Springer Finance

Tomas Björk
Mariana Khapko
Agatha Murgoci

Time-Inconsistent Control Theory with Finance Applications

Springer

Springer Finance

The Springer Finance series, launched in 1998, is addressed to students, academic researchers and practitioners working on increasingly technical approaches to the analysis of financial markets. It covers mathematical and computational finance broadly, reaching into foreign exchange, term structure, risk measure and management, portfolio theory, equity derivatives, energy finance and commodities, financial economics.

All titles in this series are peer-reviewed to the usual standards of mathematics and its applications.

More information about this series at http://www.springer.com/series/3674

Tomas Björk • Mariana Khapko • Agatha Murgoci

Time-Inconsistent Control Theory with Finance Applications

Springer

Tomas Björk
Stockholm School of Economics
Stockholm, Sweden

Mariana Khapko
University of Toronto
Toronto, ON, Canada

Agatha Murgoci
Ørsted
Hellerup, Denmark

ISSN 1616-0533 ISSN 2195-0687 (electronic)
Springer Finance
ISBN 978-3-030-81842-5 ISBN 978-3-030-81843-2 (eBook)
https://doi.org/10.1007/978-3-030-81843-2

Mathematics Subject Classification: 49L20, 49L99, 49N10, 49N90, 60J05, 60J20, 60J70, 91A10, 91A80, 91B02, 91B50, 91B70, 91G10, 91G80, 91-02

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The purpose of this book is to present an overview of, and introduction to, the time-inconsistent control theory developed by the authors during the last decade. The theory is developed for discrete as well as continuous time, and the kernel of the content is drawn mainly from our journal articles Björk and Murgoci (2014), Björk et al. (2017), and Björk et al. (2014). Starting from these articles, we have included more examples and substantially simplified the exposition of the discrete-time theory compared with that of Björk and Murgoci (2014). Moreover, we have extended our framework to study time-inconsistent stopping problems, including stopping problems with non-exponential discounting, mean-variance objective, and distorted probabilities. Alongside our own results we have included discussions of recent developments in the field. In order to make the text more self-contained we have also added a brief recapitulation of optimal control and stopping in discrete and continuous time.

It is important to recognize that the analysis of time inconsistency has a long history in economics and finance literature. The idea of time consistency was alluded to by Samuelson (1937) when introducing the most commonly used time-discounting method in economics, exponential discounting. Strotz (1955) in his seminal paper pointed out that any other choice of discounting function, apart from the exponential case, will lead to a dynamically inconsistent problem. More generally, in order to achieve dynamic consistency and analyze the agent's problem with a standard toolkit, one needs to make specific and often restrictive assumptions about the objective functional the agent is maximizing (or minimizing). If the agent's objective does not conform to these assumptions (see Remark 1.1), time consistency fails to hold and the usual concept of optimality does not apply. Loosely speaking, this means that the agent's tastes change over time, so that a plan for some future period deemed optimal today is not necessarily optimal when that future period actually arrives.

How can one handle time-inconsistent problems? One approach is to look for a solution that is optimal today, ignoring the time inconsistency. Strotz (1955) refers to an agent who fails to recognize the time inconsistency issue and adopts such an approach as "spendthrift." The term later coined in the literature is "naïve."

The naïve agent's strategies are myopic and constantly changing. Strotz (1955) also outlines two approaches for modeling a "sophisticated" agent who is aware of the time inconsistency of their tastes: the strategy of pre-commitment and that of consistent planning. In the former case, the agent decides on a plan of action that is optimal today and commits to it, ignoring the incentives to revise it in the future. In the latter case, the agent internalizes the incentives to deviate and treats them as a constraint, thus aiming to arrive at a deviation-proof solution. Importantly, in both cases the agent recognizes that their "today self" and their "future selves" may have conflicting tastes.

In our work we choose to follow the consistent planning approach of Strotz (1955) in that an agent's strategies are taken to be the outcome of an intrapersonal game whose players are successive incarnations of the same agent. Essentially, we replace the usual concept of optimality with a more general concept of intrapersonal equilibrium and look for Nash subgame-perfect equilibrium points (Selten, 1965). Using this game-theoretic approach, we present an extension of the standard dynamic programming equation, in the form of a system of nonlinear equations, for determining the intrapersonal equilibrium strategy. This extended system, loosely speaking, accounts for the incentives to deviate as time evolves and an agent's tastes change. This means that, for a general Markov process and a fairly general objective functional, we obtain a plan that the agent will actually follow. We fully acknowledge that, while our focus here is on the game-theoretic approach, the other approaches that have been studied in the literature—the problem of a naïve agent who reoptimizes as time goes by or that of a sophisticated agent who is able to pre-commit—are both interesting and relevant.

The structure of the book is as follows. Following an introductory chapter, in Part I we start by providing a brief review of optimal stochastic control in discrete time. We present the standard results of discrete-time dynamic programming theory and illustrate them by solving a standard linear quadratic regulator problem and a simple discrete-time equilibrium model. Part II contains the main results for stochastic time-inconsistent control problems in discrete time, originally developed in Björk and Murgoci (2014), together with extensions and applications. We first give an account of time-inconsistent control theory[1] and present a number of interesting extensions, including the generalization of the additively separable expected utility model. The rest of the chapters in Part II discuss concrete examples of the general theory. The applications we present include control problems with non-exponential discounting and with mean-variance objective, time-inconsistent regulator problems, and a time-inconsistent version of the simple equilibrium model. In Part III, we summarize the continuous-time optimal control theory.

