Abstract Time Consistency and Decompostion

Henri Gerard
J.-P. Chancelier, M. De Lara, P. Carpentier

ENPC - CERMICS
Labex Bézout - LIGM

Búzios, Brazil, July 01, 2016

Time consistency in a nutshell

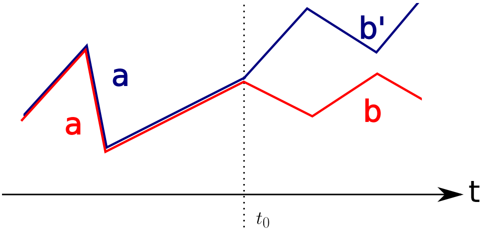
Given two processes $(\boldsymbol{X}_0, \cdots, \boldsymbol{X}_T)$ and $(\boldsymbol{Y}_0, \cdots, \boldsymbol{Y}_T)$,

- we look for numerical evaluations (risk measures) of the tails $(\boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_T)$ of the process
- that satisfy time consistency, in the same way that the mathetical expectation does in

$$\mathbb{E}_{\mathbb{P}}[\boldsymbol{X}_0 + \cdots + \boldsymbol{X}_T] = \mathbb{E}_{\mathbb{P}}\big[\boldsymbol{X}_0 + \cdots + \boldsymbol{X}_t + \underbrace{\mathbb{E}_{\mathbb{P}}[\boldsymbol{X}_{t+1} + \cdots + \boldsymbol{X}_T \mid \mathcal{F}_t]}_{\text{tail of the process}}\big]$$

Such a property is essential to establish a dynamic programming equation in dynamic optimization

Illustration of time consistency



- $(\mathcal{F}_t)_{t\in[\![0;T]\!]}$ filtration of Ω
- $\mathcal{X}_t = L^p(\Omega, \mathcal{F}_t, \mathbb{P})$
- $\mathcal{X}_{t_1,t_2} = \mathcal{X}_{t_1} \times \mathcal{X}_{t_1+1} \times \cdots \mathcal{X}_{t_2-1} \times \mathcal{X}_{t_2}$
- $\mathbb{A} = \mathcal{X}_{1,t_0}$
- $\mathbb{B} = \mathcal{X}_{t_0+1,T}$

We focus on the risk averse case

- There is an extensive literature on the subject:
  Epstein and Schneider (2003), Ruszczynski and Shapiro (2006), Artzner, Delbaen, Eber, Heath, and Ku (2007), Ruszczyński (2010), Pflug and Pichler (2012)
- We want to study time consistency for
  - a general criterion (not necessarily time-additive and with dynamic risk measures)

$$\mathbb{F}[\boldsymbol{X}_0, \cdots, \boldsymbol{X}_T] = \mathbb{F}_0\big[\boldsymbol{X}_0, \cdots, \boldsymbol{X}_t, \underbrace{\mathbb{F}_{t+1}[\boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_T \mid \mathcal{F}_t]}_{\text{tail of the process}}\big]$$

  - with general structure of information (not necessarily filtration, to account for decentralized information among agents)

Outline

1. Abstract notion of time consistency and characterization
2. Revisiting classical examples of the literature
   - Artzner, Delbaen, Eber, Heath, and Ku (2007)
   - Ruszczyński (2010)
3. Conditions for time consistency
4. Perspectives for optimization under risk and conclusion

Outline of the section

Notations and abstract notion of time consistency

We introduce the following notations

- $\mathbb{A}$ and $\mathbb{B}$ are two sets
- $\preccurlyeq_{\mathbb{B}}$ is a preorder on $\mathbb{B}$
- $\preccurlyeq_{\mathbb{A}\times\mathbb{B}}$ is a preorder on $\mathbb{A} \times \mathbb{B}$

Definition (Time consistency)

$$b \preccurlyeq_{\mathbb{B}} b' \Rightarrow (a, b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a, b')$$

Variations around time consistency

Definition (Strong time consistency)

$$\begin{matrix} b \preccurlyeq_{\mathbb{B}} b' \\ a \preccurlyeq_{\mathbb{A}} a' \end{matrix} \Rightarrow (a, b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a', b') \ , \ \ \forall (a, a') \in \mathbb{A}^2 \ , \ \ \forall (b, b') \in \mathbb{B}^2$$

Definition (Time consistency)

$$b \preccurlyeq_{\mathbb{B}} b' \Rightarrow (a, b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a, b') \ , \ \ \forall a \in \mathbb{A} \ , \ \ \forall (b, b') \in \mathbb{B}^2$$

