Mathematical Foundations for Finance

Chapter II: Arbitrage and Martingale Measures

Prof. Dr. Beatrice Acciaio
beatrice.acciaio@math.ethz.ch

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Table of contents

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

1. Arbitrage
2. The Fundamental Theorem of Asset Pricing (FTAP)
3. Equivalent Martingale Measures

Arbitrage

Goals

- Formalise the idea that a reasonable financial market model should not allow the construction of riskless yet profitable investment strategies.
- Characterise this by an equivalent mathematical property.

Setup

- Consider a financial market on $(\Omega, \mathcal{F}, P)$,
- in *finite discrete time* $\{0, \dots, T\}$ with filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0,\dots,T}$,
- where *discounted asset prices* are given by the processes $S^0 \equiv 1$ and $S = (S_k)_{k=0,1,\dots,T}$, the latter taking values in $\mathbb{R}^d$.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Recall: any pair (V_0, ϑ) with $V_0 \in L^0(\mathcal{F}_0)$ and with an $\mathbb{R}^d$-valued predictable process ϑ can be *identified* with a self-financing strategy φ. We write $\varphi \mathrel{\widehat{=}} (V_0, \vartheta)$.

The corresponding value process is then given by

$$V(\varphi) = V_0 + G(\vartheta) = V_0 + \vartheta \cdot S = V(V_0, \vartheta)\,.$$

- $G(\vartheta) = V(0, \vartheta)$ describes the *cumulative gains or losses* one can generate from initial capital 0 through *self-financing trading via* ϑ.
- We call a strategy φ *a-admissible* if $V(\varphi) \geq -a$, and *admissible* if it is a-admissible for some $a \geq 0$.

Note that these notions, except for 0-admissibility, depend on the chosen unit or numéraire (here S^0).

Definition. An *arbitrage opportunity* is an

- admissible self-financing strategy $\varphi \widehat{=} (0, \vartheta)$ with zero initial wealth
- and such that $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$.

The financial market $(\Omega, \mathcal{F}, \mathbb{F}, P, S^0, S)$, or shortly S, is called *arbitrage-free* if there exist no arbitrage opportunities. We say that *S satisfies (NA)*.

Interpretation: An arbitrage opportunity produces *something* (non-negative final wealth $V_T(\varphi) \geq 0$, with strictly positive probability of having strictly positive final wealth) *out of nothing* (zero initial capital) *without risk* (strategy is self-financing).

Exercise Argue that absence of arbitrage is a natural economic/financial requirement for a reasonable model of a financial market (explain why such "money pumps" cannot exist (for long) in a well-functioning market).

Example. If there exist an asset i_0 and a date k_0 such that $S^{i_0}_{k_0+1} \leq S^{i_0}_{k_0}$ P-a.s. and $P[S^{i_0}_{k_0+1} < S^{i_0}_{k_0}] > 0$, then S admits arbitrage.
(Analogous if "$\leq$" and "$<$" are replaced by "$\geq$" and "$>$" resp.)

Indeed, the price process S^{i_0} can only go down from time k_0 to $k_0 + 1$ and does so in some cases; so if we sell short that asset at time k_0, we run no risk and have the chance of a genuine profit.

Formally,

$$\vartheta^i_{k+1} := -I_{\{i=i_0\}} I_{\{k=k_0\}}, \qquad \text{for } k = 0, \dots, T-1\,,$$

gives an arbitrage opportunity, as one easily checks. This also illustrates the well-known wisdom that *"bad news is better than no news"*.

Variations of (NA)

There are variations of the absence of arbitrage assumption:

- *condition (NA$_+$)*: there exists no *0-admissible* self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ (forbids 0-adm. self-fin.str. that produce something out of nothing)

Variations of (NA)

There are variations of the absence of arbitrage assumption:

- *condition (NA$_+$)*: there exists no *0-admissible* self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ (forbids 0-adm. self-fin.str. that produce something out of nothing)
- *condition (NA')*: there exists no self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ (forbids the same *for all (not necessarily admissible)* self-fin.str.)

Variations of (NA)

There are variations of the absence of arbitrage assumption:

- *condition* (NA_+): there exists no *0-admissible* self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ (forbids 0-adm. self-fin.str. that produce something out of nothing)
- *condition* (NA'): there exists no self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ with $V_T(\varphi) \geq 0$ P-a.s. and $P[V_T(\varphi) > 0] > 0$ (forbids the same *for all (not necessarily admissible)* self-fin.str.)

Note: We quickly see the implications $(\mathrm{NA}') \Longrightarrow (\mathrm{NA}) \Longrightarrow (\mathrm{NA}_+)$. This distinction is important in continuous time or with an infinite time horizon.

For finite discrete time, however, these three concepts are all equivalent!

Proposition. For a financial market in finite discrete time, TFAE:

1) S satisfies (NA)
2) S satisfies (NA$'$)
3) For every (not necessarily admissible) self-financing strategy φ with $V_0(\varphi) = 0$ P-a.s. and $V_T(\varphi) \geq 0$ P-a.s., we have $V_T(\varphi) = 0$ P-a.s.
4) We have
$$\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\},$$
where $\mathcal{G}'$ is the space of all final wealths that one can generate from zero initial wealth through self-financing trading:
$$\mathcal{G}' := \big\{ G_T(\vartheta) \,\big|\, \vartheta \text{ is } \mathbb{R}^d\text{-valued and predictable} \big\},$$
and $L^0_+(\mathcal{F}_T)$ denotes the space of all (equivalence classes, for the relation of equality P-a.s., of) nonnegative $\mathcal{F}_T$-measurable r.v.
5) S satisfies (NA$_+$)

Proof of the Proposition. "2) $\Leftrightarrow$ 3)" is obvious, and "2) $\Leftrightarrow$ 4)" is a direct consequence of the parametrisation of self-financing strategies from Chapter 1. It is also clear that 2) $\Rightarrow$ 1) $\Rightarrow$ 5).