[1] The term "time-inconsistent" control was coined in the literature to emphasize the contrast with optimal control theory, which deals with time-consistent problems. This terminology may seem a bit confusing because, while the problem itself is inherently time inconsistent, the controls that we aim to find are deviation-proof, meaning that they are time consistent. To be as precise as we can be, we are studying time-consistent behavior of non-committed sophisticated agents who are maximizing (or minimizing) a time-inconsistent objective functional.

We first give a brief introduction to standard dynamic programming results in continuous time and then proceed to illustrate the theory with a number of examples. In Part IV, we build on results developed in Björk et al. (2017) for a class of continuous-time stochastic control problems that, in various ways, are time inconsistent. The structure of chapters in this part intentionally parallels that of Part II, reflecting the fact that the discrete-time setting serves as a natural starting point for the limiting arguments we use in the continuous-time case. In Part V, we briefly summarize the standard optimal stopping theory. Then, in Part VI,[2] we extend our methods in order to tackle time-inconsistent stopping problems in discrete and continuous time, including stopping problems under prospect-type distorted probabilities. Examples studied in this last part include a time-inconsistent version of the simple secretary problem, costly procrastination, and the problem of selling an asset (or investing in a project) that becomes time inconsistent if we allow for non-exponential discounting or mean-variance preferences. Finally, we review some basic concepts from arbitrage theory in the appendix.

This text is intended for graduate students and researchers in finance and economics who are interested in the issues of time inconsistency that prevail in many dynamic choice problems. In this book we aim to give the main arguments on how to handle time-inconsistent problems, outline the guiding intuition, and illustrate the general theory with a number of examples that are relevant in finance. While the continuous-time applications are likely to be the main focus of mathematical finance researchers, the discrete-time examples largely target the economics readership. Our focus on presenting main arguments and ideas means that we often go lightly on some of the more technical issues, so measurability and integrability issues are at times swept under the carpet.

Since the book is intended to be self-contained, it contains a brief summary of optimal control and stopping in discrete and continuous time. The reader comfortable with these standard results is welcome to skip the summary chapters and proceed to the more complicated time-inconsistent framework directly. We acknowledge that there are many excellent textbooks on optimal stopping and control. This is why we keep our discussion of the standard theory brief and refer the reader to the extensive literature on the subject for further information. The summary of the standard results is included for completeness as well as to allow for comparing and contrasting the "intrapersonal equilibrium" results with the standard optimal results in concrete applications.

It is also worth noting that a number of open problems remain for future research. First, we note that existence and/or uniqueness remain to be proved for solutions of the extended Bellman system in a number of settings. Second, the present theory depends critically on the Markovian structure. It is intriguing to follow the new

[2] Note that Part VI was unfortunately finalized without Tomas Björk. The results presented in this last part of the book are the product of numerous discussions between the authors over the last few years. However, any remaining errors or omissions in this part are the responsibility of Mariana Khapko and Agatha Murgoci.

developments in the literature that operate without this assumption. Third, in this book we present extensions of the standard dynamic programming results for time-inconsistent problems, and it would be very interesting to see whether there exists an efficient martingale formulation for these problems. Further open research problems are discussed in Björk and Murgoci (2014) and Björk et al. (2017).

Notes on the literature can be found at the end of most chapters. These notes provide discussions of the relevant literature, emphasizing new developments and alternative approaches. They provide the reader with opportunities to explore each topic further. We have tried to keep the reference list as complete as possible, including both the work that has influenced us and also the new papers where our methodology has been used. Any serious omission is unintentional.

Dedication

Here we would like to pay our tribute to Tomas Björk, without whom this book would not exist. Tomas was an internationally recognized figure in financial mathematics, a brilliant scholar and teacher, and a caring colleague. But to us he was so much more. He was our role model, our mentor, and a close friend. We are forever grateful for all the time, guidance, support, and encouragement he so generously bestowed upon us. We miss him, dearly, every day.

In its obituary for Tomas, the Bachelier Finance Society rightly remarked that he "was still active in his beloved mathematics up to the last day." Indeed, this book, putting together a decade of his interest in time-inconsistent problems, was among the last things Tomas was working on. We very much hope that we have been able to complete it in accordance with his vision.

Toronto, ON, Canada Mariana Khapko
Hellerup, Denmark Agatha Murgoci
April 2021

Acknowledgements

We are very grateful for helpful comments received from Suleyman Basak, Eric Böse-Wolf, Boualem Djehiche, Ivar Ekeland, Ali Lazrak, Kristoffer Lindensjö, Traian Pirvu, Martin Schweizer, Mogens Steffensen, Xunyu Zhou, and from several anonymous referees at *Finance and Stochastics*, *Mathematical Finance*, and *Springer Finance*. Finally we would like to express our gratitude to Catriona Byrne, our editor at Springer, for her invaluable advice and patience, and Tom Moss Gamblin, the copy editor who worked on the book, for helping us make the text as clear and understandable as possible.

