SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Dynamic Mean–Variance Portfolio Choice: Markowitz Foundations, Time Inconsistency, and Game-Theoretic Equilibria

av

Emre Kaplaner

2025 - No K25

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Dynamic Mean–Variance Portfolio Choice: Markowitz Foundations, Time Inconsistency, and Game-Theoretic Equilibria

Emre Kaplaner

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Yishao Zhou

2025

Abstract

This thesis investigates dynamic mean–variance portfolio optimization with a focus on the fundamental challenge of time inconsistency. In the classical single-period setting, one obtains a clear risk–return trade-off via a quadratic optimization, but extending to multi-period or continuous-time horizons reveals that variance penalties cannot be nested in the usual recursive optimization framework. To overcome this, we adopt a game-theoretic intrapersonal approach in which each date-t decision-maker is treated as a "player" sharing the same preferences but controlling only that period's choice. A time-consistent policy is then defined as a subgame-perfect equilibrium: no future self can profitably deviate when all others stick to the prescribed strategy. We derive the corresponding extended Bellman recursions in discrete time and outline the continuous-time analogue as an equilibrium HJB system. By solving these equations in representative examples, we demonstrate a tractable method for generating dynamically credible mean–variance strategies that respect real-world constraints.

Keywords: mean–variance, time inconsistency, game theory, extended Bellman system, HJB, efficient frontier.

Sammanfattning

Denna avhandling studerar dynamisk mean–variance-optimering med särskilt fokus på det centrala problemet tidsinkonsistens. I den klassiska enperiodsmodellen uppnås en tydlig avvägning mellan risk–avkastning via kvadratisk optimering, men i flerperiods- eller kontinuerlig-tid kan inte variansen bäddas in i den vanliga rekursiva optimeringsprincipen. För att hantera detta antar vi ett spelteoretiskt intrapersonellt perspektiv där varje tidpunkt betraktas som en “spelare” med gemensamma preferenser men kontroll endast över sitt eget beslut. En tidskonsistent policy definieras som en subspelsperfekt jämvikt: inga framtida spelare"har incitament att avvika, givet att de övriga följer strategin. Vi härleder de utvidgade Bellman-rekursionerna i diskret tid och skissar motsvarande ekvivalents-HJB i kontinuerlig tid. Genom att lösa dessa ekvationer i illustrativa exempel visar vi att man kan generera dynamiskt trovärdiga mean–variance strategier på ett praktiskt gångbart sätt.

Nyckelord: mean–variance, tidsinkonsistens, spelteori, utvidgat Bellmansystem, HJB, effektiv front.

Acknowledgements

I am very grateful to my supervisor, Professor Yishao Zhou, for her guidance, insightful feedback and support throughout this project. Her expertise has been invaluable in shaping the direction of this project.

I would also like to thank my friends, Zacharias Veiksaar and Alvar Mikkola, for engaging discussions that sharpened my ideas and contributed to the completion of this thesis.

Contents

Chapter 1

Introduction

Optimal portfolio selection is one of the main fields of study within modern quantitative finance, minimizing financial risk and maximizing expected return. Since Harry Markowitz's seminal work in 1952 [3], the mean-variance framework has provided a foundation for understanding and designing rational investment strategies. By quantifying risk as the variance of portfolio returns and expressing investor preferences through trade-offs between return and risk, Markowitz laid the groundwork for decades of research in mathematical finance and optimization.

While the classical Markowitz model is formulated in a single-period, static setting, real-world investment decisions unfold dynamically over time. Investors continuously receive new information, adjust their risk preferences, and rebalance their portfolios in response to evolving market conditions. This aspect introduces new layers of complexity and motivates the development of multi-period and continuous-time models that better reflect the actual decision-making environment.

However, extending the mean-variance objective to a multi-period, or dynamic, setting is not straightforward. A major challenge arises from the time-inconsistency of the variance term. The variance of terminal wealth cannot be decomposed additively across time steps, making standard dynamic programming tools inapplicable. This means that the classical Bellman equation becomes insufficient for solving dynamic mean-variance problems and alternative approaches are needed.

In this thesis, we begin by formulating the classical Markowitz problem and solving it using tools from convex optimization and quadratic programming. We then extend the framework to a multi-period setting and highlighting the time-inconsistency challenge. To address the arising issue, we adopt a game-theoretic approach in which each point in time is viewed as a separate player in a dynamic game, and we derive subgame-perfect Nash equilibria through an extended Bellman system. Finally, we consider the continuous-time limit, deriving the Hamilton–Jacobi–Bellman (HJB) equation as the PDE analogue of dynamic programming and discussing its role in continuous-time portfolio optimization.

Chapter 2

The Classical Markowitz Portfolio Optimization

The groundwork for the modern portfolio theory was laid out by Harry M. Markowitz in his seminal paper in 1952 [3]. By introducing a quantitative framework for balancing risk and return, Markwoitz fundamentally transformed portfolio management. By measuring risk as the variance of a portfolio's returns and incorporating the covariances among assets, Markowitz showed how diversification can systematically reduce overall risk even when individual assets carry high volatility. He introduced the concept of the efficient frontier, defining the set of portfolios that deliver the highest expected return for each level of risk, and provided the first practical algorithm to pinpoint these optimal combinations. This allowed investors to make decisions based on their unique risk preferences. In this chapter, we will first formulate the classical Markowitz mean-variance optimization problem, defining the objective, some standard constraints, and then explore how quadratic programming and related optimization techniques can be applied.

2.1 Single-Period Mean-Variance

2.1.1 Formulating the Mean–Variance Optimization Problem

We consider an investor who wishes to allocate their wealth among n financial assets. Let $\mathbf{w} = (w_1, w_2, \ldots, w_n)^\top \in \mathbb{R}^n$ denote the vector of portfolio weights, where each w_i represents the fraction of the total wealth invested in asset i. We also define the expected returns, which are assumed, $\mu \in \mathbb{R}^n$, ,where μ_i represents the expected return and R_i represents the return for asset i. The covariance matrix of asset returns is $\Sigma \in \mathbb{R}^{n\times n}$. where the element i, j in the matrix represents the covariance between asset i and j. The portfolio return R_P is then given by $R_P = \sum_{i=1}^{n} \mathbf{w}_i R_i$.

Proposition 2.1.1 (Expected return). *The expected return of the portfolio is given by*

$$E[R_P] = \sum_{i=1}^{n} w_i \mu_i = \boldsymbol{w}^\top \mu.$$

Proof. Linearity of expectation gives

$$E[R_P] = E[\mathbf{w}^\top R] = \mathbf{w}^\top E[R] = \mathbf{w}^\top \mu.$$

□

Proposition 2.1.2 (Variance of return)**.** *The variance of the portfoltio return is given by*

$$Var(R_P) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov(R_i, R_j) = \boldsymbol{w}^\top \Sigma \boldsymbol{w}.$$

Proof. By definition we get,

$$Var(R_P) = E[(R_P - E[R_P])^2] = E[((\mathbf{w}^\top R) - \mathbf{w}^\top \mu)^2]$$

since $\mathbf{w}^\top \mu$ is constant,

$$E[((\mathbf{w}^\top R) - \mathbf{w}^\top \mu)^2] = E[\mathbf{w}^\top (R-\mu)(R-\mu)^\top \mathbf{w}] = \mathbf{w}^\top E[(R-\mu)(R-\mu)^\top]\mathbf{w} = \mathbf{w}^\top \Sigma \mathbf{w}$$

□

We can now formulate a classical Markowitz optimization problem.

$$\begin{aligned} \min_{\mathbf{w}} \quad & \mathbf{w}^\top \Sigma \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^\top \mu = R, \\ & \mathbf{w}^\top \mathbf{1} = 1, \end{aligned}$$

where R is the desired target return. Let $\mathbf{1}$ be a n number of 1's in a vector. The full investment constraint $\mathbf{w}^\top \mathbf{1} = 1$. then ensures that the entire wealth is allocated among the assets while the expected return constraint $\mathbf{w}^\top \boldsymbol{\mu} = R$ ensures that the target return is met.

2.1.2 Convex Optimization Framework

As seen above we want to maximize return or minimize variance given some constraints, but we need to ensure that we can find a minimum or a maximum for a given optimization problem. According to Bazaraa et. al. [1, Ch. 2] a very important and widely used result is the Weierstrass's Theorem, which guarantees a maximizing (or minimizing) solution under certain conditions. The proof is taken from Bazaraa et. al.

Theorem 2.1.3 (Weierstrass's Theorem)**.** *Let S be a nonempty compact subset and let $f\colon S \longrightarrow R$ be continous on S. Then the problem min$\{f(\boldsymbol{x}) : \boldsymbol{x} \in S\}$ attains its minimum; that is, there exists a minimizing solution to this problem.*

Proof. Since f is continuous on S and S is both closed and bounded, f is bounded below on S. Consequently, since $S \neq \emptyset$, there exists a greatest lower bound $\alpha = \inf\{f(x) : x \in S\}$. Now let $0 < \epsilon_k \downarrow 0$, and consider the set $S_k = \{x \in S : \alpha \leq f(x) \leq \alpha + \epsilon_k\}$ for each $k = 1, 2, \ldots$ By the definition of an infimum, $S_k \neq \emptyset$ for each k, so we may construct a sequence of points $\{x_k\} \subseteq S$ by selecting a point $x_k \in S_k$ for each $k = 1, 2, \ldots$. Since S is bounded, there exists a convergent subsequence $\{x_k\}_k \to \bar{x}$, indexed by the set K. By the closedness of S, we have $\bar{x} \in S$; and by the continuity of f, since $\alpha \leq f(x_k) \leq \alpha + \epsilon_k$ for all k, we have that $\alpha = \lim_{k\to\infty, k\in K} f(x_k) = f(\bar{x})$. Hence, we have shown that there exists a solution $\bar{x} \in S$ such that $f(\bar{x}) = \alpha = \inf\{f(x) : x \in S\}$, so $\bar{x}$ is a minimizing solution. □

Proposition 2.1.4 (Existence of a solution for the mean-variance problem)**.** *Consider the problem*

$$\min_{\mathbf{w}\in\mathbb{R}^n} \mathbf{w}^\top \Sigma \mathbf{w} \quad \textit{subject to} \quad \mathbf{w}^\top \mu = R,\ \mathbf{w}^\top \mathbf{1} = 1,$$

where $\Sigma \succ 0$ is a positive–definite covariance matrix and the affine constraints admit at least one feasible point. Then the problem admits at least one minimizer.