Therefore, we are left to prove "5) $\Rightarrow$ 2)". We show the contraposition: if there was a self-fin.str. $\varphi = (0, \vartheta)$ with $V_T(\varphi) \geq 0$ and $P[V_T(\varphi) > 0] > 0$, then we could construct a strategy $\tilde{\varphi}$ which additionally is *0-admissible*.

Indeed, if φ is not already 0-admissible itself, then the set $A_k := \{V_k(\varphi) < 0\}$ has $P[A_k] > 0$ for some k. We let k_0 be the largest of these k and then define $\tilde{\varphi}$ as the strategy φ on A_{k_0} after time k_0.

In words, we wait until we can start on some set with a negative initial capital and transform that via φ into something non-negative. This yields the desired arbitrage opportunity (details left as exercise for you). **q.e.d.**

Equivalent notions of (NA) - remarks

- This result proves the above comment that all formulations for absence of arbitrage are equivalent in finite discrete time. (Can you spot where we used the finiteness?)
- The mathematical relevance of the Proposition is that it translates the no-arbitrage condition (NA) into the formulation in 4) which has a very useful *geometric interpretation*. We shall exploit this in the next section.

Definition. Two probability measures Q and P on $(\Omega, \mathcal{F})$ are *equivalent (on $\mathcal{F}$)*, written as $Q \approx P$ on $\mathcal{F}$, if they have the same nullsets (in $\mathcal{F}$), that is, if for each set $A \in \mathcal{F}$, we have $P[A] = 0$ if and only if $Q[A] = 0$.

Intuitively, this means that while P and Q may differ in their quantitative assessments, they qualitatively "agree on what is possible or impossible".

Example. Let us consider the *multinomial model* (see previous chapter) as an event tree on the canonical path space $(\Omega, \mathcal{F}) = (\{1, \dots, m\}^T, 2^\Omega)$.

- Any probability measure P on $(\Omega, \mathcal{F})$ can be described by its collection of one-step transition probabilities p_{ij}, which all lie in $[0, 1]$.
- We usually assume that $P[\{\omega\}] > 0$ for all $\omega \in \Omega$, that means all p_{ij} lie in $(0, 1)$. In this case, we have $Q \approx P$ if and only if all one-step transition probabilities q_{ij} for Q lie in $(0, 1)$ as well.

Exercise. Does the assumption $p_{ij} \in (0, 1)$ for all i, j restrict the multinomial model?

Existence of EMM implies (NA)

Lemma. ($\exists$ EMM $\Rightarrow$ (NA)) *If there exists a probability measure $Q \approx P$ on $\mathcal{F}_T$ such that S is a Q-martingale, then S is arbitrage-free.*

Proof. If S is a Q-martingale and $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ is an admissible self-financing strategy, then $V(\varphi) = G(\vartheta) = \vartheta \cdot S$ is a stochastic integral uniformly bounded below (by some $-a$ with $a \geq 0$). From the first chapter, we then know that $V(\varphi)$ is also a Q-martingale and so

$$E_Q[V_T(\varphi)] = E_Q[V_0(\varphi)] = 0.$$

Now suppose in addition that $Q \approx P$ on $\mathcal{F}_T$, so that Q-a.s. and P-a.s. are the same thing on $\mathcal{F}_T$. If $\varphi \mathrel{\widehat{=}} (0, \vartheta)$ is an admissible self-financing strategy with $V_T(\varphi) \geq 0$ P-a.s., then also $V_T(\varphi) \geq 0$ Q-a.s. But $E_Q[V_T(\varphi)] = 0$ by the above argument, and so we must have $V_T(\varphi) = 0$ Q-a.s., hence also $V_T(\varphi) = 0$ P-a.s. Therefore, S is arbitrage-free. **q.e.d.**

1) To have (NA) it is enough for S to be a *local* Q-martingale, because we could still use the results from the first chapter.

2) **An alternative proof:** Suppose that $Q \approx P$ is such that S is a local Q-martingale and take an admissible self-financing strategy $\varphi \mathrel{\widehat{=}} (0, \vartheta)$. Then $V(\varphi) = G(\vartheta) = \vartheta \cdot S$ is a local Q-martingale (see first chapter), with $V_0(\varphi) = 0$, and $V(\varphi)$ is bounded below because φ is admissible.
Then $V(\varphi)$ is a Q-supermartingale (this is easily argued via localising and passing to the limit with the help of Fatou's lemma, exercise for you), and so we get $E_Q[V_T(\varphi)] \leq E_Q[V_0(\varphi)] = 0$.
If in addition $V_T(\varphi) \geq 0$ P-a.s., we also get $V_T(\varphi) \geq 0$ Q-a.s., hence $V_T(\varphi) = 0$ Q-a.s., and then also $V_T(\varphi) = 0$ P-a.s. This allows us to conclude as before.

3) **Yet another proof:** We can also give a complete proof of the lemma which relies only on proved results.

We still use that with $\varphi \mathrel{\widehat{=}} (0, \vartheta)$, we have $V(\varphi) = G(\vartheta) = \vartheta \cdot S$.

Now because ϑ is predictable, the process $\vartheta^{(n)}$ defined by $\vartheta_k^{(n)} := \vartheta_k I_{\{|\vartheta_k| \leq n\}}$ is again predictable and bounded. So if S is a martingale under Q, then $\vartheta^{(n)} \cdot S$ is again a Q-martingale (this part was proved in the previous chapter).

Moreover, the definition of $\vartheta^{(n)}$ makes it clear that (see lecture notes) $(\vartheta^{(n)} \cdot S)^{-} \leq (\vartheta \cdot S)^{-}$; so as $V(\varphi)$ is bounded below by $-a$ because φ is admissible, the entire sequence $(G(\vartheta^{(n)}))_{n \in \mathbb{N}} = (\vartheta^{(n)} \cdot S)_{n \in \mathbb{N}}$ is also bounded below by $-a$.

