Abstract Time Consistency
and
Decompostion

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

ENPC - CERMICS
Labex Bézout - LIGM

June 03, 2016

Time consistency in a nutshell

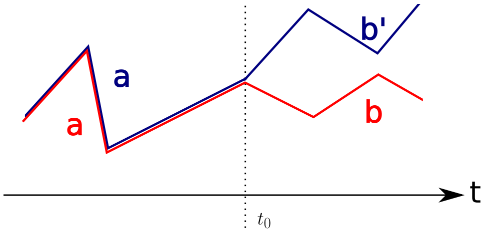
Given two processes $(\boldsymbol{X}_0, \cdots, \boldsymbol{X}_T)$ and $(\boldsymbol{Y}_0, \cdots, \boldsymbol{Y}_T)$,

- we look for numerical evaluations (risk measures) of the tails $(\boldsymbol{X}_t, \cdots, \boldsymbol{X}_T)$ 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

Example of time consistency



Example

- $(\mathcal{F}_t)_{t \in [\![0;T]\!]}$ filtration of $(\Omega, \mathcal{F}, \mathbb{P})$
- $\mathcal{X}_t = \mathcal{L}^p(\Omega, \mathcal{F}_t, \mathbb{P})$
- $\mathbb{A} = \mathcal{X}_0 \times \ldots \times \mathcal{X}_{t_0}$
- $\mathbb{B} = \mathcal{X}_{t_0+1} \times \ldots \times \mathcal{X}_T$

We focus on the risk averse case

- There is a strong 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)

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

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

Outline

1. Abstract notion of time consistency and functional representation
2. Revisiting classical examples of the literature
   - Artzner, Delbaen, Eber, Heath, and Ku (2007)
   - Ruszczyński (2010)
3. 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)

$$\left.\begin{array}{l} b \preccurlyeq_{\mathbb{B}} b' \\ a \preccurlyeq_{\mathbb{A}} a' \end{array}\right. \Rightarrow (a, b) \preccurlyeq_{\mathbb{A}\times\mathbb{B}} (a', b') \ , \ \ \forall (a, a') \in \mathbb{A}^2 \ , \ \ \forall (b, b') \in \mathbb{B}^2$$

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

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

Functional representation

Definition (Representation of a preorder)

Let $\preccurlyeq_{\mathbb{B}}$ be a preorder on the set $\mathbb{B}$
Let $f : \mathbb{B} \to \mathbb{Y}$ be a mapping, where $\mathbb{Y}$ is equipped with a preorder $\preccurlyeq_{\mathbb{Y}}$
We say that $(f, \preccurlyeq_{\mathbb{Y}})$ is a representation of $(\mathbb{B}, \preccurlyeq_{\mathbb{B}})$ if

$$b \preccurlyeq_{\mathbb{B}} b' \Leftrightarrow f(b) \preccurlyeq_{\mathbb{Y}} f(b')$$

Remark

$(f, \leq_{\mathbb{R}})$ is called a numerical representation

Remark

Let $f : \mathbb{B} \to \mathbb{Y}$ be a mapping, where $\mathbb{Y}$ is equipped with a preorder $\preccurlyeq_{\mathbb{Y}}$
Then $f(b) \preccurlyeq_{\mathbb{Y}} f(b')$ induces a preorder $\preccurlyeq_{\mathbb{B}}$ on $\mathbb{B}$

Functional definition of time consistency

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

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

Remark

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

Example

$$f\big((b_{t_0+1}, \dots, b_T)\big) = \mathbb{E}_{\mathbb{P}}[b_{t_0+1} + \cdots + b_T \mid \mathcal{F}_{t_0}]$$
$$g\big((a_0, \dots, a_{t_0}, b_{t_0+1}, \dots, b_T)\big) = \mathbb{E}_{\mathbb{P}}[a_0 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T]$$

We introduce a set-valued mapping

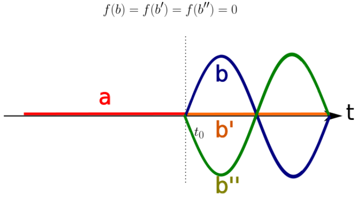
Given an aggregator $g : \mathbb{A} \times \mathbb{B} \to \mathbb{X}$ and a factor $f : \mathbb{B} \to \mathbb{Y}$, we introduce a set-valued mapping, called subaggregator

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

$$\mathbb{E}_{\mathbb{P}}\big[a_0 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T\big] = \mathbb{E}_{\mathbb{P}}\Big[a_0 + \cdots + a_{t_0} + \mathbb{E}_{\mathbb{P}}\big[b_{t_0+1} + \cdots + b_T \bigm| \mathcal{F}_{t_0}\big]\Big]$$

$$\phi^{f,g}(a,y) = \mathbb{E}_{\mathbb{P}}[a_0 + \cdots + a_{t_0} + y] \,, \quad y \in \mathcal{X}_{t_0}$$

$$g(a,b) = \phi^{f,g}(a, f(b))$$

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 hope now to cover other cases with our abstract framework
- We now 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 measure
Artzner, Delbaen, Eber, Heath, and Ku (2007)