Contents

Part II Time-Inconsistent Control in Discrete Time

Part III Optimal Control in Continuous Time

Part IV Time-Inconsistent Control in Continuous Time

Part V Optimal Stopping Theory

Chapter 1
Introduction

In this chapter we introduce the concept of time inconsistency in dynamic choice problems. We start by reviewing the key ideas of dynamic programming and listing the main reasons for the time consistency in a given problem. We then present a number of seemingly simple examples from financial economics in which time consistency fails to hold. To tackle these (and similar problems), we outline the different approaches developed in the literature for handling time inconsistency in a dynamic stochastic control setting. In this book, we take the game-theoretic approach and look for subgame-perfect equilibrium strategies. Additionally, we emphasize that, similar to control problems, a stopping problem can be time inconsistent if it does not admit a Bellman optimality principle.

1.1 A Standard Control Problem

A standard discrete-time stochastic optimal control problem is that of maximizing (or minimizing) a functional of the form

$$E\left[\sum_{n=0}^{T} H_n\left(X_n, u_n\right) + F(X_T)\right],$$

where X is some controlled Markov process and X_n is the value of this random process at time n, u_n is the control applied at time n, and H, F are given real-valued functions. A typical example is when X is a controlled scalar stochastic equation of the form

$$X_{n+1} = \mu(X_n, u_n, Y_{n+1}),$$

T. Björk et al., *Time-Inconsistent Control Theory with Finance Applications*,
Springer Finance, https://doi.org/10.1007/978-3-030-81843-2_1

where Y is the stochastic noise process, μ is a given mapping, and we have some initial condition $X_0 = x_0$. For simplicity we assume that there are no constraints on the scalar control u_n.

The object of the present text is to study problems that are similar to the one stated above, but where there is also an element of "time inconsistency." In order to understand exactly why and how our problems are different from the standard one above, and what the term "time inconsistency" really means, we need to review, very briefly, the main ideas of dynamic programming.

1.2 Dynamic Programming and the Concept of Time Consistency

A standard way of attacking a problem like the one above is by using Dynamic Programming, so we now briefly recall some of the main ideas (see Chap. 2 for details). We restrict ourselves to *control laws*, that is the control applied at time k, given that $X_k = y$, is of the form $\mathbf{u}_k(y)$; the control law $\mathbf{u}$ is a deterministic function of the variables k and y. We then embed our original control problem in a family of problems indexed by the initial point. More precisely we consider, for every (n, x), the problem $\mathcal{P}_{n,x}$ of maximizing the reward functional

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T} H_k\left(X_k, u_k\right) + F(X_T)\right]$$

given the initial condition $X_n = x$. Denoting the optimal control law for $\mathcal{P}_{n,x}$ by $\hat{\mathbf{u}}_k^{n,x}(y)$ (where $n \leq k \leq T-1$) and the corresponding optimal value function by $V_n(x)$ we see that the original problem corresponds to the problem $\mathcal{P}_{0,x_0}$.

We note that *ex ante* the optimal control law $\hat{\mathbf{u}}_k^{n,x}(y)$ for the problem $\mathcal{P}_{n,x}$ must be indexed by the initial point (n, x) but, as is well known, problems of the kind described above turn out to be *time consistent* in the sense that the *Bellman optimality principle* applies, which roughly says that the optimal control is independent of the initial point. More precisely: if a control law is optimal on the time interval $\{n, \ldots, T\}$, then it is also optimal for any subinterval $\{m, \ldots, T\}$ where $n \leq m$, or more formally

$$\hat{\mathbf{u}}_k^{n,x}(y) = \hat{\mathbf{u}}_k^{m,z}(y),$$

for all states x, y, z and for all times $n \leq m \leq k$.

Given the Bellman principle, it is easy to derive the Bellman equation

$$\begin{aligned} V_n(x) &= \sup_{u \in R} \left\{ H_n(x,u) + E_{n,x}\left[V_{n+1}(X^u_{n+1}) \right] \right\}, \\ V_T(x) &= F(x), \end{aligned}$$

for the determination of V.

We end this section by listing some important conditions concerning time consistency, and in the next section we will see some seemingly quite natural problems where these conditions do not hold, thus giving rise to time inconsistency.

Remark 1.1 Some of the main reasons for the time consistency of the indexed family of problems $\left\{\mathcal{P}_{n,x} : x \in \mathbf{R}, n = 0, 1, 2, \ldots \right\}$ are as follows.

- The term $H_k\left(X_k, u_k\right)$ in the problem $\mathcal{P}_{n,x}$ is allowed to depend on k, X_k and u_k. It is *not* allowed to depend on the initial point (n, x).
- The terminal evaluation term is allowed to be of the form $E_{n,x}\left[F(X_T)\right]$, i.e. the expected value of a nonlinear function of the terminal value X_T. We are *not* allowed to have a term of the form $G(E_{n,x}\left[X_T\right])$, which is a nonlinear function of the expected value.
- We are *not* allowed to let the terminal evaluation function F depend on the initial point (n, x).

1.3 Some Disturbing Examples

We will now consider four seemingly simple examples from financial economics in which time consistency fails to hold. In all these cases we consider a financial market with a risky asset as well as a risk-free asset with rate of return r. We denote by X the market value of a self-financing portfolio, and by c the consumption process. We now consider three indexed families of optimization problems. In all cases the (naïve) objective is to maximize the objective functional $J_n(x, \mathbf{u})$, where (n, x) is the initial point and $\mathbf{u}$ a shorthand expression for the control strategy, consisting of consumption and portfolio weights.

