Duality methods for stochastic optimal control problems in finance

Doctoral thesis

Duality Methods for Stochastic Optimal Control Problems in Finance

Thijs Kamma

2022

This book was typeset by the author using LaTeX.

Cover	Stefanie van den Herik, 2022
Publisher	Maastricht University, the Netherlands
ISBN	978-94-6469-134-4
Printing	ProefschriftMaken || www.proefschriftmaken.nl

Duality Methods for Stochastic Optimal Control Problems in Finance

DISSERTATION

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. dr. P. Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday, 12th of January, 2023, at 10:00 hours

by

Thijs Kamma

Supervisor
Prof. dr. A.A.J. Pelsser

Co-supervisor
Dr. T. Post

Assessment Committee
Prof. dr. P.C. Schotman (Chair)
Prof. dr. W.F.M. Bams
Prof. dr. G. Deelstra (Université Libre de Bruxelles)
Prof. dr. ir. M.H. Vellekoop (University of Amsterdam)

This research was financially supported by the Network for Studies on Pensions, Aging and Retirement (NETSPAR).

To my family: Corrie, Johan and Stan.

Acknowledgments

From a strictly materialist perspective, man is integrated into a complex network, an inimitably braided chain, an immense field of forces that simultaneously affect or may affect his behaviour as well as the way he thinks. Therefore, in spite of the currently prevailing liberal paradigm, we find ourselves in a continuous state of dependency, solely entitled to a considerably small, nearly negligible space of true "autonomy". Against the background of an acknowledgement section, this overshadowing degree of dependency inherent in our experienced reality can pose serious problems. While being reliant upon purely immanent stimuli as the mere basis on which our existence can – physically – be predicated, we problematically tend to dedicate our words of gratitude to living entities that, although immanently present, are acknowledged for particular intangible qualities. Or should one thank the air for providing us with oxygen for the entire duration of our PhD track? Should we explicitly thank the apple trees and and wheat fields for supplying food? What about gravity, H_2O, or general causality? In that respect, disregarding the indisputable impossibility to generate infinite lists of manifest objects fundamental to life, why do we conventionally restrict ourselves to concrete individuals?

The answer to the preceding question is simple: while the forces of nature are bound to bestow upon the earth what God commands them to in His unparalleled fidelity to the Truth, people are capable of voluntarily opting for particular actions in accordance with our Lord's respect for man's free will. For the purpose of gratitude, at least in the present case, the invisible, intentionally cultivated acts of freedom are correspondingly of great and absolute importance. In other words, although we fail to breathe in the absence of oxygen, had we not been encouraged by person X and/or Y, even in the presence of all "baseline

assets" – man-made or organically established –, we might not have have arrived at the state in which we currently reside. For this reason, specifically in view of the relatively momentous milestone constituted by this dissertation, I feel deeply indebted to numerous people. This feeling is strongly corroborated by the fact that I have never deemed myself intellectually worthy enough of entering a doctoral programme, let alone of uttering these words of gratitude unique to the final stage of the corresponding writing process. Despite people's sincere belief in my alleged capacities, my supposed skills, as well as my putative intelligence – particularly voiced by my family –, and predominantly due to a voluminous body of assertions that contrast the subjects of the aforementioned belief – largely expressed by myself – the latter feeling dissolved into personal truth, underscoring my undeniable weaknesses. However, where these weaknesses painfully came to the surface, the securing wings of heartfelt love, trust and support carried me beyond the horizon of my own abilities, into the achievement set out by this thesis. This brief introduction serves to show how sincerely grateful I am, albeit to varying degrees, to the people addressed in the sequel.

First and foremost, I would like to thank my supervisor, Prof. Antoon Pelsser. During my master studies, by virtue of an increasingly genuine love for mathematics and philosophy, the initially vague idea to pursue an academic career in one of the more fundamental fields began to take on a concrete and certainly realisable shape. There was, however, one problem: at UM, I would only be able to acquire 60 ECTS. In addition to this, as a consequence of my poor performance in the undergraduate studies, the one-year master could not be extended to a two-year research-based equivalent. Therefore, at the end of my master's first period, with the aim of ultimately being eligible for a PhD track, the plan was to enroll in a programme at KU Leuven. The structure of their curriculum offered the possibility to obtain a dual degree in pure mathematics and theoretical physics, with a minor in philosophy. Although this would have been a huge detour with respect to my desired objective – time-wise and financially –, it at least enabled me to get where I wanted to arrive, along an axis based on theoretical science that captured my profound interest. Being determined to steer my life in this fairly novel direction, shortly after essentially taking the corresponding decision, I received an e-mail from Antoon directed to all E&OR master students. It contained information on a PhD position, fully funded by *NETSPAR*, under the supervision of him and other *NETSPAR* fellows at UM and TiU. On the grounds of the "*Life Insurance I*" course, and particularly because of his paper on market-consistent valuations, under the aegis of Antoon's guidance, I was confident that I would be allowed to deal with an extraordinarily satisfying level of mathematics. With the latter in mind, not too convinced of eventual success due to my inadequate profile, I approached him for the aforesaid PhD position. He reassured me, explaining that this concerned an externally funded position, due to which no restrictions were placed upon the precise structure of the applicants' master studies, that I was formally eligible for the PhD track. No longer than half a year later, after an interesting research project excellently monitored by himself, I signed a contract with HR, making me Antoon's newest PhD student.

For making my dreams come true in this comparatively hopeless situation, for accordingly providing me with this amazing opportunity, and for his continued belief in me during the entirety of the doctoral programme – even after repeatedly handing in completely "*Thesaurus*-ised" drafts –, I am forever grateful to Antoon. Of all the things I have greatly appreciated in him over the past five years – including his mental support, human honesty, fair feedback, inspiring presentation skills, and the immense amounts of time allocated to our *weekly* meetings –, I would like to comment on his incredible wealth of knowledge and the resulting dexterity to cleverly apply it in a myriad of different situations. Antoon's understanding of extremely abstract concepts is accompanied by an abundance of intuition, alien to a great majority of scholars, allowing him to "play around" with mathematical material in an unusually flexible sense, without harming or undermining the technical meaning attached to these concepts. While my own familiarity with inverse Fourier machinery, the Clark-Ocone formula or Euler-Lagrange equations is largely restricted to the mathematical dimension of these notions, by means of an authentic passion for his field of expertise – often assisted by elaborate analyses of simple examples in Excel –, Antoon's understanding extends this form of insight into practically relevant and more tangible research domains. The former unique flexibility gives rise to a.o. novel ideas, new approaches to existing methods, and the ability to ceaselessly embed academic material in an economically meaningful context. Due to these exceptional competencies, perpetually supported by his stimulat-

ing and passionate enthusiasm, he significantly contributed to the finalisation of this doctoral dissertation. In addition to this, specifically on account of his eagerness for research, he constitutes a genuine inspiration pertaining to my career in academia, actively motivating me to grow into the best possible version of my scholar-self. Thank you for everything, Antoon.