This allows us to use Fatou's lemma and conclude from the martingale property of each $G(\vartheta^{(n)})$ that $V(\varphi) = \vartheta \cdot S$ is a Q-supermartingale:

$$\begin{aligned} E_Q[G_k(\vartheta) \,|\, \mathcal{F}_{k-1}] &= E_Q\left[\lim_{n\to\infty} G_k(\vartheta^{(n)}) \,\middle|\, \mathcal{F}_{k-1}\right] \\ &\le \liminf_{n\to\infty} E_Q[G_k(\vartheta^{(n)}) \,|\, \mathcal{F}_{k-1}] \\ &= \liminf_{n\to\infty} G_{k-1}(\vartheta^{(n)}) \;=\; G_{k-1}(\vartheta). \end{aligned}$$

Now we can finish the proof as before in 2).

This allows us to use Fatou's lemma and conclude from the martingale property of each $G(\vartheta^{(n)})$ that $V(\varphi) = \vartheta \cdot S$ is a Q-supermartingale:

$$\begin{aligned} E_Q[G_k(\vartheta) \,|\, \mathcal{F}_{k-1}] &= E_Q\Big[\lim_{n\to\infty} G_k(\vartheta^{(n)}) \,\Big|\, \mathcal{F}_{k-1}\Big] \\ &\leq \liminf_{n\to\infty} E_Q[G_k(\vartheta^{(n)}) \,|\, \mathcal{F}_{k-1}] \\ &= \liminf_{n\to\infty} G_{k-1}(\vartheta^{(n)}) \;=\; G_{k-1}(\vartheta). \end{aligned}$$

Now we can finish the proof as before in 2).

4) In continuous time, things get more complicated. It is useful and important to have the alternative route via 2). Also in discrete but infinite time, we must be careful about the behaviour at ∞.

Example. Consider the *multinomial model* on the canonical path space and suppose as usual that $P[\{\omega\}] > 0$ for all $\omega \in \Omega$.

To find $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale, we need to find one-step transition probabilities in the open interval $(0,1)$ such that

$$E_Q\left[\widetilde{S}^1_k/\widetilde{S}^0_k \,\middle|\, \mathcal{F}_{k-1}\right] = \widetilde{S}^1_{k-1}/\widetilde{S}^0_{k-1} \qquad \text{for all } k.$$

Since we have

$$\frac{\widetilde{S}^1_k/\widetilde{S}^0_k}{\widetilde{S}^1_{k-1}/\widetilde{S}^0_{k-1}} = \frac{\widetilde{S}^1_k/\widetilde{S}^1_{k-1}}{\widetilde{S}^0_k/\widetilde{S}^0_{k-1}} = \frac{Y_k}{1+r},$$

we can equivalently require

$$E_Q[Y_k/(1+r) \,|\, \mathcal{F}_{k-1}] = 1 \qquad \text{for all } k.$$

Now fix k and look at a node corresponding to an atom $A = A_{\bar{x}_1,\dots,\bar{x}_{k-1}} \in \mathcal{F}_{k-1}$ at time $k-1$ with corresponding one-step transition probabilities $q_1, \dots, q_m$.

Remark: We sometimes omit to write the indices for $q_j = q_j^{(k,A)} = q_{\bar{x}_1,\dots,\bar{x}_{k-1},j}$, *but of course the one-step transition probabilities can depend on the time k and on the atom A.*

Now fix k and look at a node corresponding to an atom $A = A_{\bar{x}_1,\dots,\bar{x}_{k-1}} \in \mathcal{F}_{k-1}$ at time $k-1$ with corresponding one-step transition probabilities $q_1, \dots, q_m$.

Remark: We sometimes omit to write the indices for $q_j = q_j^{(k,A)} = q_{\bar{x}_1,\dots,\bar{x}_{k-1},j}$, but of course the one-step transition probabilities can depend on the time k and on the atom A.

For the associated probability measure Q, the quantities $q_j = Q[Y_k = 1 + y_j \,|\, \mathcal{F}_{k-1}]$ for branch $j = 1, \dots, m$ then describe the (one-step) conditional distribution of Y_k given $\mathcal{F}_{k-1}$ at that node, and so

on A: $$E_Q[Y_k \,|\, \mathcal{F}_{k-1}] = E_Q[Y_k \,|\, A] = \sum_{j=1}^{m} q_j^{k,A}(1 + y_j) = 1 + \sum_{j=1}^{m} q_j^{k,A} y_j$$

which implies

$$E_Q[Y_k \,|\, \mathcal{F}_{k-1}] = \sum_{\text{atoms } A \in \mathcal{F}_{k-1}} I_A E_Q[Y_k \,|\, A] = 1 + \sum_{\text{atoms } A \in \mathcal{F}_{k-1}} I_A q_j^{k,A} y_j.$$

Note: *Although we have started with a particular atom A, the resulting condition always looks the same. This is due to the homogeneity in the structure of the multinomial model.*

Hence, the above conditional expectation equals $1 + r$ if and only if the equation

$$\sum_{j=1}^{m} q_j^{k,A} y_j = r$$

has a solution $q_1^{k,A}, \ldots, q_m^{k,A}$.

Note: Although we have started with a particular atom A, the resulting condition always looks the same. This is due to the homogeneity in the structure of the multinomial model.

Hence, the above conditional expectation equals $1 + r$ if and only if the equation

$$\sum_{j=1}^{m} q_j^{k,A} y_j = r$$

has a solution $q_1^{k,A}, \dots, q_m^{k,A}$.

Since we require all the $q_j^{k,A}$ to lie in $(0,1)$ and since we have $y_m > y_{m-1} > \cdots > y_1 > -1$ by the assumed labelling, a solution exists if and only if $y_m > r > y_1$, that is, if and only if the riskless interest rate r for the bank account lies strictly between the smallest and largest return values, y_1 and y_m, for the stock.

As a consequence, we obtain the following result.

Corollary. *In the multinomial model with parameters* $y_1 < \cdots < y_m$ *and* r*, there exists a probability measure* $Q \approx P$ *such that* $\widetilde{S}^1/\widetilde{S}^0$ *is a* Q*-martingale if and only if* $y_1 < r < y_m$*.*

Interpretation: The condition $y_1 < r < y_m$ is intuitive: in comparison to the riskless bank account $\widetilde{S}^0$, the stock $\widetilde{S}^1$ can have both higher and lower growth than $\widetilde{S}^0$. Hence $\widetilde{S}^1$ is genuinely more risky than $\widetilde{S}^0$.

