Dynamic Risk Measures

Lecture Notes
of a
Mini-Course at Paris IX, Dauphine,
October and December 2007

by Frank Riedel

Institute of Mathematical Economics, Bielefeld University

† I would like to thank Rose-Anne Dana and Elyès Jouini for their hospitality and support when I stayed at Dauphine. The Financial support of the Risk Foundation (Groupama chair) is gratefully acknowledged.
Introduction

These short lecture notes present an introduction to time–consistent dynamic risk measures. We quickly review the main results on static risk measures in the first chapter. The second chapter discusses time–consistent coherent risk measures and the equivalent concept of sets of probability measures that are stable under pasting of marginal and conditional distributions. Finally, we shortly present one way of extending the results to dynamic convex risk measures.

CHAPTER 1

Lecture 1: Static (Convex) Risk Measures and (Variational) Preferences

1. Value at Risk

Risk measures aim to measure the risk of a financial position in monetary terms. We want to give meaning to the statement

Your risk is 1000\(e\).

But what does such a statement mean? One answer, and the current industry standard, is Value at Risk. Your value at risk for a given confidence level \(\alpha\) is 1000\(e\) if the probability of losing more than 1000\(e\) is \(\alpha\). In other words, Value at Risk is nothing but a quantile:

\[
\text{V@R}_\alpha(X) = \inf \{ k \in \mathbb{R} : P[X + k \leq 0] \leq \alpha \}.
\]

The Value at risk at level \(\alpha\) is thus the smallest amount of money that one adds to a position such that the probability of a loss is less or equal than \(\alpha\). Let us say that V@R accepts a position \(X\) if you do not have to add any money or you can even withdraw money without having a risk higher than \(\alpha\) to lose money. We call

\(A_\alpha := \{X \text{ random variable} | \text{V@R}_\alpha(X) \leq 0\}\)

the acceptance set of Value at Risk for the level \(\alpha\). Then we can write

\[
\text{V@R}_\alpha(X) = \inf \{ k \in \mathbb{R} : X + k \in A_\alpha \}.
\]

**Proposition 1.** V@R\(_\alpha\) is monotone, cash– and law–invariant. Formally:

1. if \(X \geq Y\), then \(\text{V@R}_\alpha(X) \leq \text{V@R}_\alpha(Y)\);
2. for \(k \in \mathbb{R}\) we have \(\text{V@R}_\alpha(X + k) = \text{V@R}_\alpha(X) - k\);
3. if \(P^X = P^Y\), then \(\text{V@R}_\alpha(X) = \text{V@R}_\alpha(Y)\).

**Exercise 1.** Prove the proposition.

Monotonicity makes a lot of sense in financial contexts and needs no discussion. Cash–invariance is the defining property of monetary risk measures — in contrast to standard utility theory, we aim to measure risk in terms of money. In fact, the risk of a position is going to be the amount of money that one adds to the position in order to make us indifferent between the zero position and the risky position plus that sum. (In Economics, similar concepts are certainty equivalent and equivalent variation). Law–invariance, however, is a strong property.
It says that the risk of a position depends only on its distribution. In some contexts, the agent might not have sufficient knowledge of the distribution of the random variables under scrutiny; for this reason, we drop this property in the following.

We take the first two properties of Value at Risk to define a new concept.

**Definition 2.** Let \((\Omega, \mathcal{F}, P_0)\) be a probability space. Write \(L^\infty = L^\infty(\Omega, \mathcal{F}, P_0)\) for the space of all essentially bounded measurable random variables.

A monetary risk measure \(\rho\) is a monotone and cash–invariant mapping \(\rho : L^\infty \to \mathbb{R}\) in the sense of the following two conditions:

1. if \(X \geq Y\) a.s., then \(\rho(X) \leq \rho(Y)\);
2. for \(k \in \mathbb{R}\) we have \(\rho(X + k) = \rho(X) - k\).

**Remark 3.** We deal here with the space \(L^\infty = L^\infty(\Omega, \mathcal{F}, P_0)\) of all essentially bounded random variables on a given probability space \((\Omega, \mathcal{F}, P_0)\). The measure \(P_0\) serves only the role of fixing the sets of measure zero. Economically, this means that the decision maker has perfect knowledge about sure events. In a finite model, one can take \(P_0\) to be the uniform distribution without loss of generality. For coherent risk measures that are continuous from below, one can always construct \(P_0\) from the set of priors \(Q\), if the measurable space \((\Omega, \mathcal{F})\) has a nice topological structure, see Tutsch (2006).

**Remark 4.** All probability measures in these notes are going to be absolutely continuous with respect to \(P_0\).

For economists (and the writer!), it is usually easier to think in terms of utility than risk. Let us define the corresponding utility function.

**Definition 5.** A monetary utility function \(U\) is a mapping \(U : L^\infty \to \mathbb{R}\) such that

1. it is monotone: if \(X \geq Y\) a.s., then \(U(X) \geq U(Y)\);
2. it is cash–invariant: for \(k \in \mathbb{R}\) we have \(U(X + k) = U(X) + k\).

If \(\rho\) is a monetary risk measure, than \(U(X) = -\rho(X)\) is a monetary utility function and vice versa.

Value at Risk is a statistical measure and one is not used to see such quantile–based measures in decision theory — and for a good reason. Value at Risk violates the principle of diversification.

**Example 6.** Let us consider a model with two independent and identically distributed assets \(B_1\) and \(B_2\), and suppose that

\[
\]
Let us assume that $0 < p = 0.01 < 0.1 = \alpha$. A risk manager who invests all his money into asset 1 or 2 has no risk according to V@R:

\[ V@R_\alpha(B_1) = V@R_\alpha(B_2) = -1. \]

From the point of view of Value at Risk, the two assets are riskless! Now let us consider another investor who diversifies and has the position $1/2(B_1 + B_2)$. His Value at Risk at level 10% is now +0.5 because the probability of losing 50 Cent is now

\[ P[B_1 = 1, B_2 = -2] + P[B_1 = -2, B_2 = 1] = 2p(1 - p) = 16.38\%. \]

The diversified portfolio has a higher risk than putting all one’s eggs into one basket.

2. Convex and Coherent Risk Measures

The failure of Value at Risk in the above example has led to axiomatic research about monetary risk measures.

**Definition 7.** A convex risk measure is a monetary risk measure that is convex. A positively homogeneous convex risk measure is called a coherent risk measure.

Constructing risk measures via acceptance sets. Suppose that $\mathcal{A}$ is a subset of $L^\infty$ that contains 0. $\mathcal{A}$ is going to be our acceptance set, economically the set of all positions that one prefers weakly to the zero position. By setting

\[ \rho(X) = \inf\{k \in \mathbb{R} : X + k \in \mathcal{A}\}, \]

we get a cash–invariant mapping. It is a monetary risk measure if $X \leq Y$ a.s. and $X \in \mathcal{A}$ implies $Y \in \mathcal{A}$. It is convex if $\mathcal{A}$ is convex, and coherent if $\mathcal{A}$ is a convex cone.

**Exercise 2.** Prove these claims about acceptance sets and their risk measures. (The converse is also true. Try to prove this, or consult Föllmer and Schied (2004), Chapter 4.1).