Second, I would like to thank my co-supervisor, dr. Thomas Post, for his guidance over the last four years. I am especially appreciative of the fact that he eagerly acquainted me with the details of standard peer-review procedures, and thereby implicitly facilitated the publication of my first paper in a top-tier journal. In view of my current appointment, I would like to expressly add to this that his fruitful efforts regarding my stressful search on the academic job market certainly did not go unnoticed. Thank you for everything, Thomas. Moreover, as part of the people who were directly involved in the finalisation process of this dissertation, I would like to express my sincere gratitude to the members of the assessment committee, Prof. Peter Schotman, Prof. Dennis Bams, Prof. Griselda Deelstra and Prof. Michel Vellekoop. With the negligibly small amount of responsibilities I have as a junior scholar, especially compared to those of full-time professors, I already experience the distressing pressure of particular constraints with respect to time. In that regard, I am more than grateful to the committee for their time and effort invested in reading the dissertation, as well as for their feedback and – of course – for their acceptation of the thesis. I would like to extend my heartfelt thanks to Michel in particular, for allowing me to give my first invited talk at the Winterschool of Mathematical Finance. In a tough period of my PhD track, considerably complicated by the corona-related measures, his invitation made me feel more confident about my academic work, encouraging me to keep my scholarly chin up. Furthermore, as a result of his careful reading of one of my working papers, and due to his tremendously helpful feedback, Chapter 3 obtained the noticeably improved form in which it is available today. Thank you for all your help, Michel. Last, I would like to offer my sincere appreciation to Prof. Stan van Hoesel and Prof. Rudi Zagst for their willingness to participate in the opposition of my defense. In light of the Chair's warm welcome in Munich, and the wonderful opportunity to continue my scientific activities at a great university with an excellent team, I would like to candidly express my gratitude to Prof. Zagst in

particular, and his staff in general. Knowing that I would be able to work in such a friendly and high-level environment took away a lot of the burdensome stress added to the completion phase of this dissertation.

Third, I would like to thank all of those who are or were directly associated with Maastricht University and/or NETSPAR. My sincere gratitude goes to my former fellow PhD students, for their help, support, and kindness, but primarily for (re-)familiarising me with ordinary day-to-day life, and accordingly keeping me in touch with reality. Even though the latter may be perceived in an ironic sense, the self-induced delirious isolation in my office during the entirety of the pre-corona era, in part, led to an unconscious and surely unintentional neglect of – or indifference with respect to –, fundamentally, all that truly matters in life. For proactively encouraging me to step outside of the confines of our weirdly small office, and for his indubitable uniting influence on the coterie of PhD students, I believe that I am obliged to thank Adam. From this perspective, I also feel greatly indebted to Niloufar. Our extensive conversations, walks in the park, and her profoundly fascinating viewpoints established a direct trigger to develop myself further outside academia, in intellectual areas not necessarily related to my research. Thank you for everything. For always being willing to grab a cup of coffee, for being the mere person at KE with whom I shared the same research interests, the same supervisor, the same obligations, and for her kind willingness to frequently discuss mathematical content unique to our main domain of portfolio optimisation, I would also like to express my deep appreciation of Li. During the more difficult periods of the doctoral programme, I was able to draw great support from the relatively short conversations with Dewi, wherein we typically exchanged our experiences with regard to the pressure and stress closely linked to the process of writing a dissertation. Thank you for this. The rather bleak final year of my doctoral project – filled with a corona-induced loneliness, as well as troubling worries regarding my future and the success of the whole project – was pleasantly brightened by the presence of Arman and Ashkan. I would like to thank both of you for your kindness, your emotional aid, and our valuable contact outside the office. Ultimately, for being great colleagues, and for being part of the comfortable working environment at UM, I want to extend my appreciation to: Aditja, Alex, Benoit, Caterina, Elisa, Farzaneh, Francesco, Joep, Julian, Li,

Luca, Marie, Moritz, Niels, Rasmus, Roland, Sean, and Shash.

As for the senior staff at KE, I would first of all like to express my genuine gratitude to Dries Vermeulen. As the formal line-manager of the mathematical economics section, endowed with a tremendous amount of other obligations inherent in his full professorship, I have no illusions about the limited amount of time he has. Nevertheless, he was always approachable for questions of any sort, for some small talk about varying topics, or particular personal problems, thereby giving the department a secure and cozy character. I specifically appreciate his efforts for me and for the students concerning the course "*Actuarial Statistics*", which reassured me about my teaching skills. Thank you for everything. For his friendliness, his support on the subject of my teaching affairs, and for allowing me to be part of the great initiative formed by the Foundation Programme – from which I have been able to learn a lot –, I would like to extend my warmest thanks to Dirk Tempelaar. The outspoken trust in his teaching staff, principally based on the "hard" benchmarks outlined by the evaluation forms, strengthened the confidence in my own pedagogical skills. Thank you, Dirk. It is against this background that I also feel tremendously indebted to Joris Poecke and his staff from the Foundation Programme. The evidently serene, incomparably comfortable and enormously encouraging atmosphere surrounding the entire curriculum can completely be attributed to their terribly diligent efforts, which are permanently aimed at improving or maintaining the students' welfare. The ensuing teaching environment, was, correspondingly, very inviting, placing the primary emphasis on a genuine preparation for university education, and positively modified my perspective on pedagogy in a considerable fashion. Thank you for everything. For inspiring and implicitly helping me to (re-)appreciate classical teaching styles, for his unreserved support throughout the proportionately large number of courses I have taught under his outstanding coordination, but most importantly for his admirable commitment to education and his loyalty to the teaching staff, I want to express my sincere appreciation to Christian Kerckhoffs. Thank you for all your efforts over the last four years. As a significant part of the KE branch, I cannot ignore the invariably supportive and eternally helpful (former) employees of the secretarial office: thank you, Karin van den Boorn, Vera Hoekstra and Yolanda Paulissen. Ultimately, I would like to thank the

cleaning lady of our department. Although we unfortunately did not exchange names, I always enjoyed the breaks from work she silently enforced upon me, in which we used to have interesting conversations about daily life, politics, and the university-related working conditions. There is a fairly good chance I spent more time talking to her than to my office mate. Thank you very much.

Since the research central to this doctoral dissertation was completely funded by NETSPAR, I would like to express my gratitude to all associated researchers, staff members, and industry partners, for furnishing an invaluable platform to discuss research output, for providing an important network, and for the general opportunity to contribute to the debate on Dutch pension reforms. As a result, among other things, of my frequent participation in the pension-linked NETSPAR events, I have learnt to consistently keep in sight the practical dimension of the technical problems with which we generally deal. Of all the NETSPAR fellows, I would like to thank Anne Balter in particular, for all the inestimable effort she put into guiding me at the start of my PhD track. Thank you, Anne as well as all the other NETSPAR associates, for everything.

The ruthless solitude peculiar to the pursuit of a doctorate, thoroughly exacerbated by the pandemic, was also conducive to my spiritual well-being and intellectual life outside of academia, substantially broadening the small horizon of my original thinking, which facilitated the whole grace-induced process that made the previously closed eyes of my long-dormant soul receptive to the remnants that eternity had left in this sublunar vale of tears, and thereby contributed along non-negligible lines to my awareness of almighty God and His marvelous footprints in creation. At the outset, this awareness was predominantly emanating from a sincere appreciation of beauty at a mere terrestrial level, which, on the basis of a germinating yet small amount of upright faith, transcended from the world as an immanent phenomenon, instantaneously influencing my transformation from an agnostic cynic/liberal into a surprisingly more rational Christian, acknowledging God's omnipotence in His visible manifestation and corresponding embodiment of the Truth, the Good and the Beautiful. While the commencement of this process can be dated back to an earlier period, the entirety of which is richly endowed with meaningful and positively provoking omens – God literally suffered so as to be able to safe our