1. Non-exponential discounting:

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T-1} \beta(k-n) U(c_k) + \beta(T-n) F(X_T) \right].$$

In this problem U is the local utility of consumption, F is the utility of terminal wealth, and β is the *discounting function*. This problem differs from a standard problem by the fact that the initial point in time n enters into the discounting function (see Remark 1.1). Obviously, if β is a power function so $\beta(k-n) =$

δ^{k-n}, then we can factor out δ^{-n} and convert the problem into a standard problem with objective functional

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T-1} \delta^k U(c_k) + \delta^T F(X_T)\right].$$

One can show, however, that every choice of the discounting function β *except* the power case will lead to a time-inconsistent problem. More precisely, the Bellman optimality principle will not hold.

2. Mean-variance utility:

$$J_n(x, \mathbf{u}) = E_{n,x}\left[X_T\right] - \frac{\gamma}{2} Var_{n,x}\left(X_T\right).$$

This case is a dynamic version of a standard Markowitz investment problem where we want to maximize utility of final wealth. The utility of final wealth is basically linear in wealth, as given by the term $E_{n,x}\left[X_T\right]$, but we penalize the risk by the conditional variance $\frac{\gamma}{2} Var_{n,x}\left(X_T\right)$, where the constant γ measures the degree of risk aversion. This looks innocent enough, but we recall the elementary formula

$$Var[X] = E[X^2] - (E[X])^2.$$

Now, in a standard time-consistent problem we are allowed to have terms like $E_{n,x}\left[F(X_T)\right]$ in the objective functional, meaning that we are allowed to include the expected value of a nonlinear function of terminal wealth. In the present case, however we have the term $\left(E_{n,x}\left[X\right]\right)^2$. This is not an expected value of a nonlinear function, but instead a nonlinear function of the expected value, and we thus have a time-inconsistent problem (see Remark 1.1).

3. Endogenous habit formation:

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\ln\left(X_T - x + \beta\right)\right], \quad \beta > 0.$$

In this particular example we basically want to maximize log utility of terminal wealth. In a standard problem we would have the objective $E_{n,x}\left[\ln\left(X_T - d\right)\right]$ where $d > 0$ is the lowest acceptable level of terminal wealth. In our problem, however, the lowest acceptable level of terminal wealth is given by $x - \beta$ and it thus depends on your wealth $X_n = x$ at time n. This again leads to a time-inconsistent problem. (We remark in passing that there are other examples of endogenous habit formation that are indeed time consistent.)

4. Non-expected utility:

$$J_n(x, \mathbf{u}) = \int_0^\infty w\left(P_{n,x}\left(U(X_T) > z\right)\right) dz.$$

In this problem $U : \mathbf{R} \to \mathbf{R}_+$ is the utility function, $w : [0, 1] \to [0, 1]$ is the *probability distortion (weighting)* function with $w(0) = 0$, $w(1) = 1$, and $P_{n,x}(\cdot)$ denotes the conditional probability. Due to the probability distortion, the above payoff functional is evaluated via a nonlinear expectation, the so-called Choquet expectation or Choquet integral. The nonlinear distortion of the probability scale leads to a time-inconsistent problem. If there is no probability distortion, so that $w(x) = x$, then we recover the expected utility appearing in a standard time-consistent problem.

1.4 Approaches to Handling Time Inconsistency

In all four examples of the previous subsection we are faced with a time-inconsistent family of problems, in the sense that, if for some fixed initial point (n, x) we determine the control law $\hat{\mathbf{u}}$ that maximizes $J_n(x, \mathbf{u})$, then at some later point (k, X_k) the control law $\hat{\mathbf{u}}$ (restricted to the interval $[k, T]$) will no longer be optimal for the functional $J_k(X_k, \mathbf{u})$. It is thus conceptually unclear what we mean by "optimality" and even more unclear what we mean by "an optimal control law," so our first task is to specify more precisely exactly which problem we are trying to solve. There are then at least three different ways of handling a family of time-inconsistent problems like the ones above.

- We fix *one* initial point, for example $(0, x_0)$, and then try to find the control law $\hat{\mathbf{u}}$ that maximizes $J_0(x_0, \mathbf{u})$. We then simply disregard the fact that at a later point in time such as (n, X_n) the control law $\hat{\mathbf{u}}$ will not be optimal for the functional $J_n(X_n, \mathbf{u})$. In the economics literature, this is known as *pre-commitment*.
- At every point in time we view our problem as the pre-committed problem and compute the optimal pre-committed control for today. Tomorrow (or at $t + dt$ in continuous time) we look at a new pre-commited problem and so on. We are thus rolling over a continuously updated sequence of pre-committed optimal controls. In the economics literature, this is known as the behavior of a *naïve (or myopic)* agent.
- We take a game-theoretic perspective, viewing the problem as an intrapersonal game and look for Nash subgame-perfect equilibrium points.

All of the three strategies above may in different situations be perfectly reasonable, but in the present work we choose the last one. The basic idea is then that, when we decide on a control action at time t, we should explicitly take into account that at future times we will have a different objective functional or, loosely speaking, that "our tastes are changing over time." We can then view the entire problem as a non-cooperative game, with one player for each time n, where player n can be viewed as the future incarnation of ourselves (or rather of our preferences) at time n. Player n chooses the control law $\mathbf{u}(n, \cdot)$ so, given this point of view, it is natural to look for Nash equilibria for the game, and this is exactly our approach. For the case of a finite-time horizon, the approach works roughly as follows.