Proof. Let

$$F := \{w \in \mathbb{R}^n :\ \mu^\top w = R,\ \mathbf{1}^\top w = 1\}$$

be the feasible set. By assumption $F \neq \varnothing$. Since the constraints are linear and continuous, F is an intersection of closed hyperplanes, hence closed. The objective $f(w) = w^\top \Sigma w$ is continuous. Because $\Sigma \succ 0$, let $\lambda_{\min} > 0$ denote its smallest eigenvalue. Then for all $w \in \mathbb{R}^n$,

$$f(w) = w^\top \Sigma w \ \geq\ \lambda_{\min} \|w\|^2. \tag{1}$$

Fix any $w_0 \in F$ and set $M := f(w_0) + 1$. Consider the sublevel set

$$S := \{w \in F :\ f(w) \leq M\} = F \cap f^{-1}((-\infty, M]).$$

It is nonempty (contains w_0) and closed (intersection of two closed sets). Moreover, by (1), if $w \in S$ then $\lambda_{\min}\|w\|^2 \leq f(w) \leq M$, hence

$$\|w\| \leq \sqrt{M/\lambda_{\min}},$$

so S is bounded. Thus S is compact (Heine–Borel). By Weierstrass's theorem, the continuous function f attains its minimum on the compact set $S \subset F$. Any minimizer on S is therefore a feasible global minimizer of the original problem. □

We can now guarantee under certain conditions that we will have a solution to the optimization problem. But knowing that there is a solution, is not the same as knowing how to find that solution. Therefore, we examine some properties that Markowitz portfolio optimization problems has and how we can use those properties to solve the optimization problem. We begin by defining the feasible set and properties of convex optimization. We use standard definitions, which can be found in [1, Ch. 2].

Definition 2.1.5 (Convex set)**.** *A set $\Omega \subset \mathbb{R}^n$ is called convex if, for every pair of points $x, y \in \Omega$ and every scalar λ with $0 \leq \lambda \leq 1$, the point*

$$\lambda\, x + (1 - \lambda)\, y$$

also lies in Ω. Equivalently,

$$\forall\, x, y \in C,\ \forall\, \lambda \in [0, 1] : \quad \lambda x + (1 - \lambda) y \ \in\ C.$$

Proposition 2.1.6 (Convexity of the feasible set)**.** *Define the feasible set of portfolios with a target expected reurn R and a full investment constraint as*

$$\Omega = \{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w}^\top \boldsymbol{1} = 1, \boldsymbol{w}^\top \mu = R\}, \quad \boldsymbol{1} = (1, 1, ..., 1).$$

Ω is then a convex subset of $\mathbb{R}^n$

Proof. Take any two portfolios $\mathbf{w}_1, \mathbf{w}_2 \in \Omega$, and any $\lambda \in [0,1]$. Define

$$\mathbf{w} = \lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2.$$

Since $\mathbf{w}^\top \mathbf{1} = 1$ for any $\mathbf{w} \in \Omega$

$$\mathbf{w}^\top \mathbf{1} = \lambda\, \mathbf{w}_1^\top \mathbf{1} + (1-\lambda)\, \mathbf{w}_2^\top \mathbf{1} = \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1,$$

and

$$\mathbf{w}^\top \mu = \lambda\, (\mathbf{w}_1)^\top \mu + (1-\lambda)\, (\mathbf{w}_2)^\top \mu = \lambda\, R + (1-\lambda)\, R = R.$$

Hence $w \in \Omega$. Because every convex combination of feasible portfolios remains feasible, the set Ω is convex. □

Definition 2.1.7 (Convex function)**.** *Let $f : S \to \mathbb{R}$ where S is a nonempty convex set in $\mathbb{R}^n$. The function f is said to be convex on S if*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

for each $x_1, x_2 \in S$ and for each $\lambda \in (0,1)$. The function f is called strictly convex on S if the above inequality is strict, i.e.,

$$f(\lambda x_1 + (1-\lambda)x_2) < \lambda f(x_1) + (1-\lambda) f(x_2)$$

for each $x_1 \neq x_2 \in S$ and for each $\lambda \in (0,1)$.

Proposition 2.1.8 (Convexity of objective function)**.** *Let $\Sigma \in \mathbb{R}^{n\times n}$ be symmetric and positive semidefinite, and define*

$$f\colon \mathbb{R}^n \to \mathbb{R}, \qquad f(\mathbf{w}) = \mathbf{w}^\top \Sigma\, \mathbf{w}.$$

Then f is convex on $\mathbb{R}^n$. Moreover, if Σ is positive definite, *f is strictly convex.*

Proof. Let $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^n$ and $\lambda \in (0,1)$. Consider

$$f\big(\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2\big) = \big(\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2\big)^\top \Sigma\, \big(\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2\big).$$

Expanding the quadratic form gives

$$\lambda^2\, \mathbf{w}_1^\top \Sigma \mathbf{w}_1 + 2\lambda(1-\lambda)\, \mathbf{w}_1^\top \Sigma \mathbf{w}_2 + (1-\lambda)^2\, \mathbf{w}_2^\top \Sigma \mathbf{w}_2.$$

Meanwhile,

$$\lambda\, f(\mathbf{w}_1) + (1-\lambda)\, f(\mathbf{w}_2) = \lambda\, \mathbf{w}_1^\top \Sigma \mathbf{w}_1 + (1-\lambda)\, \mathbf{w}_2^\top \Sigma \mathbf{w}_2.$$

Subtracting, one finds

$$\lambda\, f(\mathbf{w}_1) + (1-\lambda)\, f(\mathbf{w}_2) - f\big(\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2\big) = \lambda(1-\lambda)\big(\mathbf{w}_1 - \mathbf{w}_2\big)^\top \Sigma\, \big(\mathbf{w}_1 - \mathbf{w}_2\big).$$

Since Σ is positive semidefinite, the right–hand side is ≥ 0. Hence

$$f\big(\lambda \mathbf{w}_1 + (1-\lambda)\mathbf{w}_2\big) \leq \lambda\, f(\mathbf{w}_1) + (1-\lambda)\, f(\mathbf{w}_2),$$

f is convex. If Σ is positive definite and $\mathbf{w}_1 \neq \mathbf{w}_2$, then $\big(\mathbf{w}_1 - \mathbf{w}_2\big)^\top \Sigma\, \big(\mathbf{w}_1 - \mathbf{w}_2\big) > 0$, so the above inequality is strict, and f is strictly convex. □

Lemma 2.1.9 (Uniqueness of the minimiser)**.** *Assume $\Sigma \succ 0$, then the quadratic form $f(\mathbf{w}) = \mathbf{w}^\top \Sigma \mathbf{w}$ is strictly convex. A strictly convex function over a convex feasible set possesses at most one minimiser; hence the Markowitz problem admits a unique optimal portfolio.*

Since the feasible set is convex and the objective function is convex on the convex domain, we have a convex program, which means that the optimization problem can be solved through a theorem central to optimization, called KKT-conditions. (See Bazaara et. al. [1])

Theorem 2.1.10 (Karush–Kuhn–Tucker (KKT) conditions)**.** *Let X be a nonempty open set in $\mathbb{R}^n$ and let $f, g_i : X \to \mathbb{R}$ for $i = 1, \dots, m$. Consider the problem*

$$\min f(x) \quad \textit{subject to } x \in X \textit{ and } g_i(x) \le 0 \textit{ for } i = 1, \dots, m.$$

Let $\bar{x}$ be a feasible solution and denote $g(x) = (g_1(x), \dots, g_m(x))$. Suppose that f and g_i for $i \in \mathcal{E} = \{i : g_i(\bar{x}) = 0\}$ are differentiable at $\bar{x}$. Furthermore, suppose that $\nabla g_i(\bar{x})$ for $i \in \mathcal{E}$ are linearly independent. If $\bar{x}$ solves the problem locally, there exist scalars $u_i \ge 0$ for $i = 1, \dots, m$, such that

$$\nabla f(\bar{x}) + \sum_{i \in \mathcal{E}} u_i \nabla g_i(\bar{x}) = 0, \quad u_i g_i(\bar{x}) = 0 \textit{ for } i = 1, \dots, m.$$

Remark 2.1.11. The gradients of the two affine constraints are $\nabla g_1(\mathbf{w}) = \mu$ and $\nabla g_2(\mathbf{w}) = \mathbf{1}$. Since μ and $\mathbf{1}$ are linearly independent, the Linear-Independence Constraint Qualification holds. Consequently, the Karush–Kuhn–Tucker conditions are both necessary *and sufficient* for optimality.

2.1.3 A closed form solution to the classical Markowitz optimization problem

Because the variance function is convex and the feasible region is convex, the Markowitz problem is a convex quadratic program. Hence global optimality can be certified via the Karush–Kuhn–Tucker conditions. We now derive those conditions and solve for the portfolio weights in closed form. Forming the langrangian function by incorporating the constraints with a multiplier, we get:

$$L(\mathbf{w}, \lambda, \nu) = \mathbf{w}^\top \Sigma \mathbf{w} - \lambda \Big(\mathbf{w}^\top \mu - R \Big) - \nu \Big(\mathbf{w}^\top \mathbf{1} - 1 \Big)$$

where λ, ν are the Langrange multipliers. We differentiate the Lagrangian with respect to $\mathbf{w}$ and setting the gradient to equal zero:

$$\frac{\partial L}{\partial \mathbf{w}} = 2\Sigma \mathbf{w} - \lambda \mu - \nu \mathbf{1} = \mathbf{0}.$$

where $\frac{\partial L}{\partial \mathbf{w}}$ is a compact form of the partial derivatives with respect to $\mathbf{w}_1, ..., \mathbf{w}_n$. Thus, we have

$$\Sigma \mathbf{w} = \frac{\lambda}{2} \mu + \frac{\nu}{2} \mathbf{1}.$$

Defining

$$a = \frac{\lambda}{2} \quad \text{and} \quad b = \frac{\nu}{2},$$

the solution can be written as:

$$\mathbf{w} = a\,\Sigma^{-1}\mu + b\,\Sigma^{-1}\mathbf{1}.$$

We now substitute this expression into the expected return constraint:

$$\begin{aligned}\mathbf{w}^\top\mu &= \left(a\,\Sigma^{-1}\mu + b\,\Sigma^{-1}\mathbf{1}\right)^\top \mu \\ &= a\,\mu^\top\Sigma^{-1}\mu + b\,\mathbf{1}^\top\Sigma^{-1}\mu \\ &= aB + bA = R,\end{aligned}$$

where we define

$$B = \mu^\top\Sigma^{-1}\mu \quad \text{and} \quad A = \mathbf{1}^\top\Sigma^{-1}\mu,$$

as well as the full investment constraint:

$$\begin{aligned}\mathbf{w}^\top\mathbf{1} &= \left(a\,\Sigma^{-1}\mu + b\,\Sigma^{-1}\mathbf{1}\right)^\top \mathbf{1} \\ &= a\,\mu^\top\Sigma^{-1}\mathbf{1} + b\,\mathbf{1}^\top\Sigma^{-1}\mathbf{1} \\ &= aA + bC = 1,\end{aligned}$$

with

$$C = \mathbf{1}^\top\Sigma^{-1}\mathbf{1}.$$

The solution satisfies these two equations, yielding the following system of linear equations:

$$\begin{pmatrix} B & A \\ A & C \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} R \\ 1 \end{pmatrix}.$$