souls –, the aforementioned circumstances brought me significantly closer to our Lord, Jesus Christ, and therefore to His Bride, the Catholic Church. Although the unmistakably pivotal event of my conversion and according initiation are seemingly unrelated to this doctoral dissertation, both its intrinsic meaning and ensuing eternal value exceedingly surpass the solely temporal worth of this essentially material product, on account of which it is practically impossible to decouple the former from the latter with respect to the task of genuinely expressing my heartfelt gratitude. Therefore, with a primary eye on this overarching value, I would like to thank all the people who proactively contributed to the rather lengthy process leading to my much-appreciated conversion, and thereby – knowingly or unconsciously – supported the laborious completion of the thesis at hand. In particular, I feel greatly indebted to Fr. Geraldo, for his highly estimable catechises, for both his time and unrivaled enthusiasm invested in his sharing of unique wisdom, and for his invaluable efforts with regard to my actual initiation. The final year of my doctoral programme acquired an entire new dimension of monumental proportions, intensifying the depth of my faith and my appreciation of Maastricht, primarily due to the H. Mass at the Church of St. Martin. Thank you for everything, Fr. Geraldo. For his thoroughly inspiring faith, for his unquenchable willingness to help me out with any question or issue, and for being an incredibly great friend, I would like to extend my wholehearted gratitude to my godfather, João. Without him, who is immediately responsible for my overdue familiarisation with the astonishingly vibrant parish life in Maastricht, it is more than likely that I would not have met any of the people mentioned in this paragraph. Thank you for everything, João. As a result of the appreciably large number of wonderful people who were heartily involved with my conversion, it is unfortunately not possible to dedicate special words of gratitude to each of them for his/her undeniably valuable help and effort. From this perspective, I hope that all of you are able to forgive me for expressing my warmest thanks to you by means of the following supposedly impersonal list. Fr. Dautzenberg, Fr. Nacho, Sr. Agnes, Alfred, Hélie, Oliver, Thijs, Willy, Yoshua, as well as the other valued members of Den Eker, the Church of St. Martin, and the Maastricht Student Chaplaincy – from the bottom of my heart: thank you all for everything.

Even though my gratitude to the people mentioned above is sincere and expressed on the foundation of a genuine heart, in this respect inhabited by an abundance of exclusively honest intentions, it must be stated that only a limited amount of individuals have had the implicit and silently forced obligation to deal with me and the corresponding "dark side" of this doctoral programme at a highly frequent rate or even on a daily basis: my appreciated small circle of friends and my decidedly indispensable family. Pertaining to these highly esteemed people, specifically concerning my parents and my brother, it is simply not possible to repay their utterly hard years of incessant support by means of a few words devoted to a heartfelt expression of authentic gratefulness. For this reason, I would like to avoid any appearance that might allude to a compensation of their never-ending support, according to an inappropriate and extremely insensitive "*quid pro quo*" principle, and therefore expressly declare that my true gratitude goes far beyond these sparse words and their impossibly sufficient meaning, especially – as already addressed above – where it concerns my parents and my brother, and since their persistent love cannot be "made up" by anything that I wholeheartedly say or write down, I will continue to do everything possible to show my deeply felt appreciation – wholly mindful of the fact that, with a probability bordering on certainty, in this regard, even these efforts may not be sufficient. Taking the latter into account, I would like to start by extending my utmost gratitude to my best friend, in a way my older brother, Haroun, for his continuous support of immeasurable worth over the past four years, for replacing the disbelief in myself by his absolutely indestructible trust, but most importantly for his extraordinary friendship. While there is an incalculable wealth of things for which I am more than grateful, I am particularly and exceedingly appreciative of his strengthening of my faith, as well as his guidance through the gruesome period outlined by the first "heavy" pandemic-induced lockdown. With all my heart: thank you, and your family, for everything – and, please, bear in mind that my unfeigned thankfulness surpasses, in terms of its true value, the confines of the aforementioned words. For introducing me to a fundamentally more valuable life of total abstinence from alcohol, liberated from being a miserable slave to many different vices, and for encouraging me to intellectually enrich myself in accordance with my otherwise unemployed sources of alleged ingenuity, I would like to thank Roger. Our regular encounters during the entire course of this doctoral programme were mentally stimulating and helped me to distract myself from the corresponding writing process. Thank you for everything. Ultimately, I would like to extend my sincere appreciation to my aunt, Jitty, for most likely being the mere person apart from my co-workers, parents and brother, who has repeatedly informed herself about my well-being, mental health and progress with reference to my doctoral dissertation. When the buzz surrounding the completion of this work cools down, I hope to devote myself to the wonderful books she has given me. Thank you for everything, Jitty.

Finally and most importantly, I would like to thank the people to whom this dissertation is entirely dedicated, whose years of unconditional love and never-ending support – brutally tried to the limit and beyond – impel me, in a decisively positive sense, to assert that this doctoral dissertation constitutes an achievement that can be attributed more to their undeniably admirable virtues, which permanently emerged in their warmhearted efforts on my behalf, than to my own presumably intellectual merits: my parents and my brother. In spite of the many correspondingly available grounds, rendered by their downright and evidently unmatched commitment, on the basis of which the previously mentioned dedication could be rationalised, the sole, more than well-thought-through reason underpinning its essential meaning is twofold. Following the preceding words, on one hand, the intrinsic meaning of the dedication originates from my honest conviction that this truly is as much their dissertation as it is mine, on account of which the primary corresponding reason underscoring its core value consists in the forthright willingness to hand this book over to them, to give it to them, to "dedicate" it to them, in an attempt to expressly emphasise the shared character of this achievement – as if it were their very own accomplishment, ensuing from their steadfast, backbreaking work. On the other hand, the dedication serves to demonstrate and blatantly accentuate the importance of their irreplaceable presence in my life, prominently presenting itself in their reliable grounding of the backbone corresponding to this doctoral thesis, thereby urging me to a sincere form of undivided gratitude, which establishes the remaining reason pertaining to the aforementioned core value.

For this reason, first, I would like to thank my brother, for being the continually glowing light accompanying my life's journey, particularly when darkness overshadowed every source of joy. Stan, your unnaturally pure and unrestricted love, your inspiring work ethic, your openly conveyed feelings of esteem where it concerns me, your smile, igniting a fire of inextinguishable happiness inside me – the remembrance of which bestows a great deal of power upon me in difficult times –, and your readiness to help me in literally any situation, are what dragged me through the mud of this research and the associated writing process. For all these efforts, you may consider the doctorate – after a hopefully successful defence – as yours, or, if you deem this too much in your consistent and unreservedly beautiful sense of humility, please regard yourself as the most immediate cause of what is laying in your hands right now. Thank you for everything, Stan. Second, I would like to extend my truthfully warmest gratitude to my father, for being the rock-solid, immensely trustworthy, unwavering pillar of strength, on which I could count all my life as the mere place of completely untroubled peace my distressed head was able to rest on. Jo, the stimulating consistency concerning your structured and passionate hard work, your honorable love, your unique reliability, bestowing upon your presence a caring atmosphere of faithful and unambiguous honesty, your invaluable wisdom – often formulated in sententious, lapidary phrases, which shed an undeniably illuminating light on mind-bending problems of nearly any kind –, and the countless moments in which you have actively helped me, have greatly supported me on my arduous path through the academic labyrinth of which this dissertation is the immediate consequence. As in Stan's case, for your priceless help over the years, including the laborious phase outlined by my doctoral pursuit, you can validly think of this work as your own, or, if you believe that this is a grave exaggeration, please think of yourself as the most direct cause of what is laying in your hands right now. Thank you for everything, Jo. Third and last, but most certainly not least, I would like to express my warmest thanks, my most sincere gratitude, my heartfelt appreciation to my mother, i.e. the shadow supervisor of this project, for being the indestructible cornerstone of my existence, as a never-ending source of animating serenity, on the basis of which all the typically cumbersome paths towards the most authentic form of the Good were "revealed" to me. Mom, your unsurpassed persistence and corresponding ability to cope with any gruelling situation, your natural inclination to self-sacrifice, the immeasurable amount of hours invested in our extremely valuable daily conversations, endowed with a highly reinforcing nature, your staying at home to take care of us – which, although frowned upon by modern man, generated the loving circumstances of domestic tranquility that enabled me to write and defend this dissertation at a relatively young age, and which, most importantly, acquainted me with the absolutely non-negligible significance of motherly love as the purely tellurian equivalent of Mary's mercy –, and your continuous effort, no matter what, to ensure that we lacked absolutely nothing throughout the entirety of our fortunate lives, were the driving forces behind anything that I have accomplished, specifically this doctoral dissertation. As my shadow supervisor, I believe it goes without saying that you are almost lawfully considered to be the co-author of this thesis. Thank you for everything, mom. Our Lord unequivocally uttered that "*[...] wisdom is justified of her children*" (Matt. 11:19). Interpreting these powerful words in an exceedingly literal sense, I hope that I have done a good job.