Constructing risk measures from expected utility: Shortfall Risk. A typical utility function in economics is the expected utility function of the form $EU(X) = E^Pu(X)$ for some Bernoulli utility function $u$. Typically, one assumes that $u$ is increasing and concave. If $u$ is strictly concave, then $EU(X)$ is not cash invariant, hence not a monetary utility function. But one can construct risk measures from expected utility by setting

\[ \rho(X) = \inf\{k \in \mathbb{R} : E^P[u(X + k)] \geq u_0\} \]

for some exogenous utility level $u_0$ and some probability measure $P$(usually $u_0 = u(0)$). Such a risk measure is called shortfall risk.

**Exercise 3.** Check that $\rho$ is a convex risk measure if $u$ is concave.
Example 8.  
(1) under risk–neutrality, we have $\rho(X) = E^P(-X)$;  
(2) under constant absolute risk aversion $A$, i.e. $u(x) = -\exp(-Ax)$, we have  
$$\rho(X) = 1/A \log E^P e^{-AX}.$$  
This measure is called entropic risk measure. The relation with relative entropy is explained below.  
(3) A coherent risk measure constructed from expected utility has $u(x) = ax$ for $x > 0$ and $u(x) = bx$ for $x \leq 0$ for some $0 < a \leq b$.

3. Representation Theorems

We are now going to apply the methods of convex analysis to obtain a dual representation for convex risk measures. Note that the dual space of $L^\infty$ is the space $\text{ba}(\Omega, \mathcal{F}, P_0)$ of boundedly additive measures that are absolutely continuous with respect to $P_0$. We denote by $\text{ba}^1$ the subset of all positive measures $\mu \ll P_0$ with $\mu(\Omega) = 1$. The following theorem applies the well–known Legendre–Fenchel duality from convex analysis for convex risk measures. Note that we can restrict to $\text{ba}^1$ in the variational characterization due to cash invariance and monotonicity.

Theorem 9 (Föllmer and Schied (2002a), Fritelli and Gianin (2002), Heath). Every convex risk measure can be written as  
$$\rho(X) = \max_{\mu \in \text{ba}^1} E^\mu(-X) - \alpha(\mu)$$  
for the penalty function  
$$\alpha(\mu) = \sup_{X \in L^\infty} E^\mu(-X) - \rho(X)$$  
$$= \sup_{X|\rho(X) = 0} E^\mu(-X).$$  

Every convex monetary utility function can be written as  
$$U(X) = \min_{\mu \in \text{ba}^1} E^\mu(X) + c(\mu)$$  
for the cost function  
$$c(\mu) = \sup_{X \in L^\infty} U(X) - E^\mu(X)$$  
$$= \sup_{X|U(X) = 0} -E^\mu(X).$$

Proof. Let us start with equation (2). For arbitrary $X$, set $Y = X + \rho(X)$. Note that from cash invariance, $\rho(Y) = 0$, and $E^\mu(-X) - \rho(X) = E^\mu(-Y)$. Hence, we can restrict to random variables $Y$ with $\rho(Y) = 0$. 


As a consequence, the penalty function \( \alpha \) is positively homogeneous, i.e. \( \alpha(\lambda \mu) = \lambda \alpha(\mu) \) for all \( \lambda > 0 \). Now apply the usual convex duality theory. \( \alpha \) is the Legendre–Fenchel transform of \( \rho \). By monotonicity, we can restrict to positive measures \( \mu \), and by positive homogeneity of \( \alpha \), we can restrict to measures with mass 1.

\[ \Box \]

**Corollary 10** (Artzner, Delbaen, Eber, and Heath (1999)). Every coherent risk measure can be written as
\[
\rho(X) = \max_{P \in \mathcal{P}} E_P(-X)
\]
for some set \( \mathcal{P} \subset ba^1 \).

Every coherent monetary utility function can be written as
\[
U(X) = \min_{P \in \mathcal{P}} E_P(X)
\]
for some set \( \mathcal{P} \subset ba^1 \).

**Proof.** By positive homogeneity, the penalty function takes only the values 0 and \( \infty \). Define
\[
\mathcal{P} = \{ Q : \alpha(Q) = 0 \}.
\]
\[ \Box \]

**Remark 11.** Can one have real probabilities instead of the merely finitely additive objects in \( ba^1 \)? Yes, if one adds suitable continuity properties. If \( \rho \) is continuous from above, one can write
\[
\rho(X) = \sup_{P \in \mathcal{P}} E_P(-X)
\]
for some set \( \mathcal{P} \subset M^1 \). Here, \( M^1 \) denotes the set of all probability measures on \( (\Omega, \mathcal{F}) \) that have a density with respect to our reference measure \( P \). In general, the sup is not attained as one can see from taking the essential supremum \( \rho(X) = \text{esssup}(X) \) (note that this is a coherent risk measure). It can be written as \( \rho(X) = \sup_{\delta \omega} E_{\delta \omega} X \), but the sup is not attained if the random variable \( X \) does not attain its supremum. If one wants the sup to be a max, one has to assume that the set \( \mathcal{P} \) is weakly compact in the set of all probability measures \( M^1 \). This is equivalent to continuity from below. (See Delbaen (2002a) for details).

**Remark 12.** Continuing in the spirit of the previous remark, let us note that convex risk measures can be written as
\[
(3) \quad \rho(X) = \sup_{P \in M^1} E_P(-X) - \alpha(P)
\]
if (and only if) they are continuous from above.

**Exercise 4.** Compute the penalty function for the entropic risk measure. (It is relative entropy, hence the name).
Solution for a finite probability space \( \Omega = \{1, \ldots, S\} \). You have to maximize
\[
- \sum X_s \mu_s - 1/A \log \sum P_s e^{-AX_s}
\]
over $X_s, s = 1, \ldots, S$. The first order condition (think about the second-order condition!) is

\[ \mu_s = \frac{P_s e^{-AX_s}}{\sum_t P_t e^{-AX_t}}. \]

It follows that

\[ e^{-AX_s} = \mu_s \sum_t P_t e^{-AX_t}. \]

Plug this into the above expression and you get

\[ \alpha(\mu) = \frac{1}{A} \sum \mu_s \log \frac{\mu_s}{P_s}. \]

**Exercise 5.** The penalty function for shortfall risk as defined in (1). Note that shortfall risk accepts a position if and only if $E^P u(X) \geq u_0$. From Eqn (2), we have then to solve

\[ \minimize E^Q X \quad \text{subject to} \quad E^P u(X) = u_0. \]

This is the familiar expenditure minimization problem for an expected utility maximizer when the price function is given by the probability $Q$. Assuming that the first-order condition is always binding, we get

\[ \lambda u'(X) = \frac{dQ}{dP}, \]

for a suitable multiplier $\lambda > 0$, or

\[ X = \int \frac{dQ}{dP} / \lambda. \]

The multiplier $\lambda$ is determined by the constraint

\[ E^P u \left( \int \frac{dQ}{dP} / \lambda \right) = u_0. \]

The penalty function is thus given by

\[ \alpha(Q) = -E^Q \int \frac{dQ}{dP} / \lambda. \]

Plugging in the CARA utility function, we recover the result from the previous exercise on the entropic risk measure.