Definition (Weak time consistency)

$$b \sim_{\mathbb{A}} b' \Rightarrow (a, b) \sim_{\mathbb{A}\times\mathbb{B}} (a, b') \ , \ \ \forall a \in \mathbb{A} \ , \ \ \forall (b, b') \in \mathbb{B}^2$$

We will focus on weak time consistency

Proposition

We have the following implications

$$\begin{gathered}\textit{Strong time consistency}\\ \Downarrow\\ \textit{Time consistency}\\ \Downarrow\\ \textit{Weak time consistency}\end{gathered}$$

We can work with preorders induced by mappings

Definition (Time consistency for mappings (or mapping induced orders))

Consider two mappings $g : \mathbb{A} \times \mathbb{B} \to \mathbb{X}$ and $f : \mathbb{B} \to \mathbb{Y}$, where the sets $\mathbb{X}$ and $\mathbb{Y}$ are respectively equipped with the preorders $\preccurlyeq_{\mathbb{X}}$ and $\preccurlyeq_{\mathbb{Y}}$
The quadruplet $(g, \preccurlyeq_{\mathbb{X}}, f, \preccurlyeq_{\mathbb{Y}})$ is said to satisfy time consistency when

$$f(b) \sim_{\mathbb{Y}} f(b') \Rightarrow g(a, b) \sim_{\mathbb{X}} g(a, b')$$

We say that f is a factor and that g is an aggregator

Example

$$f\big((b_{t_0+1}, \dots, b_T)\big) = \mathbb{E}_{\mathbb{P}}[b_{t_0+1} + \cdots + b_T \mid \mathcal{F}_{t_0}]$$
$$g\big((a_1, \dots, a_{t_0}, b_{t_0+1}, \dots, b_T)\big) = \mathbb{E}_{\mathbb{P}}[a_1 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T]$$

We introduce a set-valued mapping

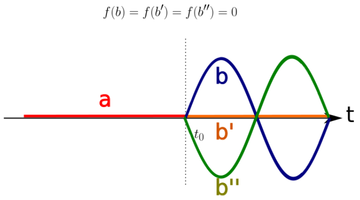
Given an aggregator $g : \mathbb{A} \times \mathbb{B} \to \mathbb{X}$ and a factor $f : \mathbb{B} \to \mathbb{Y}$, we introduce a set-valued mapping, called subaggregator

Definition

We denote by $\phi^{f,g} : \mathbb{A} \times \mathbb{Y} \rightrightarrows \mathbb{X}$ the set-valued mapping

$$\phi^{f,g}(a, y) = \{g(a, b) : b \in f^{-1}(y)\}$$

If $y \notin \mathrm{Im}(f)$ then $\phi^{f,g}(a, y) = \emptyset$



Example

$$f(b) = \mathbb{E}_{\mathbb{P}}[b_{t_0+1} + \cdots + b_T \mid \mathcal{F}_{t_0}]$$

Nested decomposition of time consistent mappings

Theorem (Weak nested decomposition)

The aggregator g and factor f are weakly time consistent if and only if the set-valued function $\phi^{f,g}$ is a mapping

Remark

We then have a nested formula $g(a,b) = \phi^{f,g}(a, f(b))$

Example

$$\mathbb{E}_{\mathbb{P}}\big[a_1 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T\big] = \mathbb{E}_{\mathbb{P}}\Big[a_1 + \cdots + a_{t_0} + \underbrace{\mathbb{E}_{\mathbb{P}}\big[b_{t_0+1} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]}_{f(b)}\Big]$$

$$\phi^{f,g}(a,y) = \mathbb{E}_{\mathbb{P}}[a_1 + \cdots + a_{t_0} + y] \,, \quad y \in \mathcal{X}_{t_0}$$

Functional characterization of three notions of time consistency

Weak	Usual	Strong
$b \sim_{\mathbb{B}} b'$ $\Downarrow$ $(a,b) \sim_{\mathbb{A}\times\mathbb{B}} (a,b')$	$b \preccurlyeq_{\mathbb{B}} b'$ $\Downarrow$ $(a,b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a,b')$	$a \preccurlyeq_{\mathbb{A}} a'$, $b \preccurlyeq_{\mathbb{B}} b'$ $\Downarrow$ $(a,b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a,b')$
$\phi^{f,g}$ is a mapping	$\phi^{f,g}$ is a mapping increasing in its second argument	$\phi^{f,g}$ is a mapping increasing in both arguments

Conclusion of the abstract section

- We have developed an abstract framework
  to deal with time consistency
- We have illustrated the notions on a (simple) running example
- We expect to cover other cases with our abstract framework
- We first show how to apply this framework
  to examples of the literature