The determinant of the coefficient matrix is:

$$\Delta = BC - A^2.$$

If $\Delta \neq 0$ then inverting the matrix, we obtain:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{\Delta}\begin{pmatrix} C & -A \\ -A & B \end{pmatrix}\begin{pmatrix} R \\ 1 \end{pmatrix}.$$

Thus, we have:

$$a = \frac{CR - A}{\Delta}, \quad b = \frac{B - AR}{\Delta}.$$

Substituting a and b back into the expression for $\mathbf{w}$, the optimal portfolio weights are given by:

$$\boxed{\mathbf{w}^* = \Sigma^{-1}\left(\frac{CR - A}{\Delta}\,\mu + \frac{B - AR}{\Delta}\,\mathbf{1}\right).}$$

where

$$A = \mathbf{1}^\top\Sigma^{-1}\mu, \quad B = \mu^\top\Sigma^{-1}\mu, \quad C = \mathbf{1}^\top\Sigma^{-1}\mathbf{1}, \quad \Delta = BC - A^2.$$

Using only initial quantities, we get the optimum:

$$\mathbf{w}^* = \Sigma^{-1}\left(\frac{(\mathbf{1}^\top\Sigma^{-1}\mathbf{1})R - (\mathbf{1}^\top\Sigma^{-1}\mu)}{(\mu^\top\Sigma^{-1}\mu)(\mathbf{1}^\top\Sigma^{-1}\mathbf{1}) - (\mathbf{1}^\top\Sigma^{-1}\mu)^2}\,\mu + \frac{(\mu^\top\Sigma^{-1}\mu) - (\mathbf{1}^\top\Sigma^{-1}\mu)R}{(\mu^\top\Sigma^{-1}\mu)(\mathbf{1}^\top\Sigma^{-1}\mathbf{1}) - (\mathbf{1}^\top\Sigma^{-1}\mu)^2}\,\mathbf{1}\right).$$

We can now formulate our result as a theorem.

Theorem 2.1.12 (Closed-form Markowitz weights)**.** *Let* $A = \mathbf{1}^\top\Sigma^{-1}\mu,\ B = \mu^\top\Sigma^{-1}\mu,\ C = \mathbf{1}^\top\Sigma^{-1}\mathbf{1},\ \Delta = BC - A^2 \neq 0$. *The unique global minimiser of* $\min\{\mathbf{w}^\top\Sigma\mathbf{w} : \ \mathbf{w}^\top\mu = R,\ \mathbf{w}^\top\mathbf{1} = 1\}$ *is*

$$\mathbf{w}^* = \Sigma^{-1}\left(\tfrac{CR-A}{\Delta}\,\mu + \tfrac{B-AR}{\Delta}\,\mathbf{1}\right).$$

2.2 Efficient Frontier

Having established existence and uniqueness of the variance-minimization problem and derived its explicit solution via the KKT conditions, we now describe how these optimal portfolios trace out the so-called efficient frontier in mean–variance space.

Definition 2.2.1 (Efficient Portfolio)**.** *A portfolio w^* is* efficient *if there exists no other portfolio w such that*

$$E[R(w)] \;\ge\; E[R(w^*)] \quad \textit{and} \quad Var(R(w)) \;\le\; Var(R(w^*)),$$

with at least one of the inequalities strict. Equivalently, w^* minimizes variance among all portfolios achieving a given expected return.

Proposition 2.2.2 (Parametric Representation of the Efficient Frontier)**.** *Let*

$$A = \mathbf{1}^\top \Sigma^{-1} \mathbf{1}, \quad B = \mathbf{1}^\top \Sigma^{-1} \mu, \quad C = \mu^\top \Sigma^{-1} \mu, \quad \Delta = AC - B^2 \neq 0.$$

Then for each target return R, the unique variance-minimizing portfolio is

$$w^*(R) = \frac{CR - B}{\Delta} \Sigma^{-1} \mathbf{1} \;+\; \frac{A - BR}{\Delta} \Sigma^{-1} \mu,$$

which satisfies $w^*(R)^\top \mu = R$ and has variance

$$Var(R(w^*(R))) = \frac{C\,R^2 - 2B\,R + A}{\Delta}.$$

As R varies in the interval $(B/A,\, C/B)$, the curve $\big(Var(R(w^*(R))),\, R\big)$ is an upward-opening parabola in the variance–return plane, known as the efficient frontier.

Proof. From the KKT derivation we have

$$w^* = \Sigma^{-1}\big(\gamma\,\mathbf{1} + \lambda\,\mu\big),$$

with γ, λ determined by

$$\mathbf{1}^\top w^* = 1, \quad {w^*}^\top \mu = R.$$

Writing these as

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} \gamma \\ \lambda \end{pmatrix} = \begin{pmatrix} 1 \\ R \end{pmatrix}, \quad \Delta = AC - B^2,$$

we invert to obtain

$$\gamma = \frac{C - BR}{\Delta}, \quad \lambda = \frac{AR - B}{\Delta}.$$

Substitution into $w^* = \Sigma^{-1}(\gamma\,\mathbf{1} + \lambda\,\mu)$ yields the stated expression. Finally,

$$Var(R(w^*)) = {w^*}^\top \Sigma\, w^* = (\gamma\,\mathbf{1} + \lambda\,\mu)^\top \Sigma^{-1} (\gamma\,\mathbf{1} + \lambda\,\mu) = \frac{CR^2 - 2BR + A}{\Delta},$$

completing the proof. □

Using the parametric representation in Proposition 2.2.2, I compute the mean vector μ and covariance matrix Σ from daily (adjusted) prices of the OMXS30 constituents over my sample period, then annualize by a factor of 252. Each stock is plotted at $(\sqrt{\Sigma_{ii}},\, \mu_i)$, and the efficient frontier is drawn as the image of

$$w^*(R) = \Sigma^{-1}\big(\gamma(R)\,\mathbf{1} + \lambda(R)\,\mu\big), \quad \gamma(R) = \tfrac{C - BR}{\Delta},\ \lambda(R) = \tfrac{AR - B}{\Delta},$$

where $A = \mathbf{1}^\top \Sigma^{-1}\mathbf{1}$, $B = \mathbf{1}^\top \Sigma^{-1}\mu$, $C = \mu^\top \Sigma^{-1}\mu$, and $\Delta = AC - B^2$. This curve corresponds to the classical, fully invested unconstrained case. If nonnegativity constraints $w \geq 0$ are imposed instead, the long-only efficient set is obtained by solving a quadratic program for each target return and lies weakly above the unconstrained frontier.

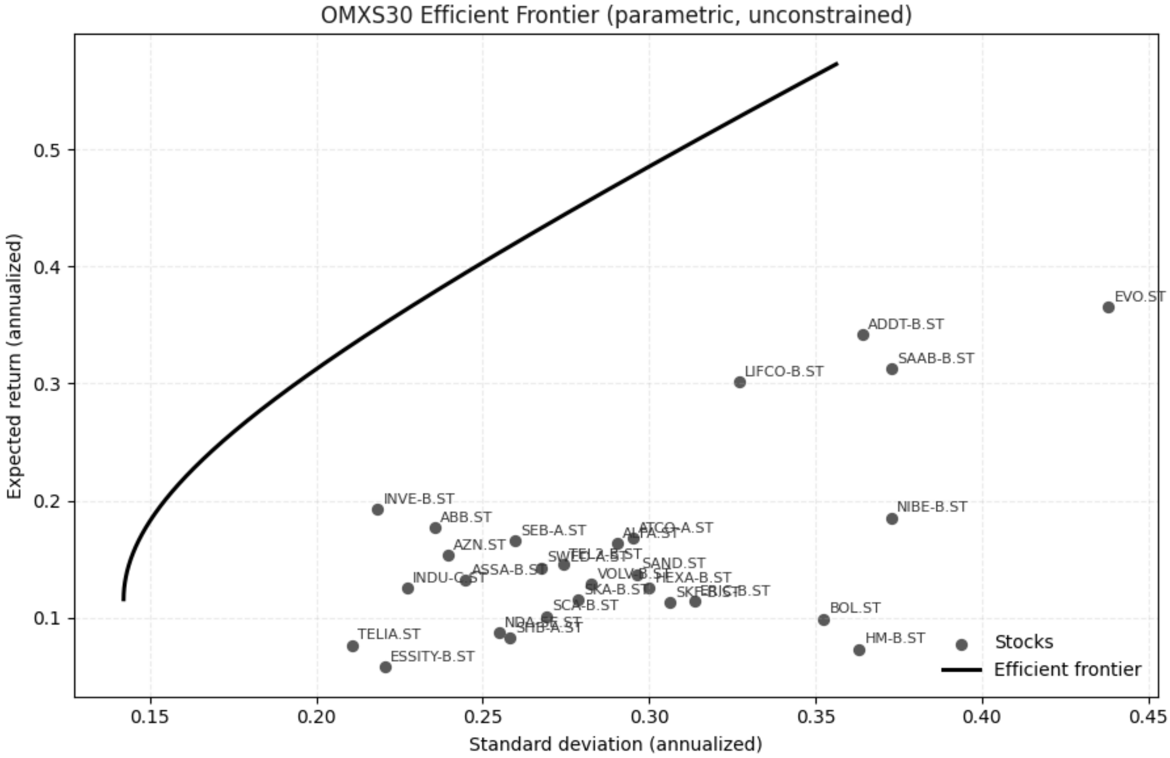

Figure 2.1: Efficient frontier for OMXS30 constituents, plotted using the parametric representation from Proposition 2.2.2. Individual stocks are shown as points at $(\sqrt{\Sigma_{ii}},\, \mu_i)$.

Having obtained explicit formulas for the efficient portfolios $w^*(R)$ as a function of the target return R, we now show that any efficent portfolio can be constructed as a convex combination of two efficient portfolios.