As a consequence, we obtain the following result.

Corollary. *In the multinomial model with parameters* $y_1 < \cdots < y_m$ *and* r*, there exists a probability measure* $Q \approx P$ *such that* $\widetilde{S}^1/\widetilde{S}^0$ *is a* Q*-martingale if and only if* $y_1 < r < y_m$*.*

Interpretation: The condition $y_1 < r < y_m$ is intuitive: in comparison to the riskless bank account $\widetilde{S}^0$, the stock $\widetilde{S}^1$ can have both higher and lower growth than $\widetilde{S}^0$. Hence $\widetilde{S}^1$ is genuinely more risky than $\widetilde{S}^0$.

One has the feeling that this should not only be sufficient to exclude arbitrage opportunities, but necessary as well. This feeling is correct, as we shall see in the next section.

Exercise. Go through the proof (the previous example), make sure you understand the concept (notation is the most difficult part here), and fill in the remaining gaps.

For the special case of the binomial model, we can even say a bit more.

Corollary. *In the binomial model with parameters $u > d$ and r, there exists a probability measure $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale if and only if $u > r > d$. In that case, Q is unique (on $\mathcal{F}_T$) and characterised by the property that $Y_1, \ldots, Y_T$ are i.i.d. under Q with parameter*

$$Q[Y_k = 1 + u] = q^* = \frac{r - d}{u - d} = 1 - Q[Y_k = 1 + d].$$

For the special case of the binomial model, we can even say a bit more.

Corollary. *In the binomial model with parameters $u > d$ and r, there exists a probability measure $Q \approx P$ such that $\widetilde{S}^1/\widetilde{S}^0$ is a Q-martingale if and only if $u > r > d$. In that case, Q is unique (on $\mathcal{F}_T$) and characterised by the property that $Y_1, \dots, Y_T$ are i.i.d. under Q with parameter*

$$Q[Y_k = 1 + u] = q^* = \frac{r-d}{u-d} = 1 - Q[Y_k = 1 + d].$$

Proof. With $m = 2$ and $q := q_2$ the martingale condition $\sum_{j=1}^{m} q_j y_j = r$ reduces to $(1-q)d + qu = r$, which has the unique solution

$$q^* = \frac{r-d}{u-d}\,.$$

Because the one-step transition probabilities are thus the same in each node throughout the tree, the i.i.d. description follows as in chapter 1. **q.e.d.**

The Fundamental Theorem of Asset Pricing (FTAP)

We have seen a *sufficient* condition for S to be arbitrage-free.

Moreover, the multinomial model has led us to suspect that this condition might be *necessary* as well. In this section, we shall prove that this is indeed so, for every financial market model in finite discrete time.

We have seen a *sufficient* condition for S to be arbitrage-free.

Moreover, the multinomial model has led us to suspect that this condition might be *necessary* as well. In this section, we shall prove that this is indeed so, for every financial market model in finite discrete time.

To formulate this result, we first introduce a very important concept.

Definition. An *equivalent (local) martingale measure (E(L)MM)* for S is a probability measure Q equivalent to P on $\mathcal{F}_T$ such that S is a (local) Q-martingale. We denote by $\mathbb{P}_e(S)$ or $\mathbb{P}_e$ the set of all EMMs for S and by $\mathbb{P}_{e,loc}$ the set of all ELMMs for S. Clearly, $\mathbb{P}_e \subseteq \mathbb{P}_{e,loc}$.

Saying that $\mathbb{P}_{e,loc}(S)$ is non-empty is the same as saying that there exists an equivalent local martingale measure Q for S. We have seen that this property implies that S satisfies (NA).

It is very remarkable and important that the converse implication also holds true.

Theorem. (Dalang-Morton-Willinger: (NA) $\Leftrightarrow$ $\exists$EMM) *A financial market model S in finite discrete time is arbitrage-free if and only if there exists an equivalent martingale measure for S:*

$$\text{(NA)} \quad \Longleftrightarrow \quad \mathbb{P}_e(S) \neq \emptyset \quad \Longleftrightarrow \quad \mathbb{P}_{e,loc}(S) \neq \emptyset.$$

1) The FTAP translates the economic/financial condition of absence of arbitrage into an equivalent, purely mathematical condition.
$\rightarrow$ use of martingale theory for the study of financial market models.

1) The FTAP translates the economic/financial condition of absence of arbitrage into an equivalent, purely mathematical condition.
$\rightarrow$ use of martingale theory for the study of financial market models.

2) The classical theorems in martingale theory on gambling state *"one cannot win in a systematic way if one bets on a martingale"*.
The FTAP can be viewed as a *converse*: *"if one cannot win by betting on a given process, then that process must be a martingale"*
— at least after an equivalent change of probability measure.

1) The FTAP translates the economic/financial condition of absence of arbitrage into an equivalent, purely mathematical condition.
$\rightarrow$ use of martingale theory for the study of financial market models.

2) The classical theorems in martingale theory on gambling state *"one cannot win in a systematic way if one bets on a martingale"*.
The FTAP can be viewed as a converse: *"if one cannot win by betting on a given process, then that process must be a martingale"*
— at least after an equivalent change of probability measure.

3) Note that we make no integrability assumptions about S (under P); so it is also noteworthy that S, being a Q-martingale, is automatically integrable under (some) Q. (However, this is a minor point. Indeed, we can always construct a probability measure R equivalent to P such that under R, S becomes as nicely integrable as we want. But of course such an R will in general not be a martingale measure for S.)

- Proving the theorem of Dalang, Morton, and Willinger is not elementary if one wants to allow models where the underlying probability space $(\Omega, \mathcal{F}, P)$ is infinite, or more precisely if one of the σ-fields $\mathcal{F}_k$, $k < T$, is infinite.
- This level of generality is needed very quickly, for instance as soon as we want to work with returns which take more than only a finite number of values.
- For now, we content ourselves here with an explanation of the *key geometric idea* behind the proof, and with the exact argument for the case where Ω (or rather $\mathcal{F}_T$) is finite (like for instance in the canonical setting for the multinomial model).