Outline of the section

2. Revisiting classical examples of the literature
   - Artzner, Delbaen, Eber, Heath, and Ku (2007)
   - Ruszczyński (2010)

2 Revisiting classical examples of the literature

- Artzner, Delbaen, Eber, Heath, and Ku (2007)
- Ruszczyński (2010)

Nested decomposition of coherent risk measures Artzner, Delbaen, Eber, Heath, and Ku (2007)

Let $\mathbb{A} = \mathcal{X}_{1,t_0}$ and $\mathbb{B} = \mathcal{X}_{t_0+1,T}$

$$\begin{aligned} g : \mathbb{A} \times \mathbb{B} &\to \mathbb{R} \\ (a, b) &\to \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[a_1 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T\big] \end{aligned}$$

$$\begin{aligned} f : \mathbb{B} &\to \mathcal{X}_{t_0} \\ b &\to \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[b_{t_0+1} + \cdots + b_T \;\big|\; \mathcal{F}_{t_0}\big] \end{aligned}$$

where $\mathcal{Q}$ is a (closed convex) set of probability distributions on $\Omega = \mathbb{R}^{t_0} \times \mathbb{R}^{T-t_0}$

A probability distribution $\mathbb{Q}$ on the product space $\Omega = \mathbb{R}^{t_0} \times \mathbb{R}^{T-t_0}$ can be naturally decomposed into

- a marginal distribution $m_{\mathbb{Q}}$
- a stochastic kernel $k_{\mathbb{Q}}$ conditional to the $\sigma-$field $\mathcal{F}_{t_0}$

Rectangularity

Definition (Epstein and Schneider (2003))

We say that a set $\mathcal{Q}$ of probability distributions is rectangular
if the image of $\mathcal{Q}$ by the mapping $\mathbb{Q} \mapsto (m_{\mathbb{Q}}, k_{\mathbb{Q}})$ is a rectangle
By an abuse of notation, we will write

$$\mathcal{Q} = \mathcal{M} \times \mathcal{K}$$

where $\mathcal{M}$ is a set of marginal distributions
and $\mathcal{K}$ is a set of stochastic kernels

Theorem

If $\mathcal{Q}$ is a rectangular set of probability distributions,
then f and g are
time consistent and we have

$$g(a,b) = \phi^{f,g}(a, f(b)) \ , \ \ \phi^{f,g}(a,y) = \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[a_1 + \cdots + a_{t_0} + y\big]$$

Sketch of proof

- First, we use a tower formula

$$\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\Big[a_1 + \cdots + a_{t_0} + \mathbb{E}_{\mathbb{Q}}\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

- Second, we use the property that $\mathcal{Q}$ is rectangular and that $a_1 + \cdots + a_{t_0}$ is $\mathcal{F}_{t_0}$ measurable

$$\sup_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathbb{E}_m\Big[a_1 + \cdots + a_{t_0} + \mathbb{E}_k\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

- Third, we take the supremum over the complete sup semilattice of $\mathcal{F}_{t_0}-$measurable random variables

$$\sup_{m\in\mathcal{M}} \mathbb{E}_m\Big[a_1 + \cdots + a_{t_0} + \sup_{k\in\mathcal{K}} \mathbb{E}_k\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

2 Revisiting classical examples of the literature

- Artzner, Delbaen, Eber, Heath, and Ku (2007)
- Ruszczyński (2010)

Ruszczyński framework

We call dynamic risk measure a sequence $\{\rho_{t,T}\}_{t=1}^{T}$
of conditional risk measures $\rho_{t,T} : \mathcal{X}_{t,T} \to \mathcal{X}_t$
(with the monotonicity property)

Definition (Ruszczyński)

A dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ is called time consistent if, for all $1 \leq t < t_0 \leq T$, and all sequences $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{X}_{t,T}$,

$$
\begin{gathered}
\boldsymbol{X}_k = \boldsymbol{Y}_k \ , \ \ k \in [\![t; t_0 - 1]\!] \\
\rho_{t_0,T}(\boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T) \leq \rho_{t_0,T}(\boldsymbol{Y}_{t_0}, \cdots, \boldsymbol{Y}_T) \\
\Downarrow \\
\rho_{t,T}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_T) \leq \rho_{t,T}(\boldsymbol{Y}_t, \cdots, \boldsymbol{Y}_T)
\end{gathered}
$$

Ruszczyński framework

Theorem (Ruszczyński (2010))