Verbalised in a mildly ironic tone, at the beginning of this rather extensive acknowledgement section, I elaborated on my perspective concerning people's natural inclination to exclusively concentrate on other individuals when expressing their gratitude with respect to certain achievements. In spite of the hopefully convincing arguments carefully brought forward, I want to avoid any appearance of ungratefulness as for the incredibly gracious gifts of an immanent or intangible nature, benevolently entrusted to mankind by our Lord. For this reason, and especially for my valuable competencies, but perhaps even more for my considerable number of shortcomings, notwithstanding the fact that this dissertation is completely dedicated to my family, through the persistent intercession of Our Sweet Lady, Mary, I offer everything revolving around this work – the entirety of which, in conformity with the Jesuits' viewpoint, may be regarded as a continuous prayer –, including my heartfelt thankfulness regarding the people addressed in this section, to God as a humble sign of ultimate gratitude and genuine worship pertaining to His unrivaled greatness. "*Now unto God and our Father be glory for ever and ever. Amen.*" (Phil. 4:20)

Thijs Kamma
Munich
14.11.2022

Contents

1

Introduction

Duality methods are essential to linear programming and convex optimisation. For a given optimisation problem, henceforth the primal, these methods are namely able to generate so-called dual problems. One can interpret the dual as a "shadow" problem that mirrors the primal formulation in a specific mathematical sense.[1] The importance of duality is entirely attributable to these problems for three main reasons. First, the dual formulation of a primal problem typically facilitates the retrieval of optimal solutions. This is particularly true for linear programs.[2] Second, the dual sheds alternative light on the dynamics underscoring the primal. For instance, within the confines of trading constraints, the latter has disclosed to us how the optimal allocation of assets depends on the subjective valuation of unhedgeable risk.[3] Third, for convex optimisation problems, the dual renders mathematically defined duality relations. These relations describe how the primal variables depend on the dual variables, and vice versa. In general, this dependency brings us a few steps closer to the identification of the optimal decision variables in closed-form.

To study an individual's optimal trading and consumption behaviour, the literature on portfolio optimisation ordinarily relies on stochastic optimal control problems. These problems can generally be subsumed under the umbrella of convex optimisation in infinite-dimensional spaces. As a result, duality plays a crucial role in portfolio and consumption theory. In this dissertation, we concentrate on duality methods in the context of investment, consumption and related retirement decisions. One can distinguish the literature on portfolio-based duality methods into roughly three categories: (i) applied studies, (ii) theoretical studies, and (iii) mixtures of the previous two. This dissertation covers all three categories. In Chapter 4, we touch upon item (i), and employ duality techniques to derive optimal policy rules for a pension fund that offers a DC scheme. In Chapter 3, covering item (ii), we derive a mathematical formulation of the dual corresponding to a problem involving non-standard

[1] We use the terms "problem" and "formulation" interchangeably.

[2] Solution techniques for linear programs can benefit from the dictionary-related duality inversion, cf. chapter 5 of Vanderbei (2014). Note that the claim in the main text also applies to convex problems, cf. Rockafellar (2015). For example, in the field of portfolio optimisation, the martingale duality method, cf. Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991), considerably simplifies a derivation of the relevant optima in closed-form. We elaborate on such methods in section 1.2.

[3] See e.g. Cvitanić and Karatzas (1992) for an extensive theoretical treatment of this result.

habit formation. In Chapter 2, which deals with item (iii), we develop an applied approximate method for constrained optimal investment-consumption problems, predicated on theoretical duality results.

We divide the remainder of this introductory chapter into three connected parts. First, using a visual approach, we intuitively elaborate on the technical details underpinning duality for convex optimisation problems. Second, we address the practical value of duality along the lines of finance-linked topics extracted from the separate chapters of this dissertation. Third and last, we analyse these chapters in more detail and comment on their contributions to the literature on convex duality and portfolio optimisation.

1.1 What is Duality?

To aid the reader's understanding of duality, we provide Figure 1.1. This figure is adapted from Chapter 2 and presents a visual illustration of the convex duality mechanism. In our explanation of the technicalities associated with duality, we employ Figure 1.1 as a handy reference point. For this purpose, it is important to note that the graph can be divided into two parts: a lower half and an upper half. The lower half corresponds to the primal problem. We stress that this problem outlines the formulation in which we are originally interested. To make this more concrete, in the spirit of portfolio optimisation, the primal can be identified as e.g. an ordinary utility-maximisation problem. In conformity with the former, the upper half of the figure corresponds to the dual problem. The placement of the dual in the upper half is not without meaning. We expand on this positioning when explaining the dual-side of the figure. To conclude this brief sketch, we observe that the preceding two halves are related to each other via the vertical arrows on the right- and left-hand sides of Figure 1.1. Both arrows pertain to the duality relations.

1.1.1 Primal Problem

We proceed with an analysis and a discussion of Figure 1.1's lower half: the primal-side entitled "Primal". Here, we assume that the primal concerns a

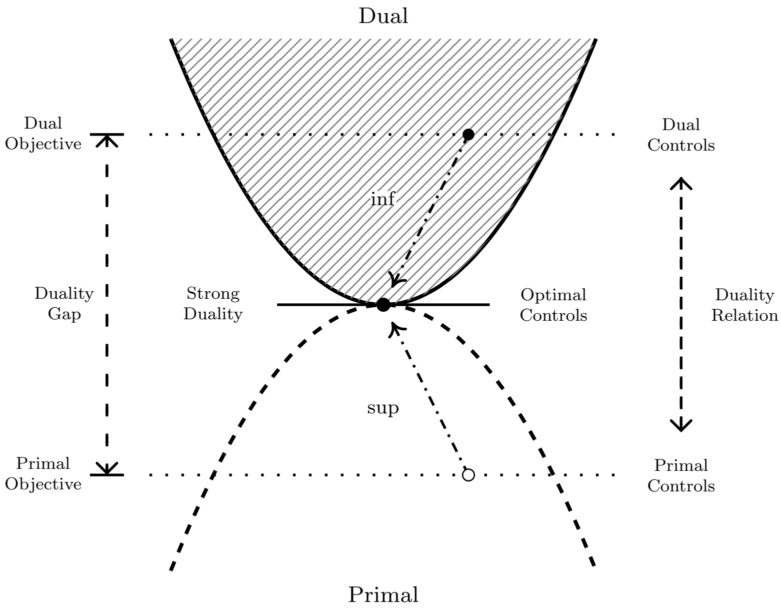

Figure 1.1. Duality for convex optimisation problems. This figure presents a graphical illustration of the core mechanism characterising duality for convex optimisation problems. For details on the meaning of this graph and its individual attributes, we refer to section 1.1 of this chapter. Likewise, one can consult Chapter 2 and the figures therein.

maximisation problem rather than a minimisation problem. Note that this postulate is by no means restrictive. Minimisation problems can easily be transformed into maximisation problems using a slight modification of the value function.[4] Let us (re-)turn to Figure 1.1. On the primal-side, we are able to discern a curve with a dashed shape. This curve represents the set of primal controls or primal decision variables. As indicated by the horizontally dotted line, each primal control implies a value for the primal objective. The objective, otherwise known as value function, identifies the target of optimisation. In line with its maximisation incentive, the primal aims to optimise the value function over all available primal controls. Concretely, it attempts to "steer"

[4] If the primal concerned a minimisation problem, Figure 1.1 would simply need to be inverted. In other words, the upper half should be identified as the primal; the lower as the dual. All other attributes included in the graph could remain unchanged. Note that by multiplying the objective of the minimisation problem with -1, the formulation is converted or transformed into an equivalent maximisation problem.

the objective and the thereby implied primal controls towards the point of optimality. In this point of optimality, the objective is maximised and the related primal controls are optimal. Note that the latter point is displayed by the black dot in the middle of Figure 1.1. The operation of maximisation is depicted by the dash-dotted arrow in the lower curve. This description concludes the visual illustration of the primal problem.

