Working Paper
Series

FINRISK
National Centre of Competence in Research
Financial Valuation and Risk Management

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 381

Co-monotonicity of optimal investments and the design of structured financial products

Marc Oliver Rieger

First version: April 2007
Current version: April 2007

This research has been carried out within the NCCR FINRISK project on "Behavioural and Evolutionary Finance"

Co-monotonicity of optimal investments and the design of structured financial products

Marc Oliver Rieger*

April 25, 2007

Abstract

We prove that under very weak conditions optimal financial products have to be co-monotone with the inverted state price density. Optimality is meant in the sense of the maximization of an arbitrary preference model, e.g. Expected Utility Theory or Prospect Theory. The proof is based on methods from transport theory. We apply the general result to specific situations, in particular the case of a market described by the Capital Asset Pricing Model, where we derive an extension of the Two-Fund-Separation Theorem. We use our results to derive a new approach to optimization in wealth management, based on a direct optimization of the return distribution of the portfolio. We provide existence and non-existence results for optimal products in this framework. Finally we apply our results to the study of down-and-out barrier options, show that they are not optimal and describe a construction of a cheaper product yielding the same return distribution.

JEL classification: G11,C61.
Keywords: Co-monotonicity, structured products, portfolio optimization, no-arbitrage condition, decision theory.

*Marc Oliver Rieger, ISB, University of Zurich, `rieger@isb.uzh.ch`

1 Introduction

Structured financial products have gained a large popularity in many countries in the last years. With a volume of 225 billion Euro of structured retail products issued worldwide in 2005 and an yearly increase of nearly 30%, they are quickly becoming a standard form of investment for private investors.[1] Their success sometimes challenges traditional financial models, but can often be explained by behavioral theories that take effects like loss aversion into account.

On the other hand, many banks have understood today that selling separate financial products is not the optimal way to achieve an overall optimal portfolio for the client, since it does not consider the correlations between the different products in the client's portfolio. Therefore an integrated wealth management that optimizes the overall portfolio of a client or offers a taylor-made collection of assets and structured products is more and more frequently offered. A systematic approach to this optimization that takes into account potential behavioral biases of the client is therefore of high importance.

In this article we develop such a systematic approach for wealth management and study properties that an optimized portfolio should have. We consider this in the framework of structured products, where we assume that the total wealth of the client is invested into this structured product and there is hence no background risk to be hedged. On the other hand we aim to impose only the mildest possible conditions on the preferences of the client. In particular we do not assume that the client will want to optimize its investment according to Mean-Variance Theory or to the rational framework of Expected Utility Theory, but we allow explicitly for other decision models, e.g. behavioral models like Prospect Theory. We also allow for "benchmarking", i.e. for variable reference points that are set, for instance, by some index (like a stock market index), as well as for decision models based on the total return.

One main result of our work is that even in this general setting certain properties of an optimal financial product are always present. In particular, we will

[1] source: BNP Paribas, Structuredretailproducts.com

show that optimal products "follow the market", i.e. they are co-monotone with the market portfolio (in the case of a CAPM market) or with the inverted state price density (in the general case). This result has immediate consequences for the design of financial products in the context of wealth management, and we will mention some of the implications later on.

Let us have a closer look at the model we are studying: We consider complete and efficient financial markets in which all market participants have homogenous beliefs and act according to a maximization of their utility. The main focus of this article lies on the question what properties a financial product on such a market has to satisfy if it is optimal in the sense that it maximizes a given utility of an investor. Before we make this question more precise, we first review some properties of such markets (compare, e.g., [4] for details):

Let the returns of an asset be given by the probability measure p on $\mathbb{R}$, and let the state price density be π and let their mean and variance be given by $\mathbb{E}(p)$, $\mathbb{E}(\pi)$ and $\mathrm{var}(p)$, $\mathrm{var}(\pi)$, respectively. Let R be the return of the risk-free asset (i.e. the interest– or risk-free rate). Then we can derive from the no-arbitrage condition that all financial products that are available for a fixed price and can be described by a joint probability measure T such that $p = \int_{\mathbb{R}} dT(\cdot, y)$ and $\pi = \int_R dT(x, \cdot)$, satisfy the constraint

$$\mathbb{E}(p) - R \quad = \quad -\beta_{p\pi}(\mathbb{E}(\pi) - R), \qquad \text{where}\, \beta_{p\pi} = \frac{\operatorname{cov} T}{\mathrm{var}(\pi)}. \tag{1}$$

We make henceforth the general assumption that the state prices are nonnegative (compare [4]) which corresponds to assuming that the preferences of the market participants are weakly monotone.

An optimal product is defined as a product that maximizes a given utility subject to condition (1). The utility could hereby be given according to Expected Utility Theory, Prospect Theory or a different model, depending on the application one has in mind.

We define the *inverted state price density* $\tilde{\pi}$ by $\tilde{\pi}(\mathbb{E}(\pi) + x) := \pi(\mathbb{E}(\pi) - x)$. It has been observed in the literature that in certain cases an optimal asset is co-monotone to $\tilde{\pi}$. (We will give a precise definition of co-monotonicity in Section 2.1.) This is a classical result in the context of Pareto-efficiency, see, e.g., [10], and has been generalized by Dybvig [5] to optimal portfolio design in the case of finite state spaces with equal probability for each state and an expected utility maximizer. Dybvig states that then "any cheapest way to achieve a lottery assigns the outcomes of the lottery to the states in reverse order of the state price density", in other words that an optimal portfolio and the inverted state price density are co-monotone. He proves this case and notes that an extension to continuous distributions should be possible. Later, this approach has been used in the context of insurance risks by Dhaene [3] and also by Carlier and Dana [2].

In this article we generalize the concept in a mathematically rigorous way to arbitrary probability measures and to arbitrary preference relations for investors without background risk. This is based on general mathematical results which we present in Section 2.1 of this article and which extend works by Gangbo and McCann [6, 12]. More information on the mathematical background and some generalizations in the context of transport plans can be found in [15]. We apply the mathematical results in Section 2.2 to obtain general conditions under which the outcome distribution of every optimal financial product is co-monotone with the inverted state price density. It turns out that this property is much more universal than had been anticipated. We discuss special cases in Section 2.3–2.5. In particular, we investigate the implications of our result for a market described by the Capital Asset Pricing Model (CAPM) and extend a variant of the Two-Fund-Separation Theorem to the case of arbitrary decision models used for the product selection.

The co-monotonicity result opens the path for a new approach to the design of optimal financial products, which is based not on the optimization of asset allocations, but instead of a direct optimization of the underlying return distribution. This approach is explained in Section 3.1 and existence results for optimal financial products are derived. We briefly sketch some numerical methods for the computation of such optimal products in Section 3.2. There are, however, strong limitations to this approach, which provide interesting insights into the shortfalls of pricing formulas based on the no-arbitrage condition. These limitations will be discussed in Section 3.3.

We conclude this article by another practical application, namely the study of barrier options (Section 4). We demonstrate that these options are not optimal and can be improved by a monotonizing procedure that can be performed explicitly. We suggest an explanation for why the non-optimal products are nevertheless quite popular on the market, based on systematic misestimation of probabilities. An experimental study confirms this idea.

2 Co-monotonicity

In this section we present some of our main results. We start out from a mathematical analysis of joint probability measures in Section 2.1. The mathematically less inclined reader might skip this section and continue with Section 2.2 at first reading, relying on the intuitive idea that joint probability measures that minimize (or maximize) certain quantities are co-monotone, which means roughly that the probability measures that are connected via the joint probability measure "follow" each other: a larger outcome of one of them always corresponds to a larger outcome of the other and vice versa. The results summarized in Section 2.1 are extensions of work by Gangbo and McCann [6, 12]. They can essentially be found in [15], however we adapt them to our application.

In Section 2.2 we apply these results to the study of optimal financial products and prove that such optimal products are co-monotone with the inverted state price density under general conditions. In the following Sections 2.3–2.5, we study special cases of this general statement which are of particular interest.

2.1 Co-monotonicity of joint probability distributions

The main mathematical tool that we apply in this article is the so-called "transport theory". This theory originally dealt with optimizing transports of soil, e.g., in construction or mining, and goes back to the 18th century where the French mathematician Monge [13] introduced the first version of this problem. Major progress on this problem has been achieved in the 1940s with the seminal work by the Russian economist and mathematician Kantorovich [8]. We state his formulation for the one-dimensional case that is of particular interest for our purpose:

Definition 2.1 (Transport problem). *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and let $c\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a lower semicontinuous function (the* cost function*). Then the* transport problem *consists of finding a joint probability measure $T \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ which minimizes*

$$C(T) := \int_{\mathbb{R}} \int_{\mathbb{R}} c(x, y) dT(x, y), \tag{2}$$

such that the marginals of T are given by μ and ν, i.e.,

$$\mathrm{pr}_1 T = \mu, \quad \mathrm{pr}_2 T = \nu,$$

where pr_1 *is the projection on the first coordinate and* pr_2 *the projection on the second coordinate, i.e.*

$$\mathrm{pr}_1 T := \int_{\mathbb{R}} dT(\cdot, y), \quad \mathrm{pr}_2 T := \int_{\mathbb{R}} dT(x, \cdot).$$

In the optimum, the mass μ is transported for least cost to ν according to the transport plan T. It is well known that the above transport problem admits a solution, even in its n-dimensional generalization, see, e.g., [1].

The central property we use in this article is co-monotonicity:

Definition 2.2 (Co-monotonicity of joint probability measures).
Let $T \in \mathcal{P}(\mathbb{R}, \mathbb{R})$ be a joint probability measure with marginals $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Then T is called co-monotone *if for all $(x_1, y_1), (x_2, y_2) \in \operatorname{supp} T$ with $x_2 > x_1$, we have $y_2 \geq y_1$. (We also sometimes say that μ and ν are co-monotone.)*

It is easy to see that this property can be expressed equivalently as a monotonicity property as follows:

Remark 2.3. *T is co-monotone if and only if it satisfies the following condition:*

For all Borel sets $A, B \subset \mathbb{R}$ with $\mu(A) > 0, \mu(B) > 0$ and $\inf\{x \in B\} > \sup\{x \in A\}$ we define

$$A' := \operatorname{supp} \mathrm{pr}_2(T|_{A \times \mathbb{R}}), \quad B' := \operatorname{supp} \mathrm{pr}_2(T|_{B \times \mathbb{R}}).$$

Then $\inf\{x \in B'\} \geq \sup\{x \in A'\}$.

We can illustrate the definition by the observation that co-monotonicity of a joint probability measure means nothing else than that its first marginal "follows" its second marginal: the larger the outcome of the second marginal, the larger the outcome of the first (and vice versa).

We define the following useful symbol:

Definition 2.4 (Push forward)**.** *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a measurable function and μ a measure on $\mathbb{R}$, then we define the push forward $f_\#\mu$ by*

$$(f_\#\mu)(B) := \mu(f^{-1}(B)).$$

The definition of co-monotonicity is a natural extension of the usual notion of monotonicity in the following sense:

Proposition 2.5. *Let T be a joint probability measure with marginals μ and ν. If T can be interpreted as a map that maps the measure μ pointwise to ν, then T is co-monotone if this map is monotone as a function on $\mathbb{R}$. More precisely, if there exists a map $\tilde{T}\colon \operatorname{supp}\mu \to \mathbb{R}$ such that $T = (Id \times \tilde{T})_\#\mu$, then T is co-monotone iff $\tilde{T}$ is μ-a.e. monotone.*[2]

Proof. It is straigthforward to prove that $\operatorname{supp} T$ satisfies the condition of Def. 2.2 if and only if $\tilde{T}$ is monotone. □

We say that a sequence $(\nu_j) \subset \mathcal{P}$ converges weakly-$\star$ to ν, i.e. $\nu_j \overset{\star}{\rightharpoonup} \nu$ if

$$\int \phi(x)\nu_j(x) \to \int \phi(x)\nu(x)$$

for all bounded continuous functions ϕ.