Suppose a dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ *satisfies, for all* $(\boldsymbol{X}_t, ..., \boldsymbol{X}_T)$*, the conditions*

$$\rho_{t,T}(\boldsymbol{X}_t, \boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_T) = \boldsymbol{X}_t + \rho_{t,T}(0, \boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_T)$$
$$\rho_{t,T}(0, ..., 0) = 0$$

Then $\{\rho_{t,T}\}_{t=1}^{T}$ *is time consistent if and only if, for all* $1 \leq t < t_0 \leq T$ *and all* $(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_T)$*, the following identity is true*

$$\rho_{t,T}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T) = \rho_{t,t_0}\big(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \rho_{t_0,T}(\boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T)\big)$$

where $\rho_{t,t_0}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0}) = \rho_{t,T}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0}, 0, \cdots, 0)$

Links between our framework and Ruszczyński's one (1)

- aggregator g is $\rho_{t,T} : \mathcal{X}_{t,t_0-1} \times \mathcal{X}_{t_0,T} \to \mathcal{X}_t$
- factor f is $\rho_{t_0,T} : \mathcal{X}_{t_0,T} \to \mathcal{X}_{t_0}$

The mappings g and f are weakly time consistent if and only if $\phi^{\rho_{t_0,T},\rho_{t,T}}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \cdot)$ is a mapping

Then we have

$$\rho_{t,T}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T) = \underbrace{\phi^{\rho_{t_0,T},\rho_{t,T}}}_{\text{subaggregator}} \Big(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \rho_{t_0,T}(\boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T)\Big)$$

Links between our framework and Ruszczyński's one (2)

If in addition

$$\rho_{t,T}(\boldsymbol{X}_t, ..., \boldsymbol{X}_T) = \boldsymbol{X}_t + \rho_{t,T}(0, \boldsymbol{X}_{t+1}, ..., \boldsymbol{X}_T)$$
$$\rho_{t,T}(0, ..., 0) = 0$$

Then the subaggregator is given by

$$\phi^{\rho_{t_0,T}, \rho_{t,T}} = \rho_{t,t_0}$$

that is

$$\rho_{t,T}(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T) = \rho_{t,t_0}\big(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_{t_0-1}, \rho_{t_0,T}(\boldsymbol{X}_{t_0}, \cdots, \boldsymbol{X}_T)\big)$$

Links between our framework and Ruszczyński's one (3)

Ruszczyński	HG, MDL, JPC, PC
Usual time consistency $b \preccurlyeq_{\mathbb{B}} b' \Rightarrow (a,b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a,b')$	Weak time consistency $b \sim_{\mathbb{B}} b' \Rightarrow (a,b) \sim_{\mathbb{A}\times\mathbb{B}} (a,b')$
Monotonicity property	$\varnothing$
Additive criterion	Any criterion
Explicit subaggregator ρ_{t,t_0}	Existence $\phi^{\rho_{t_0,T},\rho_{t,T}}$

Conclusion of the section and perspectives

So far, we have

- revisited different examples with one framework

We want to

- extend the results established for time additive cases to other time aggregators (multiplicative, maximum. . .)
- study more general Fenchel transforms than $\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\boldsymbol{X}]$
- find proper factors

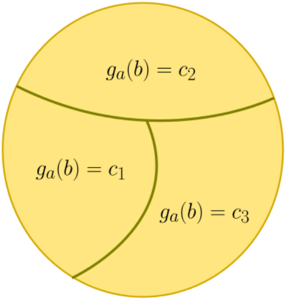
Outline of the section

Exhibiting weak time consistent factors

- Till now, aggregator g and factor f were given
- From now on, we only consider an aggregator $g : \mathbb{A} \times \mathbb{B} \to \mathbb{X}, \preccurlyeq_{\mathbb{X}}$
- We say that $f : \mathbb{B} \to \mathbb{Y}, \preccurlyeq_{\mathbb{Y}}$ is a weak time consistent factor (WTC factor) if the quadruplet $(g, \preccurlyeq_{\mathbb{X}}, f, \preccurlyeq_{\mathbb{Y}})$ is weak time consistent
- We will characterize WTC factors in term of partitions

Parametric aggregator and recall on partitions

We denote by $g_a : \mathbb{B} \to \mathbb{X}$ the parametric aggregator which depends on $a \in \mathbb{A}$
We denote by $\mathbb{X}/ \preccurlyeq_{\mathbb{X}}$ the partition of $\mathbb{X}$ induced by the equivalence relation $\sim_{\mathbb{X}}$