**Example 13.** (1) A coherent generalization of Value at Risk is Average Value at Risk given by

\[ \text{AV@R}_\lambda(X) = \frac{1}{\lambda} \int_0^\lambda \text{V@R}_\alpha(X) d\alpha. \]

Suppose that $X$ has a continuous, strictly increasing distribution function $F(x) = P_0[X \leq x]$. Then $\text{V@R}_\alpha(X) = -F^{-1}(\alpha)$.
3. REPRESENTATION THEOREMS

and $\lambda = P_0[-X \geq \text{VaR}_\lambda(X)]$. Substituting $y = F^{-1}(\alpha)$, we get

$$\text{AV@R}_\lambda(X) = \frac{1}{P_0[-X \geq \text{VaR}_\lambda(X)]} \int_{-\infty}^{F^{-1}(\lambda)} yF(dy)$$

$$= E^{P_0}[-X \mid -X \geq \text{VaR}_\lambda(X)].$$

Average Value at Risk is thus the conditional loss given that the losses exceed the Value at Risk. In particular, it dominates Value at Risk. In fact, it is the smallest law–invariant coherent risk measure that is greater or equal Value at Risk (Artzner, Delbaen, Eber, and Heath (1999) and Föllmer and Schied (2002)). For later reference, we note that Average Value at Risk has the representation

$$\text{AV@R}_\lambda(X) = \sup_{Q \ll P_0 : \frac{dQ}{dP_0} \leq \lambda^{-1}} E^Q[-X].$$

(2) Convex capacities give rise to coherent risk measures. Let $\nu$ be a convex capacity with core $\mathcal{P} = \{Q : Q(A) \geq \nu(A) \text{ for all events } A\}$. Then the Choquet expectation $E^\nu(X) = \min_{Q \in \mathcal{P}} E^Q(X)$ is a coherent monetary utility function, and

$$\rho(X) = -E^\nu(X)$$

is a coherent risk measure.

Relation with Decision Theory. In order to model uncertainty aversion, Gilboa and Schmeidler (1989) have introduced and axiomatized utility functionals of the form

$$U(X) = \min_{P \in \mathcal{P}} E^P u(x)$$

for some set of probabilities $\mathcal{P}$ and a Bernoulli utility function $u$. If we take $u$ to be linear here, we get a coherent monetary utility function. The Gilboa/Schmeidler approach has been generalized to variational preferences by Chateauneuf, Maccheroni, Marinacci, and Tallon (2005):

$$U(X) = \min_{P \in \text{ba}^1} E^P u(X) + c(P)$$

for some cost function $c$.

If we take a risk–neutral agent, we thus get coherent resp. concave monetary utility functions.

Robust Decisions in Macroeconomics. There is also a close relationship to robust control theory as it is recently developed in Macroeconomics. As institutions can never be completely sure that they use the right model, one would like to obtain robust decisions that remain almost optimal if the underlying model is varied slightly. In order to model this, Anderson, Hansen, and Sargent (2000) use, in our terms, the entropic risk measure and compute optimal policies.
CHAPTER 2

Lecture 2: Dynamic Coherent Risk Measures, Consistency and Stability under Pasting

1. Dynamic Convex Risk Measures

We now move on to a dynamic framework. Let the information flow be given by a filtration $(\mathcal{F}_t)_{t=0,...,T}$ for a finite horizon $T$ on our probability space $(\Omega, \mathcal{F}, P_0)$. To keep things simple, we focus on payments $X \in L^\infty(\mathcal{F}_T)$ at the final time $T$ only; the general case of cash flows is treated in Riedel (2004) and a different treatment of risk measures on price processes is in Artzner, Delbaen, Eber, Heath, and Ku (2002). As time evolves, we want to have a sequence of risk assessments $\rho_t(X)$ that correctly update the new information.

A risk assessment at time $t$ is thus a mapping $\rho_t: L^\infty(\mathcal{F}_T) \rightarrow L^\infty(\mathcal{F}_t)$; it maps time $T$–risks into $\mathcal{F}_t$–measurable cash requirements. It is also natural to replace the constants in the definition of cash invariance and convexity with random variables that the agent knows at time $t$. After all, at time $t$, an $\mathcal{F}_t$–measurable random variable is a constant for the agent!

**Definition 14.** A dynamic monetary risk measure is a sequence $(\rho_t)_{t=0,...,T}$ of mappings $\rho_t: L^\infty(\mathcal{F}_T) \rightarrow L^\infty(\mathcal{F}_t)$ that is

1. monotone: if $X \geq Y$ a.s., then $\rho_t(X) \leq \rho_t(Y)$;
2. and adapted cash–invariant: for $K \in L^\infty(\mathcal{F}_t)$ we have $\rho(X + K) = \rho(X) - K$.

$(\rho_t)$ is called a dynamic convex risk measure if it satisfies in addition

$$\rho_t(\Lambda X + (1 - \Lambda)Y) \leq \Lambda \rho_t(X) + (1 - \Lambda)\rho_t(Y)$$

for all $t$, all $X,Y \in L^\infty(\mathcal{F}_T)$ and all $\mathcal{F}_t$–measurable $\Lambda \in [0,1]$. If we have even

$$\rho_t(\Lambda X) = \Lambda \rho_t(X)$$

for all $t$, all $X \in L^\infty(\mathcal{F}_T)$ and all $\mathcal{F}_t$–measurable $\Lambda \geq 0$, then $(\rho_t)$ is called a dynamic coherent risk measure.

**Remark 15.** We are going to encounter frequently expressions like

$$\rho_t(X) = \text{esssup}_{Q \in \mathcal{P}} E^Q [-X | \mathcal{F}_t].$$
Let us discuss here whether these objects are well defined. The essential supremum of a family of random variables \((Z_Q)_{Q \in \mathcal{P}}\) is the smallest random variable \(Z\) such that \(P_0[Z \geq Z_Q] = 1\) for all \(Q \in \mathcal{P}\). The definition thus depends on the probability \(P_0\). To be clear, we should thus write

\[
\rho_t(X) = P_0 - \text{esssup}_{Q \in \mathcal{P}} \mathbb{E}^Q [-X | \mathcal{F}_t].
\]

The next problem then is that the conditional expectations \(\mathbb{E}^Q [\cdot | \mathcal{F}_t]\) are defined only up to \(Q\)-null sets. This might look problematic. However, as long as all measures \(Q\) are equivalent with respect to the reference measure \(P_0\), there is no problem. We are thus going to impose conditions that ensure equivalence later on, see Assumptions 16 and 17.

Let us make some assumptions on the set of priors \(\mathcal{P}\) for these lecture notes.

**Assumption 16.** All \(P \in \mathcal{Q}\) are equivalent to the reference measure \(P_0\) on \(\mathcal{F}_T\), i.e. for all \(A \in \mathcal{F}_T\) we have \(P(A) = 0\) if and only if \(P_0(A) = 0\).

By the Radon–Nikodym theorem, the set of priors \(\mathcal{P}\) can be identified with the set of densities \(\mathcal{D} = \left\{ \frac{dP}{dP_0} \mid P \in \mathcal{P} \right\}\).

**Assumption 17.** The family of densities

\[
\mathcal{D} = \left\{ \frac{dP}{dP_0} \mid P \in \mathcal{P} \right\}
\]

is weakly compact in \(L^1(\Omega, \mathcal{F}_T, P_0)\). (Equivalently, the family of densities \(\left( \frac{dP}{dP_0} \right)_{P \in \mathcal{P}}\) is uniformly integrable with respect to \(P_0\).)