The following result introduces a construction that can be used for numerical computations of co-monotone joint probability measures. This and the following results are extensions of previous work in the context of transport plans by Gangbo and McCann [6, 12] and can in slightly different form also be found in [15].

Theorem 2.6. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ then there exists a co-monotone joint probability measure with marginals μ and ν and it can be approximated by weighted sums of Dirac measures.*

[2] We remark that it is necessary to allow $\tilde{T}$ to be non-monotone on a set N with $\mu(N) = 0$, since $\tilde{T}$ can be defined arbitrarily on such sets.

Proof. Let $\varepsilon > 0$. We first choose bounded intervals $D = (d_1, d_2)$, $E = (e_1, e_2) \subset \mathbb{R}$ such that $\mu(D) = \nu(E) \geq 1-\varepsilon$ and $\mu((-\infty, d_1)) = \nu((-\infty, e_1))$. Then we decompose $D\times E \subset \mathbb{R}\times\mathbb{R}$ into squares. For simplicity we assume that $d_1 = e_1 = 0$. We define $I_i^k := [(i-1)2^{-k}, i2^{-k})$ for $1 \leq i \leq N := [1+\max(d_2, e_2)\cdot 2^k]$ and $m_i := \mu(I_i^k \cap D)/|I_i^k|$, $n_i := \nu(I_i^k \cap E)/|I_i^k|$.

Denote the midpoints of the intervals I_i^k by z_i^k. We define sequences of measures μ_k and ν_k by $\mu_k := \sum_{i=1}^N m_i \delta_{z_i^k}$ and $\nu_k := \sum_{i=1}^N n_i \delta_{z_i^k}$. We choose $\varepsilon = 1/k$ and construct a co-monotone joint probability measure T_k for μ_k and ν_k: define $T_k := \sum_{i,j=1}^N a_{i,j}^k \delta_{(z_i^k, z_j^k)}$ where the constants $a_{i,j}^k$ are determined by the following algorithm:

Set $i = j = 1$, $L = m_1$. While $i \leq N$ or $j \leq N$ do the following:

- If $L > n_j$ then set $L = L - n_j$ and $a_{i,j}^k = n_j$.
- If $L \leq n_j$ then set $L = 0$ and $a_{i,j}^k = L$.
- If $L = 0$ then increase i by one and set $L = m_i$, otherwise increase j by one.

The algorithm terminates since $\sum_{i=1}^N m_i = \mu(D) = \nu(E) = \sum_{j=1}^N n_j$. The resulting joint probability measure $T_k = \sum_{i,j=1}^N a_{i,j}^k \delta_{z_i^k}$ is co-monotone by construction.

Finally we let $k \to \infty$. The resulting co-monotone joint probability measure T_k converge weakly-$\star$ to a limit $T \in \mathcal{P}(\mathbb{R}\times\mathbb{R})$, which is still co-monotone. $\square$

Marginals determine a co-monotone joint probability measure uniquely as the following Proposition shows:

Proposition 2.7. *Let $T_1, T_2 \in \mathcal{P}(\mathbb{R}\times\mathbb{R})$ be co-monotone joint probability measures with marginals μ and ν. Then $T_1 = T_2$.*

Proof. We discretize T_1 and T_2 similarly as before: Define $a_{ij}^k := \int_{I_i^k\times I_j^k} dT_1(x, y)$, $b_{ij}^k := \int_{I_i^k\times I_j^k} dT_2(x, y)$ and $T_1^k := \sum_{i,j=1}^k a_{ij}^k \delta_{(z_i^k, z_j^k)}$ and $T_2^k := \sum_{i,j=1}^k b_{ij}^k \delta_{(z_i^k, z_j^k)}$ analogously, where I_i^k and z_i^k are defined as in the proof of Theorem 2.6.

Then for $k \to \infty$ we have again $T_1^k \overset{\star}{\rightharpoonup} T_1$ and $T_2^k \overset{\star}{\rightharpoonup} T_2$. Suppose that $T_1 \neq T_2$. Then for k sufficiently large, there are i_0, j_0 such that $a_{i_0 j_0} < b_{i_0 j_0}$. Then since $\sum_i a_{i j_0} = \sum_i b_{i j_0}$ (T_1 and T_2 have the same marginals), there exists a $i_1 \neq i_0$ such that $a_{i_1 j_0} < b_{i_1 j_0}$. Let us assume without loss of generality that $i_1 < i_0$. Since $\sum_j a_{i_1 j} = \sum_j b_{i_1 j}$, there exists a $j_1 \neq j_0$ such that $a_{i_1 j_1} > b_{i_1 j_1}$. By co-monotonicity we must hence have $j_1 < j_0$. Since $\sum_i a_{i j_1} = \sum_i b_{i j_1}$, there exists a $i_2 \neq i_1$ such that $a_{i_2 j_1} < b_{i_2 j_1}$, and by co-monotonicity we must have $i_2 < i_1$. Iterating this argument, we get an infinite sequence of i_k with $i_{k+1} < i_k$. Since i_k are indices from a finite index set, this is a contradiction. □

In the following we need this little lemma:

Lemma 2.8. *Let $A = (a_{i,j}) \in \mathbb{R}^{n\times m}$ be a matrix with nonnegative entries. Let $c\colon \{1,\dots,n\} \times \{1,\dots m\} \to \mathbb{R}$ be a function satisfying the inequality (3). Define*

$$C(A) := \sum_{i=1}^{n} \sum_{j=1}^{m} c(i,j) a_{i,j}.$$

Then there exists a matrix $B = (b_{i,j}) \in \mathbb{R}^{n\times m}$ with the following properties:

(i) $b_{i,j} \geq 0$ *for all* i, j,

(ii) $\sum_{i=1}^{n} b_{i,j} = \sum_{i=1}^{n} a_{i,j}$ *for all* j *and* $\sum_{j=1}^{m} b_{i,j} = \sum_{j=1}^{m} a_{i,j}$ *for all* i,

(iii) $C(B) \leq C(A)$,

(iv) either $b_{i,1} = 0$ *for all* $i = 2,\dots,n$
or $b_{1,j} = 0$ *for all* $j = 2,\dots,m$.

If c satisfies the stronger condition (4), then (iii) can be strengthened to $C(B) < C(A)$.

Proof. The proof is constructive, in fact we give a simple algorithm that computes B for a given matrix A. Since property (iv) is directly connected to the co-monotonicity of a corresponding joint probability measure, we say that this algorithm "monotonizes" a given matrix A.

A key feature in the monotonization will be the following construction which we call a *switch* of (i_1, j_1) and (i_2, j_2):

Take $i_1, i_2 \in \{1, \ldots, n\}$ and $j_1, j_2 \in \{1, \ldots, m\}$ with $i_1 < i_2$, $j_2 < j_1$. Define $\beta := \min\{a_{i_1,j_1}, a_{i_2,j_2}\}$ and

$$\begin{aligned}
b_{i_1,j_1} &:= a_{i_1,j_1} - \beta, \\
b_{i_2,j_2} &:= a_{i_2,j_2} - \beta, \\
b_{i_1,j_2} &:= a_{i_1,j_2} + \beta, \\
b_{i_2,j_1} &:= a_{i_2,j_1} + \beta, \\
b_{i,j} &:= a_{i,j} \quad \text{for all other pairs } (i, j).
\end{aligned}$$

A small calculation shows that the matrix $B := (b_{i,j})$ satisfies the properties (i)-(iii) in the statement of this lemma and that moreover either $b_{i_1,j_1} = 0$ or $b_{i_2,j_2} = 0$ (or both).

Now we just need to find a sequence of switches that transforms A into a matrix B satisfying property (iv) and we have proved the lemma. This can be achieved with the help of the following algorithm:

Set $i = n$ and $j = m$. While $i > 1$ and $j > 1$ do the following:

- Switch $(i, 1)$ and $(1, j)$. (The result will again be called A.)
- If $a_{1,j} = 0$ then decrease j by one.
- If $a_{i,1} = 0$ then decrease i by one.

Finally, set $B = (a_{i,j})$.

The properties that the switch satisfies ensure that the algorithm terminates and that its result B satisfies the properties (i)-(iii). A closer look at the algorithm reveals furthermore that in each processing of the while loop either $a_{i,1}$ or $a_{1,j}$ is set to zero. From this it follows in particular that B also satisfies (iv). This proves the lemma. □

We can now state and prove the main result of this section which specifies the conditions under which solutions of transport problems are co-monotone:

Theorem 2.9. *Let c be a continuous cost function satisfying for all $x_1, x_2, y_1, y_2 \in \mathbb{R}$ with $x_1 < x_2$ and $y_1 < y_2$*

$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1), \tag{3}$$

then the transport problem of Definition 2.1 admits a co-monotone minimizer (and possibly other minimizers which do not need to be monotone).

If c even satisfies the strict inequality

$$c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1), \tag{4}$$

then the transport problem admits a unique minimizer, and this minimizer is co-monotone.

We mention that the second statement of this Theorem could be proved with a simpler method using a proof by contradiction. However, we prefer to present the following, more general and constructive proof:

Proof. Let T be some minimizer of the transport problem. We approximate T by a sequence of measures which are finite sums of Dirac masses on a grid. To be more precise we define for $i, j = -4^k, \dots, 4^k$ the squares

$$Q^k_{ij} := [i2^{-k}, (i+1)2^{-k}) \times [j2^{-k}, (j+1)2^{-k}).$$

We denote the midpoint of the square Q^k_{ij} by M^k_{ij} and define T_k by

$$T_k := \sum_{i,j=-4^k}^{4^k} 2^k T(Q^k_{ij}) \delta_{M^k_{ij}}. \tag{5}$$

It follows from either a short direct computation or Lemma 7 in [12] that $T_k \overset{\star}{\rightharpoonup} T$ for $k \to \infty$. Now we can "monotonize" T_k, i.e. we can perform a finite number of manipulations on T_k which lead to a modified joint probability measure T'_k which is co-monotone and satisfies $C(T'_k) \leq C(T_k)$ (respectively, $C(T'_k) < C(T_k)$ if condition (4) holds). This is equivalent to convert the matrix a given by

$$a_{i,j} := T(Q^k_{ij})$$

(where we omit for simplicity the index k) iteratively into a matrix $a'_{i,j}$ where for all $i_1 < i_2$ and $j_1 > j_2$ either $a'_{i_1,j_1} = 0$ or $a'_{i_2,j_1} = 0$ whereby not changing the sums over rows or columns, only allowing for nonnegative entries and not increasing (respectively, decreasing) the corresponding cost $C(T'_k)$ of the joint probability measure T'_k defined by

$$T'_k := \sum_{i,j=-4^k}^{4^k} 2^k a'_{i,j} \delta_{M^k_{ij}}.$$

This can be achieved by applying Lemma 2.8 to the matrix

$$a^{(m)}_{i,j} = \begin{pmatrix} a^{(m)}_{1,1} & a^{(m)}_{1,2} & \dots & a^{(m)}_{1,q} \\ \vdots & & & \vdots \\ a^{(m)}_{p,1} & a^{(m)}_{p,2} & \dots & a^{(m)}_{p,q} \end{pmatrix}.$$

We call its monotonization $a^{(m+1)}_{ij}$.