Although it still seems a little abstract, the lower half of Figure 1.1 is what most academic researchers in finance are familiar with. In fact, we are able to include many different problems under the aegis of this lower half. To make the matter more tangible, let us visit a class of optimisation problems relevant to finance and this dissertation: utility-maximisation over terminal wealth and/or consumption. We indicate how these problems are related to Figure 1.1. In the aforementioned branch of problems, an agent attempts to maximise his/her expected utility from terminal wealth and/or consumption. To this end, the individual tries to make the most optimal decisions with regard to investment and/or savings. This operation is represented in Figure 1.1's lower half by the dash-dotted arrow. As a consequence, it is clear that the primal objective identifies a person-specific expected utility criterion. The dashed curve of primal controls accordingly contains all *admissible* portfolio weights and/or consumption rules.[5] Consistent with the horizontally dotted line, it is evident that the individual's decisions regarding investment and savings imply a specific value for the utility-related primal objective. Last, we note that that the black dot in the middle represents the point of maximal expected utility and related optimal investment-consumption decisions. The relation between this set of problems and Figure 1.1 is therefore complete. We have visualised this primal-based relation in Figure 1.2.[6]

1.1.2 Dual Problem

With the primal problem at hand, we continue with an evaluation of Figure 1.1's upper half: the dual side entitled "Dual". The principal distinguishing

[5] *Admissibility* refers to *feasibility* of control variables in the context of portfolio optimisation. We make use of this term in Chapters 2, 3 and 4. Consider e.g. Karatzas and Shreve (1998) for a mathematical definition of admissible control variables.

[6] For a more concrete identification of a graph similar to Figure 1.1, cf. Chapter 2.

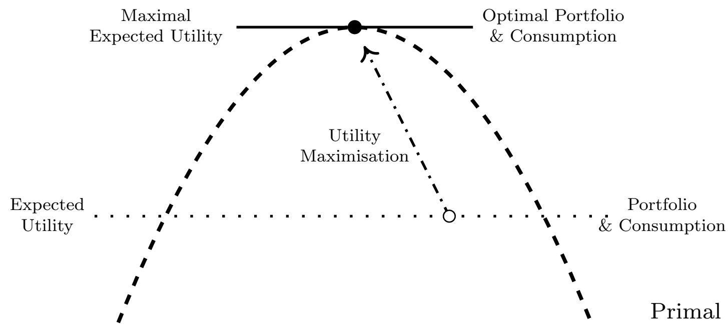

Figure 1.2. Primal for utility-maximisation problem. This figure presents a graphical illustration of the primal-side corresponding to a (constrained) utility-maximisation problem. For details on the meaning of this graph and its individual attributes, we refer to section 1.1.1 of this chapter. Likewise, one can consult Chapter 2 and the figures therein.

feature of a dual formulation is its value function. The dual value function namely always generates an upper bound on the primal value function.[7] This upper bound is the main and most prominent link from the dual to the primal, and the other way around. For this reason, we position the dual problem in the upper half of Figure 1.1. As addressed before, the dual can be interpreted as a shadow problem "mirroring" the primal. This way of mirroring can be taken quite literally with regard to the nature of optimisation. That is, whereas the primal consists in a maximisation procedure, the dual consists in a minimisation routine. To emphasise this inverse imitation, we present the primal and dual problems as mirrored shapes in Figure 1.1. Note that this idea of mirroring is corroborated by the dual-implied upper bound. More specifically, along a technical axis, one is able to argue that the primal and dual sets of control variables mirror each other in a mathematically valid manner.[8] However, for ease of exposition, we omit a corresponding argument.

Returning to Figure 1.1, we note that the dual-based upper half closely resembles the primal-based lower half. Indeed, most attributes linked to the dual abide by a similar interpretation. In more precise terms, the upper curve stands for the set of dual control variables. The horizontally dotted line signifies that

[7]In case the primal outlines a minimisation problem, the dual naturally generates a lower bound on the primal value function. The remaining features remain unaffected.
[8]That is, these sets are dual to each other.

each dual control implies a value for the dual objective. Recall that the dual objective spawns an upper bound on the primal value function. Moreover, as in the primal case, the objective simply represents the target of optimisation. In accordance with its minimisation target, the dual problem seeks to "steer" the objective and the thereby implied dual controls to the point of optimality. This operation is illustrated by the dash-dotted arrow in the upper curve. Observe that the point of optimality is given by the black dot in the middle of the plot, and *coincides* with the primal-optimal point. This phenomenon of coinciding points of optimality is referred to as strong duality. It particularly implies that the difference between the dual and primal value functions, i.e. the duality gap, is equal to zero when both problems are optimised.[9] We elaborate on the strong duality phenomenon when explaining the duality relations. This description concludes the visual illustration of the dual problem.

The upper half of Figure 1.1 is even more abstract than its counterpart in the lower half. While we were able to intuitively identify a myriad of financially relevant optimisation problems with the primal-side of the graph, for the dual-side this is substantially less straightforward. The primarily technical nature of a dual formulation encumbers such immediate identifications. Nevertheless, to make the dual-side of Figure 1.1 more concrete for a finance-oriented audience, we (re-)visit the earlier branch of utility-maximisation problems. For this illustration, we assume that the utility-maximising agent faces trading constraints. Relative to the earlier example, this solely implies that the agent maximises the same expected utility criterion, however, subject to some additional constraints with respect to his/her portfolio weights. As a result, not all risk present in the financial environment can be hedged, and the market is incomplete.

In this situation, it is a well-documented fact that the dual serves to determine the worst-case equivalent martingale measure.[10] To this end, it minimises an expected *conjugate* utility criterion over all available equivalent martingale

[9] It could be the case that the duality gap is non-zero. In that instance, the respective problems are said to satisfy weak duality. As this is not true for the problems relevant to this dissertation, we exclude an analysis of this special case.

[10] Consider e.g. Cvitanić and Karatzas (1992), Bardhan (1994), and Tepla (2000). These papers demonstrate the relation between a so-called auxiliary artificial market and the non-unique equivalent martingale measures. For the notion of an equivalent martingale measure, cf. the theoretical study by Delbaen and Schachermayer (1994).