Theorem 2.2.3 (Two-Fund Separation)**.** *Let $R_1 \neq R_2$ be two distinct target returns in $(B/A,\, C/B)$ and let*

$$w^*(R_1),\ w^*(R_2)$$

be the corresponding efficient portfolios from Proposition 1.4.2. Then for any R with $R_1 \leq R \leq R_2$, the efficient portfolio $w^*(R)$ satisfies

$$w^*(R) = \frac{R_2 - R}{R_2 - R_1}\, w^*(R_1) + \frac{R - R_1}{R_2 - R_1}\, w^*(R_2).$$

In other words, every efficient portfolio is a convex combination of the two basis portfolios $w^*(R_1)$ and $w^*(R_2)$.

Proof. Recall the parametric representation

$$w^*(R) = \alpha(R)\,\Sigma^{-1}\mathbf{1} + \beta(R)\,\Sigma^{-1}\mu,$$

where

$$\alpha(R) = \frac{C\,R - B}{\Delta}, \quad \beta(R) = \frac{A - B\,R}{\Delta}, \quad \Delta = AC - B^2.$$

Note that $\alpha(R)$ and $\beta(R)$ are affine functions of R. For $\lambda = \frac{R_2 - R}{R_2 - R_1}$, we have $\lambda R_1 + (1 - \lambda)R_2 = R$, and hence

$$\alpha(R) = \lambda\,\alpha(R_1) + (1-\lambda)\,\alpha(R_2), \quad \beta(R) = \lambda\,\beta(R_1) + (1-\lambda)\,\beta(R_2).$$

Substituting into the representation of $w^*(R)$ gives

$$\begin{aligned} w^*(R) &= \lambda\left[\alpha(R_1)\Sigma^{-1}\mathbf{1} + \beta(R_1)\Sigma^{-1}\mu\right] + (1-\lambda)\left[\alpha(R_2)\Sigma^{-1}\mathbf{1} + \beta(R_2)\Sigma^{-1}\mu\right] \\ &= \lambda\,w^*(R_1) + (1-\lambda)\,w^*(R_2). \end{aligned}$$

Since $0 \le \lambda \le 1$ when $R_1 \le R \le R_2$, this is a convex combination of $w^*(R_1)$ and $w^*(R_2)$, as claimed. □

This result shows that, in practice, an investor need only invest in two efficient "funds" to span the entire efficient frontier: one can vary the target return (and hence overall risk level) simply by mixing these two portfolios in appropriate proportions.

Chapter 3

Mean-Variance in multi-period Portfolio Optimization

One of the key limitation of the Markowtiz model is that it optimizes allocation for one period. In practice, an investor receives new information through out the investment period, changes in market conditions, forecasts and risk preferences, and would therefore want to adjust the allocation dynamically. In the 1950s and 1960s, mathematicians and engineers quickly found a tool to tackle this issue called dynamic programming, which allows for optimizing over multiple periods and adds a control term for adjustments in allocation. Instead of solving just one quadratic program, the investor now at each date asks "Given where I stand today, what decision today will maximize my ultimate objective when combined with pest-possible future actions?" This recursive viewpoint is based on the Bellman's Principle of Optimality, which is the central theorem to find a optimal portfolio in a multi-period time setting.

Attempting to apply Bellman's principle to a mean-variance objective encounters a fundamental obstacle: the variance of terminal wealth cannot be decomposed into multiple periods variances because it depends on covariances across all future returns. Consequently, the risk profile at the initial date is determined by interactions among returns over the entire horizon, making it impossible to isolate today's decision from its impact on subsequent risk. A strategy deemed optimal when planned from the outset will generally lose its optimality upon reaching an intermediate date, since the conditional distribution of remaining returns-and thus the variance to be minimized-has shifted. This breakdown of dynamic consistency prevents straightforward dynamic-programming solutions, necessitating specialized techniques or alternative risk measures to restore time-consistent decision making. In this chapter, mean–variance optimization in a multi-period framework is examined, along with practical solution methods for achieving consistent portfolio strategies.

3.1 Mean-Variance in multi-period time setting

3.1.1 Dynamic Programming

To be able to analyze the problem mentioned above we have to understand what the Bellman optimality is and how the dynamic programming is derived in this section. We introduce the dynamic programming framework as a way to decompose the multi-period mean–variance problem into a sequence of stage-wise decisions. The optimization horizon is split into discrete time steps, each solved conditionally on the previous one. In principle, a Bellman recursion then links today's choice to its impact on tomorrow's expected return and risk, ensuring that each stage accounts for how interim allocations shape the variance of terminal wealth.

For a finite-horizon decision problem the cumulative reward of any particular control sequence $\{u_s\}_{s=0}^{T-1}$ starting from an initial state x_0 is defined as

$$J\big(x_0; u_0, \ldots, u_{T-1}\big) \;=\; \mathbb{E}_{0,x_0}\left[\sum_{s=0}^{T-1} \ell_s\big(x_s, u_s\big) \;+\; \phi\big(x_T\big)\right],$$

where the state trajectory $\{x_s\}$ evolves according to $x_{s+1} = f_s\big(x_s, u_s\big)$. Here, $\ell_s(x_s, u_s)$ denotes the immediate reward collected at stage s when action u_s is taken in state x_s, and $\phi(x_T)$ represents any residual reward contingent on the terminal state. The function J therefore quantifies the total payoff accrued by following that exact sequence of decisions; it does not involve any maximization or optimization, but rather serves as the performance index against which different control paths can be evaluated.

The performance index $J(x_0; u_0, \ldots, u_{T-1})$ measures the total reward obtained by a specific control sequence. To evaluate the best possible outcome from any intermediate stage, we introduce the value function: for each time t and state x, the value function $V_t(x)$ assigns the maximal cumulative reward achievable from that point onward under an optimal policy.

Definition 3.1.1 (Value Function)**.** *For each stage t and state x, the* value function *$V_t(x)$ is the maximal cumulative reward attainable from t, with the state x, to the terminal horizon T. It is defined by*

$$V_t(x) \;=\; \max_{u_t,\ldots,u_{T-1}} \left[\sum_{s=t}^{T-1} \ell_s\big(X_s, u_s\big) \;+\; \phi\big(X_T\big)\right],$$

subject to

$$X_t = x, \qquad X_{s+1} = f\big(X_s, u_s\big) \quad \textit{for } s = t, \ldots, T-1.$$

Remark 3.1.2. To make the exposition easy to follow we assume in this text that the maximum can be reached, instead of involving supremum.

The value function $V_t(x)$ summarizes, for each state x at time t, the greatest reward one can secure going forward. Rather than evaluating every conceivable sequence of actions, Bellman's principle of optimality tells us that an optimal decision today must coincide with an optimal strategy tomorrow. Consequently, $V_t(x)$ satisfies a recursive relationship: the best reward achievable now equals the immediate payoff plus the best reward from the subsequent state. This insight transforms a sprawling multi-stage problem into a chain of single-step optimizations.

For the remainder of this chapter, we consider discrete times $t = 0, 1, \ldots, T$. When conditioning on the information available at time t and current wealth $X_t = x$, we write $E_{t,x}[\cdot]$ for the corresponding conditional expectation and $\mathrm{Var}_{t,x}(\cdot)$ for the conditional variance.

Before we state and prove the optimality results, we recall two key properties of our controlled state process that we will use repeatedly: the Markov property and the law of iterated expectations (tower property).

Definition 3.1.3 (Markov property)**.** *Under any feedback policy $u = \{u_t\}_{t=0}^{T-1}$, the state process $\{X_t\}_{t=0}^{T}$ evolves according to*

$$X_{t+1} \;=\; f_t\big(X_t,\, u_t(X_t)\big), \quad t = 0, \dots, T-1.$$

We say $\{X_t\}$ is Markov *if, for every $t < T$ and any measurable set A,*

$$\Pr\big(X_{t+1} \in A \mid X_0, \dots, X_t\big) \;=\; \Pr\big(X_{t+1} \in A \mid X_t\big).$$

Equivalently, for any integrable random variable Z,

$$\mathbb{E}\big[Z \mid X_0, \dots, X_t\big] \;=\; \mathbb{E}\big[Z \mid X_t\big].$$

Definition 3.1.4 (Law of iterated expectations/Tower Property)**.** *For any integers $0 \le s < t \le T$ and any integrable random variable Y,*

$$\mathbb{E}\big[\,\mathbb{E}\big[Y \mid X_t\big] \;\big|\; X_s\big] \;=\; \mathbb{E}\big[Y \mid X_s\big].$$

That is, conditioning in two steps (first on X_t, then on X_s) is the same as conditioning directly on the earlier state.

Now we can state and prove the fundamental results of dynamic programming, based on the Bellman Principle of Optimality.

Theorem 3.1.5 (Bellman's Principle of Optimality)**.** *Fix an initial stage t and state x, and let $\{u_s^{t,x,*}\}_{s=t}^{T-1}$ denote an optimal control law for the problem starting at (t, x). Then for any subinterval $[m, T]$ with $m \ge t$, the restriction $\{u_s^{t,x,*}\}_{s=m}^{T-1}$ remains optimal for the subproblem beginning at stage m in state $X_m^{t,x}$. Equivalently, for all $s \ge m$ and all y,*

$$u_s^{t,x,*}(y) \;=\; u_s^{m,\, X_m^{t,x},*}(y),$$

where $X_m^{t,x}$ is the state reached at time m under the policy $\{u_s^{t,x,*}\}$.