1. Given that $X_{T-1} = x$, player $T-1$ has a standard optimization problem to solve, namely that of maximizing
$$J_n(x, u_{T-1})$$
over u_{T-1}. We denote the optimal u by $\hat{u}_{T-1}(x)$.
2. Given that $X_{T-2} = x$, and that player $T-1$ is using $\hat{u}_{T-1}$, player $T-2$ now maximizes
$$J_n(x, u_{T-2}, \hat{u}_{T-1})$$
over u_{T-2}. We denote the optimal u by $\hat{u}_{T-2}(x)$.
3. We then proceed by induction.

It is fairly easy to formalize these ideas in discrete time (see Chap. 5 for precise definitions) but it is far from trivial in continuous time (see Chap. 15).

1.5 Stopping Problems and Time Inconsistency

A standard discrete-time optimal stopping problem consists in maximizing (or minimizing) an objective functional of the form
$$\max_{0\le\tau\le T} E\left[F(\tau, X_\tau)\right],$$
where τ is the stopping time. The goal here is to determine the best time to intervene and stop a process in order to maximize expected rewards or minimize expected costs.

Similar to control problems, a stopping problem can become time inconsistent if conditions outlined in Remark 1.1 fail to hold. This means that the examples discussed in Sect. 1.3 can be extended to cover corresponding time-inconsistent stopping problems. For example, we can study the problem of determining the best time to sell an asset with price process X and mean-variance objective of the form
$$E\left[X_\tau\right] - \frac{\gamma}{2} Var\left[X_\tau\right].$$

Our approach to handling a time-inconsistent stopping problem will again be a game-theoretic one. We will view the problem as a non-cooperative game, where we have one player at each point in time. This player can only choose the stopping decision at that particular time. Then, instead of looking for an "optimal" stopping rule, we aim to find subgame-perfect Nash equilibrium stopping strategies.

1.6 The Time-Inconsistent Framework

The objective of the present text is to present a theory for time-inconsistent control and stopping problems in a reasonably general Markovian framework. We do this for discrete-time as well as continuous-time models. The discrete-time setup is roughly as follows, and the continuous-time theory is very similar.

- We consider a general controlled Markov process X, living on some suitable space (details are given below). It is important to notice that we do not make any structural assumptions whatsoever about X, and we note that the setup obviously includes the case when X is determined by a system of stochastic difference equations.
- For the case of time-inconsistent control, we consider a general reward functional of the form

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T-1} H_k\left(n, x, X_k^{\mathbf{u}}, \mathbf{u}_k(X_k^{\mathbf{u}})\right) + F_n(x, X_T^{\mathbf{u}})\right] + G_n\left(x, E_{n,x}\left[X_T^{\mathbf{u}}\right]\right),$$

  where we also allow the case $T = \infty$.
- For the case of time-inconsistent stopping, we consider a reward functional of the form

$$J_n(x, s) = E_{n,x}\left[F_n(x, \tau, X_\tau)\right] + G_n\left(x, E_{n,x}\left[X_\tau\right]\right),$$

  where s is a stopping strategy, which is a function of time and the process value prescribing whether to stop or to continue at any given point, and τ is the corresponding stopping time induced by s. See Chap. 23 for precise definitions.

Referring to the discussion in Remark 1.1 we see that with the choice of functionals above, time inconsistency will enter at several points:

- The shape of the utility functionals depends explicitly on the initial position (n, x) in time and space, as can be seen in the appearance of n and x in the expressions $F_n(x, X_T)$ and $F_n(x, \tau, X_\tau)$, and similarly for the other terms. In other words, as the X process moves around, our utility function changes.
- We have expressions of the form $G_n\left(x, E_{n,x}\left[X_T^{\mathbf{u}}\right]\right)$ and $G_n\left(x, E_{n,x}[X_\tau]\right)$. Each of these, even apart from the appearance of n and x in the function G, is not the expectation of a nonlinear function, but a nonlinear function of the expected value. We thus do not have access to iterated expectations, so the problem becomes time inconsistent.

This setup is studied in some detail, both in discrete and in continuous time, and the main outcomes are as follows.

- For time-inconsistent control problems, we derive an extension of the standard Bellman equation (Hamilton–Jacobi–Bellman equation in continuous time) to a non-standard system of equations for the determination of the equilibrium value function and the equilibrium control.
- We study a number of concrete examples of time-inconsistent control problems. In particular, we study non-exponential discounting, mean-variance optimal portfolios, and a dynamic equilibrium model with time-inconsistent preferences.
- We show that our methodology can be extended from time-inconsistent control to a fairly general class of time-inconsistent stopping models. In particular, we present an extension of the Wald–Bellman equation (variational inequalities in continuous time) to a non-standard extended system that allows for the determination of the equilibrium value function and the equilibrium stopping strategy.
- We consider concrete applications that include stopping problems with non-exponential discounting, mean-variance objective, and distorted probabilities.