These two assumptions are stronger than actually necessary. However, they simplify the presentation because we now have strictly positive densities, which implies that the conditional expectations are \(P_0\)-a.s. well defined, and the supremum in

\[
\rho_t(X) = \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P [-X | \mathcal{F}_t]
\]

is always attained for some worst case measure \(P^* \in \mathcal{P}\). For the general case, refer, e.g., to Föllmer and Penner (2006). Note that Assumption 17 is satisfied without loss of generality when the densities in \(\mathcal{D}_t\) are bounded by a \(P_0\)-integrable random variable. In particular, the assumption is satisfied whenever the state space \(\Omega\) is finite. On the other hand, there are economic reasons that suggest to impose weak compactness. It is equivalent to continuity from below for the risk measure, see Corollary 4.35 in Föllmer and Schied (2004) or Chateauneuf, Maccheroni, Marinacci, and Tallon (2005). A behavioral justification for such kind of continuity has already been given by Arrow (1971).
2. Inconsistency: An Example

In this lecture, we focus on dynamic coherent risk measures. It is natural to suppose that these can be represented via

\[ \rho_t(X) = \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P[-X | \mathcal{F}_t] \]

for a (convex, weak*-closed) class of priors \( \mathcal{P} \subset \mathcal{M}^1(\mathcal{P}_0) \) that are equivalent to the reference measure \( \mathcal{P}_0 \). In other words, we react to new information by computing the conditional expectation separately for every prior in \( \mathcal{P} \), and apply then the worst-case principle. Of course, they are several other possible choices. For example, one could eliminate some priors, or one could use updating rules that have been proposed in the context of nonlinear integration like the Dempster–Shafer rule. Riedel (2004) shows, however, that only the full Bayesian updating used here can be time-consistent. Therefore, we focus on this representation in this lecture.

Even the representation (5) can lead to inconsistencies as the next example shows.

**Example 18.** Consider Figure 1.

- Figure 1. Tree for Example 18. The decision maker uses two probabilistic models. In Model 1, the (conditional) probability of moving up is 1/3 in every node; in Model 2, this probability is 2/3.

We have a two-period binomial tree. The agent uses two priors. Under prior 1, the up–movements are independent and occur with probability 1/3. Similarly, under prior 2, the up–movements are again independent and occur with probability 2/3. The corresponding probabilities...
on the state space are marked to the right of the tree. The agents evaluates the riskiness of a position that pays off +4 after two ups or two downs. Else, the agent loses 5.

Ex ante, the expected risk using either prior is \( 2 \cdot \frac{2}{9} \cdot 5 - (1/9 + 4/9) \cdot 4 = 0 \). Hence, the agent accepts the risk at time 0. At time 1, the agent reevaluates the riskiness using the conditional probabilities. In the upper part of the tree, the worst case is clearly to go down. Hence, the agent places 2/3 on that branch. The risk is then \( 2/3 \cdot 5 - 1/3 \cdot 4 = 2 \). In the lower node, the agent places probability 2/3 on going up. The risk here is also 2. Whatever happens, the agent rejects the position at time 1. Summing up, the agent accepts the position at time 0 knowing that he is going to reject it at time 1 whatever happens. You do not want to have a risk manager like this in your bank.

Let us try to understand why the naive updating leads to contradictory decisions in the preceding example. If the agent knows that he is going to reject the position \( X \) at time 1 in any state of the world, he should reject it today. In other words, if \( \rho_1(X) = 2 \) in all states of the world, then we would expect also \( \rho_0(X) = 2 \). But we have \( \rho_0(X) = 0 \).

The reason for this is that the agent does not take the worst case measure into account at time 0. In fact, he uses only two probabilities under which the up–movements are independent and identically distributed. However, the conditional probabilities that are important are path–dependent. In the upper node, the worst case is to go down. Hence, the worst case conditional probability is 1/3 for the up–movement. In the lower node, the worst case is to go up. Here, the worst case probability for up is 2/3. The agent should use these conditional probabilities when evaluating the risk at time 0. The corresponding distribution is depicted in Figure 2. (The agent is indifferent about the marginal probabilities for going up in the first period in this example.)

The worst case probability is obtained by pasting together different conditional probabilities in the different branches of the tree with some marginal probabilities for the first period. The problem in the example is that the pasted probability measure is not in the set of priors that the agent uses to evaluate the ex ante riskiness. We thus guess that the set of priors should be closed under this pasting operation in order to get dynamically consistent decisions. And we shall see that this is indeed a necessary and sufficient condition for consistency.

### 3. Stability under Pasting

We now discuss the operation of pasting that we described in the above Example 18 in full generality. Several definitions have been used, and as they are all useful, we state them here, and prove their equivalence.
Following Epstein and Schneider (2003), we call $Q$ rectangular if for all stopping times $t \leq T$ and all $P, Q \in \mathcal{P}$ the measure $PQ_t$ given by

$$PQ_t(B) = \mathbb{E}^Q P(B|\mathcal{F}_t) \quad (B \in \mathcal{F})$$

belongs to $Q$ as well. We call $PQ_t$ the pasting of $Q$ after $P$ at time $t$. This operation is not symmetric, of course.

Föllmer and Schied (2002) call $Q$ stable under pasting if for all stopping times $\tau$, sets $A \in \mathcal{F}_\tau$, and priors $P, Q \in Q$, there exists a measure $R \in Q$ such that $R = P$ on $\mathcal{F}_\tau$ and for all random variables $Z \geq 0$ one has

$$\mathbb{E}^R [Z|\mathcal{F}_\tau] = \mathbb{E}^P [Z|\mathcal{F}_\tau] 1_{A^c} + \mathbb{E}^Q [Z|\mathcal{F}_\tau] 1_A. \tag{6}$$

In the spirit of Delbaen (2002b), call the set of priors $Q$ is time-consistent if the following holds true. For $P$ and $Q$ in $\mathcal{P}$, let $(p_t)_{t=0,\ldots,T}$ and $(q_t)_{t=0,\ldots,T}$ be the density processes of $P$ resp. $Q$ with respect to $P_0$, i.e.

$$p_t = \frac{dP}{dP_0}|_{\mathcal{F}_t}, \quad q_t = \frac{dQ}{dP_0}|_{\mathcal{F}_t} \quad (t \leq T).$$

Fix some stopping time $t \in \{0,\ldots,T\}$. Define a new probability measure $R$ by setting for $s \in \{0,\ldots,T\}$

$$\frac{dR}{dP_0}|_{\mathcal{F}_s} = \left\{ \begin{array}{ll} \frac{p_t}{q_s} & \text{if } s \leq t \\ q_t & \text{else} \end{array} \right.. \tag{7}$$

Then $R$ belongs to $Q$ as well.

**Lemma 19.** The following assertions are equivalent:
2. Dynamic Coherent Risk Measures

(1) \( Q \) is time-consistent,
(2) \( Q \) is stable under pasting,
(3) \( Q \) is rectangular.