We can now apply the same lemma in the next iteration step to proceed from $(m+1)$ to $(m+2)$ where we apply it only for all but the first row of $a^{(m+1)}$ (if $a^{(m+1)}_{1,2} = \cdots = a^{(m+1)}_{1,q} = 0$) or all but the first column (if $a^{(m+1)}_{2,1} = \cdots = a^{(m+1)}_{p,1} = 0$).

Starting with $a^{(0)} := a_{i,j}$, the iteration stops after finitely many steps when the remaining matrix has been reduced to a vector (since in every step the matrix gets reduced by either a row or a column). The result a' of this iteration is of the desired form: It satisfies by construction the condition that for all $i_1 < i_2$ and $j_1 > j_2$ either $a'_{i_1,j_1} = 0$ or $a'_{i_2,j_1} = 0$, and hence its associated joint probability measure T'_k is co-monotone. Moreover, since in every iteration step the sums over rows and columns of $a^{(m)}$ are preserved, the joint probability measure T'_k has the same marginals as T_k, and finally $C(T'_k) \leq C(T_k)$ or, in the case of condition (4), even $C(T'_k) < C(T_k)$.

We now take the limit $k \to \infty$. There exists a T' such that (at least for a subsequence) $T'_k \overset{\star}{\rightharpoonup} T'$. Since (T'_k) is tight and $\|T'_k\| \to 1$, we obtain from Prokhorov's Theorem [14] that $T' \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$. Since $\mathrm{pr}_1 T - \mathrm{pr}_1 T'_k = \mathrm{pr}_1 T - \mathrm{pr}_1 T_k \to 0$ when $k \to \infty$, and the same holds for pr_2, we have constructed a joint probability measure T' with marginals μ and ν. Due to the weak-$\star$ convergence we also have $C(T) - C(T') = \lim_{k\to\infty} C(T_k) - C(T'_k) \geq 0$ (respectively > 0). Therefore T' is a minimizing joint probability measure. It remains to show that T' is co-monotone. Suppose that it is not, then there must be sets $D, E \subset \mathbb{R} \times \mathbb{R}$ with $T'(D), T'(E) > 0$ and such that for all $(x_1, y_1) \in D$ and $(x_2, y_2) \in E$ we have $x_1 < x_2$ and $y_1 > y_2$. We can assume that D and E are such that we can choose squares S_D, S_E in $\{Q_{ij}^k\}$ with $S_D \subset D$ and $S_E \subset E$ and $T'(S_D), T'(S_E) > 0$, but this leads to a contradiction: By the weak-$\star$ convergence we would have that also $T_k'(S_D), T_k'(S_E) > 0$ for k large enough. This would be a contradiction to the co-monotonicity of T_k'. Hence T' is a co-monotone minimizer. □

In the next section we will see how Theorem 2.9 can be applied to the study of financial products.

We conclude this section with a useful approximation lemma that essentially states that co-monotone distributions can be approximated by functions:

Lemma 2.10. *Let T be a co-monotone joint probability measure with marginals μ and ν. Then there is a sequence of co-monotone joint probability measures $T_\varepsilon = (Id \times \psi_\varepsilon)_\# \mu$ where $\psi_\varepsilon \colon \operatorname{supp} \mu \to \mathbb{R}$ such that $T_\varepsilon \overset{\star}{\rightharpoonup} T$.*

Proof. The approximating joint probability measures T_ε can be chosen as

$$T_\varepsilon(x, y) := T(x + \varepsilon \tanh(y), y),$$

where tanh denotes the hyperbolic tangent, but could be replaced by any smooth increasing function which is bounded from above and below and is zero at zero.

Define $\psi_\varepsilon(x) := (\operatorname{supp} T_\varepsilon) \cap (\{x\} \times \mathbb{R})$. Suppose that ψ_ε is not a function, then it must be set-valued, i.e. there exists some $x \in \mathbb{R}$ with $\operatorname{card}(\psi_\varepsilon(x)) > 1$. Take $y_1, y_2 \in \psi_\varepsilon(x)$ with $y_1 > y_2$ then $x_1 := x - \varepsilon \tanh(y_1) > x_2 := x - \varepsilon \tanh(y_2)$ and hence the points (x_1, y_1) and (x_2, y_2) violate the co-monotonicity condition.

It remains to prove that $T_\varepsilon \overset{\star}{\rightharpoonup} T$ as $\varepsilon \to 0$: let $\phi \in C^1(\mathbb{R}^2)$ be a bounded function. Then we have

$$\begin{aligned}
&\left| \int_{\mathbb{R}^2} \phi(x, y)\, T_\varepsilon(x, y) - \int_{\mathbb{R}^2} \phi(x, y)\, T(x, y) \right| \\
= \;& \left| \int_{\mathbb{R}^2} (\phi(x, y) - \phi(x + \varepsilon \tanh(y), y))\ T(x, y) \right| \\
\leq \;& \int_{\mathbb{R}^2} \left| \phi(x, y) - (\phi(x, y) + \varepsilon \tanh(y) |\nabla \phi(x, y)|) \right| T(x, y) \\
\leq \;& \int_{\mathbb{R}^2} \varepsilon \sup |\tanh(y)|\, \|\nabla \phi\|_\infty T(x, y) \;=\; \|\nabla \phi\|_\infty \varepsilon,
\end{aligned}$$

which converges to zero as $\varepsilon \to 0$. Using a standard approximation argument for ϕ, we obtain $T_\varepsilon \overset{\star}{\rightharpoonup} T$. □

2.2 Optimal investments – the general case

We have now all mathematical tools at hand to study co-monotonicity of financial products. We first define the optimization problem we want to study:

Definition 2.11 (Optimal financial products)**.** *Let $T \in \mathcal{P}(\mathbb{R}, \mathbb{R})$ be a joint probability measure with the marginals $\mathrm{pr}_1 T = p$ and $\mathrm{pr}_2 T = \pi$, where p is the return distribution of a financial product and π the state price density. Let $U : \mathcal{P}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ be a function that assigns to every joint probability distribution a utility. Then we call T* optimal *if it maximizes $U : \mathcal{P}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ under the no-arbitrage condition (1).*

In the following we restrict the class of admissible utility functions. The main underlying assumptions are a "positive attitude" regarding additional returns and that there is no background risk involved into the investment decision. These two assumptions are made rigorous in the following definition:

Definition 2.12 (Admissible utility functions)**.** *We call $U : \mathcal{P}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$* admissible *if the following two conditions hold:*

(i) $U(T) \le U(T(\cdot, \cdot - c))$ for all $c > 0$,

(ii) There is a non-decreasing function $h : \mathbb{R} \to \mathbb{R}$ and a function $\tilde{U} : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ such that $U(T) = \tilde{U}(p_h)$, where $p_h(y) := \mathrm{pr}_1 T(x, y + h(x)) = \int T(x, y + h(x))\, dy$.

Condition (i) can be summarized as "more money is better": if a financial product is described by T, then an alternative product that is only different in that it yields an *additional* sure return of $c > 0$, is always at least weakly preferred. Condition (ii) seems at first glance quite technical, however, it becomes very natural, once we have seen how our main result (Theorem 2.13) fails if (ii) is violated, see Remark 2.14 below. In Section 2.4 and Section 2.5 we will moreover discuss two important special cases of (ii): in the first case $h \equiv 0$ and hence the utility depends only on the return distribution, in the second case $h(y) = y$ and hence the utility depends on the difference between return and state price density.

We formulate our main result:

Theorem 2.13 (Co-monotonicity of general optimal financial products)**.** *Let $T \in \mathcal{P}(\mathbb{R}, \mathbb{R})$ be a joint probability measure describing a financial product, where the marginals of T, i.e. $p := \mathrm{pr}_1 T$ and $\pi := \mathrm{pr}_2 T$ are the return distribution and the state price density. Let T be optimal with respect to an admissible utility function (compare Def. 2.11 and Def. 2.12).*

Then T is co-monotone with the inverted state price density $\tilde{\pi}$, i.e. $\hat{T}$ defined by $\hat{T}(x, y) := T(-x, y)$ is co-monotone.

It is important to notice here that we have made no assumptions on the precise form of T and its regularity. T could for instance be an absolutely continuous measure, i.e. a probability distribution, or it could be a finite weighted sum of Dirac measures, but it does not need to be: it could be a combination of both or generally *any* probability measure. Moreover, we have made no precise assumptions on the decision model underlying the utility function U. In fact, it is easy to check that standard decision models as Expected Utility Theory, Prospect Theory and Mean-Variance Theory could all be used.

Let us now see why the second condition for an admissible utility was so important:

Remark 2.14. *If condition (ii) of Def. 2.12 is violated, the utility U could be chosen such that joint probability measures which fail to satisfy co-monotonicity have particularly large utility: as a trivial example, we simply define $U(T) = 1$ for all T which are not co-monotone with $\tilde{\pi}$, and $U(T) = 0$ otherwise. This would obviously satisfy condition (i), but an optimal product for U could not be co-monotone with $\tilde{\pi}$.*

We will now turn our attention to the general proof of Theorem 2.13, before we discuss important special cases of this result in more detail in the following sections. The main idea of the proof is to apply the general Theorem 2.9 to prove that the covariance of T is maximized when T is co-monotone. For given p we can then monotonize T in a way which leaves the utility unchanged, but at the same time decreases the price of the product according to the no-arbitrage condition (1). The price reduction can then be used to improve the product by adding a sure return. Condition (i) implies that this new product has a larger utility. We conclude that an optimal T is co-monotone. To make this idea work for the general case of condition (ii) will be the main task of the proof:

Proof. Let $p \in \mathcal{P}$ and let $T \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ be a joint probability measure with marginals p and $\tilde{\pi}$ that maximizes the covariance, i.e.

$$\operatorname{cov} T = \max\{\operatorname{cov} T | T \in \mathcal{P}(\mathbb{R} \times \mathbb{R}),\ \operatorname{pr}_1 T = \mathbb{E}(p),\ \operatorname{pr}_2 T = \mathbb{E}(\tilde{\pi})\}. \tag{6}$$

We prove that T is co-monotone: first, we reformulate the problem (6) as a transport problem, i.e. we want to find $T \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ minimizing

$$C(T) := \int_{\mathbb{R}} \int_{\mathbb{R}} c(x, y)\, dT(x, y),$$

such that $\operatorname{pr}_1 T = p$, $\operatorname{pr}_2 T = \tilde{\pi}$, where

$$c(x, y) := -(x - \mathbb{E}(p))(y - \mathbb{E}(\tilde{\pi})). \tag{7}$$

From Theorem 2.9 we know that T is co-monotone if for all $x_1 < x_2$ and $y_1 < y_2$

$$c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1). \tag{8}$$

We prove that c defined as in (8) satisfies this inequality: without loss of generality we can assume that $\mathbb{E}(p) = \mathbb{E}(\tilde{\pi}) = 0$. Hence it is sufficient to check condition (8) for $c(x, y) = -xy$.

If $x_1 < x_2$ and $y_1 < y_2$ we have $(x_1 - x_2)(y_2 - y_1) < 0$. A short computation gives

$$\begin{aligned} 0 &> (x_1 - x_2)(y_2 - y_1) \\ &= -x_1 y_1 + x_1 y_2 - x_2 y_2 + x_2 y_1 \\ &= c(x_1, y_1) + c(x_2, y_2) - c(x_1, y_2) - c(x_2, y_1). \end{aligned}$$

Therefore condition (8) holds, and we can apply Theorem 2.9 to prove the co-monotonicity of T.

In other words: any joint probability measure that maximizes the covariance, given its marginals, is co-monotone.