1.7 Notes on the Literature

The game-theoretic approach to time inconsistency using Nash equilibrium points as outlined above has a long history starting with Strotz (1955) which studied a deterministic Ramsay problem with non-exponential discounting. Subsequent works by Pollak (1968), Peleg and Yaari (1973), and Goldman (1980) helped to provide a more formal definition of Strotz's strategy of "consistent planning" in discrete time and show that it exists under fairly general conditions. Since these results, the implications of non-exponential discounting, often referred to as the present bias, have received a lot of attention in economics. Some of the most important works along these lines in discrete and continuous time are Barro (1999), Harris and Laibson (2001), Krusell and Smith (2003), Ekeland and Pirvu (2008), Vieille and Weibull (2009), Ekeland and Lazrak (2010), Ekeland et al. (2010), Marín-Solano and Navas (2010), Harris and Laibson (2013), and Pirvu and Zhang (2014). Similarly to studies of consumer behavior under non-exponential discounting, the work of Basak and Chabakauri (2010) provides an important contribution to the finance literature, applying the game-theoretic approach to the portfolio problem of a mean-variance investor. Building on the results in Basak and Chabakauri (2010) and Björk and Murgoci (2014), Czichowsky (2013) studies the mean-variance problem in a general semi-martingale setting.

In all the papers cited above, the various authors have studied particular models and/or objective functionals, and the methodology has been tailor-made for the particular problem under study. The present text, which is based on the journal articles Björk and Murgoci (2014), Björk et al. (2017), and Björk et al. (2014), is an attempt to derive a reasonably *general* (albeit Markovian) theory of time-inconsistent control, and we do this by using dynamic programming arguments. We would like to acknowledge that in this research project we have been much inspired by Basak and Chabakauri (2010) and Ekeland and Lazrak (2010). In recent years, there have also been a number of interesting developments for time inconsistency and optimal stopping. We review these in Part VI, where we extend our framework to study time-inconsistent stopping problems.

Besides the game-theoretic approach to time inconsistency, which is the main focus of the present work, the literature has been investigating alternative approaches to handling time-inconsistent problems. For example, the mean-variance portfolio choice problem has been solved using the pre-commitment approach by Richardson (1989), Bajeux-Besnainou and Portait (1998), Zhou and Li (2000), and Li and Ng (2000). In addition to considering the fully pre-committed agent, the literature has also studied the strategy of an agent who is rolling over a continuously updated sequence of pre-committed optimal controls. Pedersen and Peskir (2016, 2017) formalize such a strategy in terms of "dynamic optimality" for the mean-variance optimal stopping (2016) and mean-variance portfolio selection (2017) problems respectively. The "dynamically optimal" individual is similar to the continuous version of Strotz's "spendthrift" or the naïve individual as later described by Pollak (1968), who at any given point looks for an optimal solution for that point in time only. Lioui (2013) and Vigna (2020) present interesting comparisons between the different approaches to time inconsistency described above for the mean-variance asset allocation problem. A further alternative approach is to study time inconsistency using the methodology of the stochastic maximum principle, and there has recently been very active research in this area. See for example Djehiche and Huang (2016) and Hu et al. (2012).

Among papers using the results originally developed in Björk and Murgoci (2014) and Björk et al. (2017) (and now presented in this book), we can identify two large groups: papers applying our theory to a particular finance problem and papers developing numerical techniques that allow for the study of how various constraints (leverage constraints, no-shorting constraint, etc.) affect both the game-theoretic and the naïve solution.

In the first category, we mention Kryger and Steffensen (2010), Dong and Sircar (2014), Kronborg and Steffensen (2015), Li et al. (2015b), Landriault et al. (2018), Zhou et al. (2019), and Dai et al. (2021). Kryger and Steffensen (2010) study investment problems with endogenous habit formation and group utility. Dong and Sircar (2014) present an extensive study of time-inconsistent problems related to portfolio optimization. They demonstrate that, when the time-inconsistent problem is close to a time-consistent one, asymptotic approximation methods allow for the obtainment of tractable solutions. Kronborg and Steffensen (2015) use the game-theoretic approach to solve the mean-variance and the mean-standard deviation problems that include consumption and labor income. Li et al. (2015b) derive the reinsurance strategy for a mean-variance objective with stochastic interest rates and inflation risk. Landriault et al. (2018) extend the mean-variance results for a random investment horizon. Zhou et al. (2019) solve for the time-consistent insurance and reinvestment strategies in a mean-variance problem with generalized correlated returns. Dai et al. (2021) study the mean-variance model for log returns in complete and incomplete markets, linking the results to the standard relative risk aversion (CRRA) utility maximization in complete markets. The authors use the same equilibrium concepts as those presented in Chap. 15 but solve the problem with the help of backward stochastic differential equations (BSDEs). Their twist on the utility function appears to be particularly suited to study long-term allocation problems.

Among the papers concentrating on numerical techniques, we note Wang and Forsyth (2012) and van Staden et al. (2019, 2021). Wang and Forsyth (2012) develop a numerical scheme to solve the mean-variance problem in a time-consistent way and analyze the impact of realistic investments constraints, such as leverage constraints, discrete rebalancing, transaction costs. They compare these results to those obtained by solving a mean–quadratic variation problem and show that while both problems result in very similar trade-offs from the terminal-wealth perspective, the optimal controls are different, and hence the equivalence between the two problems is not necessarily present in scenarios with investment constraints. In models with jumps, van Staden et al. (2019) revisit the problem of comparing the mean–quadratic variation and mean-variance problems. van Staden et al. (2021) study the robustness to model misspecification of both subgame-perfect equilibrium and pre-commitment solutions to the mean-variance problem. The game-theoretic approach is shown to be more robust when there are no investment constraints or when rebalancing is continuous. However, in scenarios with multiple investment constraints and discrete rebalancing, the pre-commitment solution can be more robust to model misspecification errors.