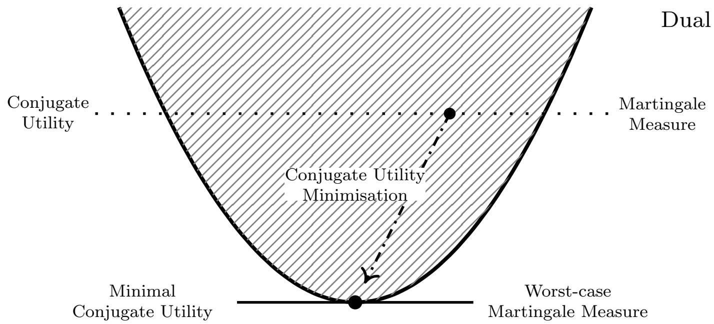

Figure 1.3. Dual for utility-maximisation problem. This figure presents a graphical illustration of the dual-side corresponding to a constrained utility-maximisation problem. For details on the meaning of this graph and its individual attributes, we refer to section 1.1.2 of this chapter. Likewise, one can consult Chapter 2 and the figures therein.

measures. This operation is represented in Figure 1.1's upper half by the dash-dotted arrow. Straightforwardly, the dual objective must be identified as the conjugate expected utility criterion. The upper curve correspondingly comprises of all feasible equivalent martingale measures. In compliance with the horizontally dotted line, as in the primal, we have that the equivalent martingale measures generate particular values for the agent's conjugate utility criterion. Note that the conjugate criterion can be regarded as the mirrored analog of the agent's expected utility function. In a way, it quantifies the agent's attitude towards the unhedgeable uncertainty and the associated martingale measures. Ultimately, we observe that the black dot in the middle of the graph represents the point of minimal expected conjugate utility where the equivalent martingale measure is optimal, i.e. least-favourable. In the same point, the primal and dual value functions bind, satisfying strong duality. The relation between this set of constrained problems and Figure 1.1 is therefore complete. We have visualised this dual-based relation in Figure 1.3.

1.1.3 Duality Relations

To finalise our visually oriented explanation of convex duality, we turn to the right- and left-hand sides of Figure 1.1. Apart from the phenomenon of strong duality and some vague geometric mirroring, it is at this stage not clear how the primal and dual problems are related to each other. To bridge this provisional “gap”, the aforementioned sides of the graph prove useful. As addressed in the beginning of this section, both sides (in)directly pertain to the duality relations. The duality relations stipulate how the primal control variables analytically depend on the dual control variables, and vice versa. In particular, if one is in possession of set of primal control variables, via the duality relations, one is also in possession of a set of dual control variables. The relations consequently ensure that maximisation of the primal is equivalent to minimisation of the dual. This concretely means that acquiring the optimal primal control variables is equivalent to solving both the primal problem and the dual problem.[11] The relations accordingly “glue together” the primal and dual problems. Note that the preceding set of claims still holds true for the other direction (dual-primal instead of primal-dual). The former connection reveals that the duality gap is an immediate consequence of the duality relations. By means of these relations, the duality gap namely measures the extent up to which the control variables are different from the optimal ones.[12] In particular, the duality gap grows as the control variables divert away from the point of optimality. The moment that one obtains the optimal controls, strong duality holds. This description concludes the visual illustration of the duality relations.

Due to the analyses around Figures 1.2 and 1.3, we have developed a more practical understanding of the primal- and dual-sides of Figure 1.1. To make the duality relations more explicit, we correspondingly (re-)visit the earlier collection of constrained utility-maximisation problems. For this class of problems, we have already argued that the primal consists in a maximisation procedure of expected utility over all consumption and portfolio rules. Related to this primal, the dual minimises an analogous expected conjugate utility criterion over all equivalent martingale measures. The duality relations should therefore specify how the investment-consumption choices depend on the martingale measures, and vice versa. In the context of constrained utility-maximisation,

[11] Although this is an immediate result of the claim in the main text, we stress the following. Insertion of the *optimal* primal control variables into the duality relations results in the *optimal* dual control variables. This is also true for the other direction.

[12] This observation is central to the approximate method developed in Chapter 2. In section 1.3, we provide a brief overview of this method.

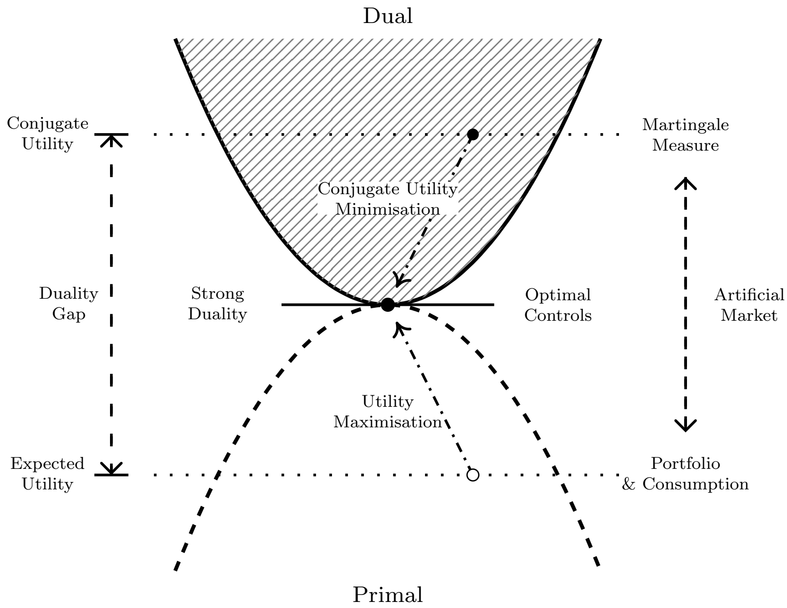

Figure 1.4. Duality for utility-maximisation problem. This figure presents a graphical illustration of the duality mechanism corresponding to a constrained utility-maximisation problem. For details on the meaning of this graph and its individual attributes, we refer to section 1.1 of this chapter. Likewise, one can consult Chapter 2 and the figures therein.

this dependency is established by a so-called auxiliary or artificial market.[13] This artificial market directly ensues from the duality relations and plays a similar role: linking the primal-side to the dual-side. The auxiliary market relaxes all trading constraints and is uniquely defined by a dual-feasible martingale measure. In this market, one is therefore able to derive analytical expressions for the primal controls. Note that these controls depend by construction on the dual-linked martingale measures. Put differently, the former analytical expressions outline duality relations. Through a worst-case characterisation of the martingale measure defining the auxiliary market, the dual attempts to make the analytical expressions primal-optimal. Similarly, via an optimal specification of the same expressions, the primal tries to find the least-favourable

[13]The concept of an artificial market is developed in the pioneering paper on portfolio-related convex duality by Cvitanić and Karatzas (1992). In Chapter 2, we make use of this concept to develop an approximate method.

martingale measure. The interplay between the preceding two mechanisms demonstrates how the duality relations operate in practice. This explicitly shows how Figures 1.2 and 1.3 are linked to each other. For the constrained utility-maximisation, we are consequently able to "glue" the latter graphs together and disclose the complete duality mechanism. This is done in Figure 1.4, which spells out a concretised variant of Figure 1.1.

1.1.4 Concluding Remarks

In spite of the overview's stylised character, we have touched upon all duality-linked features relevant to a proper understanding of this dissertation in a general sense. The main purpose of this section was to acquaint the reader with duality and explain what this technically entails. In doing so, we relied on visual machinery and primarily aimed to answer the following question: *What is duality?* For this purpose, we divided the explanation into two parts. One part concentrated on the purely technical aspects; and one part analysed duality in an applied utility-maximisation setup. Based on the ingredients at hand, our answer to the question central to this section is as follows: *Duality constitutes the technical mechanism revolving around the interplay between a convex optimisation problem and its corresponding dual formulation.* In the sequel, we attempt to make this interplay more explicit by highlighting the practical use of duality. We conclude this section by noting that the separate chapters of this dissertation can all be related to the illustration in Figure 1.1 or particular parts thereof. In fact, all topics concern portfolio-based utility-maximisation and can therefore be identified with the more tangible illustration in Figure 1.4. In section 1.3, we explicitly indicate where and how the topics of these chapters can be situated in both of these illustrations.