Due to the previous results, we only need to prove that absence of arbitrage implies the existence of an equivalent martingale measure for S.

- We know that (NA) is equivalent to $\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\}$, where

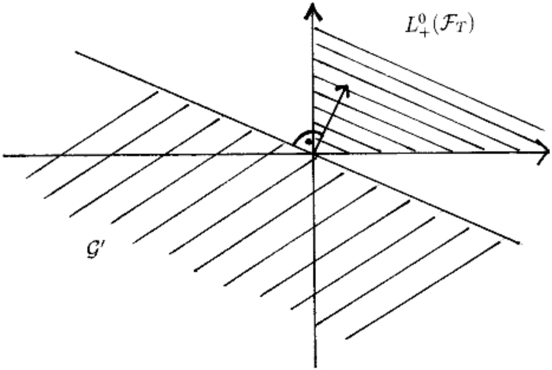
$$\mathcal{G}' = \big\{ G_T(\vartheta) \,\big|\, \vartheta \text{ is } \mathbb{R}^d\text{-valued and predictable} \big\}$$

  is the space of all final positions one can generate from initial wealth 0 by self-financing (but not necessarily admissible) trading.

- In geometric terms: the upper-right quadrant of non-negative random variables, $L^0_+(\mathcal{F}_T)$, intersects the linear subspace $\mathcal{G}'$ only at 0.

- As a consequence, the two sets $L^0_+(\mathcal{F}_T)$ and $\mathcal{G}'$ can be *separated by a hyperplane*, and the normal vector defining that hyperplane then yields (after suitable normalisation) the (density of the) desired EMM.

FTAP - idea of the proof



Graphical illustration of the condition $\mathcal{G}' \cap L^0_+(\mathcal{F}_T) = \{0\}$

FTAP - idea of the proof

- We see that the existence of an EMM follows from the existence of a separating hyperplane between two sets.
- In that sense, the proof is (not surprisingly) not constructive, and it is also clear that we cannot expect uniqueness of an EMM in general.
- The latter fact can also easily be seen directly:
  Because the set $\mathbb{P}_e(S)$ is obviously convex, it is either empty, or contains exactly one element, or contains infinitely (uncountably) many elements.

Proof of the FTAP for finite Ω (or $\mathcal{F}_T$).

- If Ω (or $\mathcal{F}_T$) is finite, then every random variable on $(\Omega, \mathcal{F}_T)$ can take only a finite number of values, say $n \in \mathbb{N}$.
- So we can identify $L^0(\mathcal{F}_T)$ with the finite-dimensional space $\mathbb{R}^n$ and $L^0_+(\mathcal{F}_T)$ with $\mathbb{R}^n_+$. The set $\mathcal{G}' \subseteq L^0(\mathcal{F}_T)$ can then be identified with a linear subspace $\mathcal{H}$ of $\mathbb{R}^n$, and so (NA) translates into $\mathcal{H} \cap \mathbb{R}^n_+ = \{0\}$.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Proof of the FTAP for finite Ω (or $\mathcal{F}_T$).

- If Ω (or $\mathcal{F}_T$) is finite, then every random variable on $(\Omega, \mathcal{F}_T)$ can take only a finite number of values, say $n \in \mathbb{N}$.
- So we can identify $L^0(\mathcal{F}_T)$ with the finite-dimensional space $\mathbb{R}^n$ and $L^0_+(\mathcal{F}_T)$ with $\mathbb{R}^n_+$. The set $\mathcal{G}' \subseteq L^0(\mathcal{F}_T)$ can then be identified with a linear subspace $\mathcal{H}$ of $\mathbb{R}^n$, and so (NA) translates into $\mathcal{H} \cap \mathbb{R}^n_+ = \{0\}$.
- Recall that a set $A \in \mathcal{F}_T$ is an atom in $\mathcal{F}_T$ if $P[A] > 0$ and if any $B \in \mathcal{F}_T$ with $B \subseteq A$ has either $P[B] = 0$ or $P[B] = P[A]$.
  Then any $\mathcal{F}_T$-measurable random variable Z has the form

$$Z = \sum_{A \text{ atom in } \mathcal{F}_T} Z I_A = \sum_{A \text{ atom in } \mathcal{F}_T} z_A I_A \text{ with } z_A \in \mathbb{R}.$$

- We consider the set of all $\mathcal{F}_T$-measurable $Z \geq 0$ with $\sum_{A \text{ atom in } \mathcal{F}_T} z_A = 1$ and identify this with the subset

$$\mathcal{K} = \left\{ z \in \mathbb{R}^n_+ \,\middle|\, \sum_{i=1}^n z_i = 1 \right\} \subset \mathbb{R}^n_+ \,,$$

where n denotes the (finite, by assumption) number of atoms in $\mathcal{F}_T$.

- Notice that $\mathcal{K}$ does not contain the vector 0, so that we must have $\mathcal{H} \cap \mathcal{K} = \emptyset$. Moreover, $\mathcal{K}$ is convex and compact

FTAP - proof for finite Ω (or $\mathcal{F}_T$)

- We consider the set of all $\mathcal{F}_T$-measurable $Z \geq 0$ with $\sum_{A \text{ atom in } \mathcal{F}_T} z_A = 1$ and identify this with the subset

$$\mathcal{K} = \left\{ z \in \mathbb{R}^n_+ \,\middle|\, \sum_{i=1}^n z_i = 1 \right\} \subset \mathbb{R}^n_+ ,$$

where n denotes the (finite, by assumption) number of atoms in $\mathcal{F}_T$.

- Notice that $\mathcal{K}$ does not contain the vector 0, so that we must have $\mathcal{H} \cap \mathcal{K} = \emptyset$. Moreover, $\mathcal{K}$ is convex and compact

- So a classical separation theorem for sets in $\mathbb{R}^n$ implies that there exists a vector $\lambda \in \mathbb{R}^n$, $\lambda \neq 0$ such that

$$\forall h \in \mathcal{H} : \lambda^{\mathrm{tr}} h = 0 \quad \text{and} \quad \forall z \in \mathcal{K} : \lambda^{\mathrm{tr}} z > 0 .$$

- Notice that all unit coordinate vectors in $\mathbb{R}^n$ lie in $\mathcal{K}$. Hence, going through all of them, the second property implies that all coordinates of λ must be strictly positive.