Suppose now T is a joint probability measure that maximizes the utility, but is not co-monotone. We compute, denoting $T^h(x,y) := T(x, y + h(x))$:

$$\begin{aligned}\operatorname{cov} T &= \int_{\mathbb{R}}\int_{\mathbb{R}} (x - \mathbb{E}(\tilde{\pi}))(y - \mathbb{E}(p))\, dT(x,y) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} (x - \mathbb{E}(\tilde{\pi}))(y + h(x) - \mathbb{E}(p))\, dT^h(x,y) \\ &= \operatorname{cov} T^h + \int_{\mathbb{R}}\int_{\mathbb{R}} (x - \mathbb{E}(\tilde{\pi}))h(x)\, dT^h(x,y) \\ &= \operatorname{cov} T^h + \int_{\mathbb{R}} (x - \mathbb{E}(\tilde{\pi}))h(x)\, d\tilde{\pi}(x).\end{aligned}$$

We use $p^h = \mathrm{pr}_1 T^h$ and $\mathrm{pr}_2 T = \mathrm{pr}_2 T^h = \tilde{\pi}$. We can maximize $\operatorname{cov} T$ without changing the marginals of T^h by maximizing $\operatorname{cov} T^h$, since only the first term in the above equation depends on T^h, whereas the second term only depends on $\tilde{\pi}$. Applying our above derivation, we see that $\operatorname{cov} T^h$ can be maximized by monotonizing T^h. We call the resulting co-monotone joint probability measure $\tilde{T}^h$ and denote $\tilde{T}^h(x, y - h(x))$ by $\tilde{T}$. Since $U(T)$ depends only on $\mathrm{pr}_1 T^h$, the utility is unchanged, i.e. $U(T) = U(\tilde{T})$. The covariance, however, has increased, i.e. $\operatorname{cov} \tilde{T} > \operatorname{cov} T$, since otherwise T^h would have been already co-monotone, but then T would have been co-monotone as well, since h is by assumption a monotone function. We define

$$d := R + \frac{\mathbb{E}(\tilde{\pi}) - R}{\operatorname{var}(\tilde{\pi})} \operatorname{cov} \tilde{T} - \mathbb{E}(p).$$

Since p satisfies the no-arbitrage condition (1), we have

$$R + \frac{\mathbb{E}(\tilde{\pi}) - R}{\operatorname{var}(\tilde{\pi})} \operatorname{cov} T - \mathbb{E}(p) = 0$$

and therefore $d > 0$. Now define a new product $S \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ by

$$S(x,y) := \tilde{T}(x, y - d).$$

Then S satisfies the no-arbitrage condition (1) and its utility is by assumption (i) larger than the utility of $\tilde{T}$. Hence $U(S) > U(\tilde{T}) = U(T)$ and therefore T cannot be an optimal financial product. Thus every maximizer has to be co-monotone. □

2.3 The special case of CAPM

A simplifying assumption on financial markets which is often done in applications is to equal the inverted state price density with the market return. This is based on the Capital Asset Pricing Model (CAPM) which follows from the Mean-Variance approach introduced by Markowitz [11]. Its fundamental assumption is that every investor in the market selects his portfolio according to the Mean-Variance preferences, i.e. considers only mean and variance of the assets. In such a market every investor would hold only assets of a market portfolio and the risk free asset as the Two-Fund Separation Theorem shows.

Although in his work Markowitz recommended the use of a mean-variance rule, "...both as a hypothesis to explain well-established investment behavior and as a maxim to guide one's own action" [11], it is in general not a rational decision model as we can see, e.g., by considering the Mean-Variance Paradox. The suggestion to use Mean-Variance as a descriptive model for market behavior seems more promising: many practitioners are following this approach to some extend and therefore CAPM might work as a descriptive model of a real financial market. The natural question is now, how to invest in such a market in order to maximize a utility function that is not necessarily of Mean-Variance type, but, for instance, follows Expected Utility Theory. Obviously, the Two-Fund Separation Theorem will not hold in this case, but can we find some other general results describing optimal investments?

Based on the general results of the previous sections, we can at first state the following variant of Theorem 2.13:

Proposition 2.15 (Co-monotonicity in CAPM markets)**.** *Every financial product on a CAPM market which is optimal for an arbitrary admissible utility has a return distribution that is co-monotone with the market return.*

Proof. To prove this, we just note that the no-arbitrage condition (1) in the case of a CAPM market becomes

$$\mathbb{E}(p) - R \quad = \quad \beta_{pm}(\mathbb{E}(m) - R), \tag{9}$$

where $\beta_{pm} = \operatorname{cov}(p, m)/\operatorname{var}(m)$ and m is the market return. The market return takes therefore the role of the inverted state price density in Theorem 2.13. □

We can now extend the Two-Fund Separation to the case of arbitrary admissible decision models. Note that the Two-Fund Separation Theorem implies that every product that is optimal in the Mean-Variance framework, has a return (adjusted by a constant depending on the risk-free rate) which depends linearly on the market return. A product which is optimal for an arbitrary admissible decision model does not necessarily have a return depending linearly on the market return, but its return depends monotone on the market return. In other words, we have the following result:

Theorem 2.16 (Generalized "Two-Fund-Separation")**.** *Consider a CAPM market. An optimal investment for an investor with admissible utility gives a return which is co-monotone with the market return. If the joint probability measure can be described by a function R_p, this function is monotone. If the investor is Mean-Variance maximizer, this function is affine and $R_p(x) = (1 - \lambda)R + \lambda x$ with $\lambda \in \mathbb{R}$.*

Proof. The first part of this result is a reformulation of Proposition 2.15.

The second part simply follows from Two-Fund Separation: every optimal product has a return of the form $\lambda x + (1 - \lambda)R$, where x is the market return, R the risk-free rate and $\lambda \in \mathbb{R}$ the shares invested into the market portfolio. □

An immediate consequence of this result is that an optimal financial product should never speculate on falling prices, since this would violate co-monotonicity.

At this point it is important to keep in mind our two fundamental assumptions: homogenous beliefs and no background risk: in a market situation where an investor has an information advantage he might have reason to believe in falling prices and therefore it might be very wise for him to speculate on them. If beliefs are homogenous, however, there is no reason in a CAPM market to invest in a way that is not co-monotone. Similarly, an investor might need to hedge a background risk and is therefore investing into an additional investment which is not co-monotone. On the other hand, even when we assume homogenous beliefs and no background risk, there might still be many reasons for a non-Mean-Variance investor to violate the Two-Fund Separation and instead to construct rather complex portfolios taylor-made for his preferences.

2.4 Performance based on the outcome

In this section we study an important special case of the general results of Section 2.2, namely when utility is only based on the outcome. We call an investor with such a utility a "private investor", since prototypical private investors would fall into this category.

We start with a precise definition what we understand by a "private investor":

Definition 2.17 (Private investor)**.** *A* private investor *is described by a utility functional U satisfying the following conditions:*

1. *The utility functional depends only on the return distribution of the investment, i.e. $U = U(p)\colon \mathcal{P}(\mathbb{R}) \to \mathbb{R}$. ("Only the result matters.")*
2. *If we shift the return distribution to the right, the utility increases, i.e. if the return distribution is given by $p \in \mathcal{P}(\mathbb{R})$ and $p_c := p(\cdot - c)$ for $c > 0$ then $U(p) < U(p_c)$. ("The more, the better.")*

It is easy to see that a private investor's utility function is admissible, we just have to set $h := 0$ in Def. 2.12. Therefore, we can apply Theorem 2.13 and obtain the following result:

Proposition 2.18. *An optimal financial product for a private investor is co-monotone with the inverted state price density.*

This implies, e.g., that for all private investors in a CAPM market the results of the previous section, in particular Proposition 2.15, apply, i.e. an optimal investment should "follow the market".

At this point it might be worthwhile to discuss how these results relate to the usual portfolio optimization strategy that tries to identify investments which are uncorrelated with the market return to improve the overall performance of the portfolio. Did we prove that such strategies have a substantial flaw? This is of course not the case for the following reasons:

- Our first assumption implies the absence of background risk. This means that we consider the *overall investment* of a person, rather than an additional position that he might or might not add to his portfolio. Our results do not say anything about the structure of such additional positions. It might even be useful for the investor to take an additional position into his portfolio which is anti-co-monotone with the market (e.g. by going short in an asset) in order to hedge a certain risk induced by a different part of his portfolio.

- Another misunderstanding may arise from the word "market". This means of course the entirety of all possible investments, not only stocks. In particular, an investment which is uncorrelated with the stock market is usually still correlated with the "market" in this general sense.

- We have assumed that beliefs are homogenous and (to derive Prop. 2.15) that the market can be described by the CAPM model. In reality we might profit from anomalies of the market that are not described by the models of classical finance. For such situations, Prop. 2.15 is not applicable.

Although these limitations set a caveat on applications of our results, co-monotonicity with the inverted state prices (or the market return, if we can describe the market by CAPM) should still hold in practice if we do not aim to exploit market anomalies and if we consider our investment portfolio as a whole.

In the next section we will see that this is even the case if we do not think in absolute returns, but instead measure returns with respect to a benchmark index.

2.5 Performance based on a benchmark

Let us now consider the somehow opposite case of a private investor: an investor whose utility only depends on the return of his investment *relative to the state price*. In the case of a CAPM market, this would correspond to an investor who is only interested in the excess return of his investment compared to the market return. (Since this is the practical relevant case, we assume in this Section a CAPM market.)

We call such an investor a "fund manager", imagining a fund manager who is paid depending on the performance of his fund *with respect to the market return*. More precisely, we define for the case of a CAPM market:

Definition 2.19 (Fund manager)**.** *A* fund manager *is described by a utility U satisfying the following conditions:*

1. *The utility depends only on the difference between the return distribution of the investment and the market, i.e.* $U = U(p_m)\colon \mathcal{P}(\mathbb{R}) \to \mathbb{R}$*, where* $p_m := \mathrm{pr}_2\, T(x, y - x)$*. (In other words, the market index is used as benchmark.)*
2. *If we shift the return distribution to the right, the utility increases, i.e. if the return distribution is given by* $p \in \mathcal{P}(\mathbb{R})$ *and* $p_c := p(\cdot - c)$ *for* $c > 0$ *then* $U(p - m) < U(p_c - m)$*.*

We can now easily see that this is just another special case of the general Theorem 2.13, in fact a fund manager's utility is admissible, we just have to choose h as identity in Def. 2.12. Therefore, we obtain:

Proposition 2.20. *In a CAPM market, an optimal portfolio for a fund manager has a return distribution that is co-monotone with the market return.*

There is another interesting consequence of the assumptions on the preferences of a fund manager:

Proposition 2.21 (Benchmarking leads to risky products)**.** *An optimal product for a fund manager in a CAPM market is at least as risky as the market portfolio, i.e. the difference between the return distribution and the market return is a non-decreasing function of the return.*

Proof. This follows immediately from the co-monotonicity of the optimal T^h in the proof of Theorem 2.13. □

In the case of an investor with expected utility preferences with respect to the market return, i.e. an utility of the form

$$U = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(x - y)\, dT(x, y), \tag{10}$$

we can strengthen the last proposition:

Proposition 2.22. *A product that maximizes the utility function U in (10) with $u \in C^1$, $u'(0) > 0$ in a CAPM market must be riskier than the market portfolio, i.e. the difference between the return of the product and the market return is an non-decreasing and non-constant function of the market return.*

Proof. Due to the co-monotonicity of any optimal joint probability distribution and Lemma 2.10, it is again sufficient to consider the case where $\operatorname{supp} T$ is of the form $\{(x, R_p(x))\}$.

