P_B))], \;\; \theta > 0.$$

- θ is called the multiplier.
  - For any $\theta > 0$

$$U_\theta(X) \leq \mathbb{E}_{P_B}[u(X)], \;\; \text{pessimism effect}$$

  - As $\theta \to \infty$ then $U_\theta(X) \to \mathbb{E}_{P_B}[u(X)]$

The following questions arise:

1. Which is a suitable distance in the space of probability measures?
2. In general it seems that

$$U(X) = \mathbb{E}_{P_B}[u(X)] + C(X)$$

   where $C(X) \leq 0$ is a correction term to the barycentric expected utility

   - Can we calculate the correction term C explicitly so as to use the utility functional U for concrete valuations?
   - Can we understand the interplay between risk and uncertainty?

Optimal transportation enters : The Wasserstein distance

A very convenient metric (distance) in the space of probability measures is the Wasserstein distance from optimal transport.

Given two probability measures P and Q their 2-Wasserstein distance can be defined as

$$d_2(P,Q) = W_2(P,Q) = \left\{ \inf_{\gamma \in \Pi(P,Q)} \mathbb{E}_\gamma[(X-Y)^2] \right\}^{1/2}, \\ X \sim P, \;\; Y \sim Q, \;\; (X,Y) \sim \gamma,$$

where $\Pi(P,Q)$ is the set of all probability measures on $\Omega \times \Omega$ with marginals P and Q.

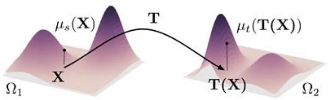


Figure: $\mu_s = P$, $\mu_t = Q$, T is the transport map (source: http://people.irisa.fr/Nicolas.Courty/OATMIL/)

The calculation of the Wasserstein metric is in general a difficult computational problem.

However, in certain cases explicit answers can be found:

- For measures on $\mathbb{R}$ a closed form result exists in terms of the quantile functions

$$d_2(P,Q) = \left\{ \int_0^1 (F^{-1}(s) - G^{-1}(s))^2 ds \right\}^{1/2},$$
$$F(x) = P((-\infty, x]), \;\; G(x) = G((-\infty, x]).$$

  This leads to a closed form solution for the barycenter as quantile average

$$F_B^{-1}(s) = \sum_{i=1}^{N} w_i F_i^{-1}(s),$$
$$F_B(x) = P_B((-\infty, x]), \;\; F_i(x) = P_i((-\infty, x]), \;\; i = 1, \cdots, N.$$

- For normal families $X \sim N(\mu_1, S_1)$, $Y \sim N(\mu_2, S_2)$ we have that

$$d_2(P,Q) = \left\{ \|\mu_1 - \mu_2\|^2 + Tr(S_1 + S_2 - 2(S_1^{1/2} S_2 S_1^{1/2})^{1/2}) \right\}^{1/2}$$

  This leads to a "closed" form solution for the barycenter $P_B \sim N(\mu_B, S_B)$

$$\mu_B = \sum_{i=1}^{N} w_i \mu_i,$$
$$S_B \text{ solves } S_B = \sum_{i=1}^{N} w_i (S_B^{1/2} S_i S_B^{1/2})^{1/2}$$

Why the Wasserstein metric?

- $d_2(\cdot, \cdot)$ is a true metric in the space of probability measures - unlike KL divergence - compatible with the weak-* topology
- It allows us to extend one of the most desirable properties of the Kullback-Leibler divergence - that of reducing robust decision problems within the exponential family of distributions to quadratic optimization problems - to any family
- It can be used to establish an upper bound for the difference in expected utility associated with using different probability measures for determining the expected utility associated with a random variable

$$|\mathbb{E}_P[u(X)] - \mathbb{E}_Q[u(X)]| \leq C\, d_2(P, Q).$$

Fréchet-Wasserstein mean utilities

We will now consider that $X \in \mathbb{R}$ (a natural assumption for a risk), focus on the multiplier case, and metrize $\mathbb{P}$ with the 2-Wasserstein metric.

Definition 2 (F-W multiplier preferences)

$$U_\theta(X) = \min_{P \in \mathbb{P}} [\mathbb{E}_P[u(X)] + \frac{\theta}{2}(F_{\mathbb{M}}(P) - F_{\mathbb{M}}(P_B))], \;\; \theta > 0,$$

$$F_{\mathbb{M}}(P) = \sum_{i=1}^{N} w_i W_2(P, P_i)$$

Theorem 3 (PeXY2021)

There exists a positive constant θ_c such that for $\theta > \theta_c$:

(a) *A minimizer P_θ of problem (2) i is expressed in terms of the quantile F_θ^{-1}, which is the solution of*

$$\frac{1}{\theta}u'(z) + z = F_{\mathcal{B}}^{-1}(s), \;\; s \in [0,1], \;\; \theta > 0, \tag{3}$$

and

$$U_\theta(X) = \int_0^1 u(F_\theta^{-1}(s))ds + \frac{1}{2\theta}\int_0^1 (u'(F_\theta^{-1}(s)))^2 ds. \tag{4}$$