- So we can define a probability measure Q on $\mathcal{F}_T$ by

$$q_i := \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}, \quad 1 \leq i \leq n.$$

Indeed, all q_i lie in $(0, 1)$ and sum to 1.

- Notice that all unit coordinate vectors in $\mathbb{R}^n$ lie in $\mathcal{K}$. Hence, going through all of them, the second property implies that all coordinates of λ must be strictly positive.

- So we can define a probability measure Q on $\mathcal{F}_T$ by

$$q_i := \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}, \quad 1 \leq i \leq n.$$

Indeed, all q_i lie in $(0,1)$ and sum to 1.

- Because $P[A] > 0$ for all atoms $A \in \mathcal{F}_T$, it is clear that Q is equivalent to P on $\mathcal{F}_T$. The property that $\lambda^{\mathrm{tr}} h = 0$ for all $h \in \mathcal{H}$ translates via the identification of $\mathcal{H}$ and $\mathcal{G}'$ into

$$E_Q[G_T(\vartheta)] = 0 \qquad \text{for all } \mathbb{R}^d\text{-valued predictable } \vartheta.$$

FTAP - proof for finite Ω (or $\mathcal{F}_T$)

- For any time k, asset number i and set $A \in \mathcal{F}_{k-1}$, we can consider

$$\vartheta := I_{\{\text{time} = k\}} I_{\{\text{asset number} = i\}} I_A,$$

which gives $G_T(\vartheta) = I_A(S_k^i - S_{k-1}^i)$. The fact that this has Q-expectation 0 for arbitrary $A \in \mathcal{F}_{k-1}$ means that $E_Q[S_k^i - S_{k-1}^i \mid \mathcal{F}_{k-1}] = 0$ for all k. Hence, S is a Q-martingale. Note that integrability is not an issue here because Ω (or $\mathcal{F}_T$) is finite.

q.e.d.

FTAP - generalisations

- In continuous time or with an infinite time horizon, existence of an EMM still implies (NA), *but the converse is not true*. One needs a sort of topological strengthening that not only forbids arbitrage from each single strategy, but also excludes the possibility of creating "arbitrage in the limit by using a sequence of strategies".
- The resulting condition is called *(NFLVR)* for *"no free lunch with vanishing risk"*, and the corresponding equivalence theorem, due to Freddy Delbaen and Walter Schachermayer in its most general form, is called the *fundamental theorem of asset pricing (FTAP)*.
- The basic idea for proving the FTAP is still the same as in our above proof, but the techniques and arguments are much more advanced. (To be accurate: also the concept of EMM must be generalised a little to obtain that theorem.)

Combining our results for bi- and multinomial models with the latter theorem now immediately yields the following corollaries.

Corollary. *The multinomial model with parameters* $y_1 < \cdots < y_m$ *and* r *is arbitrage-free if and only if* $y_1 < r < y_m$.

Corollary. *The binomial model with parameters* $u > d$ *and* r *is arbitrage-free if and only if* $u > r > d$. *In that case, the EMM* Q^* *for* $\widetilde{S}^1/\widetilde{S}^0$ *is unique (on* $\mathcal{F}_T$*) and is given by*

$$Q^*[Y_k = 1 + u] = \frac{r - d}{u - d} = 1 - Q^*[Y_k = 1 + d] .$$

Equivalent Martingale Measures

We begin with $(\Omega, \mathcal{F})$ and a filtration $\mathbb{F} = (\mathcal{F}_k)_{k=0,1,\dots,T}$ in finite discrete time. On $(\Omega, \mathcal{F})$, we have two probability measures Q and P, and we assume that $Q \approx P$.

The *Radon–Nikodým theorem* tells us that there exists a *density* $\frac{dQ}{dP} =: \mathcal{D}$. This is a random variable $\mathcal{D} > 0$ P-a.s. (because $Q \approx P$) such that $Q[A] = E_P[\mathcal{D} I_A]$ for all $A \in \mathcal{F}$, or more generally

$$\int_\Omega Y\,dQ = E_Q[Y] = E_P[Y\mathcal{D}] = \int_\Omega Y\mathcal{D}\,dP$$

for all random variables $Y \geq 0$, which explains the notation to some extent.

The point of these formulae is that they tell us how to compute Q-expectations in terms of P-expectations and vice versa.

To get similar transformation rules for conditional expectations, we define the P-martingale Z by

$$Z_k := E_P[\mathcal{D} \,|\, \mathcal{F}_k] = E_P\left[\frac{dQ}{dP} \,\middle|\, \mathcal{F}_k\right] \qquad \text{for } k = 0, 1, \dots, T.$$

Because $\mathcal{D} > 0$ P-a.s., the process $Z = (Z_k)_{k=0,1,\dots,T}$ is strictly positive in the sense that $Z_k > 0$ P-a.s. for each k, or also $P[Z_k > 0 \text{ for all } k] = 1$.

Z is called the *density process (of Q, with respect to P)*. The following lemma makes it clear why. We will not give a proof of this lemma, but strongly recommend to do this as an exercise.

Lemma.

1) *For every* $k \in \{0, 1, \ldots, T\}$*, any* $A \in \mathcal{F}_k$*, and any* $\mathcal{F}_k$*-measurable random variable* Y *which is* ≥ 0 *or in* $L^1(Q)$*, we have*

$$Q[A] = E_P[Z_k I_A] \quad \textit{and} \quad E_Q[Y] = E_P[Z_k Y], \; \textit{respectively.}$$

This means that Z_k *is the density of* Q *with respect to* P *on* $\mathcal{F}_k$*.*

Lemma.