Proof. Let $u^* = \{u_s^{t,x,*}\}_{s=0}^{T-1}$ be an optimal feedback policy on $[0, T]$. For each $t \ge 0$ and state x, define the reward-to-go functional

$$J_t(x; u) \;=\; \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_k\big(X_s, u_s(X_s)\big) \;+\; \phi(X_T)\Big].$$

Suppose, to the contrary, that there exists t and an alternative feedback law $\tilde{u} = \{\tilde{u}_s\}_{s=t}^{T-1}$ on $[t, T]$ such that

$$\mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(X_s, \tilde{u}_s(X_s)\big) + \phi(X_T)\Big] \ge \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(X_s, u_s^*(X_s)\big) + \phi(X_T)\Big]$$

for some x. Define a "patched" policy $\bar{u}$ on $[0, T]$ by

$$\bar{u}_s(y) = \begin{cases} u_s^*(y), & 0 \le s < t, \\ \tilde{u}_s(y), & t \le s \le T-1. \end{cases}$$

Then by the law of iterated expectations and the Markov property,

$$\begin{aligned} J_0(x_0;\bar{u}) &= \mathbb{E}_{0,x_0}\Big[\sum_{k=0}^{t-1}\ell_s\big(X_s,u_s^*(X_s)\big)\Big] + \mathbb{E}_{0,x_0}\Big[\mathbb{E}_{t,X_t}\big[\sum_{s=t}^{T-1}\ell_s(X_s,\tilde{u}_s(X_s)) + \phi(X_T)\big]\Big] \\ &\geq \mathbb{E}_{0,x_0}\Big[\sum_{s=0}^{t-1}\ell_s\big(X_s,u_s^*(X_s)\big)\Big] + \mathbb{E}_{0,x_0}\Big[\mathbb{E}_{t,X_t}\big[\sum_{s=t}^{T-1}\ell_s(X_s,u_s^*(X_s)) + \phi(X_T)\big]\Big] \\ &= J_0(x_0;u^*), \end{aligned}$$

contradicting the optimality of u^*. Hence the restriction of u^* to $[t,T]$ is optimal. □

Building on Bellman's principle of optimality, we recognize that the maximal reward from (t,x) is decomposed into two parts: the immediate reward obtained by choosing the best control at time t, and the maximal reward-to-go from the resulting next state. Concretely, if u_t^* is the optimal decision at (t,x) and leads to state $x' = f_t(x,u_t^*)$, then the remaining sequence $\{u_{t+1}^*,\dots,u_{T-1}^*\}$ must itself be optimal for the subproblem beginning at $(t+1,x')$. Note that the maximum of the reward-to-go functional is the value function by definition. Hence we get the dynamic programming equation, or the Bellman equation.

Theorem 3.1.6 (The Bellman equation)**.** *The optimal value funciton satisifies the recursive equation*

$$V_t(x) \;=\; \max_{u_t}\Big\{\ell_t\big(x,\,u_t\big) \;+\; V_{t+1}\big(f_t(x,u_t)\big)\Big\}, \qquad t = 0,1,\dots,T-1,$$

with the terminal condition

$$V_T(x) \;=\; \phi(x).$$

Furthermore, the maximum in the equation is realized by the optimal control law $u_t^*(x)$.

Proof. By definition,

$$V_t(x) = \max_{u_t,\dots,u_{T-1}} \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1}\ell_s(X_s,u_s(X_s)) + \phi(X_T)\Big].$$

Fix t and x, and let $u^* = \{u_s^*\}$ achieve the maximum. For any u, define $\tilde{u}$ as

$$\tilde{u}_s = \begin{cases} u, & s = t, \\ u_s^*, & s > t. \end{cases}$$

Then

$$J_t(x;\tilde{u}) = \mathbb{E}_{t,x}\Big[\ell_t(x,u) + \sum_{s=t+1}^{T-1}\ell_s(X_s,u_s^*(X_s)) + \phi(X_T)\Big] = \ell_t(x,u) + \mathbb{E}_{t,x}\big[V_{t+1}(X_{t+1})\big],$$

Using iterated expectations and the Markow property. Since u^* is optimal,

$$V_t(x) = J_t(x;u^*) \geq J_t(x;\tilde{u}) = \ell_t(x,u) + \mathbb{E}_{t,x}[V_{t+1}(X_{t+1})],$$

for all u, and equality holds at $u = u_t^*(x)$. Therefore

$$V_t(x) = \max_u\Big\{\ell_t(x,u) + \mathbb{E}_{t,x}[V_{t+1}(X_{t+1})]\Big\},$$

with terminal condition $V_T(x) = \phi(x)$. □

3.1.2 Mean-variance Problem Setup

Let the weights w_i in the portfolio $w_t \in \mathbb{R}^n$ change per time period $t = 0, 1, ..., T$ according to some matrix A. Then the weights evolve each period according to $w_{t+1} = A\,w_t$. The matrix A captures the passive evolution of weights. To be able to rebalance the portfolio at each time t we add a control vector $u_t \in \mathbb{R}^m$ with a matrix $B \in \mathbb{R}^{n\times m}$ that maps the control to a change in weights. This gives us the dynamic: $w_{t+1} = A\,w_t + B\,u_t$. Let the vector $R_t \in \mathbb{R}^n$ be the return vector with $\mathbb{E}[R_t] = \mu_t$ and $\mathrm{Cov}[R_t] = \Sigma$, we then get:

$$X_{t+1} = X_t + (1 \;+\; w_t^\top R_{t+1}), \quad \mathbb{E}\big[X_t\big] = \mu_t^\top w_t, \quad \mathrm{Var}\big[X_t\big] = w_t^\top \Sigma w_t.$$

At the end of the time period $t \in [0, T]$ we want the trade-off between expected return and variance to be maximized. We can now construct the optimization problem as such:

$$\max_{\{\mathbf{u}_t\}_{t=0}^{T-1}} \quad J\big(\{\mathbf{u}_t\}\big) \;=\; \mathbb{E}\big[\mathbf{X}_T\big] \;-\; \frac{\gamma}{2}\,\mathrm{Var}\big[\mathbf{X}_T\big]$$

where $\gamma > 0$ reflects the investor's risk-aversion parameter. In dynamic programming, we would like to decompose our objective function J into a stage-wise reward function.

3.1.3 The Time-inconsistency of-mean variance problem

Consider the family of reward functionals as before

$$J_t(x; u) \;=\; \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(X_s, u_s(X_k)\big) \;+\; \phi(X_T)\Big].$$

According to Björk et al.[2] , this family is time-consistent (i.e. admits a Bellman recursion) if:

1. Stage rewards depend only on the current time, state, and control: $\ell_s(X_s, u_s)$ may depend on s, X_s, and u_s, but *not* on the initial pair (t, x).
2. The terminal payoff enters solely as $\mathbb{E}_{t,x}[\phi(X_T)]$, with no term of the form $G\big(\mathbb{E}_{t,x}[X_T]\big)$.
3. The terminal reward function ϕ itself does not depend on (t, x).

Under these conditions, the Bellman equation holds for all t and x:

$$V_t(x) \;=\; \max_{u_t}\Big\{\,\ell_t(x, u_t) \;+\; \mathbb{E}_{t,x}\left[V_{t+1}(x_{t+1})\right]\Big\}.$$

Write the mean–variance criterion as

$$J_t(x; u) = \mathbb{E}_{t,x}[X_T] - \frac{\gamma}{2}\Big(\mathbb{E}_{t,x}[X_T^2] - (\mathbb{E}_{t,x}[X_T])^2\Big) = \mathbb{E}_{t,x}[X_T] - \frac{\gamma}{2}\mathbb{E}_{t,x}[X_T^2] + \frac{\gamma}{2}\big(\mathbb{E}_{t,x}[X_T]\big)^2.$$

The first two terms can be absorbed into stagewise, but they depend on the initial (t,x) on which the expectation is conditioned, the final term $\frac{\gamma}{2}(\mathbb{E}_{t,x}[X_T])^2$ is of the forbidden form $G(\mathbb{E}_{t,x}[X_T])$, since it depends nonlinearly on $\mathbb{E}_{t,x}[X_T]$.

Therfore we can't apply Bellaman's Principle of Optimality here, because of the problems time-inconsistency. Hence, we will need to explore another approach that to develop an alternative recursion characterizing a subgame-perfect equilibrium policy.

3.2 The Game-Theoretic solution

3.2.1 The idea of Game Theory

Consider a reward functional

$$J_t(x, u) = \mathbb{E}_{t,x}\left[\sum_{s=t}^{T-1} \ell_s\big(z,\, X_s^u,\, u_s(X_s^u)\big)\right] \;+\; \phi\big(x,\, X_T^u\big) \;+\; G\Big(x,\, \mathbb{E}_{t,x}[X_T^u]\Big),$$

where $(t, x) \in \mathbb{R}^N$ is a fixed initial point, $u = \{u_1, \ldots, u_{T-1}\}$ is a feedback control and T is a fixed time horizon. As shown earlier, this is a time-inconsistent problem.

In a time-inconsistent mean–variance problem, each decision date can be viewed as a separate "player" in a non-cooperative game. At date t, the "self" or player t observes the current state x and selects a control $u_t(x)$, knowing that future selves $t+1, \ldots, T-1$ will follow some predetermined strategy. By treating the family of selves $\{0, 1, \ldots, T-1\}$ as players in a dynamic game, one seeks a profile of feedback controls $\{u_t^*(x)\}$ that no single self has an incentive to deviate from, this is a subgame-perfect Nash equilibrium.

Definition 3.2.1 (Subgame-Perfect Nash Equilibrium)**.** *Let $\{u_s^*\}_{s=0}^{T-1}$ be a candidate feedback law. For any stage t and state x, define the* deviation law $u^{t,x,u}$ *by*

$$u_t^{t,x,u}(x) = u, \quad u_s^{t,x,u}(y) = u_s^*(y) \quad (s = t+1, \ldots, T-1).$$

We say $\{u_s^*\}$ is a subgame-perfect Nash equilibrium if, for every t and x,

$$\sup_u J_t\big(x;\, u^{t,x,u}\big) \;=\; J_t\big(x;\, u^*\big),$$

where $J_t(x; \cdot)$ is the reward functional from (t, x). In that case the equilibrium value function *is*

$$V_t(x) = J_t\big(x;\, u^*\big).$$

No self t can improve the continuation payoff by unilaterally replacing $u_t^*(x)$ with any other u, given that all later selves adhere to the equilibrium strategy.

Remark 3.2.2. An equivalent, more concrete way to obtain the subgame-perfect Nash equilibrium $\{u_t^*\}_{t=0}^{T-1}$ is by backward induction:

1. *Final stage ($n = T-1$)*:

$$u_{T-1}^*(x) = \arg\max_u J_{T-1}(x;\, u)\,.$$

2. *Inductive step*: For $t = T-2, T-3, \ldots, 0$, assume $u_{t+1}^*, \ldots, u_{T-1}^*$ are fixed. Then

$$u_t^*(x) = \arg\max_u J_t\big(x;\, u,\, u_{t+1}^*, \ldots, u_{T-1}^*\big).$$

Proceeding recursively from $T-1$ down to 0 yields the equilibrium law $\{u_t^*\}$ and the associated equilibrium value functions $v_t(x) = J_t(x; u^*)$.

3.2.2 The Extended Bellman System

The main result of the game theoretic approach, which shows that any subgame-perfect equilibrium in the time-inconsistent mean-variance problem can be characterized by a coupled backward recursion involving the value function V_t and the auxiliary sequences f_t and g_t.