(b) *Assuming that $u \in C^3$, the following perturbative expansion for sufficiently large θ holds:*

$$\begin{aligned} F_\theta^{-1} &= F_{\mathcal{B}}^{-1} - \frac{1}{\theta}u'(F_{\mathcal{B}}^{-1}) + \frac{1}{\theta^2}u''(F_{\mathcal{B}}^{-1})u'(F_{\mathcal{B}}^{-1}) + O\left(\frac{1}{\theta^3}\right), \\ U_\theta(X) &= \mathbb{E}_{P_B}[V_\theta(X)], \\ V_\theta(x) &:= u(x)\left(1 - \frac{1}{2\theta}\frac{(u'(x))^2}{u(x)} + \frac{1}{2\theta^2}\frac{(u'(x))^2 u''(x)}{u(x)}\right) + O\left(\frac{1}{\theta^3}\right). \end{aligned} \tag{5}$$

Comments

1. It holds that $F_\theta^{-1} \leq F_{\mathcal{B}}^{-1}$ and $U_\theta(X) \leq U_{\mathcal{B}}(X) := \mathbb{E}_{P_B}[u(X)]$ for all $\theta > \theta_c$, i.e., the DM underestimates X at all confidence levels as compared to the Wasserstein barycenter.
2. $F_\theta^{-1} \to F_{\mathcal{B}}^{-1}$ and $U_\theta(X) \to U_\beta(X) = \mathbb{E}_{P_B}[u(X)]$ as $\theta \to \infty$.
3. If u' is a decreasing function we get the sharper estimate $F_\theta^{-1} \leq F_{\mathcal{B}}^{-1} - \frac{1}{\theta} u'(F_{\mathcal{B}}^{-1})$ for all $\theta > \theta_c$.
4. The expansion in (5) looks like an expected utility representation, but note that the "utility function" V_θ is not in general an increasing function of x unless θ is large enough.
5. The above expressions are interesting as they present the interplay between risk aversion (as quantified by the derivatives of u) and ambiguity aversion (as quantified by θ).

6. For the case of CRRA utilities of the form

$$u_\gamma(x) = \begin{cases} \ln x, & \text{for} \quad \gamma = 1, \\ \frac{x^{1-\gamma}}{1-\gamma}, & \text{for} \quad \gamma > 1. \end{cases}$$

we obtain - upon defining $\gamma_1 = 1 + 2\gamma$, $\gamma_2 = 2 + 3\gamma$ - that

$$U_\theta(X) = \mathbb{E}_{P_B}[u_\gamma(X)] - \frac{1-\gamma_1}{2\theta}\mathbb{E}_{P_B}[u_{\gamma_1}(X)] - \frac{\gamma(1-\gamma_2)}{2\theta^2}\mathbb{E}_{P_B}[u_{\gamma_2}(X)] + O\left(\frac{1}{\theta^3}\right), \tag{6}$$

with the leading order in the expansion for $U_\theta(X)$ is the barycentric expected utility, while the corrections can be interpreted again as barycentric expected utilities albeit corresponding to other members of the CRRA family, but importantly with larger risk aversion coefficients.

7. A similar result can be obtained for CARA utility functions $u(x) = \frac{1}{\lambda}(1 - e^{-\lambda x})$, as

$$V_\theta(x) = \frac{1}{\lambda}(1 - e^{-\lambda x}) - \frac{\lambda}{2\theta}e^{-2\lambda x} - \frac{\lambda^2}{2\theta^2}e^{-3\lambda x} + O\left(\frac{1}{\theta^3}\right). \tag{7}$$

Marginal utility

An important tool in valuation is marginal utility (it can provide estimates for prices).

Definition 4 (Marginal ambiguity-averse utility)

The marginal utility of X is defined as

$$\mathcal{M}_\theta(X) := \lim_{\epsilon \to 0} \frac{1}{\epsilon}(U_\theta(X+\epsilon) - U_\theta(X)),$$

(provided that the limit exists) where $\epsilon > 0$ is a non-random infinitesimal endowment

Proposition 5 (PeXY2021)

For $\theta > \theta_c$ it holds that

(a) *The marginal utility $\mathcal{M}_\theta$ is represented as*

$$\mathcal{M}_\theta(X) = \int_0^1 u'(F_\theta^{-1}(s))ds = \mathbb{E}_{Q_\theta}[u'(X)]. \tag{8}$$

Morever, it holds that

$$\mathcal{M}_\theta(X) \geq \mathcal{M}_\mathcal{B}(X) = \mathbb{E}_{P_B}[u'(X)], \ \ \forall\, \theta > \theta_c,$$

while $\mathcal{M}_\theta(X) \to \mathcal{M}_\mathcal{B}(X)$ for $\theta \to \infty$.