**Proof.** Time-consistency implies stability: Suppose that \( Q \) is time-consistent. Fix a stopping time \( \tau \), sets \( A \in \mathcal{F}_\tau \), and priors \( P, Q \in Q \). Let \((p_t)\) and \((q_t)\) be the density processes of \( P \) and \( Q \) with respect to \( P_0 \). Define a new stopping time \( \sigma = \tau 1_A + T 1_{A^c} \). By time-consistency, the measure \( R \) given by

\[
\frac{dR}{dP_0} = \frac{p_\sigma}{q_\sigma} q_T \in Q.
\]

Note that

\[
\frac{dR}{dP_0} = \frac{p_\tau}{q_\tau} q_T 1_A + p_T 1_{A^c}.
\]

Taking conditional expectations, we get

\[
\frac{dR}{dP_0} \bigg|_{\mathcal{F}_\tau} = p_\tau.
\]

Hence, \( R = P \) on \( \mathcal{F}_\tau \). Application of Bayes' formula yields (6):

\[
\mathbb{E}^R[Z|\mathcal{F}_\tau]^{Bayes} = \mathbb{E}^P_\tau \left[ Z \frac{p_\tau}{q_\tau} q_T | \mathcal{F}_\tau \right] \left( \frac{dR}{dP_0} |_{\mathcal{F}_\tau} \right)^{-1} = 1_A \mathbb{E}^P_\tau \left[ Z \frac{p_\tau}{q_\tau} q_T | \mathcal{F}_\tau \right] p_\tau^{-1} + 1_A \mathbb{E}^P_\tau [Zp_T | \mathcal{F}_\tau] p_\tau^{-1} \]

\[
\mathbb{E}^P_\tau = 1_A \mathbb{E}^Q[Z|\mathcal{F}_\tau] + 1_A \mathbb{E}^P[Z|\mathcal{F}_\tau].
\]

Stability implies Rectangularity: Fix a stopping time \( \tau \) and \( P, Q \in Q \). Take \( A = \Omega \). By stability, there exists a measure \( R \in Q \) with \( R = P \) on \( \mathcal{F}_\tau \) and (6). Take \( Z = 1_B \) for \( B \in \mathcal{F} \). (6) yields \( R(B|\mathcal{F}_\tau) = Q(B|\mathcal{F}_\tau) \). As \( R = P \) on \( \mathcal{F}_\tau \), we obtain

\[
R(B) = \mathbb{E}^R R(B|\mathcal{F}_\tau) = \mathbb{E}^P R(B|\mathcal{F}_\tau) = \mathbb{E}^P Q(B|\mathcal{F}_\tau).
\]

Rectangularity implies Time–Consistency: Let \( P, Q \in Q \) and \( \tau \) be a stopping time. Define \( R \) by setting

\[
\frac{dR}{dP_0} = \frac{p_\tau}{q_\tau} q_T.
\]

For \( B \in \mathcal{F} \), we obtain by conditioning and using Bayes' formula

\[
R(B) = \mathbb{E}^P_\tau \left[ 1_B \frac{p_\tau}{q_\tau} q_T \right] = \mathbb{E}^P_\tau \left[ \frac{p_\tau}{q_\tau} \mathbb{E}^P_{\tau_\tau} [1_B q_T | \mathcal{F}_\tau] \right] = \mathbb{E}^P_\tau \left[ p_\tau Q(B|\mathcal{F}_\tau) \right] = \mathbb{E}^P Q(B|\mathcal{F}_\tau) .
\]

Rectangularity yields \( R \in Q \). \( \square \)
4. TIME–CONSISTENCY

We conclude here with an easy yet important lemma on taking expectations under the pasting of $Q$ after $P$ at $t$.

**Lemma 20.** Let $P, Q$ be two equivalent probability measures, and $X \in L^\infty(F_T)$. Then
\[ E^{PQ}_{t+1}[X|F_t] = E^P[E^Q[X|F_{t+1}]|F_t] \]
More generally, we have for $u \leq t$
\[ E^{PQ}_{t+1}[X|F_u] = E^P[E^Q[X|F_{t+1}]|F_u] \]
and for $u > t$
\[ E^{PQ}_{t+1}[X|F_u] = E^Q[X|F_u] \]

**Exercise 6.** Use the generalized Bayes’ rule (Karatzas and Shreve (1991), Lemma 3.5.3) to prove this lemma.

4. Time–Consistency

We call a dynamic risk measure $(\rho_t)_{t=0,\ldots,T}$ time–consistent if and only if for all $X, Y \in L^\infty(F_T)$ we have
\[ \rho_{t+1}(X) = \rho_{t+1}(Y) \implies \rho_t(X) = \rho_t(Y). \]

**Remark 21.** The usual definition of time–consistency in decision theory would use the formally stronger statement
\[ \rho_{t+1}(X) \geq \rho_{t+1}(Y) \implies \rho_t(X) \geq \rho_t(Y). \]
With monotonicity and cash invariance the two definitions are equivalent. To see this, suppose that the weak form of time–consistency (8) holds true. Choose $X, Y$ with $\rho_{t+1}(X) \geq \rho_{t+1}(Y)$. Let $K := \rho_{t+1}(X) - \rho_{t+1}(Y) \geq 0$ and set $Z := X + K \geq X$. Then cash invariance yields $\rho_{t+1}(Z) = \rho_{t+1}(Y)$. Apply time consistency in the sense of (8) to obtain $\rho_t(Z) = \rho_t(Y)$. By monotonicity, $\rho_t(X) \geq \rho_t(Z) = \rho_t(Y)$. Here, the weak definition of time–consistency that uses only the indifference relation is equivalent to the strong notion of time–consistency (9) that uses the ”is riskier than”–relation.

Our aim is to establish equivalence between time–consistency and the pasting stability of the set of priors. As a first step in that direction, let us note first that time–consistency is nothing but a version of the law of iterated expectations for the nonlinear expectation defined by a risk measure.

**Lemma 22.** The sequence $(\rho_t)$ given by (5) is time–consistent if and only if we have for all $t < T$ the law of iterated expectations
\[ \rho_t(-\rho_{t+1}(X)) = \rho_t(X) \]
or equivalently
\[ \essinf_{P \in \mathcal{P}} E^P[X|F_t] = \essinf_{P \in \mathcal{P}} E^P[\essinf_{Q \in \mathcal{P}} E^Q[X|F_{t+1}]|F_t] \]
for all $X \in L^\infty(F_T)$. 

EXERCISE 7. Prove the lemma.

We show now that an agent who uses the full Bayesian updating rule as defined in (5) acts in a dynamically consistent way if and only if his set of priors is stable under pasting.

THEOREM 23. The dynamic coherent risk measure given by
\[ \rho_t(X) = \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P [-X | \mathcal{F}_t] \]
is time–consistent if and only if $\mathcal{P}$ is stable under pasting.

PROOF. Suppose that $\mathcal{P}$ is stable under pasting. We have to show
\[ \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P [X | \mathcal{F}_t] = \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P [\text{essinf}_{Q \in \mathcal{P}} \mathbb{E}^Q [X | \mathcal{F}_{t+1}] | \mathcal{F}_t] . \]
In general, of course, the right hand side is smaller or equal to the left hand side; just take $Q = P$ in the inner essential infimum on the right hand side and use the law of iterated expectations. We thus have to show that the left hand side is greater or equal the right hand side.