Theorem 3.2.3 (Extended Bellman System)**.** *Let $\hat{u} = \{\hat{u}_s\}_{s=0}^{T-1}$ be a subgame-perfect Nash equilibrium. Define for each t and x:*

$$V_t(x) \;=\; J_t\big(x;\hat{u}\big), \quad f_t(x,z) \;=\; \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(z, X_s^{\hat{u}}, \hat{u}_s(X_s s^{\hat{u}})\big) + \phi\big(z, X_T^{\hat{u}}\big)\Big], \quad g_t(x) \;=\; \mathbb{E}_{t,x}\big[X_T^{\hat{u}}\big].$$

Then for $t = 0, 1, \ldots, T-1$ and all x the following recursions hold:

$$V_t(x) = \sup_u \Big\{ \ell_t(x,u) \;+\; \mathbb{E}_{t,x}\big[f_{t+1}(X_{t+1}^u,\, x)\big] \;+ G\big(x, \mathbb{E}_{t,x}[g_{t+1}(X_{t+1}^u)]\big) \Big\} \tag{3.1}$$

$$V_T(x) = \phi(x) + G(x), \tag{3.2}$$

$$f_t(x,z) = \ell_t\big(z, x, \hat{u}_t(x)\big) \;+\; \mathbb{E}_{t,x}\big[f_{t+1}(X_{t+1}^{\hat{u}},\, z)\big],, \tag{3.3}$$

$$f_T(x,z) = \phi(z,x), \tag{3.4}$$

$$g_t(x) = \mathbb{E}_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big], \tag{3.5}$$

$$g_T(x) = x, \tag{3.6}$$

$$V_t(x) = f_t(x,x) + G\big(g_t(x)\big). \tag{3.7}$$

Moreover, one recovers the coupling identity

$$V_t(x) = \; f_t(x,x) \;+\; G\big(x, g_t(x)\big).$$

We will now state and prove propositions and definitions in subsection 3.2.3, which will then be used to prove the Extended Bellman System in subsection 3.2.4. To this end, we consider a more general optimization problem. The reward functional in this case is of the form,

$$f_t(x,z) \;=\; \mathbb{E}_{t,x}\left[\sum_{s=t}^{T-1} \ell_s\big(z,\, X_s^{\hat{u}},\, \hat{u}_s(X_s^{\hat{u}})\big) \;+\; \phi\big(z,\, X_T^{\hat{u}}\big)\right] \;+\; G\Big(t,\, x,\, \mathbb{E}_{t,x}[X_T^u]\Big).$$

Our aim is to derive the recursions that solves the optimization problems stated in remark 3.2.2.

3.2.3 Auxiliary Recursions

To capture how the "self" at time t evaluates both immediate rewards and all future payoffs under the candidate equilibrium, we introduce an auxiliary recursion that bundles the running costs from t onward while allowing a reset of the initial state. This construction is the key to expressing the deviation payoff in a one-step recursive form.

Definition 3.2.4 (Auxiliary Function Sequence)**.** *For each $t = 0, \ldots, T$ and any two states x, z, define*

$$f_t(x,z) \;=\; \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(z,\, X_s^{\hat{u}},\, \hat{u}_s(X_s^{\hat{u}})\big) \;+\; \phi\big(z,\, X_T^{\hat{u}}\big)\Big].$$

In particular, $f_t(x,x) = J_t(x;\hat{u})$.

Note that by definition, $f_t(x,z)$ represents the total expected reward from time t onward when the true state is x but all payoffs are evaluated as if the state were z. In particular, choosing $z = x$ yields the equilibrium value $J_t(x;\hat{u})$. The following proposition demonstrates that f_t admits a one-step backward recursion, paving the way for our extended Bellman system.

Proposition 3.2.5 (Backward Recursion for f_t). *Under the equilibrium policy $\hat{u}_t$, for each $t = 0, \ldots, T-1$ and all x, z:*

$$f_t(x,z) = \ell_t\big(z,\ x,\ \hat{u}_t(x)\big)\ +\ \mathbb{E}_{t,x}\big[f_{t+1}(X^{\hat{u}}_{t+1},\ z)\big],$$

with terminal condition

$$f_T(x,z) = \phi(z,x).$$

Proof. By definition,

$$\begin{aligned} f_t(x,z) =&\ \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(z,\ X^{\hat{u}}_s,\ \hat{u}_s(X^{\hat{u}}_s)\big)\ +\ \phi\big(z,\ X^{\hat{u}}_T\big)\Big] \\ =&\ \mathbb{E}_{t,x}\Big[\ell_t\big(z, X^u_t, u_t(X^u_t)\big) + \sum_{s=t+1}^{T-1} \ell_s\big(z, X^u_s, u_s(X^u_s)\big) + \phi\big(z, X^u_T\big)\Big]. \end{aligned}$$

Since this expectation is taken conditional on $X^u_t = x$, we have

$$\ell_t\big(z, X^u_t, u_t(X^u_t)\big) = \ell_t(z, x, u_t(x)).$$

Thus we can split the expectation:

$$f^u_t(x,z) = \ell_t(z,x,u_t(x))\ +\ \mathbb{E}_{t,x}\Big[\sum_{s=t+1}^{T-1} \ell_s\big(z, X^u_s, u_s(X^u_s)\big) + \phi\big(z, X^u_T\big)\Big].$$

But by definition the right term is exactly $\mathbb{E}_{t,x}[\,f^u_{t+1}(X^u_{t+1}, z)\,]$. Therefore

$$f^u_t(x,z) = \ell_t(z,x,u_t(x)) + \mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1}, z)\big],$$

as claimed. Finally, at $t = T$ the sum over stages is empty, so

$$f^u_T(x,z) = \mathbb{E}_{T,x}\big[\phi(z, X^u_T)\big] = \phi(z,x),$$

completing the proof. □

With the backward recursion for f_t now established, we turn to two additional auxiliary sequences that isolate the mean evolution of the state and the nonlinear penalty. First, define the mean-state function

$$g_t(x) = \mathbb{E}_{t,x}\big[X^{\hat{u}}_T\big],$$

which tracks the expected terminal state under the candidate equilibrium and likewise satisfies a one-step backward recursion. Next, to account for the variance penalty encoded by G, let

$$h_t(x) = G\big(g_t(x)\big),$$

and observe that h_t also propagates backward in a similar fashion. These three recursions for f_t, g_t, and h_t will be combined in Section 3.2.4 to form the Extended Bellman System.

Definition 3.2.6 (Mean Auxiliary Sequence). *For each* $t = 0, \dots, T$ *and state* x*, define*

$$g_t(x) \;=\; \mathbb{E}_{t,x}\big[X_T^{\hat{u}}\big].$$

In particular, $g_T(x) = x$*.*

Lemma 3.2.7 (Backward Recursion for g_t). *Under the equilibrium policy* $\hat{u}$*, for each* $t = 0, \dots, T-1$*:*

$$g_t(x) = \mathbb{E}_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big], \quad g_T(x) = x.$$

Proof. By the law of iterated expectations,

$$g_t(x) = \mathbb{E}_{t,x}[\,X_T^{\hat{u}}\,] = \mathbb{E}_{t,x}\big[\mathbb{E}_{t+1,X_{t+1}^{\hat{u}}}[\,X_T^{\hat{u}}\,]\big] = \mathbb{E}_{t,x}[\,g_{t+1}(X_{t+1}^{\hat{u}})\,].$$

The terminal condition $g_T(x) = x$ follows since $\mathbb{E}_{T,x}[X_T] = x$. □

Definition 3.2.8 (G-Auxiliary Sequence). *Under the equilibrium feedback law* $\hat{u}$*, define for each* $t = 0, 1, \dots, T$ *and state* x *the G-auxiliary function*

$$h_t(x) \;=\; G\big(g_t(x)\big),$$

where g_t *is the mean auxiliary sequence from Definition 3.2.6. In particular,* $h_T(x) = G(x)$*.*

Lemma 3.2.9 (Backward Recursion for h_t). *The sequence* $\{h_t\}$ *satisfies, for each* $t = 0, 1, \dots, T-1$*,*

$$h_t(x) \;=\; G\Big(\mathbb{E}_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big]\Big),$$

with terminal condition $h_T(x) = G(x)$*.*

Proof. By Definition 3.2.8,

$$h_t(x) = G\big(g_t(x)\big) = G\big(\mathbb{E}_{t,x}[X_T^{\hat{u}}]\big).$$

Applying the tower property to $g_t(x) = \mathbb{E}_{t,x}[X_T^{\hat{u}}]$ gives

$$g_t(x) = \mathbb{E}_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big],$$

so

$$h_t(x) = G\Big(\mathbb{E}_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big]\Big).$$

At $t = T$ we have $g_T(x) = x$, hence $h_T(x) = G(x)$. This completes the proof. □

Proposition 3.2.10 (Recursion for $J_t(x; u)$). *Let* $u = \{u_s\}$ *be any feedback policy, and define*

$$J_t(x; u) = \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(X_s^u, u_s(X_s^u)\big) \;+\; \phi\big(X_T^u\big)\Big],$$

and recall the auxiliary functions

$$f_t^u(x, z) = \mathbb{E}_{t,x}\Big[\sum_{s=t}^{T-1} \ell_s\big(z, X_s^u, u_s(X_s^u)\big) \;+\; \phi\big(z, X_T^u\big)\Big].$$

Then for $t = 0, 1, \dots, T-1$ and all x,

$$J_t(x;u) = \ell_t\big(x, x, u_t(x)\big) \ + \ \mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1},\, x)\big],$$

with boundary condition

$$J_T(x;u) = \phi(x,x).$$

Equivalently, since $J_{t+1}(X^u_{t+1};u) = f^u_{t+1}(X^u_{t+1}, X^u_{t+1})$, one may write

$$J_t(x;u) = \ell_t(x,x,u_t(x)) + \mathbb{E}_{t,x}\big[J_{t+1}(X^u_{t+1};u)\big] + \mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1},x)\big] - \mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1},X^u_{t+1})\big].$$

Proof. By definition $J_t(x;u) = f^u_t(x,x)$. Applying the backward recursion of Proposition 3.2.5 with $z = x$ gives

$$f^u_t(x,x) = \ell_t\big(x,x,u_t(x)\big) + \mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1},x)\big],$$

which is the first displayed formula, and $f^u_T(x,x) = \phi(x,x)$ yields $J_T(x;u) = \phi(x,x)$.