By Prop. 2.21 we know already that $R_p(x) - x$ must be non-decreasing. We want to show that it $R_p(x) - x$ is non-constant. The pricing constraint implies that $R_p(x) - x$ could only be constant if $R_p(x) = x$, i.e. if the optimal product were the market portfolio itself. Let us suppose that $R_p(x) = x$ is optimal. Our goal is to construct an improved product that satisfies the pricing constraint and has a higher utility for the fund manager.

For simplicity, we assume $u(0) = 0$. Moreover, we assume that m is absolutely continuous. (Otherwise we could approximate m by a sequence of absolutely continuous measures.)

Let $\varepsilon \geq 0$ and define

$$R_{p_\varepsilon}(x) := \begin{cases} x - \varepsilon + \delta, & x < m_0 \\ x + \varepsilon + \delta, & x \geq m_0, \end{cases}$$

where m_0 is defined by

$$\int_{-\infty}^{m_0} dm(x) = \int_{m_0}^{+\infty} dm(x) = \frac{1}{2}$$

and $\delta > 0$ is given by the no-arbitrage condition and will be computed below. (δ is in some sense the "risk-premium" that we get for taking the additional risk expressed by ε.)

Let p_ε be the joint probability measure induced by R_{p_ε}. Then its mean value is given by

$$\begin{aligned} \mathbb{E}(p_\varepsilon) &= \int_{-\infty}^{+\infty} R_{p_\varepsilon}(x)\, dm(x) \\ &= \int_{-\infty}^{+\infty} x + \delta\, dm(x) + \varepsilon \int_{m_0}^{+\infty} dm(x) - \varepsilon \int_{-\infty}^{m_0} dm(x) \\ &= \mathbb{E}(m) + \delta. \end{aligned}$$

Therefore the covariance of the co-monotone joint probability measure with marginals m and p_ε can be computed as follows:

$$\begin{aligned}
\operatorname{cov}(p_\varepsilon, m) &= \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(x - \varepsilon + \delta - \mathbb{E}(p_\varepsilon))\, dm(x) \\
&\quad + \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(x + \varepsilon + \delta - \mathbb{E}(p_\varepsilon))\, dm(x) \\
&= \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(x - \mathbb{E}(m) - \varepsilon + \delta - \mathbb{E}(p_\varepsilon) + \mathbb{E}(m))\, dm(x) \\
&\quad + \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(x - \mathbb{E}(m) + \varepsilon + \delta - \mathbb{E}(p_\varepsilon) + \mathbb{E}(m))\, dm(x) \\
&= \operatorname{var}(m) + \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(-\varepsilon + \delta - \mathbb{E}(p_\varepsilon) + \mathbb{E}(m))\, dm(x) \\
&\quad + \int_{-\infty}^{m_0} (x - \mathbb{E}(m))(\varepsilon + \delta - \mathbb{E}(p_\varepsilon) + \mathbb{E}(m))\, dm(x).
\end{aligned}$$

We insert this and the formula for $\mathbb{E}(p_\varepsilon)$ into the no-arbitrage condition and obtain

$$\begin{aligned}
\mathbb{E}(m) + \delta - R &= (\mathbb{E}(m) - R)\Bigg(1 + \frac{1}{\operatorname{var}(m)} \int_{-\infty}^{m_0} -(x - \mathbb{E}(m))\varepsilon\, dm(x) \\
&\qquad + \frac{1}{\operatorname{var}(m)} \int_{-\infty}^{m_0} (x - \mathbb{E}(m))\varepsilon\, dm(x)\Bigg) \\
&= \frac{\mathbb{E}(m) - R}{\operatorname{var}(m)} \left(\operatorname{var}(m) + \int_{-\infty}^{m_0} x\, dm(x) - \int_{m_0}^{+\infty} x\, dm(x)\right)\varepsilon.
\end{aligned}$$

We can resolve this to obtain a formula for δ:

$$\delta(\varepsilon) = \frac{\mathbb{E}(m) - R}{\operatorname{var}(m)} \left(\int_{-\infty}^{m_0} x\, dm(x) - \int_{m_0}^{+\infty} x\, dm(x)\right)\varepsilon.$$

We see from this that $\delta(0) = 0$ and $\delta'(0) > 0$. We use a Taylor expansion to compute the utility difference of p and p_ε:

$$\begin{aligned}
U(p_\varepsilon) - U(p) &= \frac{1}{2}u(\delta - \varepsilon) + \frac{1}{2}u(\delta + \varepsilon) \\
&= u'(0)\delta(\varepsilon) + O((\delta(\varepsilon) - \varepsilon)^2, (\delta(\varepsilon) + \varepsilon)^2) \\
&= u'(0)(\delta'(0)\varepsilon) + O(\varepsilon^2).
\end{aligned}$$

Therefore, for $\varepsilon > 0$ sufficiently small, this difference is positive, i.e. $U(p_\varepsilon) > U(p)$ which shows that p defined by $R_p(x) = x$ cannot be optimal. $\square$

We want to stress that a crucial condition for the derivation of this result was the differentiability of u at zero. If we replace the utility function u by a value function that has a kink at zero ("loss aversion") as suggested, e.g., by Kahneman and Tversky [7], our result does not apply, and it is conceivable that an investor would indeed stick exactly to the market portfolio.[3]

3 Designing optimal financial products

The usual way in which financial investments are optimized is by finding an optimal allocation between several assets. This is the obvious "bottom-up" approach which starts from the constituents and builds portfolios and finally financial products based on them. The alternative "top-down" approach is first to find optimal financial products and then to design the underlying structure of assets from which the product can be built.

Obviously, there are advantages to the latter method, in that it is independent of the "educated guess" that would lead to a pre-selection of assets in the asset allocation approach. The difficulty with the alternative approach, on the other hand, is two-fold: first, the design of the asset allocation is more difficult, and second it is at first glance not clear how to find an optimal financial product or even how to define what this would be.[4]

Applying the results of the previous section, we can now solve these problems to some extend which may open a new route to the design of financial products. Our key observation so far was that optimal investments are co-monotone with the inverted state price density $\tilde{\pi}$ (or, if we want to use the CAPM model, with the market return m). We have seen that this result holds quite generally which makes now our task of finding optimal investments a lot easier: we know from Prop. 2.7 that a co-monotone joint probability measure is unique. Therefore, it is sufficient

[3]This effect might explain the popularity of index funds that would otherwise be in contradiction to our above result.

[4]This approach is similar to the martingale method in continuous-time market models, where first an optimal payoff-/consumption-stream is found and then the replicating portfolio process is computed, compare [9, Chapter 5] for an overview.

to optimize the return distribution p, assuming that the joint probability measure T is the unique co-monotone joint probability measure with marginals p and $\tilde{\pi}$. We are therefore only left with maximizing the utility over all probability measures that satisfy the no-arbitrage condition. We will outline in Section 3.2 how this could be done numerically, but before that, in Section 3.1, we study existence and some properties of solutions for the resulting optimization problem. In Section 3.3 we finally consider some cases where existence fails.

3.1 Existence of optimal financial products

In this section we mainly deal with rational investors in the sense of von Neumann and Morgenstern [17], i.e. we assume that the utility U can be expressed by

$$U(T) = \int_0^\infty u(x)\,dp(x),$$

where x is here the final wealth of the investor who invests in a product with the return p, and u is the von Neumann-Morgenstern utility function. We assume that u is continuous and increasing.

From now on, we implicitly assume that, for given p and π, T is the joint probability measure with marginals p and $\tilde{\pi}$ which is co-monotone with $\tilde{\pi}$. Thus we can define the *maximum covariance* between p and $\tilde{\pi}$ as

$$\operatorname{mcov}(p, \tilde{\pi}) := \operatorname{cov} T.$$

Our optimization problem can now be stated as finding the $p \in \mathcal{P}$ that maximizes

$$U(p) := \int_0^\infty u(x)\,dp(x),$$

subject to

$$\mathbb{E}(p) - R = \frac{\operatorname{mcov}(p, \tilde{\pi})}{\operatorname{var}(\tilde{\pi})}(\mathbb{E}(\tilde{\pi}) - R). \tag{11}$$

We formulate now the following existence result:

Theorem 3.1 (Existence of optimal financial products)**.** *Let the preferences of the investor be given by Expected Utility Theory with utility function u. Assume that u is continuous, increasing and sublinear, i.e. that $u(x)/x \to 0$ as $x \to \infty$. Assume furthermore that the state price density $\tilde{\pi}$ vanishes outside the interval $[0, M]$ and that $\mathrm{var}(\pi) > M(\mathbb{E}(\tilde{\pi}) - R)$. Then there exists an optimal financial product, i.e. the above problem admits a maximizer $p \in \mathcal{P}$. Moreover, $\mathbb{E}(p) < \infty$.*

We will see in Section 3.3 that some of the conditions made in this theorem are indeed necessary.

Proof. The proof consists of the following steps:

1. There is a constant $C > 0$ such that all $p \in \mathcal{P}$ that satisfy (11) fulfill $\mathbb{E}(p) \leq C$.
2. For every sequence (p_n) satisfying (11) there exists a $p \in \mathcal{P}$ such that a subsequence $(p_{n'})$ of (p_n) converges weakly-$\star$ to p, i.e. $p_{n'} \overset{\star}{\rightharpoonup} p$.
3. The supremum of $U(p)$ over all $p \in \mathcal{P}$ that satisfy (11) is finite.
4. The maximization problem admits a maximizer.

Step 1:
First, we use the approximation result, Lemma 2.10, and assume therefore without loss of generality that all joint probability measures can be expressed by functions. We denote the function corresponding to a return p by R_p.

We observe the following useful identity:

$$\mathrm{mcov}(p, \tilde{\pi}) = \int_0^\infty (x - \mathbb{E}(\tilde{\pi}))(R_p(x) - \mathbb{E}(p))\, d\pi(x) = \int_0^\infty x R_p(x)\, d\pi(x) - \mathbb{E}(\tilde{\pi}).$$

To simplify notation, we write $\sigma^2 := \mathrm{var}(\tilde{\pi})$. Let $p \in \mathcal{P}$ and assume that p satisfies the no-arbitrage condition (11). We want to obtain an estimate on $\mathbb{E}(p)$ using the estimate on the support of $\tilde{\pi}$:

$$\begin{aligned}
\mathbb{E}(p) &= R + \frac{\mathrm{mcov}(p, \tilde{\pi})}{\sigma^2}(\mathbb{E}(\tilde{\pi}) - R) \\
&= R + \frac{\int_0^\infty x R_p(x)\, d\pi(x) - \mathbb{E}(\tilde{\pi})}{\sigma^2}(\mathbb{E}(\tilde{\pi}) - R) \\
&\leq R + M\mathbb{E}(p) - \mathbb{E}(\tilde{\pi})\sigma^2(\mathbb{E}(\tilde{\pi}) - R).
\end{aligned}$$

Resolving this, while using the assumption $\sigma^2 = \mathrm{var}(\pi) > M(\mathbb{E}(\tilde{\pi}) - R)$, we get

$$\mathbb{E}(p) \leq \frac{R - \frac{\mathbb{E}(\tilde{\pi})}{\sigma^2}(\mathbb{E}(\tilde{\pi}) - R)}{1 - \frac{M}{\sigma^2}} < \infty, \tag{12}$$

thus arriving at the desired uniform bound for p.

Step 2:

Let (p_n) be a sequence of probability measures satisfying (11). We want to prove that we can select a subsequence $(p_{n'})$ which is converging weakly-$\star$ to a probability measure p. By Prokhorov's Theorem [14], it is sufficient to prove that (p_n) is tight, i.e. that for all $\eta > 0$ there is a compact subset K_η of $\mathbb{R}_+$ such that $p_n(K_\eta) > 1 - \eta$.