(b) *For large θ, and assuming sufficient smoothness for u, marginal utility admits the expansion*

$$\mathcal{M}_\theta(X) = \mathbb{E}_{P_B}[u'(X)\mathcal{C}_\theta(X)], \tag{9}$$

where $C_\theta(X)$ is a correction factor of the form

$$\mathcal{C}_\theta(X) = 1 - \frac{1}{\theta}u''(X) + \frac{1}{\theta^2}\left[(u''(X))^2 + \frac{1}{2}u'''(X)u'(X)\right] + O\left(\frac{1}{\theta^3}\right). \tag{10}$$

Comment

The statement of the above proposition admits a very intuitive interpretation.

Since $F_\theta^{-1} \le F_{\mathcal{B}}^{-1}$ for all $\theta > 0$, which means that the uncertainty averse agent using the probability model with quantile function F_θ^{-1} provides statistical estimates for X, which at all confidence levels are lower than the corresponding estimates provided by the model related to the Wasserstein barycenter of $\mathbb{M}$.

However, by Proposition 5(a), the marginal utility of the uncertainty averse agent admits a representation under the model with quantile function F_θ^{-1}.

Since F_θ^{-1} underestimates $F_{\mathcal{B}}^{-1}$, the marginal utility corresponding to model F_θ^{-1} will be higher than the marginal utility corresponding to model $F_{\mathcal{B}}^{-1}$ (since u' is a decreasing function so that $u'(F_\theta^{-1}) \ge u'(F_{\mathcal{B}}^{-1})$).

Fréchet risk measures

Analogous concepts can be used for the definition of convex risk measures that can be used for risk assessment and management in the presence of multiple priors [PaY2018].

Here we expand the concept in the following sense:

The risk X depends on a vector of risk factors $Z = (Z_1, \cdots, Z_d)$ and there are N models available for their probability laws,

$$\mathbb{M} = \{P_1, \cdots, P_N\}.$$

The risk X is connected with the risk factors through the risk mapping

$$Z = (Z_1, \cdots, Z_d) \mapsto -X = \Phi_0(Z).$$

Convex risk measures in a nutshell

Convex risk measures (Fölmer) are important concepts in modern risk management which serve to define a capital requirement (on behalf of the regulatory authority).

Definition 6 (Convex risk measures)

Let $\mathcal{L}$ is a space of random variables containing the relevant financial positions. A convex risk measure $\rho : \mathcal{L} \to \mathbb{R}$ is a mapping that satisfies the following properties

1. Monotonicity: If $X \leq Y$ then $\rho(X) \geq \rho(Y)$
2. Cash invariance: If $m \in \mathbb{R}$ then $\rho(X + m) = \rho(X) - m$
3. Convexity: $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y)$, $0 \leq \lambda \leq 1$.

The following robust representation of convex risk measures due to Fölmer is extremely important:

$$\rho(X) = \max_{P \in \mathbb{P}} [\mathbb{E}_P[-X] - \alpha(P)],$$

where $\alpha : \mathbb{P} \to \mathbb{R}$ is a convex penalty function.

Definition 7 (**Fréchet risk measure**)

Let $\alpha : \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\}$ be an increasing function, such that $\alpha(0) = 0$, and let $\Phi_0 : \mathbb{R}^d \to \mathbb{R}_+$ be the risk mapping connecting the stochastic factors Z to the risk position X of the firm. We define the *Fréchet risk measure* for any $\gamma \in (0, \infty)$ as

$$\rho(X) := \sup_{P \in \mathbb{P}(\mathbb{R}^d)} \left\{ \mathbb{E}_P[-X] - \frac{1}{2\gamma} \alpha(F_{\mathbb{M}}(\mu)) \right\}, \tag{11}$$

where $\mathbb{M}$ is the set of priors for Z, $-X = \Phi(Z)$ and $F_{\mathbb{M}}$ the normalized Fréchet function.

Proposition 8 (Properties of Fréchet risk measures)

Consider the measurable space $(\Omega, \mathcal{F})$ *where here* $\Omega = \mathbb{R}^d$ *and* $\mathcal{F} \subseteq \mathcal{B}(\mathbb{R}^d)$ *a given* σ*-algebra.*

(i) *The Fréchet risk measures considered as mappings* $\rho : \mathcal{L} \to \mathbb{R}_+$, *where* $\mathcal{L} := \{X : \Omega \to \mathbb{R} : X \ \mathcal{F} - \text{measurable} \}$ *are convex risk measures, such that* $\rho(X;\gamma) \geq \mathbb{E}_{P_B}[-X]$, *for all* $\gamma \geq 0$, *where* P_B *is the Fréchet barycenter of* $\mathbb{M}$.

(ii) *For any fixed* $-X = \Phi_0(Z)$, *it holds that* $\rho(X;\gamma_1) \leq \rho(X;\gamma_2)$ *for any* $0 \leq \gamma_1 \leq \gamma_2$, *while*

$$\lim_{\gamma \to 0^+} \rho(X;\gamma) = \mathbb{E}_{P_B}[\Phi_0(Z)] \leq \rho(X;\gamma) \leq \operatorname{ess} \sup_{Z \in \Omega = \mathbb{R}^d} \Phi_0(Z),$$

where P_B *is the Fréchet barycenter of* $\mathbb{M}$.