The proof consists now in two steps. In the first step, we show that the essential infimum $\mathbb{E}^Q [X | \mathcal{F}_{t+1}]$ is attained by some $Q^* \in \mathcal{P}$— this requires already stability in order to show that the essential infimum is the limit of a monotone sequence $\mathbb{E}^{Q^n} [X | \mathcal{F}_{t+1}]$ for $(Q^n) \subset \mathcal{P}$. Weak compactness allows then to find the minimizer $Q^*$. We postpone this step to the following lemma. The second step is then a quick thing. With stability, we have $PQ_{t+1} \in \mathcal{P}$ for all $P, Q \in \mathcal{P}$. Lemma 20 then yields
\[ \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P [\text{essinf}_{Q \in \mathcal{P}} \mathbb{E}^Q [X | \mathcal{F}_{t+1}] | \mathcal{F}_t] = \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P [\mathbb{E}^{Q_{t+1}} [X | \mathcal{F}_t] | \mathcal{F}_t] \leq \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P [X | \mathcal{F}_t] , \]
and the desired inequality is proved.

For the reverse inequality, suppose that $\tilde{P} \tilde{Q}_{t+1} \notin \mathcal{P}$ for some $P, Q \in \mathcal{P}$ and some $t < T$. With the help of a separation theorem, we can then find $X \in L^\infty$ such that
\[ \min_{P \in \mathcal{P}} \mathbb{E}^P [\text{essinf}_{Q \in \mathcal{P}} \mathbb{E}^Q [X | \mathcal{F}_{t+1}]] \leq \mathbb{E}^{\tilde{P} \tilde{Q}_{t+1}} X < \min_{P \in \mathcal{P}} \mathbb{E}^P (X) . \]
Thus, time–consistency is violated. □

We state now the lemma needed in the above proof.

LEMMA 24. Let $T > 0$, $Z \in L^\infty (\Omega, \mathcal{F}_T, P_0)$ and $\tau \leq T$ a stopping time. Under Assumption 17, there exists a measure $P^{Z,\tau} \in \mathcal{Q}$ that coincides with $P_0$ on the $\sigma$–field $\mathcal{F}_\tau$ and
\[ \text{essinf}_{P \in \mathcal{Q}} \mathbb{E}^P [Z | \mathcal{F}_\tau] = \mathbb{E}^{P^{Z,\tau}} [Z | \mathcal{F}_\tau] . \]
Proof. We show below that there exists a sequence \((P^m) \subset \mathcal{Q}\) with \(P^m = P_0\) on \(\mathcal{F}_\tau\) such that
\[
\operatorname{essinf}_{P \in \mathcal{Q}} \mathbb{E}^P [Z|\mathcal{F}_\tau] = \lim_{m \to \infty} \mathbb{E}^{P^m} [Z|\mathcal{F}_\tau].
\]

By Assumption 17, the sequence has a weak limit point \(P^{Z,\tau} \in \mathcal{Q}\) and
\[
\operatorname{essinf}_{P \in \mathcal{Q}} \mathbb{E}^P [Z|\mathcal{F}_\tau] = \mathbb{E}^{P^{Z,\tau}} [Z|\mathcal{F}_\tau]
\]
follows.

It remains to establish the existence of the minimizing sequence \((P^m) \subset \mathcal{Q}\) with the desired properties follows if we can show that \(\Phi\) is downward directed. Hence, let \(P, \hat{P} \in \mathcal{Q}\) with \(P = \hat{P} = P_0\) on \(\mathcal{F}_\tau\). Then
\[
\min \left\{ \mathbb{E}^P [Z|\mathcal{F}_\tau], \mathbb{E}^{\hat{P}} [Z|\mathcal{F}_\tau] \right\} = \mathbb{E}^P [Z|\mathcal{F}_\tau] 1_A + \mathbb{E}^{\hat{P}} [Z|\mathcal{F}_\tau] 1_{A^c}
\]
for \(A = \left\{ \mathbb{E}^P [Z|\mathcal{F}_\tau] < \mathbb{E}^{\hat{P}} [Z|\mathcal{F}_\tau] \right\}\). We have to show that there exists \(R \in \mathcal{Q}\) with \(R = P_0\) on \(\mathcal{F}_\tau\) and
\[
\mathbb{E}^P [Z|\mathcal{F}_\tau] 1_A + \mathbb{E}^{\hat{P}} [Z|\mathcal{F}_\tau] 1_{A^c} = \mathbb{E}^R [Z|\mathcal{F}_\tau].
\]
This follows from the equivalent characterization of time–consistency in Lemma 19, 3. \(\square\)

5. Examples

Before discussing general classes of priors that are stable under pasting, let us consider some specific examples.

Example 25. Average Value at Risk is not time–consistent. For Average Value at Risk, the set of priors is given by
\[
\mathcal{P} = \left\{ Q \ll P_0 : \frac{dQ}{dP_0} \leq \lambda^{-1} \right\}.
\]
It is plausible that one can violate the uniform upper bound $\lambda^{-1}$ by pasting two measures. Indeed, in this case, we have, e.g.,

$$\frac{dPQ_1}{dP_0} \bigg| \mathcal{F}_2 = \frac{dP}{dP_0} \bigg| \mathcal{F}_1 \frac{dQ}{dP_0} \bigg| \mathcal{F}_2 / \frac{dQ}{dP_0} \bigg| \mathcal{F}_1$$

and there is no reason why this should be bounded by $\lambda^{-1}$ in general.

We give a concrete example. Consider the two-period binomial tree in Figure 3. The reference measure is the uniform distribution that puts weight $\frac{1}{4}$ on every path. Take $\lambda = \frac{3}{4}$. Then the set

$$\mathcal{P} = \left\{ Q : \frac{dQ}{dP_0} \leq \lambda^{-1} \right\} = \left\{ (q_1, q_2, q_3, q_4) \geq 0 : \sum q_i = 1, q_i \leq \frac{1}{3} \right\}$$

is not stable under pasting. Consider the pasting of $q = (1/3, 0, 1/3, 1/3)$ after $p_0 = (1/4, 1/4, 1/4, 1/4)$ at time 1. The conditional probabilities for going upwards under $q$ are $1$ and $1/2$, resp. Hence,

$$p_0 q_1 = (1/2, 0, 1/4, 1/4) \notin \mathcal{P}.$$ 

**Figure 3.** Average Value at Risk is not time-consistent.

**Exercise 8.** Let $\mathcal{P}$ be a convex set of priors. Define the stable hull $\mathcal{P}^s$ of $\mathcal{P}$ as the smallest set of priors that is stable under pasting.

- Convince yourself that the stable hull is well-defined. (The intersection of stable sets is stable).
- Show that $\mathcal{P}^s$ is convex.
- Compute the stable hull for the above example concerning Average Value at Risk.
5. EXAMPLES

5.1. Dynamic Exponential Families. We show here how to construct a general class of priors which is stable under pasting and captures the idea of independent and "identically distributed" random variables in a multiple prior setting.

Let us start with a probability space \((S, \mathcal{S}, \nu_0)\) with \(S \subset \mathbb{R}^d\). Let \((\Omega, \mathcal{B}, P_0) = \bigotimes_{t \in \mathbb{N}} (S, \mathcal{S}, \nu_0)\) be the infinite product and denote by \(X_t\) the \(t\)-th projection. Of course, the sequence \((X_t)\) is independent and identically distributed with distribution \(\nu_0\) under \(P_0\). Let \((\mathcal{F}_t)\) be the filtration generated by the sequence \((X_t)\).