Let us suppose that (p_n) is not tight. Then there exists for all $L > 0$ an $n_0(L) \in \mathbb{N}$ such that $p_{n_0(L)}((L, +\infty)) > \eta_0$.

Our strategy is now to estimate $\mathrm{mcov}(p_n, \tilde{\pi})$ from below to show that under this assumption, it would diverge. This would then imply, via the no-arbitrage condition (11), that also $\mathbb{E}(p_n)$ diverges, in contradiction to the uniform bound we have derived in step 1.

To this aim, we define M_1,M_2 such that $(x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n)) =: h(x)$ is positive on $(0, M_1)$ and (M_2, M) and negative on (M_1, M_2). This is possible since $x - \mathbb{E}(\tilde{\pi})$ and $R_{p_n}(x) - \mathbb{E}(p_n)$ are non-decreasing functions with sign changes in $\mathbb{E}(\tilde{\pi})$ and $r := R^{-1}_{p_n}(\mathbb{E}(p_n))$, respectively. Then we estimate, using the bound on the support of $\tilde{\pi}$:

$$\begin{aligned}
\mathrm{mcov}(p_n, \tilde{\pi}) &= \int_0^M (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\
&= \int_0^{M_1} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\
&\quad + \int_{M_1}^{M_2} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\
&= \int_{M_2}^{M} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\
&=: I_0 + I_1 + I_2.
\end{aligned}$$

We have to distinguish two cases, depending on whether or not $r \geq \mathbb{E}(\tilde{\pi})$:

CASE A: $r \geq \mathbb{E}(\tilde{\pi})$, i.e. $M_1 = \mathbb{E}(\tilde{\pi})$, $M_2 = r$.

In this case, we have:

$$\begin{aligned} I_0 + I_1 &\geq \int_0^{M_1} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(\mathbb{E}(\tilde{\pi})) - \mathbb{E}(p_n))\, d\pi(x) \\ &\quad + \int_{M_1}^{M_2} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(\mathbb{E}(\tilde{\pi})) - \mathbb{E}(p_n))\, d\pi(x) \;\geq\; 0. \end{aligned}$$

Now we use the assumption that (p_n) is not tight. It implies for all $L > R_{p_n}(r)$ and $n \geq n_0(L)$:

$$\begin{aligned} I_2 &= \int_{M_2}^{M} (x - \mathbb{E}(\tilde{\pi})(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\ &\geq \int_{R_{p_n}^{-1}(L)}^{M} (R_{p_n}(L) - \mathbb{E}(\tilde{\pi}))(L - \mathbb{E}(p_n))\, d\pi(x) \\ &\geq \int_{R_{p_n}^{-1}(L)}^{M} \left(R^{-1}_{p_{n_0(L)}}(L) - \mathbb{E}(\tilde{\pi})\right)(L - \mathbb{E}(p_n))\, d\pi(x) \\ &= p_{n_0(L)}((L, +\infty)) \left(R^{-1}_{p_{n_0(L)}}(L) - \mathbb{E}(\tilde{\pi})\right)(L - \mathbb{E}(p_n)) \\ &> \text{const.}(L - \mathbb{E}(p_n)) \;\to\; +\infty, \text{ as } L \to \infty. \end{aligned}$$

Taking both estimates together, we have proved in this case that $\operatorname{mcov}(p_n, m) \to +\infty$ as $n \to \infty$.

CASE B: $r < \mathbb{E}(\tilde{\pi})$, i.e. $M_1 = r$, $M_2 = \mathbb{E}(\tilde{\pi})$.

We decompose the maximum covariance analogously to step A and estimate:

$$\begin{aligned} I_0 + I_1 &= \int_0^r (x - \mathbb{E}(\tilde{\pi}))(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\ &\quad + \int_r^{\mathbb{E}(\tilde{\pi})} (x - \mathbb{E}(\tilde{\pi}))(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\ &\geq \int_0^r (r - \mathbb{E}(\tilde{\pi}))(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\ &\quad + \int_r^{\mathbb{E}(\tilde{\pi})} (r - \mathbb{E}(\tilde{\pi}))(R_{p_n}(x) - \mathbb{E}(p_n))\, d\pi(x) \\ &= (r - \mathbb{E}(\tilde{\pi})) \int_0^{\mathbb{E}(\tilde{\pi})} R_{p_n}(x)\, d\pi(x), \end{aligned}$$

which is positive, since R_{p_n} is non-decreasing, $\mathbb{E}(p) = \int_0^M R_{p_n}(x)\, d\pi(x)$ and hence $\int_0^{\mathbb{E}(\tilde{\pi})} R_{p_n}(x)\, d\pi(x) \leq \mathbb{E}(p) \int_0^{\mathbb{E}(\tilde{\pi})} d\pi(x)$.

For I_2 we can now use essentially the same estimate as in step A which proves that also in this case $\mathrm{mcov}(p_n, \tilde{\pi}) \to \infty$ as $n \to \infty$.

From the no-arbitrage condition (11) we can now immediately see that since $\mathrm{mcov}(p_n, \tilde{\pi}) \to \infty$, also $\mathbb{E}(p_n) \to \infty$ as $n \to \infty$ which contradicts the uniform bound of $\mathbb{E}(p_n)$ in (12).

Therefore (p_n) is tight and we can apply Prokhorov's Theorem to show the existence of a weak-$\star$ limit $p \in \mathcal{P}$ for a subsequence of (p_n).

Step 3:

We denote the concave envelope of the utility function u by u^c. For every $p \in \mathcal{P}$ which satisfies (11) and therefore also (12), we can estimate with the help of Jensen's inequality:

$$U(p) \;\leq\; \int_0^\infty u^c(x)\,dp(x) \;\leq\; u^c(\mathbb{E}(p)) \;\leq\; u^c\left(\frac{R - \frac{\mathbb{E}(\tilde{\pi})}{\sigma^2}(\mathbb{E}(\tilde{\pi}) - R)}{1 - \frac{M}{\sigma^2}}\right) \;<\; +\infty.$$

Due to this uniform bound we can find a maximizing sequence (p_n) of probability measures satisfying (11) and using the results of step 2 we can extract a subsequence that converges weakly-$\star$ to a limit $p \in \mathcal{P}$. It remains to prove that this limit is indeed a solution of our maximization problem. In the remaining part of the proof we write for simplicity (p_n) for the subsequence $(p_{n'})$ of (p_n).

Step 4:

Here we use the sublinearity of u and estimate for arbitrary $L > 0$:

$$\begin{aligned}\left|\int_0^\infty u(x)\,dp - \int_0^\infty u(x)\,dp_n\right| \;&\leq\; \left|\int_0^L u(x)\,dp - \int_0^L u(x)\,dp_n\right| \\ &\quad + \left|\int_L^\infty \frac{u(x)}{x} x\,dp - \int_L^\infty \frac{u(x)}{x} x\,dp_n\right|.\end{aligned}$$

Whereas the first term converges to zero as $n \to \infty$, since $p_n \overset{\star}{\rightharpoonup} p$, the second can be estimated as follows:

$$\begin{aligned}\left|\int_L^\infty \frac{u(x)}{x} x\,dp - \int_L^\infty \frac{u(x)}{x} x\,dp_n\right| \;&\leq\; \frac{u(L)}{L}\int_L^\infty x\,d(p - p_n) \\ &\leq\; \frac{u(L)}{L}|\mathbb{E}(p) - \mathbb{E}(p_n)|.\end{aligned}$$

Using again estimate (12), the integral is bounded. If we consider the limit

$L \to \infty$, this expression becomes arbitrarily small, therefore, using an appropriate sequence of $L = L(n)$, we have proved that $U(p_n) \to U(p)$.

It only remains to be proved that p satisfies the no-arbitrage condition (11). We first show that

$$\mathbb{E}(p) \leq R + \frac{\mathrm{mcov}(\tilde{\pi}, m)}{\sigma^2}(\mathbb{E}(\tilde{\pi}) - R). \tag{13}$$

To this end, we use the no-arbitrage condition for p_n and (12) and estimate for any sufficiently small $\varepsilon > 0$:

$$\begin{aligned}
\mathbb{E}(p) &= (\mathbb{E}(p) - \mathbb{E}(p_n) + \mathbb{E}(p_n)) \\
&= \int_0^M R_p(x) - R_{p_n}(x)\,d\pi + R \\
&\quad + \frac{\mathbb{E}(\tilde{\pi}) - R}{\sigma^2}\left(\int_0^{M-\varepsilon} (R_{p_n}(x) - \mathbb{E}(p_n))(x - \mathbb{E}(\tilde{\pi}))\,d\pi \right. \\
&\qquad\qquad \left. + \int_{M-\varepsilon}^M (R_{p_n}(x) - \mathbb{E}(p_n))(x - \mathbb{E}(\tilde{\pi}))\,d\pi \right) \\
&\leq \int_0^{M-\varepsilon} R_p(x) - R_{p_n}(x)\,d\pi + R \\
&\quad + \frac{\mathbb{E}(\tilde{\pi}) - R}{\sigma^2} \int_0^{M-\varepsilon} (R_{p_n}(x) - \mathbb{E}(p_n))(x - \mathbb{E}(\tilde{\pi}))\,d\pi \\
&\quad + \int_{M-\varepsilon}^M R_p(x) - R_{p_n}(x)\,d\pi + \int_{M-\varepsilon}^M R_{p_n}(x) - \mathbb{E}(p_n)(x)\,d\pi.
\end{aligned}$$

Since $\mathbb{E}(p_n)$ is uniformly bounded, it converges at least for a subsequence to some constant E. We can therefore pass to the limit as $n \to \infty$ and obtain:

$$\begin{aligned}
\mathbb{E}(p) \leq & \int_{M-\varepsilon}^M R_p(x)\,d\pi + R + \frac{\mathbb{E}(\tilde{\pi}) - R}{\sigma^2}\left(\int_0^{M-\varepsilon} (R_p(x)(x - \mathbb{E}(\tilde{\pi}))\,d\pi \right. \\
& \left. - E \int_0^{M-\varepsilon} (x - \mathbb{E}(\tilde{\pi}))\,d\pi \right) + E \int_{M-\varepsilon}^M d\pi.
\end{aligned}$$

We apply the identity

$$\int_{M-\varepsilon}^M (x - \mathbb{E}(\tilde{\pi}))\,d\pi = -\int_0^{M-\varepsilon} (x - \mathbb{E}(\tilde{\pi}))\,d\pi$$

to derive

$$\begin{aligned}\mathbb{E}(p) \quad \leq \quad & \int_{M-\varepsilon}^{M} R_p(x)\, d\pi + R + \frac{\mathbb{E}(\tilde{\pi}) - R}{\sigma^2}\left(\int_0^{M-\varepsilon} (R_p(x)(x - \mathbb{E}(\tilde{\pi}))\, d\pi\right. \\ & \left. + E\int_{M-\varepsilon}^{M} (x - \mathbb{E}(\tilde{\pi}))\, d\pi\right) + E\int_{M-\varepsilon}^{M} d\pi.\end{aligned}$$

This inequality holds for all sufficiently small $\varepsilon > 0$. We can therefore pass to the limit and obtain, as $\varepsilon \to 0$ the no-arbitrage condition (13).