1) *For every* $k \in \{0, 1, \dots, T\}$*, any* $A \in \mathcal{F}_k$*, and any* $\mathcal{F}_k$*-measurable random variable* Y *which is* ≥ 0 *or in* $L^1(Q)$*, we have*

$$Q[A] = E_P[Z_k I_A] \quad \text{and} \quad E_Q[Y] = E_P[Z_k Y], \ \text{respectively.}$$

This means that Z_k *is the density of* Q *with respect to* P *on* $\mathcal{F}_k$*.*

2) *If* $j \leq k$ *and* U_k *is* $\mathcal{F}_k$*-measurable and either* ≥ 0 *or in* $L^1(Q)$*, then we have the Bayes formula*

$$E_Q[U_k \mid \mathcal{F}_j] = \frac{1}{Z_j} E_P[Z_k U_k \mid \mathcal{F}_j] \qquad Q\text{-a.s.} \tag{1}$$

This equation relates conditional expectations under Q *and* P*.*

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Lemma.

1) *For every $k \in \{0, 1, \dots, T\}$, any $A \in \mathcal{F}_k$, and any $\mathcal{F}_k$-measurable random variable Y which is ≥ 0 or in $L^1(Q)$, we have*

$$Q[A] = E_P[Z_k I_A] \quad \text{and} \quad E_Q[Y] = E_P[Z_k Y], \text{ respectively.}$$

This means that Z_k is the density of Q with respect to P on $\mathcal{F}_k$.

2) *If $j \leq k$ and U_k is $\mathcal{F}_k$-measurable and either ≥ 0 or in $L^1(Q)$, then we have the Bayes formula*

$$E_Q[U_k \,|\, \mathcal{F}_j] = \frac{1}{Z_j} E_P[Z_k U_k \,|\, \mathcal{F}_j] \qquad Q\text{-a.s.} \tag{1}$$

This equation relates conditional expectations under Q and P.

3) *A process $N = (N_k)_{k=0,1,\dots,T}$ which is adapted to $\mathbb{F}$ is a Q-martingale if and only if the product ZN is a P-martingale. This tells us how martingale properties under P and Q are related to each other.*

- Because Z is strictly positive, we can define

$$D_k := \frac{Z_k}{Z_{k-1}} \qquad \text{for } k = 1, \dots, T.$$

The process D is adapted, strictly positive and satisfies by its definition

$$E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1,$$

because Z is a P-martingale.

- Because Z is strictly positive, we can define

$$D_k := \frac{Z_k}{Z_{k-1}} \qquad \text{for } k = 1, \dots, T.$$

The process D is adapted, strictly positive and satisfies by its definition

$$E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1,$$

because Z is a P-martingale.

- Again because Z is a martingale and by the above lemma,

$$E_P[Z_0] = E_P[Z_T] = E_P[Z_T I_\Omega] = Q[\Omega] = 1,$$

and we can of course recover Z from Z_0 and D via

$$Z_k = Z_0 \prod_{j=1}^{k} D_j \qquad \text{for } k = 0, 1, \dots, T.$$

$\rightarrow$ So every $Q \approx P$ induces via Z a pair (Z_0, D).

- If we conversely start with a pair (Z_0, D) with the above properties (i.e. Z_0 is $\mathcal{F}_0$-measurable, $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$, and D is adapted and strictly positive with $E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1$ for all k), we can define a probability measure $Q \approx P$ via

$$\frac{dQ}{dP} := Z_0 \prod_{j=1}^{T} D_j.$$

- If we conversely start with a pair (Z_0, D) with the above properties (i.e. Z_0 is $\mathcal{F}_0$-measurable, $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$, and D is adapted and strictly positive with $E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1$ for all k), we can define a probability measure $Q \approx P$ via

$$\frac{dQ}{dP} := Z_0 \prod_{j=1}^{T} D_j .$$

- Written in terms of D, the Bayes formula (1) for $j = k - 1$ becomes

$$E_Q[U_k \,|\, \mathcal{F}_{k-1}] = E_P[D_k U_k \,|\, \mathcal{F}_{k-1}]. \qquad (2)$$

This shows that the ratios D_k play the role of *"one-step conditional densities"* of Q with respect to P.

- The above parametrisation is very simple and yet very useful when we want to *construct an equivalent martingale measure* for a given process S. All we need are
  1) an $\mathcal{F}_0$-measurable random variable $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$, and
  2) an adapted strictly positive process $D = (D_k)_{k=1,\dots,T}$ satisfying $E_P[D_k \mid \mathcal{F}_{k-1}] = 1$ and $E_P[D_k(S_k - S_{k-1}) \mid \mathcal{F}_{k-1}] = 0$ for all k.

- The above parametrisation is very simple and yet very useful when we want to *construct an equivalent martingale measure* for a given process S. All we need are
  1) an $\mathcal{F}_0$-measurable random variable $Z_0 > 0$ P-a.s. with $E_P[Z_0] = 1$, and
  2) an adapted strictly positive process $D = (D_k)_{k=1,\dots,T}$ satisfying $E_P[D_k \,|\, \mathcal{F}_{k-1}] = 1$ and $E_P[D_k(S_k - S_{k-1}) \,|\, \mathcal{F}_{k-1}] = 0$ for all k.
- The simplest choice for Z_0 is clearly the constant $Z_0 \equiv 1$. This amounts to saying that Q and P should coincide on $\mathcal{F}_0$.
- If $\mathcal{F}_0$ is P-trivial (i.e. $P[A] \in \{0, 1\}$ for all $A \in \mathcal{F}_0$) as is often the case, then every $\mathcal{F}_0$-measurable random variable is P-a.s. constant, and then $Z_0 \equiv 1$ is actually the only possible choice (because we must have $E_P[Z_0] = 1$).

Concerning the D_k, *not much can be said in this generality*.

To get more structure for our model, we therefore consider a setting with *i.i.d. returns*. This means that

$$\widetilde{S}_k^1 = S_0^1 \prod_{j=1}^{k} Y_j, \qquad \widetilde{S}_k^0 = (1+r)^k,$$

where $Y_1, \dots, Y_T$ are strictly positive and i.i.d. under P.

The filtration we use is generated by $(\widetilde{S}^0, \widetilde{S}^1)$ or equivalently by $\widetilde{S}^1$ or by Y; so $\mathcal{F}_0$ is P-trivial and Y_k is under P independent of $\mathcal{F}_{k-1}$ for each k.