The above proposition implies that the parameter γ plays the role of an uncertainty aversion parameter

When metrizing the space of probability measures on $\mathbb{R}^d$ using the Wasserstein metric explicit calculations are possible.

Proposition 9

Assume that $\mathbb{M} = \{\mu_i, \;\; i = 1, \dots, n\}$ *with* $\mu_i = LS(m_i, S_i)$ *with* $m_i \in \mathbb{R}^d$ *and* $S_i \in \mathbb{P}(d)$*, where* $\mathbb{P}(d) \subset \mathbb{R}^{d\times d}$ *is the set of positive definite symmetric matrices. If the position of the firm is provided by the risk mapping* $-X = \Phi_0(Z)$ *then* $\rho(X)$ *is calculated as the solution of the matrix optimization problem*

$$\rho(X) = \sup_{(m,S)\in\mathbb{R}^d\times\mathbb{P}(d)} \left\{ \underbrace{\int_{\mathbb{R}^d} \Phi_0(\mu + S^{1/2}z)dP(z)}_{\Phi} \right. \tag{12}$$
$$\left. -\frac{1}{2\gamma}\left(\sum_{i=1}^{n} w_i \|m - m_i\|^2 + \sum_{i=1} w_i Tr(S + S_i - 2(S_i^{1/2} S S_i^{1/2})^{1/2}) \right) \right\}$$

The maximizer (m, S) *to the above problem can be found as the solution of the set of matrix equations (first order conditions):*

$$\begin{aligned} \gamma D_m \Phi(m, S) - (m - \sum_{i=1}^{n} w_i m_i) = 0, \\ 2\gamma S^{1/2} D_S \Phi(m, S) S^{1/2} - (S - \sum_{i=1}^{n} w_i (S^{1/2} S_i S^{1/2})^{1/2}) = 0, \end{aligned} \tag{13}$$

Applications

We illustrate the use of the proposed utility functionals and risk measures using 3 indicative applications:

1. Application in estimating the social discount rate
2. Application group decision making - CAT bond pricing
3. Application in estimation of risk premia under model uncertainty.

The social discount rate

The social discount rate (SDR) is one of the most fundamental but also controversial parameters in cost-benefit analysis.

In the absence of model uncertainty the SDR is determined by the classical consumption-based Ramsey discounting formula

$$r(t) = \delta - \frac{1}{t} \ln \frac{\mathbb{E}[u'(C(t))]}{u'(C(0))},$$

where δ is the utility discount rate, $C(t)$ is consumption at time t (which is a random variable) and $C(0)$ is today's consumption. Intuitively, if expected marginal utility in the future is higher, then the future is discounted less.

This formula provides a term structure for r and is a crucial parameter in standard cost-benefit analysis (Gollier 2013).

Assume now that a regulator seeks to calculate the SDR in the presence of model uncertainty.

Assume that for any fixed $t > 0$, the random variable $X = C(t)$ is the unknown consumption at this instant in time, and that there is a set of models $\mathbb{M}_t$ of probability measures concerning the distribution of $C(t)$, described in terms of the quantiles $F_{t,i}^{-1}$

For any fixed $t > 0$, a direct application of Theorem 3 for $X = C(t)$ allows the calculation of the utility functional $U_\theta(C(t))$, whereas by repeating the arguments that led to the results in Proposition 5, we can see that the relevant SDR formula now assumes the form

$$r(t) = \delta - \frac{1}{t} \ln \frac{[\mathcal{M}(C(t))]}{u'(C(0))} = \delta - \frac{1}{t} \ln \frac{\mathbb{E}_{P_t^*}[u'(C(t)]}{u'(C(0))}, \tag{14}$$

where the standard expected marginal utility is now replaced by $\mathcal{M}(C(t))$ (see Definition 4 and Proposition 5) and P_t^* is the probability measure corresponding to the quantile function $(F_t^*)^{-1}$.

With regard to formula (14), it should be noted that:

(a) This seemingly simple formula takes model uncertainty fully into consideration since the effects of uncertainty are included in the minimizing quantile $(F_t^*)^{-1}$.
(b) In the limit as $\theta \to \infty$, $r(t) \to r_{\mathcal{B}}(t)$, and the barycentric SDR $r_{\mathcal{B}}$ is obtained.
(c) Since by Proposition 5 it holds that $\mathbb{E}_{P_t^*}[u'(C(t)] > \mathbb{E}_{P_{t,\mathcal{B}}}[u'(C(t)]$, and keeping in mind that $u'(C(0)) > 0$, we conclude that

$$r(t) < r_{\mathcal{B}}(t), \;\; t \in \mathbb{R}_+, \;\; \theta > 0, \tag{15}$$

which implies that the effect of uncertainty aversion is to decrease the SDR relative to the SDR obtained under risk aversion with expected utility defined in terms of the Wasserstein barycenter model $P_{\mathcal{B}}$. This can be regarded as a second-order precautionary effect.
(d) The perturbative expansions obtained can be used to analytically approximate the SDR using formula (14) to provide information on the dependence of the SDR on various parameters of interest (such as θ or in the case of CRRA utilities the risk aversion coefficient γ).
The effect of uncertainty aversion, at least to first order in $\frac{1}{\theta}$, is to decrease the SDR as compared to the barycentric one (ambiguity prudence effect).