Part I
Optimal Control in Discrete Time

Chapter 2
Dynamic Programming Theory

Although our objective is to study time-inconsistent control problems, we will in fact make use of ideas from dynamic programming in our study. In this chapter we therefore give a brief summary of standard discrete-time dynamic programming. We will give the main arguments while going lightly on some of the more technical issues, sweeping measurability and integrability issues under the carpet.

2.1 Setup

The basic setup is that we have a filtered probability space $(\Omega, \mathcal{F}, P, \mathbf{F})$, where $\mathbf{F} = \{\mathcal{F}_n\}_{n=0}^{\infty}$. On this space we consider a discrete-time controlled Markov process X living on the state space $\mathbf{X}$, with controls u in some control space $\mathcal{U}$. Time is indexed by the set of natural numbers $\mathbf{N}$, and we use the notation $[n, m]$ to denote a discrete time interval, so $[n, m] = \{n, n+1, \dots, m-1, m\}$, where $n < m$. We also have some exogenously given objects:

- A reward functional of the form

$$E\left[\sum_{n=0}^{T-1} H_n\left(X_n, u_n\right) + F(X_T)\right].$$

- An indexed family $\{U_n(x) : x \in \mathbf{X}, n \in \mathbf{N}\}$ of subsets of $\mathcal{U}$, so $U_n(x) \subseteq \mathcal{U}$ for all $x \in \mathbf{X}$ and all $n = 0, 1, 2, \dots, T$. This family provides us with *control restrictions* in the sense that if $X_n = x$ then we must choose the control u_n such that $u_n \in U_n(x)$.

T. Björk et al., *Time-Inconsistent Control Theory with Finance Applications*,
Springer Finance, https://doi.org/10.1007/978-3-030-81843-2_2

Note 2.1 With almost no loss of understanding of the main ideas of the arguments below, the reader can informally assume that the X process is scalar (i.e. $\mathbf{X} = \mathbf{R}$), that the control u is scalar (so $\mathcal{U} = \mathbf{R}$), and that the constraints are not present (so $U_n(x) = \mathbf{R}$).

We can now state our main problem.

Problem 2.1 The problem to be solved is to choose an adapted control process u that maximizes

$$E\left[\sum_{n=0}^{T-1} H_n\left(X_n, u_n\right) + F(X_T)\right]$$

subject to the constraints

$$u_n \in U_n(X_n), \quad n = 0, 1, 2, \ldots$$

In principle the control process u is allowed to be any adapted process satisfying the constraints above, but we will restrict ourselves to the case of so-called *feedback control laws*.

Definition 2.1 A **feedback control law** is a mapping $\mathbf{u} : \mathbf{N} \times \mathbf{X} \to \mathcal{U}$.

The interpretation of this is that, given the control law $\mathbf{u}$, the control process u will be of the form

$$u_n = \mathbf{u}_n(X_n).$$

The class of feedback control laws is of course smaller than the class of adapted controls. It is however possible to prove that the optimal control is always realized by a feedback law, so from an optimality point of view there is no restriction to limiting ourselves to feedback laws.

Definition 2.2 The class $\mathbf{U}$ of **admissible feedback laws** is defined as the class of feedback laws $\mathbf{u}$ satisfying the constraints

$$u_n \in U_n(X_n), \quad n = 0, 1, 2, \ldots$$

2.2 Embedding the Problem

The way to approach our optimization problem is to embed it in a family of problems indexed by time and space. Then we can connect all these problems by a recursive equation known as the *Bellman equation*. We will see that solving the Bellman equation is equivalent to solving the optimal control problem.

Definition 2.3 For each fixed initial point (n, x) we define the problem $\mathcal{P}_{n,x}$ as the problem of maximizing

$$E_{n,x}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \mathbf{u}_k(X_k)\right) + F(X_T)\right]$$

over the class of feedback laws $\mathbf{u}$ satisfying the constraints

$$\mathbf{u}_k(x) \in U_k(x), \quad \text{for all } k \geq n, \ x \in \mathbf{X}.$$

We now proceed to define the *value function* and the *optimal value function*. Recall that $\mathbf{U}$ is the class of admissible feedback laws.

Definition 2.4

- The **value function**

  $$J : \mathbf{N} \times \mathbf{X} \times \mathbf{U} \to \mathbf{R}$$

  is defined by

  $$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \mathbf{u}_k(X_k)\right) + F(X_T)\right].$$

- The **optimal value function**

  $$V : \mathbf{N} \times \mathbf{X} \to \mathbf{R}$$

  is defined by

  $$V_n(x) = \sup_{\mathbf{u} \in \mathbf{U}} J_n(x, \mathbf{u}).$$

The interpretation is that $J_n(x, \mathbf{u})$ yields the expected utility of employing the control law $\mathbf{u}$ for the time interval $[n, T]$ if you start in state x at time n. The optimal value function $V_n(x)$ gives you the optimal utility over $[n, T]$ if you start in state x at time n.