1.2 Why is Duality Useful?

In order to answer the question in the header of this section, we visit two optimisation problems related to this dissertation. Thereby, we predominantly aim to emphasise how duality techniques may facilitate a retrieval of optimal solutions. This simultaneously helps us to better explain what we do in the remaining chapters. The first formulation is a utility-maximisation problem over terminal wealth alone. In this problem, we assume that the investor has to deal with constraints imposed upon his/her portfolio weights. In section 1.1, we have addressed a more general version of this problem. This constrained terminal wealth framework bears direct relevance to the topics of Chapter 2 and Chapter 4. The second formulation is, again, a utility-maximisation problem, however, over consumption alone. Instead of trading constraints, we incorporate a habit formation component into the problem specification. The agent is accordingly assumed to derive utility from the *difference* between consumption and this habit component. This additive habit problem bears immediate relevance to the topic of Chapter 3.[14]

1.2.1 Trading Constraints

In a terminal wealth problem, an agent maximises expected utility from the terminal value of his/her wealth process. This problem has gained significant interest since the seminal contributions by Samuelson (1969) and Merton (1969, 1971). Due to its sole dependence on a terminal wealth variable, the framework can easily be employed to model a broad variety of financially relevant issues. For instance, in a pension-related setup, the terminal variable can be identified as retirement wealth. In that case, the problem may correspond to one in which a pension fund aims to maximise the total expected utility of all its participants. Similarly, assuming that the fund operates according to a DC scheme, the problem can be utilised by a pension fund to calculate its optimal participant-specific investment policies.[15] Outside of the pension context, the configuration can also be used by individual investors to optimise their portfolio strategies over the life-cycle. This list of three examples is by far not exhaustive, see e.g. Bodie et al. (2004), Lim and Wong (2010) and Han and Hung (2012), and references therein, for additional interpretations in similar setups. As

[14] In the sequel, we refrain ourselves from providing thorough mathematical arguments and/or derivations. These are not relevant for the rudimentary overview of this section. For the mathematics, one can consult the separate chapters of this dissertation. Likewise, one may consider any of the papers cited in the main text for additional technicalities.

[15] In Chapter 4, we focus on such a problem.

individual agents and institutional investors alike are typically required to deal with solvency and/or liquidity requirements, it is sensible to account for trading constraints. We therefore include trading constraints in the framework of interest. The ensuing constrained problem is studied by a.o. Kim and Omberg (1996), Wachter (2002), and Liu (2006). The problem reads as follows:

$$\sup_{\{\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} \mathbb{E}\left[U\left(X_T\right)\right]. \tag{1.2.1}$$

Note that this problem corresponds to the lower halves of Figures 1.1 and 1.4. Without expanding too extensively on the mathematics related to (1.2.1), we note the following. The investment strategy is denoted by π_t, and the corresponding terminal wealth variable is given by X_T. That is, π_t over the trading interval, $[0,T]$, i.e. $\{\pi_t\}_{t\in[0,T]}$, generates a specific value for X_T. One may accordingly interpret $T > 0$ as the retirement age or some terminal target date. To define the agent's preferences, $U : \mathbb{R} \to \mathbb{R}$ models his/her utility function. Last, we observe that $\mathcal{A}_{X_0}$ outlines the admissibility set, which accommodates the above-mentioned trading constraints. Hence, given some starting capital, $X_0 > 0$, the agent in (1.2.1) tries to select $\{\pi_t\}_{t\in[0,T]}$ in such a manner that $\mathbb{E}\left[U\left(X_T\right)\right]$ is maximised. Even though (1.2.1) seems like a simple and mathematically elegant problem, the inclusion of trading constraints poses serious technical problems. These problems are related to the constraint-induced non-uniqueness of the equivalent martingale measure, cf. Karatzas et al. (1991a), Xu and Shreve (1992), and Detemple (2014). In fact, for general trading constraints, it is not possible to solve (1.2.1) in closed-form. This is where duality techniques start to play an important role. In particular, the dual formulation corresponding to (1.2.1) is given by:

$$\inf_{Z_T\in\mathcal{Q},\eta\in\mathbb{R}_+} \mathbb{E}\left[V\left(\eta Z_T\right)\right] + \eta X_0. \tag{1.2.2}$$

Observe that this problem can be identified with the upper halves of Figures 1.1 and 1.4. As in (1.2.1), we do not elaborate on the mathematics underscoring (1.2.2). For technical accounts of this result in more general setups, we refer to Cvitanić and Karatzas (1992), Klein and Rogers (2007), Rogers (2003, 2013). We emphasise that the problems in (1.2.1) and (1.2.2) are binding in the optimum, i.e. satisfy strong duality. Notation-wise, it is crucial to note that Z_T represents the so-called state-price price density (SPD). In the absence of interest rate risk, the SPD process spells out the non-unique equivalent martingale measure. The set denoted by $\mathcal{Q}$ therefore contains all dual-feasible martingale measures. In (1.2.2), $\eta > 0$ stands for the Lagrange multiplier, which scales the dual-optima towards the regions of feasibility. Due to the technical nature of $\eta > 0$, we ignore its precise function in the duality mechanism. Last, as the counterpart of U, we have that the function $V : \mathbb{R} \to \mathbb{R}$ specifies the conjugate utility criterion.[16] Following our explanation around Figure 1.3, the dual in (1.2.2) attempts to select Z_T in such a way that $\mathbb{E}\left[V\left(\eta Z_T\right)\right] + \eta X_0$ is minimised. This operation generally results in the worst-case martingale measure from the agent's point view. At this stage, it is not clear how the interplay between (1.2.1) and (1.2.2) may facilitate the search for optimal solutions. To this end, we provide the duality relation(s):

$$X_T = I\left(\mathcal{H}^{-1}\left(X_0\right) Z_T\right). \tag{1.2.3}$$

Here, $x \mapsto I(x)$ represents the inverse of marginal utility. Note that this equation corresponds to the left- and right-hand sides of Figures 1.1 and 1.4. As pointed out in section 1.1.3, this identity is an immediate consequence of the artificial or auxiliary market. More precisely, one is able to disentangle the auxiliary environment from the dual formulation in (1.2.2). In this environment, X_T identifies the optimal terminal wealth process. The $\mathcal{H}^{-1}\left(X_0\right)$ term follows from the Lagrange multiplier and ensures that X_T is budget-consistent with the agent's initial endowment, X_0. By means of hedging arguments, we are in turn able to recover the optimal trading strategy, $\{\pi_t\}_{t\in[0,T]}$. In keeping with the nature of a duality relation, (1.2.3) states how optimal wealth depends on the martingale measure, and vice versa. Due to the intimate link between X_T and $\{\pi_t\}_{t\in[0,T]}$, it is clear that the same relation holds with respect to the portfolio

[16] Formally speaking, the definition of $V : \mathbb{R} \to \mathbb{R}$ is given by the celebrated Legendre-Fenchel transform: $V(x) = \sup_{z\in\mathbb{R}} \{U(z) - xz\}$, for all $x \in \mathbb{R}$. The inequality inherent in this transform, $U(z) \leq V(x) + xz$ for all $x, z \in \mathbb{R}$, generally manages to generate the desired dual formulations. Note that $V(x) + xz$ supplies an upper bound $U(z)$. In Chapter 3, we visit a problem for which this inequality falls short in identifying a dual.