A numerical experiment

Following Gollier (2013, Ch. 4), we assume that the consumption process $C(t)$ follows a single factor (autoregressive) model of the form

$$\begin{aligned} C(t+1) &= C(t)\exp(x(t)), \\ x(t+1) &= \mu + y(t) + \varepsilon_x(t), \\ y(t) &= \phi y(t-1) + \varepsilon_y(t), \end{aligned} \tag{16}$$

where $\varepsilon_x(t), \varepsilon_y(t)$ are independent and serially independent with $\mathbb{E}[\varepsilon_x(t)] = \mathbb{E}[\varepsilon_y(t)] = 0$ and $Var(\varepsilon_x(t)) = \sigma_x^2$, $Var(\varepsilon_y(t)) = \sigma_y^2$, y_{-1} is some initial state, and $\phi \in [0,1]$ is a parameter representing the degree of persistency (mean reversion) of y.

The choice $\phi = 0$ reduces the model to a standard random walk model which is a discretization of a Wiener process.

The case where $\phi \neq 0$ corresponds to a discretization of an Ornstein-Uhlenbeck process.

Typically, $\{y(t)\}$ is an unobserved stochastic factor, which has an effect on the observed growth rate $\{x(t)\}$ of the consumption process $\{C(t)\}$.

A straightforward induction procedure shows that, given ϕ and y_{-1}, the stochastic consumption process $\{C(t)\}$ is lognormally distributed and in particular

$$\ln C(t) - \ln C(0) \sim N(\mu_t, \sigma_t^2),$$

where

$$\mu_t = \mu\, t + y_{-1}\frac{1-\phi^t}{1-\phi},$$

$$\sigma_t^2 = \frac{\sigma_y^2}{(1-\phi)^2}\left[t - 2\phi\frac{\phi^t - 1}{\phi - 1} + \phi^2\frac{\phi^{2t}-1}{\phi^2-1}\right] + \sigma_x^2 t.$$

When all the parameters and the distributions of noise terms concerning model (16) are fully known, i.e., when we are in a world of a single model, the Ramsey formula can be used to produce a term structure for the SDR (Gollier 2013).

Using the general class of CRRA utilities, Gollier produces an analytic formula for the term structure of the discount rate as

$$r(t) = \delta + \gamma\frac{1}{t}\mu_t - \frac{1}{2}\gamma^2\frac{1}{t}\sigma_t^2.$$

Bansal and Yaron (2004) calibrated the factor model for consumption to data from the period 1929-1998 using annual data from the USA, producing estimates for the monthly mean return and volatilities of $\mu = 0.0015$, $\sigma_x = 0.0078$, $\sigma_y = 0.00034$, and estimated the reversion parameter as $\phi = 0.979$.

Using these parameter values, Gollier implemented formula (16) to produce a term structure which is increasing or decreasing depending on the sign of y_{-1}.

In particular Gollier used a range of values for $y_{-1} \in [-0.001, 0.001]$ for his numerical experiments for the term structure. In the case where $\phi = 0$, the term structure is flat.

Even if we trust the autoregressive model for the evolution of consumption, there are parameters related to the hidden variables included in the model, the value of which can be doubted.

For the sake of illustration consider the two parameters ϕ and y_{-1}.

Different estimations or opinions regarding the parameters $p = (\phi, y_{-1})$ produce different parameters μ_t and σ_t^2, therefore different models for the distribution of $C(t)$.

The important question that arises is how the emergence of multiple models affects the SDR relative to the single model case (16).
We study this case when each model in the set $\mathbb{M}_t$ is a lognormal model of the type presented in (16).

Each has parameters with values $\mu = 0.0015$, $\sigma_x = 0.0078$, $\sigma_y = 0.00034$ (see Bansal and Yaron 2004), but has a different choice for the parameters ϕ or y_{-1} or both.

It is quite natural to allow some uncertainty for these parameters and in particular for y_{-1} as these are the ones whose estimation is more delicate.

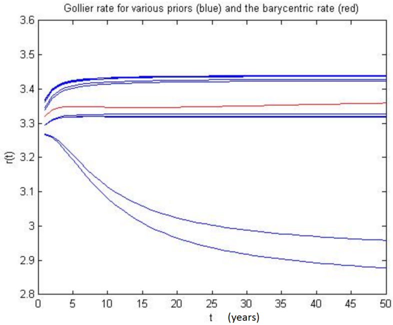


Figure: The individual term structure curves (blue) and the barycentric term structure (red).

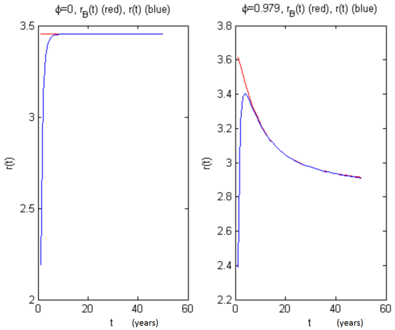


Figure: Comparison of the term structure predicted by the barycentric model $r_{\mathcal{B}}(t)$ (red line) and the uncertainty averse Fréchet-Wasserstein model $r(t)$ (blue line) for two different choices of the parameters in the Gollier model (16).