We introduce now densities that model stepwise small deviations from the distribution \(\nu_0\). The construction is the discrete time version of Girsanov transformations that readers might be familiar with. Alternatively, one can view the following construction as a dynamic exponential family — a model frequently used in Statistics. Let us first assume that the Laplace transform exists, or

\[
\int_S e^{\lambda \cdot x} \nu_0(dx) < \infty
\]

for all \(\lambda \in \mathbb{R}^d\). The dot product is the scalar product in \(\mathbb{R}^d\). The log–Laplace function

\[
L(\lambda) = \log \int_S e^{\lambda \cdot x} \nu_0(dx)
\]

is then well defined. Note that, by definition,

\[
P_0^\mathbb{R} \exp (\lambda \cdot X_t - L(\lambda)) = 1.
\]

We use this property to define densities on \((\Omega, \mathcal{B}, P_0)\).

Let \((\alpha_t)\) be a predictable process with values in \(\mathbb{R}^d\). Set

\[
D^\alpha_t = \exp \left( \sum_{s=1}^t \alpha_s \cdot X_s - \sum_{s=1}^t L(\alpha_s) \right).
\]

\(D^\alpha\) is a density process that defines a probability measure \(P^\alpha \sim P_0\). Fix two bounded, predictable processes \((a_t), (b_t)\) with values in \(\mathbb{R}^d\) and \(a_t \leq b_t\) (for every component). We denote by \(\mathcal{P}^{a,b}\) the set of all probability measures \(P \sim P_0\) whose density processes satisfy (11) for a predictable process \((a_t)\) with values \(a_t \leq \alpha_t \leq b_t\).

We record for our purposes the following lemma.

LEMMA 26. The set \(\mathcal{P}^{a,b}\) is stable under pasting.

PROOF. The proof is straightforward if one uses the the characterization of stability by densities, or time–consistency, see Eqn. (7). □

REMARK 27. In joint work with Monika Bier, we show that the sets \(\mathcal{P}^{a,b}\) are the only time consistent multiple priors in finite trees under a suitable martingale generator assumption.
Remark 28. In a continuous time diffusion setting, the analog density is
\[
\exp \left( \int_0^t \alpha_s dW_s - \frac{1}{2} \int_0^t \alpha_s^2 ds \right)
\]
for a Brownian motion \( W \). Note that the log–Laplace transform of a normal distribution is \( L(\lambda) = \lambda^2 / 2 \). The corresponding family for \( |\alpha_t| \leq \kappa \) appears in Epstein and Chen (2002). The martingale representation theorem allows to show that all equivalent sets of priors that are stable under pasting have such density processes, see Delbaen (2002b).

5.2. Equivalent (Super)Martingale Measures. Let \((S_t)\) be a bounded process on the filtered probability space \((\Omega, \mathcal{F}, P_0, (\mathcal{F}_t))\) and denote by \( \mathcal{P}^* \) the set of all equivalent probability measures \( P^* \) under which \((S_t)\) is a martingale (or supermartingale).

Lemma 29. The set \( \mathcal{P}^* \) is stable under pasting.

Proof. We prove rectangularity of \( \mathcal{P}^* \). Let \( P, Q \in \mathcal{P}^* \). We have to show that \( PQ_\tau \in \mathcal{P}^* \) for all \( \tau \). If \( t < \tau \), Lemma 20 and the fact that \( M_{t+1} \) is \( \mathcal{F}_\tau \)-measurable yield
\[
\mathbb{E}^{PQ_\tau} [M_{t+1} | \mathcal{F}_t] = \mathbb{E}^P [\mathbb{E}^Q [M_{t+1} | \mathcal{F}_\tau] | \mathcal{F}_t] = \mathbb{E}^P [M_{t+1} | \mathcal{F}_t] = M_t.
\]
Here, we used that \( P \) is an equivalent martingale measure. If \( t \geq \tau \), we use the same lemma and the fact that \( Q \) is a martingale measure:
\[
\mathbb{E}^{PQ_\tau} [M_{t+1} | \mathcal{F}_t] = \mathbb{E}^Q [M_{t+1} | \mathcal{F}_t] = M_t.
\]
The reasoning applies equally well to the set of super– or submartingale measures. \( \square \)
CHAPTER 3

Lecture 3: Dynamic Convex Risk Measures, Consistency and a Supermartingale Property

Let us now discuss dynamic versions of convex risk measures. In view of the representation (3), it seems natural to consider risk measures that can be represented as

\[ \rho_t(X) = \text{esssup}_{P \in \mathcal{M}^e(P_0)} \mathbb{E}^P[-X|\mathcal{F}_t] - \alpha_t(P) \]

for

\[ \alpha_t(P) = \text{esssup}_{\mathcal{X} \in \mathcal{X}(P) \leq 0} \mathbb{E}^P[-X|\mathcal{F}_t] \]

where \( \mathcal{M}^e(P) \) consists of all probability measures that are equivalent to \( P \). Let us also introduce

\[ Q = \text{dom} \alpha_0 \cap \mathcal{M}^e(P_0) = \{ Q \in \mathcal{M}^e(P_0) | \alpha_0(Q) < \infty \} . \]

We assume \( Q \neq \emptyset \).

**Remark 30.**

- Restricting to equivalent probability measures has the advantage that \( \alpha_t(P) \) is \( P_0 \)-a.s. well defined.
- We do not discuss here under what conditions on \( (\rho_t) \) one has the representation (12). It is sufficient that the mappings \( \rho_t \) are continuous from above and \( \alpha_t(P) \) is a.s. finite for some \( P \) equivalent to \( P_0 \), see Föllmer and Penner (2006), Lemma 3.5.

Our aim is to understand and characterize time-consistency for such objects; in light of Lemma 22, we thus want to characterize the relationship

\[ \rho_t(-\rho_{t+1}(X)) = \rho_t(X), \quad (X \in L^\infty, t = 0, \ldots, T - 1). \]

As there is no such nice property as stability under pasting in the general case, let us start with one property that generalizes from the coherent case. For coherent dynamic risk measures we have the following supermartingale characterization of time consistency.

**Theorem 31.** Let \( \mathcal{P} \) satisfy Assumptions 16 and 17. The coherent risk measures

\[ \rho_t = \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P[-X|\mathcal{F}_t] \]

are time-consistent if and only if the sequence of risk assessments \( (\rho_t(X)) \) is a \( P \)-supermartingale for all \( P \in \mathcal{P} \).
3. DYNAMIC CONVEX RISK MEASURES

Proof. Suppose that \((\rho_t)\) is time–consistent. By recursivity,
\[
\rho_t(X) = \rho_t(-\rho_{t+1}(X)).
\]
By the representation (13) \(\rho_t(-\rho_{t+1}(X)) \geq \mathbb{E}^P [\rho_{t+1}(X) | \mathcal{F}_t] \). The supermartingale property follows.

Now suppose that we have the supermartingale property. Then we get
\[
\text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P [\rho_{t+1}(X) | \mathcal{F}_t] \leq \rho_t(X)
\]
which is equivalent to
\[
\rho_t(-\rho_{t+1}(X)) \leq \rho_t(X).
\]
The inequality \(\rho_t(-\rho_{t+1}(X)) \geq \rho_t(X)\) is obvious from
\[
\rho_t(-\rho_{t+1}(X)) = \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P [\text{esssup}_{Q \in \mathcal{P}} \mathbb{E}^Q [-X | \mathcal{F}_{t+1}] | \mathcal{F}_t]
\]
\[
\geq \text{esssup}_{P \in \mathcal{P}} \mathbb{E}^P [\mathbb{E}^P [-X | \mathcal{F}_{t+1}] | \mathcal{F}_t]
\]
and the law of iterated expectations.