Let $\mathbb{A} = \mathcal{X}_0 \times \cdots \times \mathcal{X}_{t_0}$ and $\mathbb{B} = \mathcal{X}_{t_0+1} \times \cdots \times \mathcal{X}_T$

$$
\begin{aligned}
g : \mathbb{A} \times \mathbb{B} &\to \mathbb{R} \\
(a, b) &\to \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[a_0 + \cdots + a_{t_0} + b_{t_0+1} + \cdots + b_T\big]
\end{aligned}
$$

$$
\begin{aligned}
f : \mathbb{B} &\to L^p(\Omega, \mathcal{F}_{t_0}, \mathbb{P}) \\
b &\to \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[b_{t_0+1} + \cdots + b_T \ \big|\ \mathcal{F}_{t_0}\big]
\end{aligned}
$$

where $\mathcal{Q}$ is a (closed convex) set of probability distributions on $\Omega = \mathbb{R}^{t_0+1} \times \mathbb{R}^{T-t_0}$

A probability distribution $\mathbb{Q}$ on the product space $\Omega = \mathbb{R}^{t_0+1} \times \mathbb{R}^{T-t_0}$ 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

$$g(a, b) = \phi^{f,g}(a, f(b)) \ , \ \ \phi^{f,g}(a, y) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\big[a_0 + \cdots + a_{t_0} + y\big]$$

Sketch of proof

- First, we use a tower formula

$$\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}\Big[a_0 + \cdots + a_{t_0} + \mathbb{E}_{\mathbb{Q}}\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

- Second, we use the property that $\mathcal{Q}$ is rectangular and that $a_0 + \cdots + a_{t_0}$ is $\mathcal{F}_{t_0}$ measurable

$$\sup_{(m,k)\in\mathcal{M}\times\mathcal{K}} \mathbb{E}_m\Big[a_0 + \cdots + a_{t_0} + \mathbb{E}_k\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

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

$$\sup_{m\in\mathcal{M}} \mathbb{E}_m\Big[a_0 + \cdots + a_{t_0} + \sup_{k\in\mathcal{K}} \mathbb{E}_k\big[b_{t_0} + \cdots + b_T \mid \mathcal{F}_{t_0}\big]\Big]$$

2 Revisiting classical examples of the literature

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

Ruszczyński framework

We consider

- the spaces $\mathcal{X}_t = L^p(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{X}_{t,T} = \mathcal{X}_t \times \mathcal{X}_{t+1} \times ... \times \mathcal{X}_T$
- a sequence of conditional risk measures $\rho_{t,T} = \mathcal{X}_{t,T} \to \mathcal{X}_t$ (with the monotonicity property) called dynamic risk measure

Definition

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

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

Ruszczyński framework

Theorem (Ruszczyński (2010))

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

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

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

$$\rho_{\tau,T}(\boldsymbol{X}_\tau, \cdots, \boldsymbol{X}_{\theta-1}, \boldsymbol{X}_\theta, \cdots, \boldsymbol{X}_T) = \rho_{\tau,\theta}(\boldsymbol{X}_\tau, \cdots, \boldsymbol{X}_{\theta-1}, \rho_{\theta,T}(\boldsymbol{X}_\theta, \cdots, \boldsymbol{X}_T))$$

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

Result

The dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ is weakly time consistent if and only if $\phi^{\rho_{\theta,T},\rho_{\tau,T}}(\boldsymbol{X}_\tau,\cdots,\boldsymbol{X}_{\theta-1},\cdot)$ is a mapping, for all $1 \leq \tau \leq \theta \leq T$ and for all $(\boldsymbol{X}_\tau,\cdots,\boldsymbol{X}_T)$

Then we have

$$\rho_{\tau,T}(\boldsymbol{X}_\tau,\cdots,\boldsymbol{X}_{\theta-1},\boldsymbol{X}_\theta,\cdots,\boldsymbol{X}_T) = \underbrace{\phi^{\rho_{\theta,T},\rho_{\tau,T}}}_{\text{subaggregator}}\big(\boldsymbol{X}_\tau,\cdots,\boldsymbol{X}_{\theta-1},\rho_{\theta,T}(\boldsymbol{X}_\theta,\cdots,\boldsymbol{X}_T)\big)$$

We put $\mathbb{A} = \mathcal{X}_\tau \times \cdots \times \mathcal{X}_\theta$ and $\mathbb{B} = \mathcal{X}_{\theta+1} \times \cdots \times \mathcal{X}_T$, then use our abstract result

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

Suppose that, in addition, the dynamic risk measure $\{\rho_{t,T}\}_{t=1}^{T}$ satisfies, for all $(\boldsymbol{X}_\tau, \cdots, \boldsymbol{X}_T)$, $t \in [\![1, T]\!]$, the conditions

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

Then

$$\phi^{\rho_{\theta,T},\rho_{\tau,T}} = \rho_{\tau,\theta}$$

that is

$$\rho_{\tau,T}(\boldsymbol{X}_\tau, \cdots, \boldsymbol{X}_{\theta-1}, \boldsymbol{X}_\theta, \cdots, \boldsymbol{X}_T) = \rho_{\tau,\theta}(\boldsymbol{X}_\tau, \cdots, \boldsymbol{X}_{\theta-1}, \rho_{\theta,T}(\boldsymbol{X}_\theta, \cdots, \boldsymbol{X}_T))$$

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

Ruszczyński	Us
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 $\rho_{\tau,\theta}$	Existence $\phi^{\rho_{\theta,T},\rho_{\tau,T}}$

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 $\sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[\boldsymbol{X}]$

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 established inheritance properties that are useful for optimization

Ongoing work

- How can we identify factors f that yield to time consistency, for a given aggregator g?
- 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.