The Wasserstein barycenter in consensus group decision making: Application to CAT bond pricing

We consider the use of the Wasserstein barycenter in group decision making and show that it can be used in the formulation of proposals that are most likely to lead to consensus of a group of DMs with varying models concerning an unknown risk.

In particular we show that the use of the Wasserstein barycenter as a commonly acceptable probability model upon which a decision is made is the one which maximizes the probability of reaching consensus.

While the above framework can be encountered in a wide variety of situations, we choose as a motivating example the issuance and pricing of a CAT bond.

CAT bonds are risk sharing instruments, used by a group of firms in the management of extreme risks such as natural and climate-change related disaster risks (e.g., earthquakes, hurricanes, floods, fires) and cybersecurity risks.

CAT bonds are issued by insurance and reinsurance firms, corporations, government bodies and others.

The important effects of ambiguity on the spreads of CAT bonds have been noted quite early in the literature (see e.g. Bantwal and Kunreuther 2000).

Another important aspect of CAT bonds is the involvement of multiple agents in their design and pricing (see e.g. Edesess 2015), a fact that affects the calculation of the expected loss, which according to econometric studies is one of the key drivers in CAT bonds prices (see e.g. Galeotti et al. 2013).

The application in this section focuses on these two last aspects, and in particular on the determination of a model for the risk which maximizes the probability of acceptance by all parties involved, and on the resulting pricing of the CAT bond. For a detailed review on CAT bonds, including their design and mechanics, see Cummins and Barrieu (2013) and the references therein.

As a consequence of their nature, more than one agent is involved in the issuance of a CAT bond, such as insurers; reinsurers; corporations; pension funds; structuring agents who assist the issuer in selecting trigger type and are involved in placing the bond with investors (investment banks or brokers); modelling agents who estimate the risk based on models and simulations (e.g., Risk Management Solution, Inc, or Eqcat); ratings agencies and others (see for example Edesess 2015).

All, or the majority, of these different actors must agree on some common characteristics concerning the contract structure of the CAT bond, which are related to a common agreement concerning the estimation of the extreme risk. Since extreme risks are by nature rare events, the lack of sufficient historical data places them within the realm of model uncertainty, as it is not possible on the basis of statistical evidence to single out a unique probabilistic model for the random variable L corresponding to the risk.

On the other hand, in order for all parties to agree upon the issue and the actual contract terms, a commonly agreed model for the distribution of the extreme risk must be adopted. The agreement is a necessity, as the issuance of a risk sharing instrument is of mutual benefit to all parties.

Since in principle each agent involved may have a different prior for the risk, the valuation has to be effected by a commonly agreed probability model for the risk, or at least by a model which the agents involved have the maximum probability of agreeing upon.

The issue of identifying such a model of common acceptance is important for the construction and pricing of the CAT bond.

CAT bond fundamentals

This type of instrument has become very popular in recent years, as a vehicle for transferring extreme risks from insurers and re-insurers to investors.

It constitutes a tool that enables:

(a) extreme risks to be covered more efficiently, providing solvency to those involved in the insurance business; and

(b) attractive investment opportunities with potentially high returns to be provided to investors, which are largely uncorrelated to other market indices, hence offering at the same time a useful hedging tool.

The basic structure of a CAT bond is as follows.

- A sponsor or group of sponsors, typically a reinsurer, contacts a special purpose vehicle (SPV) in order to enter an alternative reinsurance contract which will guarantee solvency in case of occurrence of extreme losses.
- The sponsor, at the cost of some premium ρ, will receive insurance coverage up to some level h, in the case of extreme losses.
- The SPV for its own coverage, and in order to guarantee the possibility of covering amount h for the sponsor, issues a CAT bond, which is a standard defaultable bond, with the default triggered by the event of extreme losses of the sponsor.

Several payoff structures are possible: if the amount h is issued in bonds, then
(a) in the absence of a triggering event, the bond provides coupons to the investors corresponding to interest $r + \rho$, where r is a standard interest rate (e.g., LIBOR) and a principal h; while
(b) in the presence of the triggering event, coupons are reduced to $(r+\rho)(1-d_1)$ and the principal is reduced to $h(1-d_2)$, for suitable d_1, d_2.
Other payoff structures are possible, and there exists a variety of CAT-based derivatives such as CAT swaps which provide a multitude of opportunities for risk sharing and efficient risk management.

However, the success of such instruments, especially in the primary market, crucially depends on the choice of the premium ρ, which in turn is related to the spread of the CAT bond.

Numerous theoretical and empirical studies have shown that the most important quantity is the expected loss $EL = \mathbb{E}[G(L)]$ where L is the random variable corresponding to the catastrophic risk and $G : \mathbb{R}_+ \to \mathbb{R}_+$ is an appropriate function related to the cover agreement between the sponsor(s) and the SPV.