The Q-martingale condition for $S^1 = \widetilde{S}^1/\widetilde{S}^0$ in multiplicative form is then

$$1 = E_Q\left[\frac{S_k^1}{S_{k-1}^1}\,\middle|\,\mathcal{F}_{k-1}\right] = E_Q\left[\frac{\widetilde{S}_k^1/\widetilde{S}_k^0}{\widetilde{S}_{k-1}^1/\widetilde{S}_{k-1}^0}\,\middle|\,\mathcal{F}_{k-1}\right] = E_P\left[\frac{D_k Y_k}{1+r}\,\middle|\,\mathcal{F}_{k-1}\right].$$

To keep things as simple as possible, we now might try to choose D_k like Y_k independent of $\mathcal{F}_{k-1}$.

- ▷ Then [one can prove that] we must have $D_k = g_k(Y_k)$ for some measurable function g_k, and we have to choose g_k in such a way that we get

$$1 = E_P[D_k \mid \mathcal{F}_{k-1}] = E_P[g_k(Y_k)]$$

and

$$1 + r = E_P[D_k Y_k \mid \mathcal{F}_{k-1}] = E_P[Y_k g_k(Y_k)].$$

- ▷ Both these calculations exploit the P-independence of Y_k from $\mathcal{F}_{k-1}$. If this choice is possible, we can then choose all the $g_k \equiv g_1$, because the Y_k are (assumed) i.i.d. under P. To ensure that $D_k > 0$, we can impose $g_k > 0$

Example. We still assume that we have i.i.d. returns.

- If the Y_k are *discrete random variables* with values $(1 + y_j)_{j \in \mathbb{N}}$ and $P[Y_k = 1 + y_j] = p_j$, then g_1 is determined by its values $g_1(1 + y_j)$.

Example. We still assume that we have i.i.d. returns.

- If the Y_k are *discrete random variables* with values $(1+y_j)_{j\in\mathbb{N}}$ and $P[Y_k = 1+y_j] = p_j$, then g_1 is determined by its values $g_1(1+y_j)$.
- $Q \approx P$ means that we need $q_j := Q[Y_k = 1+y_j] > 0$ for all the j's with $p_j > 0$. Setting $q_j := p_j g_1(1+y_j)$, we are looking for q_j having $q_j > 0$ whenever $p_j > 0$ and satisfying

$$1 = E_P[g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j g_1(1+y_j) = \sum_{j\in\mathbb{N}} q_j$$

and

$$1 + r = E_P[Y_1 g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j(1+y_j)g_1(1+y_j)$$

$$= \sum_{j\in\mathbb{N}} q_j(1+y_j) = 1 + \sum_{j\in\mathbb{N}} q_j y_j .$$

Example. We still assume that we have i.i.d. returns.

- If the Y_k are *discrete random variables* with values $(1+y_j)_{j\in\mathbb{N}}$ and $P[Y_k = 1+y_j] = p_j$, then g_1 is determined by its values $g_1(1+y_j)$.
- $Q \approx P$ means that we need $q_j := Q[Y_k = 1+y_j] > 0$ for all the j's with $p_j > 0$. Setting $q_j := p_j g_1(1+y_j)$, we are looking for q_j having $q_j > 0$ whenever $p_j > 0$ and satisfying

$$1 = E_P[g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j g_1(1+y_j) = \sum_{j\in\mathbb{N}} q_j$$

and

$$1+r = E_P[Y_1 g_1(Y_1)] = \sum_{j\in\mathbb{N}} p_j(1+y_j)g_1(1+y_j)$$

$$= \sum_{j\in\mathbb{N}} q_j(1+y_j) = 1 + \sum_{j\in\mathbb{N}} q_j y_j.$$

- Note that the actual values of the p_j are not relevant here. It only matters which of them are strictly positive.

Example. In the *multinomial model* with parameters $y_1, \ldots, y_m$ and r, the above recipe boils down to finding $q_1, \ldots, q_m > 0$ with

$$\sum_{j=1}^{m} q_j = 1 \quad \text{and} \quad \sum_{j=1}^{m} q_j y_j = r\,.$$

If $m > 2$ and the y_j are as usual all distinct, there is clearly an *infinite number of solutions*.

Example. If we have *i.i.d. lognormal returns*, then

$$Y_i = e^{\sigma U_i + b} \quad \text{with} \quad U_1, \ldots, U_T \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)\,.$$

Instead of $D_i = g_1(Y_i)$, we try (equivalently) with $D_i = \tilde{g}_1(U_i)$, and more specifically with $D_i = e^{\alpha U_i + \beta}$. Then we have

$$E_P[D_i] = e^{\beta + \frac{1}{2}\alpha^2} = 1 \qquad \text{for } \beta = -\tfrac{1}{2}\alpha^2,$$

and we get

$$E_P[D_i Y_i] = e^{b + \beta + \frac{1}{2}(\alpha+\sigma)^2} = 1 + r, \quad \text{so that}$$
$$\log(1+r) = b + \beta + \frac{1}{2}(\alpha+\sigma)^2 = b + \frac{1}{2}\sigma^2 + \alpha\sigma,$$

and thus,

$$\alpha = \frac{1}{\sigma}\left(\log(1+r) - b - \frac{1}{2}\sigma^2\right).$$

So we could for instance take

$$D_k = \exp\left(-\gamma U_k - \frac{1}{2}\gamma^2\right)$$

with

$$\gamma = -\alpha = \frac{b + \frac{1}{2}\sigma^2 - \log(1+r)}{\sigma}.$$

So we could for instance take

$$D_k = \exp\left(-\gamma U_k - \frac{1}{2}\gamma^2\right)$$

with

$$\gamma = -\alpha = \frac{b + \frac{1}{2}\sigma^2 - \log(1+r)}{\sigma}.$$

Remark. If we start with a model of i.i.d. returns $Y_1, \dots, Y_T$ *under P*, then the above construction produces via the $D_k = g_1(Y_k)$ an EMM Q with the property that the returns are i.i.d. *under Q* as well. However, this is only due to the particular construction of Q; there are in general many other EMMs under which the returns are no longer i.i.d..