We denote by $\pi(g_a, \preccurlyeq_{\mathbb{X}})$ the partition of $\mathbb{B}$ induced by g_a

$$\pi(g_a, \preccurlyeq_{\mathbb{X}}) = g_a^{-1}(\mathbb{X}/ \preccurlyeq_{\mathbb{X}})$$

The time consistency property for mappings only has to do with partitions

Definition

We denote by $\underline{\pi} \preccurlyeq \overline{\pi}$ the property that a partition $\overline{\pi}$ is finer than $\underline{\pi}$
The atoms of $\overline{\pi}$ are thus subsets of atoms of $\underline{\pi}$

Proposition

A pair $(f, \preccurlyeq_{\mathbb{Y}})$ is a WTC factor for the pair $(g, \preccurlyeq_{\mathbb{X}})$ if and only if

$$\bigvee_{a \in \mathbb{A}} \pi(g_a, \preccurlyeq_{\mathbb{X}}) \preccurlyeq \pi(f, \preccurlyeq_{\mathbb{Y}})$$

We call minimal factor any mapping f such that

$$\pi(f, \preccurlyeq_{\mathbb{Y}}) = \bigvee_{a \in \mathbb{A}} \pi(g_a, \preccurlyeq_{\mathbb{X}})$$

Illustration of minimal factors in the risk neutral case

$$g(a,b) = \mathbb{E}_{\mathbb{P}}[a+b]$$

- The level lines of the parametric aggregator g_a are the random variables with the same mean

$$g(a,b) = x \Leftrightarrow \mathbb{E}_{\mathbb{P}}[b] = x$$

- Two possible minimal factors are

$$\begin{aligned} f_1(b) &= \mathbb{E}_{\mathbb{P}}[b] \ \ \Rightarrow \phi^{f,g}(a,y) = \mathbb{E}_{\mathbb{P}}[a+y] \\ f_2(b) &= \mathbb{E}_{\mathbb{P}}[b]^3 \Rightarrow \phi^{f,g}(a,y) = \mathbb{E}_{\mathbb{P}}[a+y^{\frac{1}{3}}] \end{aligned}$$

- The conditional expectation $\mathbb{E}_{\mathbb{P}}[b \mid \mathcal{F}]$ is a factor but not a minimal one

Conclusion of the section

So far, we have

- characterized all possible WTC factors f for a given aggregator g
- identified the minimal ones

Open question

- measurability, continuity, etc of factors
- extension to strong and normal time consistency

We end now with considerations about optimization and time consistency

Outline of the section

Towards dynamic programming

- We want to mix optimization with our framework to obtain dynamic programming equations of the form

$$\inf_{a\in\mathbb{A},b\in\mathbb{B}} g(a,b) = \inf_{a\in\mathbb{A}} \phi^{f,g}\big(a, \inf_{b\in\mathbb{B}} f(b)\big)$$

- For this purpose, we establish results useful for optimization

Inheritance of properties

We assume that factor f and aggregator g are weakly time consistent

Theorem (Monotonicity)

If the aggregator g is monotonous in its second argument, then the subaggregator $\phi^{f,g}$ is monotonous in its second argument

Theorem (Continuity)

If the aggregator g is continuous with a compact image, if the factor f is continuous with compact domain and image, then the subaggregator $\phi^{f,g}$ is continuous

Theorem (Convexity)

If there exists $\bar{\mathbb{B}} \subset \mathbb{B}$ such that $f(\bar{\mathbb{B}}) = \mathbb{Y}$ and such that $f_{|\bar{\mathbb{B}}}$ is linear, and if the aggregator g is convex, then the subaggregator $\phi^{f,g}$ is convex

Conclusion and ongoing work

Conclusion

- We have developed a general abstract framework for time consistency and have applied it to classic examples of the literature
- We have identified proper factors f that yield to time consistency, for a given aggregator g
- We have established inheritance properties that are useful for optimization

Ongoing work

- Develop the theory with lattices instead of preorders for optimization purpose
- Switching from time consistency to non nested consistency (using multi-agent framework "à la Witsenhausen")

References

P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and bellman's principle. *Annals of Operations Research*, 152(1):5–22, 2007.

L. G. Epstein and M. Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31, 2003.

G. Pflug and A. Pichler. On dynamic decomposition of multistage stochastic programs. *Optimization Online preprint*, 2(11), 2012.

A. Ruszczyński. Risk-averse dynamic programming for markov decision processes. *Mathematical programming*, 125(2):235–261, 2010.

A. Ruszczynski and A. Shapiro. Conditional risk mappings. *Mathematics of operations research*, 31(3):544–561, 2006.