One commonly used model for the spread is a linear model of the form $\rho = c_1 + c_2 EL$ (Galeotti et al. 2013) for appropriate constants c_1, c_2 which are determined by linear regression and may incorporate geographic or seasonal effects.

Other models are based upon utility pricing arguments and result in nonlinear models of the form $\rho = EL + \gamma (PFL)^{\alpha} (CEL)^{\beta}$, where $PFL = P(L > a)$ is the probability of first loss and $CEL = \frac{\mathbb{E}[L_{(a,a+h)} | L > a]}{h} = \frac{\mathbb{E}[L_{(a,a+h)}]}{hP(L>a)}$ is the conditional expected loss.

Consensus achievement and pricing of CAT bonds

The above discussion clearly indicates the need for agreement on the probability $P(L \geq x)$, no matter which pricing methodology is adopted for the CAT bond.

This requires the development of a scheme which allows the parties involved in the design of the CAT bond to reach consensus concerning the probability model for the extreme risk.

Even though each agent may have a different prior concerning the probability $P(L \geq x)$, it is in their common interest to agree on a common model that will favor the issuance of the bond, hence each agent will be willing to change her/his initial prior and accept a new probability model for L as long as the uneasiness caused by this change is not too high.

It is reasonable to assume that this uneasiness is an increasing function of the distance $d(Q, Q_i)$ between the prior Q_i of agent i, concerning the risk, and the adopted probability measure Q.

Put differently, the probability p_i of agent i accepting the probability measure Q can be expressed as $p_i = \varphi_i(d^2(Q, Q_i))$, where $\varphi_i : \mathbb{R}_+ \to [0, 1]$ is a decreasing function characterizing the strength of each agent's belief in the prior and her/his willingness to move.

If the agents are independent, the probability of all of them agreeing with the mediator's proposal is equal to $p = p_1 \cdots p_n$.

The choice of the probability measure Q that satisfies as many agents as possible can then be expressed as the problem of choosing Q so as to maximize probability p.

Since the problem of maximizing p is equivalent to the problem of maximizing $\ln(p)$ under the above assumptions, it can be seen that the probability measure which will maximize the probability of attaining a consensus is the solution to

$$\max_{Q \in \mathcal{P}} \sum_{i=1}^{n} \ln(\varphi_i(d^2(Q, Q_i)). \tag{17}$$

We adopt the metrization of the space of probability measures in terms of the Wasserstein metric $d(Q, Q_i) = W_2(Q, Q_i)$ and consider the corresponding problem (17).

We will show that the probability measure Q which maximizes the probability of all agents agreeing to it, is the Fréchet barycenter of the set of models, with a choice of weights which corresponds to the functions φ_i.

Proposition 10

Assume that the models Q_i, $i = 1, \cdots, n$, for risk L are expressed in terms of the probability distributions F_i and the corresponding quantiles F_i^{-1} and define the quantities $M_{ij} := \int_0^1 F_i^{-1}(s) F_j^{-1}(s) ds$, $i, j = 1, \cdots, n$. Moreover, assume sufficient smoothness and integrability conditions for the decreasing functions $\varphi_i : \mathbb{R}_+ \to [0,1]$ and let $\Phi_i := -\ln(\varphi_i)$, $i = 1, \cdots, n$, be increasing and convex.
A probability measure $Q \in \mathcal{P}(\mathbb{R})$ that maximizes the probability of agreement of all agents coincides with the Wasserstein barycenter P_B represented by the distribution function F_B given by the quantile average

$$F_B^{-1} = \sum_{i=1}^{n} w_i F_i^{-1},$$

Proposition 10 shows that the probability measure for L that maximizes the probability of agreement of all agents corresponds to the Wasserstein barycenter with a particular choice of weights, which are endogenously determined in terms of the elasticities of the functions φ_i which model the rigidity of the various agents to their priors.

In some sense these reflect the bargaining power or authority of each agent in the group.

Note that the problem studied in Proposition 10 is formally similar to (and in fact inspired by) a Nash bargaining game in the space of probability models (measures).

If agents are symmetric with $\varphi_i(z) = \exp(c_i - c\,z)$, $i = 1, \cdots, n$ for c_i, c appropriate constants and with c_i possibly varying from agent to agent but c being the same for all agents, the resulting weights will be $1/n$.

This interpretation of the Wasserstein barycenter provides a further argument in favor of its use as a decision-making tool under model uncertainty.

We are now ready to proceed with the pricing of the CAT bond. A

class of suitable models for the extreme risk is the class of generalized extreme value (GEV) distributions, described by the probability distribution functions

$$F_i(x) = \begin{cases} \exp\left(-\left(1+\xi_i\left(\frac{x-\mu_i}{\sigma_i}\right)\right)^{-1/\xi_i}\right), & \xi_i \neq 0, \\ \exp\left(-e^{-\frac{(x-\mu_i)}{\sigma_i}}\right), & \xi_i = 0, \end{cases}$$

Given a set of weights $w = (w_1, \cdots, w_n)$ determined in the context of Proposition 10, the corresponding Wasserstein barycenter is $F_{\mathcal{B}}^{-1} = \sum_{i=1}^{n} w_i F_i^{-1}$.