2.3 Time Consistency and the Bellman Principle

We now proceed to state and prove the Bellman optimality principle. Going back to the problem $\mathcal{P}_{n,x}$ introduced in Definition 2.3, we make the following simplifying assumption.

Assumption 2.1 We assume that for every initial point (n, x) there exists an optimal control law for problem $\mathcal{P}_{n,x}$. This control law is denoted by $\hat{\mathbf{u}}^{n,x}$.

The object $\hat{\mathbf{u}}^{n,x}$ is a mapping $\hat{\mathbf{u}}^{n,x} : [n, T] \times \mathbf{X} \to \mathbf{R}$, where the upper index (n, x) denotes the fixed initial point for problem $\mathcal{P}_{n,x}$. Consequently, the control applied at some time $k \geq n$ will be given by the expression

$$\hat{\mathbf{u}}_k^{n,x}(X_k).$$

We would like to stress that a priori the optimal law for the problem $\mathcal{P}_{n,x}$ could very well depend on the choice of the starting point (n, x). However, it turns out that the optimal law is *independent* of this choice. The formalization and proof of this statement is as follows.

Theorem 2.1 (The Bellman Optimality Principle) *Fix an initial point* (n, x) *and consider the corresponding optimal law* $\hat{\mathbf{u}}^{n,x}$. *Then the law* $\hat{\mathbf{u}}^{n,x}$ *is also optimal for any subinterval of the form* $[m, T]$ *where* $m \geq n$. *In other words,*

$$\hat{\mathbf{u}}_k^{n,x}(y) = \hat{\mathbf{u}}_k^{m,X_m}(y)$$

for all $k \geq m$ *and all* $y \in \mathbf{X}$. *In particular, the optimal law for the initial point* $n = 0$ *will be optimal for all subintervals. This law will be denoted by* $\hat{\mathbf{u}}$.

Put in more colloquial terms, Bellman optimality principle says that a plan for the future deemed optimal at an earlier point in time will also remain optimal. Suppose that you optimize at time $n = 0$ and follow control law $\hat{\mathbf{u}}$ up to time n, where you now have reached the state X_n. At time n you reconsider, and now decide to forget your original problem and instead solve problem $\mathcal{P}_{n,X_n}$. What the Bellman Principle tells you is that the law $\hat{\mathbf{u}}$ (restricted to the time interval $[n, T]$) is optimal, not only for your original problem, but also for your new problem. In decision-theoretic jargon we could say that our family of problems is *time consistent*, and in particular this implies that the expression "the optimal law" has a well-defined meaning—it does not depend on your choice of starting point.

Proof The proof is by contradiction. Let us assume that for some $n > 0$ there exists a law $\bar{\mathbf{u}}$ on the interval $[n, T]$ such that

$$E_{n,x}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \bar{\mathbf{u}}_k(X_k)\right) + F(X_T)\right] \geq E_{n,x}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \hat{\mathbf{u}}_k(X_k)\right) + F(X_T)\right]$$

for all $x \in \mathbf{X}$ with strict inequality for some $x \in \mathbf{X}$. We can then construct a new law $\mathbf{u}^\star$ on $[0, T]$ by the following formula

$$\mathbf{u}_k^\star(y) = \begin{cases} \hat{\mathbf{u}}_k(y) \text{ for } 0 \leq k < n-1, \\ \bar{\mathbf{u}}_k(y) \text{ for } n \leq k < T-1. \end{cases}$$

We then have

$$J_0(x_0, \mathbf{u}^\star) = E_{0,x_0}\left[\sum_{k=0}^{T-1} H_k\left(X_k, \mathbf{u}_k^\star\right) + F(X_T)\right]$$

$$= E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right] + E_{0,x_0}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \bar{\mathbf{u}}_k\right) + F(X_T)\right]$$

$$= E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right] + E_{0,x_0}\left[E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \bar{\mathbf{u}}_k\right) + F(X_T)\right]\right],$$

where we have used iterated expectations and the Markov property to obtain the last term. It now follows from the assumption concerning $\bar{\mathbf{u}}$ that we have

$$E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \bar{\mathbf{u}}_k\right) + F(X_T)\right] \geq E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right) + F(X_T)\right].$$

with strict inequality with positive probability so, again using iterated expectations and the Markov property, we obtain

$$J_0(x_0, \mathbf{u}^\star) > E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right] + E_{0,x_0}\left[E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right] + F(X_T)\right]$$

$$= E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right] + E_{0,x_0}\left[\sum_{k=n}^{T} H_k\left(X_k, \hat{\mathbf{u}}_k\right)\right]$$

$$= E_{0,x_0}\left[\sum_{k=0}^{T} H_k\left(X_k, \hat{\mathbf{u}}_k\right) + F(X_T)\right] = J_0(x_0, \hat{\mathbf{u}}).$$

We have thus obtained the inequality

$$J_0(x_0, \mathbf{u}^\star) > J_0(x_0, \hat{\mathbf{u}}),$$

which contradicts the optimality of $\hat{\mathbf{u}}$ on the interval $[0, T]$. ■

2.4 The Bellman Equation

In this section we proceed to derive the Bellman equation, which is the recursive relation for the optimal value function. We fix an arbitrary initial point (n, x) and consider the control law $\mathbf{u}^\star$ that deviates from the optimal control only at time n.