Now let us suppose this inequality were a *strict* inequality. Then we could improve p by adding a certain outcome for sure while at the same time satisfying the no-arbitrage condition (11) exactly. (We have seen in the proof of Theorem 2.13 how to do this.) This improved product p' would by assumption have a larger utility than p, i.e. $U(p') > U(p)$, but $U(p_n) \to U(p)$ as $n \to \infty$ and p_n was defined as a maximizing sequence for U. Therefore inequality (13) must in fact be an equality and (11) holds for p.

Thus p is indeed a solution of our maximization problem, and we have proved the existence result. □

It is also possible to characterize some features of solutions. In particular, we have the following result:

Proposition 3.2. *If $u'(x) \to 0$ as $x \to \infty$, and $\tilde{\pi}$ satisfies the assumptions from Theorem 3.1, then an optimal financial product has a finite maximum return, i.e.* $\operatorname{supp} p$ *is bounded.*

We briefly sketch the proof:

Proof. Assuming that there exists an optimal $p \in \mathcal{P}$ with unbounded support that can be represented by a function R_p, we define for $\varepsilon > 0$:

$$R_{p'}(x) := \begin{cases} R_p(x) + \delta, & \text{for } x \in (0, M - \varepsilon) \\ R_p(x) - \varepsilon, & \text{for } x \in (M - \varepsilon, M), \end{cases}$$

where δ is chosen such that p' satisfies the no-arbitrage condition (11). We define $\kappa := \int_{M-\varepsilon}^{M} d\pi(x)$. A lengthy, but straightforward computation shows then that $\delta = O(\varepsilon\kappa)$ and that

$$U(p') - U(p) \quad \geq \quad u'(\mathbb{E}(\tilde{\pi}))(1 - \kappa)\delta - u'(R_p(M - \varepsilon))\kappa\varepsilon + O(\varepsilon^2\kappa).$$

This is positive if $\varepsilon > 0$ is chosen small enough, since in this case $R_p(M-\varepsilon) \to +\infty$ and therefore $u'(R_p(M-\varepsilon)) \to 0$. Thus p cannot be optimal. □

What kind of problems may arise for the existence result if one replaces Expected Utility Theory by Prospect Theory? One main difference is the probability weighting that will overweight extreme small probability events. This might compensate for the diminishing marginal utility of large outcomes. It is, e.g., possible to construct a $p \in \mathcal{P}$ with finite expected value but infinite PT-utility. This phenomenon is essentially a new variant of the St. Petersburg Paradox, compare [16] for details.

3.2 Numerical approximation

In order to compute optimal financial products numerically, the existence proof of the previous section can give some rough guidance: the main idea is to optimize in p rather than in T and to compute T for every given p and $\tilde{\pi}$ as the unique co-monotone joint probability measure with marginals p and $\tilde{\pi}$. This approach is much more efficient than an optimization in T, since the number of necessary variables for an approximation is much smaller, as we will see in a moment.

We formulate this method for the case of finitely many states, since in a numerical approximation T would be replaced by a matrix and p and $\tilde{\pi}$ would be approximated by vectors. This corresponds mathematically to approximating p and $\tilde{\pi}$ by sums of weighted Dirac measures:

Let $x_1, \dots, x_N$ be the set of possible outcomes, where $x_1 < x_2 < \cdots < x_N$. We want to find the optimal vector $(p_1, \dots, p_N)$ of probabilities for these outcomes, where $p_i \geq 0$, such that:

(i) The total probability is one: $p_1 + \cdots + p_N = 1$.

(ii) The probability measure $p = \sum_{i=1}^N p_i \delta_{x_i}$ maximizes (among all probability measures of this form) a given utility $U(p)$ subject to the constraint implied by (12).

It is now clear why our approach is more efficient than a direct optimization of T: if we approximate p and $\tilde{\pi}$ by N weighted Dirac measures each, then T is a

$N \times N$-matrix. A direct optimization of T would therefore be an optimization in N^2 rather than in N variables.

However, we also have to compute the co-monotone T (or at least its covariance), given its marginals p and $\tilde{\pi}$ in an efficient way to make this idea working. Such an algorithm could be obtained from the construction of Theorem 2.9 by starting from an arbitrary joint probability measure with given marginals. It is, however, possible to compute the covariance of the co-monotone joint probability measure directly and at the same time more efficiently applying a simple algorithm used in [15] in the context of transport plans:

Set $i = j = 1$, $L = \tilde{\pi}_1$ and C=0.

Then, as long as $i \leq n$ or $j \leq N$ do the following:

- If $L > p_j$ then $L = L - p_j$, $C = C + p_j(x_i - \mathbb{E}(p))(x_j - \mathbb{E}(\tilde{\pi}))$.
- If $L \leq p_j$ then $L = 0$, $C = C + L(x_i - \mathbb{E}(p))(x_j - \mathbb{E}(\tilde{\pi}))$.
- If $L = 0$ then $i = i + 1$, $L = \tilde{\pi}_i$, otherwise $j = j + 1$.

This algorithm terminates since $\sum_{i=1}^N \tilde{\pi}_i = 1 = \sum_{j=1}^N p_j$. The variable C returns the maximum covariance of p and $\tilde{\pi}$, i.e. $C = \mathrm{mcov}(p, \tilde{\pi})$.

Using this algorithm, the constraint (11) can be computed without explicitly knowing the joint probability measure T. The resulting finite constrained maximization problem can be solved with standard algorithms for nonconcave maximization.

This rough sketch of ideas should be sufficient to demonstrate the possibility of solving this optimization problem also in a practical application, but naturally there are still interesting open questions, e.g. regarding the quality of convergence of this approximation.

3.3 Potential non-existence

In this section we come briefly back to the existence theorem of Section 3.1 and sketch an example that demonstrates how one could get non-existence in certain situations.

The source of potential problems is most easily seen when relaxing the constrained of positive final wealth (i.e. optimizing in $\mathcal{P}(\mathbb{R})$ rather than in $\mathcal{P}(\mathbb{R}_+)$). For simplicity, we set

$$\tilde{\pi} := \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_2, \quad R = 0.$$

In this case, we can construct a sequence of probability measures

$$p_n := \frac{1}{2}\delta_{-n+c(n)} + \frac{1}{2}\delta_{n+c(n)},$$

where $c(n)$ is chosen such that p_n satisfies the no-arbitrage condition (11). based on this condition, we compute $c(n)$, where we use that $\mathbb{E}(p_n) = c(n)$, $\mathbb{E}(\tilde{\pi}) = 1/2$, $\mathrm{var}(\tilde{\pi}) = 3/2$ and that $R = 0$:

$$\begin{aligned} c(n) &= \frac{\mathrm{mcov}(p_n, \tilde{\pi})}{\mathrm{var}(\tilde{\pi})}\mathbb{E}(\tilde{\pi}) \\ &= \frac{1}{3}\left((-n + c(n) - c(n))(-3/2) + (n + c(n) - c(n))(3/2)\right) \\ &= n. \end{aligned}$$

Therefore if u is unbounded, $U(p_n) \to +\infty$ as $n \to \infty$, and we have obviously non-existence. But even if u is bounded, but strictly increasing, $U(p_n)$ is strictly increasing for $n \to \infty$, but converges weakly-$\star$ to $p := \frac{1}{2}\delta_0$, thus $U(p) < \lim_{n\to\infty} U(p_n)$. Hence also in this case, an existence proof is not possible.

We see that the fundamental problem is that $\mathrm{mcov}(p_n, \tilde{\pi})$ and $\mathbb{E}(p_n)$ simultaneously tend to infinity. This problematic phenomenon that can, as we have just seen, lead to non-existence, can only be excluded under additional conditions like the one we introduced in our existence result. It is an interesting question for future work to what extend these conditions can be relaxed.

4 Applications to barrier products

Are there any successful financial products on the market that are not co-monotone? In fact there are some examples, and in this section we want to have a look on one of them which became quite popular in the last years, the so-called *down-and-out barrier products*. A barrier product, in its simplest form, is an investment in a certain underlying that guarantees capital protection as long as the prize of the underlying does not fall below a certain threshold, the “barrier level”. In the case of a “down-and-out” barrier product, once the underlying is below the threshold, the capital protection is gone and is also not recovered by future increases above the barrier level. The payoff diagram of such a product at maturity is schematically illustrated in Figure 1.

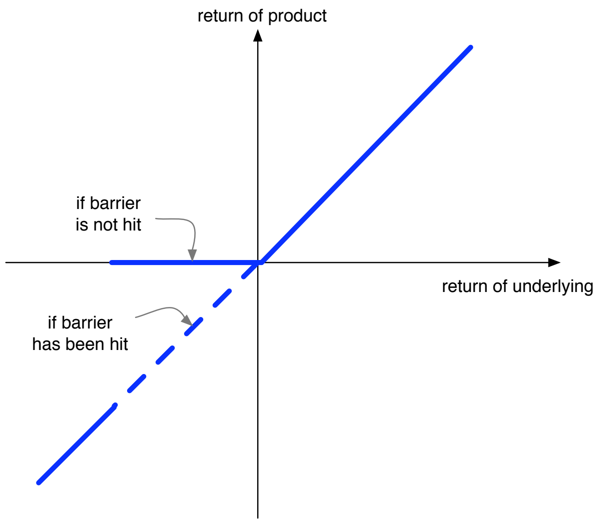

Figure 1: Payoff diagram of a down-and-out barrier product.

The joint probability distribution of this product is obviously not co-monotone if we assume a CAPM market, since its support contains the whole diagonal, but also parts of the x-axis of the above diagram. Therefore, independently of the stochastic process of the underlying, the product is not optimal, neither regarding the zero reference point nor the reference point of the underlying (compare Prop. 2.18 and Prop. 2.20).

To compute the “optimized” variant, e.g. with respect to the zero return, we need to know the stochastic process determining the underlying, in order to calculate the probability distribution of the product. Once this is done, the monotonization could be done numerically with the method of Theorem 2.6. In this special case, however, we can even perform the monotonizing analytically:

Proposition 4.1 (Monotonized barrier product)**.** *Let $T \in \mathcal{P}(\mathbb{R}^2)$ be the joint probability measure of the market return $m \in \mathcal{P}$ and the return $p \in \mathcal{P}$ of the down-and-out barrier product at maturity. Assume for simplicity that m is absolutely continuous. Let B be the barrier level. Let R_p be defined by*

$$R_p(x) := \begin{cases} 0 & \textit{for } x \in [x_0, 0], \\ P^{-1}(M(x)) & \textit{for } x \in [B, x_0), \\ x & \textit{otherwise,} \end{cases}$$

where $x_0 := \sup\{x \in \mathbb{R} \,|\, m([x_0, 0]) \geq T([B, 0) \times \{0\})\}$, $P(x) := p((-\infty, x])$ and $M(x) := m((-\infty, x])$.

Then the alternative product $T_{\mathrm{opt}} := (Id \times R_p)_{\#}m$ that yields the payoff $R_p(x)$ when the market return is x, generates the same return p, i.e. $\mathrm{pr}_2(T_{\mathrm{opt}}) = p$, but has a lower price.