In the case where all models correspond to the same shape parameter $\xi_i = \xi$, the Wasserstein barycenter corresponds to a member of the GEV family with $\mu_B = \sum_{i=1}^{n} w_i \mu_i$, $\sigma_B = \sum_{i=1}^{n} w_i \sigma_i$ and $\xi_B = \xi$.

In the general case, the quantile function $F_{\mathcal{B}}^{-1}$ is explicitly known and its inversion is therefore an easy numerical task (though not feasible in closed form).

The family of GEV distributions, which encompasses in one family the three types of extreme value distributions (Gumbel, Weibull and Fréchet), has been successfully used in the literature to model the distribution of extreme risks such as earthquakes and floods.

The premium will be determined in terms of the quantity $\mathbb{E}[L_{(a,a+h)}]$ which will now be calculated under the Wasserstein barycenter, in terms of

$$EL = \mathbb{E}_{Q_{\mathcal{B}}}[L_{(a,a+h)}] = \int_0^1 G(F_B^{-1}(s))ds = \int_{F_B(a)}^{F_B(a+h)} (F_B^{-1}(s) - a)ds + h(1 - F_B(a+h)). \quad (20)$$

easily computed numerically for the Wasserstein barycenter.

For certain special cases, such as when all the agents have models with the same shape parameter $\xi_i = \xi$ (which is a reasonable assumption for certain types of extreme risks which are modelled by the Gumbel type), the calculation can be performed analytically. In such cases,

$$F_{\mathcal{B}}^{-1}(s) = \begin{cases} \mu_B + \frac{\sigma_B}{\xi}[(-\ln s)^{-\xi} - 1], & \xi \neq 0, \\ \mu_B - \sigma_B \ln(-\ln s), & \xi = 0, \end{cases} \quad , \text{ and } \; F_B(x) = \begin{cases} \exp\left(-\left(1 + \xi\left(\frac{x-\mu_B}{\sigma_B}\right)\right)^{-1/\xi}\right), \\ \exp\left(-e^{-\frac{(x-\mu_B)}{\sigma_B}}\right), \end{cases}$$

so that EL can be approximated in terms of the exponential integral function E_1 or an appropriate series expansion.

Risk premia estimation under model uncertainty

Consider the standard risk model

$$X = \sum_{i=1}^{N} C_i, \quad C_i \text{ i.i.d. } N \sim Pois(\lambda).$$

Assume that

- N depends on a set of risk factors Z_1 in terms of the risk mapping $\lambda \mapsto \Phi_1(Z_1)$.
- C_i depends on a set of risk factors Z_2 in terms of the risk mapping $\lambda \mapsto \Phi_2(Z_2)$.

Then, conditioning on Z_1

$$E[-X] = \mathbb{E}[\Phi_1(Z_1)]\,\mathbb{E}[\Phi_2(Z_2)]$$

Assuming $Z = (Z_1, Z_2) \sim N(\mu, S)$ a direct application of Proposition 9 can be used to obtain $\rho(X)$ and through that risk premia estimations.

Extensive numerical experiments indicate that the proposed risk measure performs in a satisfactory manner.

Conclusion

- We have explored the possibility of using tools from the theory of optimal transportation and in particular the concept of Wasserstein distance in decision theory
- By quantifying the dis-similarity between various models in terms of the Wasserstein distance we propose
  - A class of variational utilities suitable for multiple priors
  - A class of risk measures suitable for multiple priors
- Both are amenable to almost closed form solutions which provide interesting insights between risk and uncertainty
- The proposed approach is illustrated in terms of selected applications.

Selected References (see also references therein)

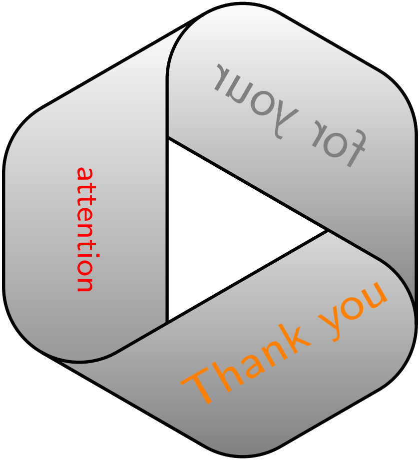
Marinacci, Massimo. Model uncertainty. Journal of the European Economic Association 13 (6) (2015): 1022-1100.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. Econometrica 74 (6) (2006): 1447-1498.

Föllmer, H., and A. Shied. Stochastic finance, De Gruyter (2004)

Petracou, Electra V., Anastasios Xepapadeas, and Athanasios N. Yannacopoulos. Decision Making Under Model Uncertainty: Fréchet - Wasserstein Mean Preferences. Management Science (2021).

Papayiannis, G. I., and A. N. Yannacopoulos. Convex risk measures for the aggregation of multiple information sources and applications in insurance. Scandinavian Actuarial Journal 2018 (9) (2018): 792-822.


Thank you
for your
attention