For the alternative form, note that by definition $J_{t+1}(X^u_{t+1};u) = f^u_{t+1}(X^u_{t+1}, X^u_{t+1})$. Hence

$$\mathbb{E}_{t,x}\big[f^u_{t+1}(X^u_{t+1},x)\big] = \mathbb{E}_{t,x}\big[J_{t+1}(X^u_{t+1};u)\big] + \mathbb{E}_{t,x}\Big[f^u_{t+1}(X^u_{t+1},x) - \ f^u_{t+1}(X^u_{t+1},X^u_{t+1})\Big],$$

and substituting this into the first recursion produces the second. □

3.2.4 Proof of Theorem 3.2.3

Let $\hat{u} = \{\hat{u}_k\}_{k=0}^{T-1}$ be a subgame-perfect equilibrium and set

$$V_t(x) := J_t(x;\hat{u}), \qquad f_t(x,z) := \mathbb{E}_{t,x}\Big[\sum_{s=n}^{T-1} \ell_k\big(z, X^{\hat{u}}_s, \hat{u}_s(X^{\hat{u}}_s)\big) + \phi\big(z, X^{\hat{u}}_T\big)\Big],$$

$$g_t(x) := \mathbb{E}_{t,x}\big[X^{\hat{u}}_T\big], \qquad h_t(x) := G\big(g_t(x)\big).$$

By the definition of subgame-perfect equilibrium (Definition 3.2.1),

$$V_t(x) = J_t(x;\hat{u}) = \sup_u J_t\big(x; u^{t,x,u}\big),$$

where $u^{n,x,u}$ uses u at time t and $\hat{u}$ thereafter. For such a deviation law, from time $t+1$ onward the policy equals $\hat{u}$, so we may apply the auxiliary recursions proved earlier. Using Proposition 3.2.5 for f_{t+1} and Lemma 3.2.7 for g_{t+1}, we compute

$$\begin{aligned} J_t\big(x;u^{t,x,u}\big) &= \ell_t(x,u) + \mathbb{E}_{t,x}\Big[\sum_{s=t+1}^{T-1} \ell_s\big(x^{\hat{u}}_s, \hat{u}_s(x^{\hat{u}}_s)\big) + \phi(x^{\hat{u}}_T)\Big] + G\Big(\mathbb{E}_{t,x}\big[X_T^{u^{n,x,u}}\big]\Big) \\ &= \ell_t(x,u) + \mathbb{E}_{t,x}\big[f_{t+1}(X^u_{t+1},\, x)\big] + G\Big(\mathbb{E}_{t,x}\big[g_{t+1}(X^u_{t+1})\big]\Big). \end{aligned}$$

Taking the maxiumum over u yields

$$V_t(x) = \sup_u \Big\{\ell_t(x,u) + \mathbb{E}_{t,x}\big[f_{t+1}(X^u_{t+1},x)\big] + G\big(\mathbb{E}_{t,x}[g_{t+1}(X^u_{t+1})]\big)\Big\},$$

and since $J_T(x;\hat{u}) = \phi(x)$ while $h_T(x) = G(x)$ by Definition 3.2.8 and Lemma 3.2.9, we have $V_T(x) = \phi(x) + G(x)$. The auxiliary recursions themselves are exactly Proposition 3.2.5:

$$f_t(x,z) = \ell_t\big(z,x,\hat{u}_n(x)\big) + \mathbb{E}_{t,x}\left[f_{t+1}(X_{t+1}^{\hat{u}},z)\right], \qquad f_T(x,z) = \phi(z,x),$$

and Lemma 3.2.7:

$$g_t(x) = \mathbb{E}_{t,x}\left[g_{t+1}(X_{n+1}^{\hat{u}})\right], \qquad g_T(x) = x,$$

together with Definition 3.2.8 and Lemma 3.2.9 for h_n:

$$h_t(x) = G\Big(\mathbb{E}_{t,x}[g_{t+1}(X_{t+1}^{\hat{u}})]\Big), \qquad h_T(x) = G(x).$$

Finally, by the definitions of V_t, f_t, and g_t,

$$V_t(x) = J_t(x;\hat{u}) = f_t(x) + G\big(g_t(x)\big),$$

which is the coupling identity. Collecting these displays gives precisely the extended Bellman system stated in Theorem 3.2.3.

3.2.5 Solution to a game-theoretic problem

In this section we formulate a multi-period portfolio optimization problem and solve it using the extended Bellman thoerem 3.2.3 to illustrate how it can be used in practice. The problem is taken from Björk et al. (see [2] Section 8.1).

We consider discrete times $t = 0, 1, \ldots, T$, where we assume there is one risky asset with price process S and a risk-free bank account with price process B. Using the increment notation $\Delta X_{t+1} := X_{t+1} - X_t$, assume

$$\Delta S_{t+1} = S_t\, Y_{t+1}, \qquad \Delta B_{t+1} = B_t\, r,$$

where $\{Y_t\}_{t=1}^T$ are i.i.d. random returns and r is a constant short rate. Define $R := 1 + r$ and the excess return $Z_{t+1} := Y_{t+1} - r$. Let u_t denote the dollar amount invested in the risky asset at time t. For a self-financing portfolio with no consumption,

$$\Delta X_{t+1} = rX_t + u_t\,(Y_{t+1} - r), \quad \text{so that} \quad X_{t+1} = R\,X_t + u_t\,Z_{t+1}.$$

We impose no constraints on the control u. The mean–variance reward from (t,x) under a strategy u is

$$J_t(x;u) = E_{t,x}[X_T^u] - \frac{\gamma}{2}\,\mathrm{Var}_{t,x}\big(X_T^u\big),$$

with risk-aversion parameter $\gamma > 0$.

With $F(x) = x - \frac{\gamma}{2}x^2$, $G(x) = \frac{\gamma}{2}x^2$, and $\ell_t \equiv 0$, the extended Bellman system becomes

$$V_t(x) = \sup_u \left\{ E_{t,x}\big[V_{t+1}(X_{t+1}^u)\big] + \frac{\gamma}{2}\Big(E_{t,x}\big[g_{t+1}(X_{t+1}^u)\big]\Big)^2 - \frac{\gamma}{2}E_{t,x}\big[g_{t+1}(X_{t+1}^u)^2\big] \right\},$$
$$V_T(x) = x, \qquad g_t(x) = E_{t,x}\big[g_{t+1}(X_{t+1}^{\hat{u}})\big], \quad g_T(x) = x,$$

where $X_{t+1} = RX_t + u_tZ_{t+1}$, $\mu := E[Z_1]$, and $\sigma^2 := \mathrm{Var}(Z_1)$. We use the linear–affine Ansatz

$$V_t(x) = A_t x + B_t, \qquad g_t(x) = a_t x + b_t.$$

Since $g_{t+1}(X^u_{t+1}) = a_{t+1}(Rx + u\mu) + b_{t+1}$ and $\mathrm{Var}(X^u_{t+1}) = \sigma^2 u^2$, the Bellman equation reduces to

$$A_t x + B_t \;=\; A_{t+1} R x + B_{t+1} + \sup_u \Big\{ -\frac{\gamma}{2}\, a_{t+1}^2 \sigma^2\, u^2 + A_{t+1}\mu\, u \Big\}. \qquad (*)$$

Maximizing the quadratic in $(*)$ yields the equilibrium control

$$\hat{u}_t(x) = \frac{\mu}{\gamma\sigma^2} \cdot \frac{A_{t+1}}{a_{t+1}^2},$$

and plugging back gives the recursions

$$A_t = R A_{t+1}, \qquad B_t = B_{t+1} + \frac{\mu^2}{2\gamma\sigma^2} \frac{A_{t+1}^2}{a_{t+1}^2}, \qquad A_T = 1,\; B_T = 0.$$

From $g_t(x) = E_{t,x}[g_{t+1}(X^{\hat{u}}_{t+1})]$ we obtain

$$a_t = R a_{t+1}, \qquad b_t = b_{t+1} + \frac{\mu^2}{\gamma\sigma^2} \frac{A_{t+1}}{a_{t+1}^2}, \qquad a_T = 1,\; b_T = 0.$$

These linear recursions solve to

$$A_t = a_t = R^{T-t}, \qquad B_t = (T-t)\,\frac{\mu^2}{2\gamma\sigma^2}, \qquad b_t = \frac{\mu^2}{\gamma\sigma^2} \sum_{k=t+1}^{T} R^{-(T-k)}.$$

We can now formulate this as a result,

Proposition 3.2.11. *For the mean-variance problem stated above, we get the*

$$V_t(x) = R^{T-t} x + (T-t)\,\frac{\mu^2}{2\gamma\sigma^2}$$

and the equilibrium control is given by

$$\hat{u}_t(x) = \frac{\mu}{\gamma\sigma^2}\, R^{-(T-t-1)}.$$

3.3 Mean-variance in continous time setting

In the discrete-time setting we allowed the investor to rebalance only at finitely many dates, slicing the horizon $[0,T]$ into T periods. In practice, however, markets trade continuously and information arrives without regard to our calendar. It therefore makes sense to let the rebalancing interval shrink to zero and model portfolio choice in continuous time. We will in this section first reproduce a short, general derivation of the continuous-time HJB deterministic and stochastic. We then discuss time inconsistency in continuous time.

3.3.1 Deterministic Hamilton-Jacobi-Bellman Equation (HJB)

In the discrete-time framework, Bellman's Principle of Optimality allowed us to peel off one rebalancing decision at a time and express the value function via a backward recursion. In moving to continuous time, we replace increments of fixed length by infinitesimal intervals, and the backward-recursion identity transmutes into a partial differential equation. This equation, the Hamilton–Jacobi–Bellman (HJB) equation, encodes in a single relation the trade-off between immediate gains and the evolution of the continuation value.

Consider the deterministic control system $\dot{x}(t) = f(x(t), u(t))$ with $x(0) = x_0$ and $u(t) \in U$, and objective

$$J(u, t) = \Phi(x(T)) + \int_0^T \ell(x(t), u(t))\, dt.$$

For $t \in [0, T]$ and $x \in \mathbb{R}^n$ define the value function

$$V(x, t) = \sup_u \left\{ \int_t^T \ell(x(s), u(s))\, ds + \Phi(x(T)) \;\middle|\; x(t) = x \right\}.$$

Fix $\Delta > 0$ small. Keep the control constant at $u \in U$ on $[t, t + \Delta)$. Then

$$x(t + \Delta) = x + f(x, u)\Delta + o(\Delta), \qquad \int_t^{t+\Delta} \ell(x(s), u)\, ds = \ell(x, u)\Delta + o(\Delta).$$

By optimality from $t + \Delta$ onward, the dynamic programming relation holds:

$$V(x, t) = \sup_u \Big\{ \ell(x, u)\Delta + V\big(x + f(x, u)\Delta,\ t + \Delta\big) \Big\} + o(\Delta).$$

Assuming V is C^1 in (x, t), a first–order Taylor expansion gives

$$V\big(x + f(x, u)\Delta,\ t + \Delta\big) = V(x, t) + \frac{\partial V}{\partial t}(x, t)\Delta + \nabla_x V(x, t)\cdot f(x, u)\,\Delta + o(\Delta).$$

Substituting into the previous display, cancelling $V(x, t)$, dividing by Δ, and letting $\Delta \to 0$ yields

$$0 = \frac{\partial V}{\partial t}(x, t) + \sup_u \big\{ \ell(x, u) + \nabla_x V(x, t) \cdot f(x, u) \big\}.$$

The terminal condition is $V(x, T) = \Phi(x)$. This is the Hamilton–Jacobi–Bellman equation for the deterministic problem. We can now state this in a theorem.