Proof. We need to verify that $\mathrm{pr}_2(T_{\mathrm{opt}}) = p$. It is clear that $(\mathrm{pr}_2(T_{\mathrm{opt}}))(\{0\}) = m([x_0, 0]) = T([B, 0] \times \{0\}) = p(\{0\})$. Hence, we only need to check the condition for $y \in (B, 0)$. There, we have:

$$(\mathrm{pr}_2(T_{\mathrm{opt}}))(y) = \frac{m(R_p^{-1}(y))}{R_p'(R_p^{-1}(y))}.$$

We compute

$$R_p'(x) = \frac{d}{dx}(P^{-1}(M(x))) = (P^{-1})'(M(x)) \cdot m(x),$$

thus

$$(\mathrm{pr}_2(T_{\mathrm{opt}}))(y) = \frac{m(R_p^{-1}(y))}{(P^{-1})'(M(R_p^{-1}(y))) \cdot m(R_p^{-1}(y))}.$$

Now since $(P^{-1})' = 1/P'(P^{-1})$ and $P^{-1}(M(x)) = R_p(x)$, this simplifies to

$$(\mathrm{pr}_2(T_{\mathrm{opt}}))(y) = P'(P^{-1}(M(R_p^{-1}(y))) = p(y).$$

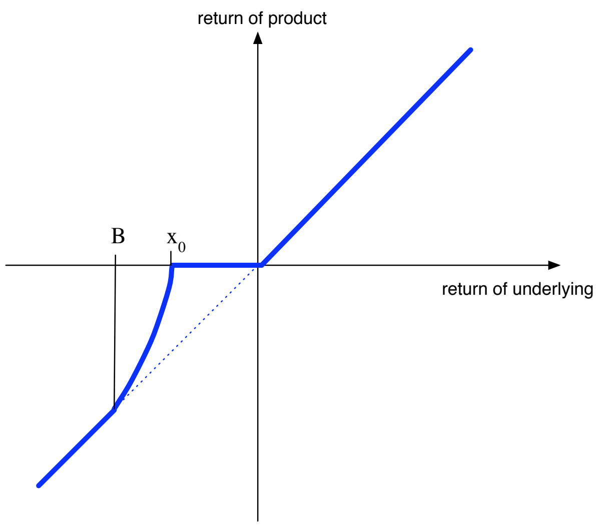

Figure 2: Payoff diagram of the optimized product, yielding the same return distribution for a lower price.

Since T_{opt} is co-monotone by construction (R_p is non-decreasing) and T is not co-monotone (as we have seen), but $U(T_{\text{opt}}) = U(p) = U(T)$, we can apply Prop. 2.18 to see that T cannot be optimal. □

The optimized product (see Figure 2) yields precisely the same return distribution for the customer, but will be cheaper to hedge for the bank.[5] This raises the question why barrier products are nevertheless so successful. What is wrong with our theory in this case?

One possibility is that we assumed here that the market can be described by CAPM. This strong assumption could be the reason for the discrepancy between their theoretical non-optimality of down-and-out barrier products and their high popularity. Another possible explanation is that the optimized product might be more difficult to hedge then the down-and-out barrier product.

The most likely answer, however, might be a very different one which is based on the perception of the customer: it seems likely that customers underestimate

[5]There is no arbitrage opportunity here, since the optimized product will yield a lower return in *some* states!

the probability that the barrier is hit at *some point in time* before maturity with respect to the probability that the prize is below the barrier level *at maturity*. How different are these two probabilities in reality? We measured this using the Dow Jones Industrial Index and assuming an arbitrary issue date between January 1, 1985 and December 31, 2004 and a maturity of one year. Figure 3 shows the probability that the barrier was hit *at some point within a year* and the probability that the prize was below the barrier level *after one year* for various barrier levels. The quotient between these probabilities is depicted in Figure 4. In particular for relatively low barrier levels between 70-80% as they are also quite frequently used in real barrier products, this quotient is quite big.

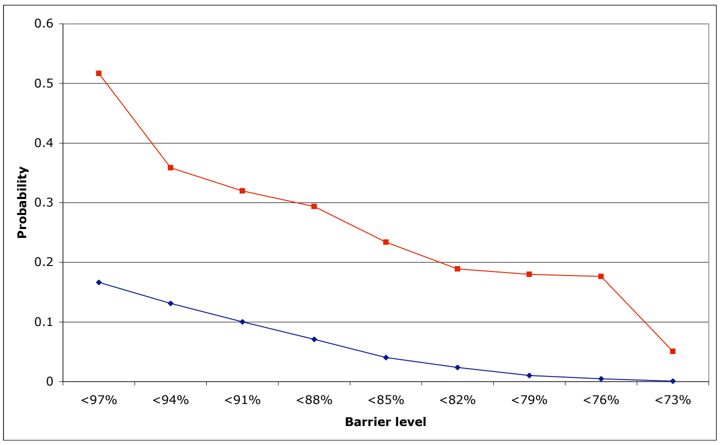

Figure 3: Probability to hit the barrier at maturity (lower line) or at some point before (upper line). The latter corresponds to the probability to lose capital protection in a down-and-out barrier product.

How do customers estimate these probabilities? Do they see the big difference between the two variants or do they underestimate the difference and could this explain why customers choose the (sub-optimal) down-and-out barrier product?

We performed an anonymous classroom experiment with a sample of $N = 109$

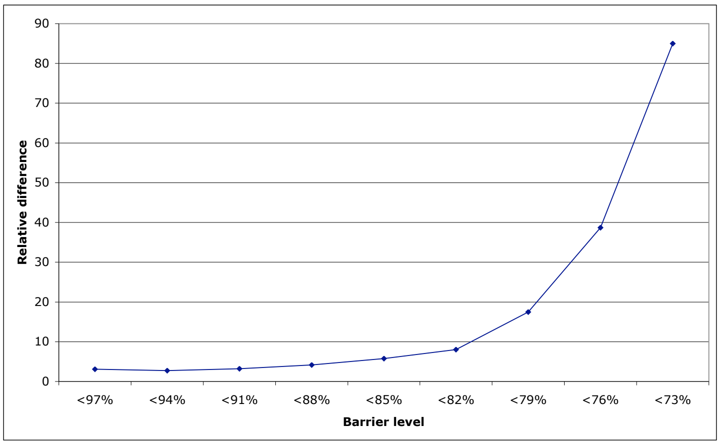

Figure 4: The quotient of the two probabilities is increasing steeply for low barrier levels. This effect is likely to be underestimated by investors.

undergraduate students of economics and finance from the University of Zurich. We asked the students to estimate the probabilities of the events "Dow Jones Index is X% lower after one year" and "Dow Jones Index is X% lower *at some point* during one year" for $X = 10$ and $X = 20$, where we specified that we are asking based on the historical data from 1985 to 2004. In other words, the participants had to estimate the probabilities from Figure 3. The choice of the sample implied a fundamental knowledge on stock market development that was needed to comprehend and perform the task. Moreover we controlled for order effects.

The results demonstrate that although the overall estimation of the probabilities was quite good, the relative difference between the two different probabilities was systematically underestimated. 42% of the participants were not even estimating an increase in this relative difference when increasing X from 10 to 20. This leads to a systematic underestimation of the risks involved in down-and-out barrier products and a systematic over-estimation of the risks involved in barrier products without down-and-out feature. In this way we can understand the high

	Real probabilities (historical data)		Estimates by participants Median		Average	
Probability that DJI is...	-10%	-20%	-10%	-20%	-10%	-20%
...lower at maturity	10.0%	1.0%	15%	5%	21.8%	14.2%
...lower at some point	32.0%	18.0%	30%	15%	41.4%	27.1%
Relative difference	3.2	17.5	2.0	3.0	1.9	1.9

Table 1: Real and estimated probabilities for the Dow Jones Index data. The results show a large misestimation of the relative difference between the two probabilities.

popularity of down-and-out barrier products and why optimized products as we have constructed them in Proposition 4.1 are not visible in the market.

5 Conclusions

We have seen that every optimal financial product is co-monotone with the inverted state price density. This holds regardless of the preferences under consideration, as long as they measure performance relative to an index which is a non-decreasing function of the inverted state price density. The result is applicable particularly for an investor who only considers absolute returns of his investment.

In the special case of a market that can (at least to some extend) be described by the Capital Asset Pricing Model, this implies that optimal portfolios for an investor who is only interested in the absolute returns of his portfolio "follow the market", i.e. their return is the better, the better the return of the market portfolio. This monotonicity also holds for an investor who is only interested in relative performance with respect to the market return. Again, in both instances the underlying decision model can be chosen arbitrarily (e.g. Expected Utility Theory, Mean-Variance or Prospect Theory). Moreover, we have showed that for an investor whose utility only depends on the relative return of his investment only financial products that are at least as risky as the market portfolio can be optimal. In the case of a concave von Neumann-Morgenstern utility investor, an optimal investment even has to be riskier than the market portfolio.

The main fundamental assumptions of these results were homogeneous beliefs and no background risk. Moreover we assumed complete and arbitrage-free markets. The assumptions on the investor's preferences were, however, very weak. In this way, we have extended previous work by Dybvig [5] in several directions: allowing for arbitrary state spaces, general probability measures, general preferences, and even preferences that are depending on the relative performance with respect to a benchmark (e.g. the state price density or the market return).

These results are certainly of theoretical interest, but we also showed some practical applications: first, we studied a new method for the construction of optimal financial products, based on the idea of finding the optimal return distribution among all probability measures satisfying the no-arbitrage condition. This approach makes it necessary to study existence of optimal financial products. We proved an existence result using ideas from calculus of variations and outlined a numerical algorithm for obtaining optimal financial products based on the investor's preferences. Some remarks on situations where existence fails, underline the role of some of the assumptions we had made in the existence theorem.

The second application was in the context of a special, currently very popular financial product, the so-called down-and-out barrier options. Under the simplifying assumption of a CAPM market, we could show that such products cannot be optimal. Given their widespread use, this is certainly a puzzle. Its solutions probably lies less in the well-known deficits of the CAPM, but likely they are instead based on a wrong perception that investors have regarding the probabilities involved in the set-up of the product. This hypothesis is supported by the experimental data we collected that demonstrated a systematic misestimation of the relevant probabilities.

Implications of our results to design and evaluation of structured products are an interesting field for further research.

Acknowledgement

The author thanks Thorsten Hens for his steady and very valuable contributions to this project and Mei Wang for many helpful discussions. Financial support by the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), Project 3, “Evolution and Foundations of Financial Markets”, and by the University Priority Program “Finance and Financial Markets” of the University of Zürich is gratefully acknowledged.

References

[1] Luigi Ambrosio. Lecture notes on optimal transport problems. In P. Colli and C.F. Rodrigues, editors, *CIME summer school in Madeira, Lecture Notes in Mathematics*, volume 1813, pages 1–52. Springer Verlag, 2003.

[2] G. Carlier and R.A. Dana. Core of convex distortions of a probability. *Journal of Economic Theory*, 113:199–222, 2003.

[3] J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31:3–33, 2002.

[4] Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, 2001.

[5] Philip H. Dybvig. Distributional analysis of portfolio choice. *The Journal of Business*, 61(3):369–393, 1988.

[6] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[7] Daniel Kahneman and Amos Tversky. Prospect Theory: An analysis of decision under risk. *Econometrica*, 47:263–291, 1979.

[8] L. V. Kantorovich. On the transfer of masses. *Dokl. Akad. Nauk. SSSR*, 37:227—229, 1942.

[9] Ralf Korn and Elke Korn. *Option Procing and Portfolio Optimization*, volume 31 of *Graduate Studies in Mathematics*. American Mathematical Society, 2001.

[10] Stephen F. LeRoy and Jan Werner. *Principles of Financial Economics*. Cambridge University Press, 2001.

[11] Harry M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[12] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[13] Gaspard Monge. *Géométrie descriptive*. Éditions Jacques Gabay, Sceaux, 1989 (reprint of the 1799 original).

[14] Yuri V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Prob. and Appl.*, I(2):157—214, 1956.

[15] Marc Oliver Rieger. Monotonicity of transport plans and applications. CVGMT preprint, Pisa, 2006.

[16] Marc Oliver Rieger and Mei Wang. Prospect Theory for continuous distributions. NCCR – Financial Valuation and Risk Management, 273, 2006.

[17] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.