The following theorem generalizes the supermartingale characterization of time consistency to convex risk measures.

Theorem 32 (Föllmer/Penner). A dynamic convex risk measure \((\rho_t)\) of the form (12) is time–consistent if and only if \(\rho_t(X) + \alpha_t(Q)\) is a \(Q\)–supermartingale for all \(Q \in \mathcal{M}^c(P_0)\) with \(\alpha_0(Q) < \infty\).

As \(\alpha_0\) can take only the values 0 and \(\infty\) for coherent risk measures, the above theorem generalizes Theorem 31 from the coherent to the convex case. We postpone its proof to the end of the lecture when we have developed other characterizations of time–consistency.

1. Characterization of Time–Consistency via Acceptance Sets

Recall the set of positions that we accept at time \(t\) given by
\[
\mathcal{A}_t = \{X \in L^\infty(\mathcal{F}_T) | \rho_t(X) \leq 0\}.
\]

Let us now consider financial positions that are \(\mathcal{F}_{t+1}\)–measurable. As we have set the interest rate to 0 throughout, these can be identified with positions that mature at time \(t+1\). Let
\[
\mathcal{A}_{t,t+1} = \{X \in L^\infty(\mathcal{F}_{t+1}) | \rho_t(X) \leq 0\}
\]
be the set of positions with maturity \(t+1\) that are accepted at time \(t\).

Now look at an investor who has to positions in her portfolio. Position A has maturity \(t+1\) and she accepts it at time \(t\), maybe because she can perfectly hedge it without risk. Position B has maturity \(T\), and the agent knows how to hedge that position at time \(t+1\). Is there any reason to reject the sum of the two positions at time \(t\)? No!
For the converse, note that an acceptable position $X$ with maturity $T$ can be decomposed as

$$X = -\rho_{t+1}(X) + Z$$

for $Z = X + \rho_{t+1}(X)$. From recursivity, $-\rho_{t+1}(X) \in \mathcal{A}_{t,t+1}$ and from cash invariance $Z \in \mathcal{A}_{t+1}$. Every position that is accepted at $t$ can thus be decomposed into an acceptable position with maturity $t+1$ and a remainder that is acceptable at time $t+1$. We thus have

**Theorem 33.** A dynamic convex risk measure $(\rho_t)$ is time-consistent if and only if for $t = 0, \ldots, T-1$

$$\mathcal{A}_t = \mathcal{A}_{t,t+1} + \mathcal{A}_{t+1}.$$  

**Exercise 9.** Give the formal proof.

### 2. Time-consistent Penalty Functions

Let us now explore what time consistency means for the penalty functions. As the penalty functions are derived from the acceptance sets, let us introduce the "one-step" penalty functions

$$\alpha_{t,t+1}(P) = \text{esssup}_{X \in \mathcal{A}_{t,t+1}} \mathbb{E}^P[-X|\mathcal{F}_t].$$

Transfer Theorem 33 to the penalty functions, and you get

**Theorem 34.** A dynamic convex risk measure $(\rho_t)$ is time-consistent if and only if for $t = 0, \ldots, T-1$

$$\alpha_t(P) = \alpha_{t,t+1}(P) + \mathbb{E}^P[\alpha_{t+1}|\mathcal{F}_t].$$  

### 3. Proof of the Supermartingale Property

The characterization (14) of time consistency in terms of penalty functions allows us to write

$$\mathbb{E}^P[\rho_{t+1}(X) + \alpha_{t+1}(P) - \alpha_t(P)|\mathcal{F}_t]$$

$$\quad = \mathbb{E}^P[\rho_{t+1}(X) + \alpha_{t+1}(P) - \alpha_{t,t+1}(P) - \mathbb{E}^P[\alpha_{t+1}(P)|\mathcal{F}_t]|\mathcal{F}_t]$$

$$\quad = \mathbb{E}^P[\rho_{t+1}(X) - \alpha_{t,t+1}(P)|\mathcal{F}_t].$$

From the representation of convex risk measures (12), we get then

$$\mathbb{E}^P[\text{esssup}_{Q \in \mathcal{M}^e(P_0)} \mathbb{E}^Q[-X|\mathcal{F}_{t+1}] - \alpha_{t+1}(Q) - \alpha_{t,t+1}(P)|\mathcal{F}_t]$$

The set

$$\{\mathbb{E}^Q[-X|\mathcal{F}_{t+1}] : Q \in \mathcal{M}^e(P_0)\}$$

is upward directed (copy of the argument at the end of the proof of Lemma 24). Moreover, we can restrict to $Q$ with $Q = P$ on $\mathcal{F}_{t+1}$ in taking the essential supremum. Note that for such $Q$ we have
\[ \alpha_{t,t+1}(Q) = \alpha_{t,t+1}(P). \] Hence we can choose a sequence \((P^n) \subset \mathcal{M}(P_0)\) with \(P^n = P\) such that
\[
= \lim \mathbb{E}^P \left[ \mathbb{E}^{P_n}[-X|\mathcal{F}_{t+1}] - \alpha_{t+1}(P^n) - \alpha_{t,t+1}(P)|\mathcal{F}_t\right] \\
= \lim \mathbb{E}^{P_n} \left[ \mathbb{E}^P[-X|\mathcal{F}_{t+1}] - \alpha_{t+1}(P^n) - \alpha_{t,t+1}(P)|\mathcal{F}_t\right] \\
\]
From time–consistency for penalty functions,
\[
= \lim \mathbb{E}^{P_n} \left[ \mathbb{E}^P[-X|\mathcal{F}_{t+1}] - \alpha_{t}(P^n)|\mathcal{F}_t\right] \leq \rho_t(P).
\]
This proves the supermartingale property.

4. Examples

Example 35. The dynamic entropic risk measure
\[ \rho_t(X) = \log \mathbb{E}^P \left[ e^{-X} | \mathcal{F}_t \right] \]
is time–consistent. This follows directly from the law of iterated expectations:
\[
\rho_t(-\rho_{t+1}(X)) = \log \mathbb{E}^P \left[ \mathbb{E}^{\log \mathbb{E}^P[e^{-X}|\mathcal{F}_{t+1}]} | \mathcal{F}_t \right] \\
= \log \mathbb{E}^P \left[ \mathbb{E}^P \left[ e^{-X} | \mathcal{F}_{t+1} \right] | \mathcal{F}_t \right] = \rho_t(X).
\]

Example 36. In the spirit of the above example, we introduce dynamic robust entropic risk as
\[ \rho_t(X) = \log \text{essinf}_{P \in \mathcal{P}} \mathbb{E}^P \left[ e^{-X} | \mathcal{F}_t \right] \]
for a class of priors \(\mathcal{P}\). As time–consistency is equivalent to the law of iterated expectations for the nonlinear expectation induced by \(\mathcal{P}\), we understand that dynamic robust entropic risk is time–consistent if and only if \(\mathcal{P}\) is stable under pasting.
Bibliography