Theorem 3.3.1 (Hamilton–Jacobi–Bellman equation, deterministic case)**.** *Let $\dot{x}(t) = f(x(t), u(t))$, $x(0) = x_0$, with controls u taking values in a compact set $U \subset \mathbb{R}^m$. Given a running reward $\ell : \mathbb{R}^n \times U \to \mathbb{R}$ and a terminal reward $\Phi : \mathbb{R}^n \to \mathbb{R}$, define*

$$V(x, t) = \sup_{u(\cdot)} \Big\{ J(u, t) \ :\ \dot{x}(s) = f(x(s), u(s)),\ t \le s \le T,\ x(t) = x \Big\}.$$

Assume f, ℓ, Φ are continuous and $V \in C^1((0, T) \times \mathbb{R}^n)$. Then V satisfies the Hamilton–Jacobi–Bellman partial differential equation

$$\frac{\partial V}{\partial t}(x, t) + \sup_{u \in U} \big\{ \ell(x, u) + \nabla_x V(x, t)\cdot f(x, u) \big\} = 0, \qquad (x, t) \in \mathbb{R}^n \times [0, T),$$

with terminal condition

$$V(x,T) = \Phi(x), \qquad x \in \mathbb{R}^n.$$

Conversely, if a C^1 function V solves this boundary-value problem, any measurable selector

$$u^*(t) \in \arg\max_{u\in U}\big\{\ell(x^*(t),u) + \nabla_x V(x^*(t),t)\cdot f(x^*(t),u)\big\}$$

is an optimal control and $V(x_0,0) = \sup_u J(u,0)$.

3.3.2 Stochastic Hamilton-Jacobi-Bellman equation (HJB)

We now sketch the stochastic analogue and emphasize what changes relative to the deterministic case since the topics on the stochastic calculus is beyond the scope of this thesis. Let the state evolve as the Itô SDE

$$dX(t) = f\big(X(t),u(t)\big)\,dt + \sigma\,dW(t), \qquad t \geq 0, \qquad X(0) = x_0,$$

where $W(t)$ is a Brownian motion (see, e.g., [4]). As before, for a fixed (x,t) we consider the problem started at (x,t):

$$\begin{cases} dX(s) = f\big(X(s),u(s)\big)\,ds + \sigma\,dW(s), & t \leq s \leq T, \\ X(t) = x. \end{cases}$$

Hence for any $\tau \in [t,T]$,

$$X(\tau) = x + \int_t^\tau f\big(X(s),u(s)\big)\,ds + \sigma\big(W(\tau) - W(t)\big).$$

Define the expected objective

$$P_{x,t}(u(\cdot)) = \mathbb{E}\left[\int_t^T \ell\big(X(s),u(s)\big)\,ds + \Phi\big(X(T)\big)\right],$$

and the value function

$$V(x,t) = \sup_u P_{x,t}(u(\cdot)).$$

To use dynamic programming we (i) identify a PDE satisfied by V and (ii) use it to construct an optimal control. With the same small-step heuristic as before, for $\Delta > 0$ small we write

$$V(x,t) \;\approx\; \sup_u \Big\{\ell(x,u)\,\Delta + V\big(x + f(x,u)\Delta + \sigma\big(W(t+\Delta) - W(t)\big),\, t+\Delta\big)\Big\}. \tag{3.8}$$

By Itô's formula assuming V is C^2 (see, e.g., [4]) and a first-order Taylor expansion in t and second order in x,

$$V\big(x + f(x,u)\Delta + \sigma\big(W(t+\Delta) - W(t)\big),\, t+\Delta\big) = V(x,t) + \frac{\partial V}{\partial t}(x,t)\,\Delta + \nabla_x V(x,t)\,f(x,u)\,\Delta + \frac{1}{2}\,\sigma^2 \sum_{i=1}^n \frac{\partial^2 V}{\partial x_i^2}(x,t)\,\Delta + o(\Delta).$$

Substituting this in (3.8), cancelling $V(x,t)$ on both sides, dividing by Δ, and letting $\Delta \to 0$ yields the HJB equation

$$0 = V_t(x,t) + \frac{1}{2}\,\sigma^2 \sum_{i=1}^{n} \frac{\partial^2 V}{\partial x_i^2}(x,t) \;+\; \sup_u \Big\{ \ell(x,u) + \nabla_x V(x,t)\, f(x,u) \Big\},$$

with terminal condition

$$\Phi(x) = V(x,T).$$

The σ^2-term arises from the quadratic variation of Brownian motion; heuristically, for components W^i, W^j one has

$$dW^i\, dW^j = \begin{cases} dt, & i = j, \\ 0, & i \neq j. \end{cases}$$

For a rigorous derivation, see [4]. The same logic can be extended to the *equilibrium* (time-consistent) version by augmenting the state with the auxiliary quantities and deriving the corresponding extended HJB, in parallel with the discrete-time extended Bellman recursion.

3.3.3 Time inconsistency in classical portfolio optimization

We consider an investor who allocates wealth among n assets with (possibly stochastic) instantaneous return vector $r(t) = (r_1(t), \ldots, r_n(t))^\top$. Let $w(t) = (w_1(t), \ldots, w_n(t))^\top$ denote the portfolio weights at time t. The wealth-weighted return is

$$\bar{r}(t) \;=\; w(t)^\top r(t).$$

Even in the absence of trading, differences in asset returns cause the weights to drift. Writing $M(t) := r(t) - \bar{r}(t) I_n$, the no-trade drift is $\dot{w}(t) = M(t)w(t)$. The controlled weight dynamics are

$$\dot{w}(t) = M(t)w(t) + B\,u(t).$$

With self-financing and no frictions, total wealth $X(t)$ evolves at the portfolio return:

$$\dot{X}(t) = \bar{r}(t)\, X(t).$$

The investor evaluates terminal wealth by the continuous-time mean–variance criterion with risk aversion $\gamma > 0$. The control problem is

$$\max_{u(\cdot)} \quad J(u(\cdot)) \;=\; \mathbb{E}[X(T)] \;-\; \frac{\gamma}{2}\,\mathrm{Var}(X(T)),$$

$$\text{subject to} \quad \begin{cases} \dot{w}(t) = M(t)w(t) + B\,u(t), \\ \dot{X}(t) = \bar{r}(t)\, X(t), \qquad \bar{r}(t) = w(t)^\top r(t), \end{cases}$$

where expectations are taken with respect to the randomness in the return process $r(\cdot)$.

The classical Hamilton–Jacobi–Bellman equation is built on the Bellman principle of optimality, which requires that an investor's strategy, once chosen, remains optimal at every subsequent instant. In particular, one assumes that the criterion can be written as the sum of an instantaneous reward and the value-to-go from the next state. This time–consistency underpins the one-step-ahead decomposition and allows the HJB PDE to characterize the true optimum.

In the continuous-time mean–variance framework, however, the objective

$$J(u(\cdot)) \;=\; \mathbb{E}[X(T)] \;-\; \frac{\gamma}{2}\,\mathrm{Var}(X(T)) \;=\; \mathbb{E}[X(T)] \;-\; \frac{\gamma}{2}\Big(\mathbb{E}[X(T)^2] - (\mathbb{E}[X(T)])^2\Big)$$

cannot be written in the form described by the HJB equations, because the term $(\mathbb{E}[X(T)])^2$ depends nonlinearly on the full expectation of terminal wealth. There is no "infinitesimal" reward L whose integral reproduces the variance penalty, and so the Bellman recursion breaks down.

As a consequence, any control law obtained by solving the standard HJB is only a pre-commitment solution: it maximizes the mean–variance criterion from the initial time, but once the process evolves, the investor has an incentive to renege on that plan. Future selves will reoptimize based on the updated conditional distribution of $X(T)$, and the original HJB strategy is no longer optimal. This time-inconsistency motivates the search for an alternative solution concept that remains enforceable at every instant; see the game-theoretic equilibrium approach in continuous time [2].

Chapter 4

Conclusion

This thesis revisited mean-variance portfolio optimization from the single-period Markowitz program to multi-period and continuous-time settings, with an emphasis on the central obstacle that emerges the moment time enters the picture: time inconsistency. In single-period, the efficient frontier follows transparently from a convex quadratic program and admits a compact parametric description. None of that changes the moment we allow the investor to revise decisions over time, but the objective does: the variance term violates the conditions set forth by Björk et al. [2] for time-consistency.

The game-theoretic viewpoint provides a clean resolution of this conflict. The multi-period investor faces an intrapersonal game in which each date-t self is a player, sharing the same preferences but controlling different segments of the policy. A pre-commitment solution, obtained by treating the problem as if the initial self could bind all future selves, is generally not credible: when time comes, a later self can do strictly better for that self by deviating. The correct dynamic solution concept is therefore an equilibrium policy that is optimal against itself at every subsequent date. Formally, this is a subgame-perfect equilibrium of the intrapersonal game: given the continuation strategy, no single self has an incentive to deviate locally.

Operationally, the equilibrium perspective replaces the standard Bellman equation by an extended system that tracks, in addition to the usual value function, certain auxiliary objects that propagate the non-separable parts of the objective. In discrete time this appears as a coupled pair of recursions: one for the equilibrium value and one (or more) for the auxiliary continuation terms.

Two broader lessons emerge. First, mean-variance is not "wrong" dynamically; it is incomplete without an equilibrium notion. Once the problem is framed as an intrapersonal game, the apparent contradiction with dynamic programming disappears and is replaced by a well-posed fixed-point problem. Second, the equilibrium perspective is not a mere technicality: it has quantitative consequences for portfolio paths, hedging motives, and the value of information, and it provides the right language for comparing policies that are implementable over time.

Bibliography

[1] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley-Interscience, Hoboken, NJ, 3 edition, 2006.

[2] Tomas Björk, Mariana Khapko, and Agatha Murgoci. *Time-Inconsistent Control Theory with Finance Applications*. Springer Finance. Springer, Cham, Switzerland, 2021. doi:10.1007/978-3-030-81843-2.

[3] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi:10.2307/2975974.

[4] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer, 7 edition, 2013.