weight(s). To make sure that the auxiliary-optimal control variables maximise (1.2.1), the worst-case Z_T must be determined. The latter process coincides with the dual-optimal martingale measure. By virtue of the duality mechanism, we have therefore been able to reduce the intricate dynamic problem in (1.2.1) to a search for martingale measures. Observe that this search hinges on *analytical* expressions for the auxiliary primal controls. Although the characterisation of Z_T may pose different difficulties, the dual is in most cases easier to solve than the primal. We refer to Brennan and Xia (2002) and Sangvinatsos and Wachter (2005) for concrete examples. In addition to this, we underline that the closed-form expressions are tremendously valuable and informative. The identity in (1.2.3) explicitly demonstrates what the optimal controls look like up to the exact specification of Z_T. Irrespective of Z_T's specification, the duality relations consequently disclose the true dynamics of the optimal controls. As such, the auxiliary solutions bring us much closer to an identification of the optima in closed-form. This demonstrates the value of duality and how it may facilitate a retrieval of optimal solutions.[17]

1.2.2 Habit Formation

In the additive habit model, an agent maximises expected utility from the difference between consumption and the habit level. The habit level depends on the agent's past consumption decisions and is compatible with different interpretations. As a result, relative to (1.2.1), terminal wealth can be ignored in outlining the target of optimisation. This problem is pioneered by a.o. Constantinides (1990), and Detemple and Zapatero (1991). The additive habit problem can be employed to model a variety of economically relevant situations. To this end, it is important to note that most conventional utility functions only admit strictly positive arguments. The additive agent is consequently

[17] In Chapter 2, we analyse a similar problem involving consumption. For the formulation central to Chapter 4, we relax all trading constraints. In the absence of such constraints, duality techniques suggest that the *dynamic* problem in (1.2.1) is identical to the following *static* one: $\sup_{X_T \text{ s.t. } \mathbb{E}[X_T Z_T] \leq X_0} \mathbb{E}\left[U\left(X_T\right)\right]$. Here, due to the fact that all available risk is traded, Z_T is uniquely defined. This static variational formulation is significantly easier to solve than its dynamic counterpart. The former dynamic-static transformation can be attributed to the martingale duality method developed by Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991).

required to keep his/her consumption levels *above* the habit component at all times. On the grounds of this lower bound, the literature on additive models interprets the habit component as a subsistence level. From macro-related perspectives, interpretations of this form are straightforward and plausible. It enables economists to analyse the society-wide savings/consumption patterns needed to maximise a population's "happiness" whilst satisfying a minimum wage requirement. Alternatively, some studies identify the habit component as a standard of living. In a micro-related setup, this requires a household to continuously consume more than the level to which they have become accustomed. While this requirement is practically unrealistic, it can tell us something about the behaviour necessary to preserve such a situation. For studies on both interpretations, see e.g. Campbell and Cochrane (1999), Muraviev (2011), and Yu (2015). The additive habit problem is characterised by:

$$\sup_{\{c_t,\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} \mathbb{E}\left[\int_0^T U\left(c_t-h_t\right)\mathrm{d}t\right]. \tag{1.2.4}$$

In the spirit of (1.2.1), this problem can be aligned with the lower halves of Figures 1.1 and 1.4. Refraining ourselves from providing rigorous mathematical arguments, we note the following. The agent's consumption behaviour is represented by c_t, and π_t denotes the corresponding investment strategy. Despite the exclusion of X_T from (1.2.4), the agent is still required to invest so as to finance his/her consumption patterns. For this reason, maximisation takes place with respect to both c_t and π_t. As before, $T > 0$ can be interpreted as the investor's retirement age or some pre-fixed target date. The preferences are described by the same utility function $U : \mathbb{R} \to \mathbb{R}$, and $\mathcal{A}_{X_0}$ denotes the admissibility set adjusted for consumption. Hence, given some endowment, $X_0 > 0$, the agent in (1.2.4) attempts to select $\{c_t,\pi_t\}_{t\in[0,T]}$ in such a way that $\mathbb{E}[\int_0^T U\left(c_t-h_t\right)\mathrm{d}t]$ is maximised. Solving this seemingly easy problem is considerably complicated by the inclusion of h_t. The habit component namely depends on past consumption decisions: $h_t = e^{-\alpha t}h_0 + \beta\int_0^t e^{-\alpha[t-s]}c_s\mathrm{d}s$, for all $t \in [0,T]$. Here, $\alpha, \beta > 0$ characterise person-specific parameters related to the agent's "degree" of habit formation. Due to the aforementioned dependence, rather than optimally choosing c_t at each point in time, one has to account for the ensuing effect on *future* values of h_t. This phenomenon is known as path-dependency. It encumbers an immediate retrieval of the optimal solutions to (1.2.4), cf. Detemple and Karatzas (2003), Bodie et al. (2004), and Polkovnichenko (2007). Hence, it may be clever to resort to duality applications instead. The dual formulation corresponding to (1.2.4) lives by:

$$\inf_{\eta \in \mathbb{R}_+} \mathbb{E}\left[\int_0^T V\left(\eta \widehat{M}_t\right) \mathrm{d}t\right] + \eta \widehat{X}_0. \tag{1.2.5}$$

We stress that this problem corresponds to the upper halves of Figures 1.1 and 1.4. Following our approach to the primal, we do not expand on the technicalities relevant to (1.2.5). For mathematical proofs of this duality result, we refer the reader to Schroder and Skiadas (2002) and Yu (2015). As in section 1.2.1, we underline that the value functions of (1.2.4) and (1.2.5) bind in the optimum, i.e. satisfy strong duality. In this dual problem, $\widehat{M}_t$ and $\widehat{X}_0$ are defined as follows: $\widehat{M}_t = M_t + \beta \mathbb{E}[\int_t^T e^{-(\alpha-\beta)[s-t]} M_s \mathrm{d}s]$ and $\widehat{X}_0 = X_0 - h_0 \mathbb{E}[\int_0^T e^{-(\alpha-\beta)s} M_s \mathrm{d}s]$, for all $t \in [0,T]$. Here, M_t represents the SPD process. Note that this process is uniquely defined, due to the implied market completeness. Optimisation in (1.2.5) takes place over $\eta > 0$ alone. This parameter identifies the Lagrange multiplier and serves the same scaling purpose as its "twin" in (1.2.2). Likewise, the conjugate utility function, $V \to \mathbb{R} \to \mathbb{R}$, is identical to the one in the aforementioned problem. Observe the similarities between the analytical structures of the different dual formulations in (1.2.2) and (1.2.5).[18] Nevertheless, whereas we were able to couple financial intuition to the formulation in (1.2.2), for the dual at hand this is notably harder. The technical function of η makes it difficult to examine (1.2.5) along well-defined economic lines. Notwithstanding, on the basis of the duality relations, we are able to derive an auxiliary market from (1.2.5). In this auxiliary market, the optimal consumption problem is specified according to:

$$\sup_{\left\{\widehat{c}_t\right\}_{t\in[0,T]} \text{ s.t. } \mathbb{E}\left[\int_0^T \widehat{c}_t \widehat{M}_t\right] \leq \widehat{X}_0} \mathbb{E}\left[\int_0^T U\left(\widehat{c}_t\right) \mathrm{d}t\right]. \tag{1.2.6}$$

[18] These similarities follow from the nature of the Legendre-Fenchel transform, $V(x) = \sup_{z\in\mathbb{R}} \{U(z) - xz\}$, for all $x \in \mathbb{R}$. The analytical structures of both formulations are direct consequences of the same inequality inherent in this transform.

We define $\widehat{c}_t$ as: $\widehat{c}_t = c_t - h_t$, for all $t \in [0,T]$. In conformity with the artificial market from section 1.2.1, the present auxiliary environment implicitly links the primal to the dual. The problem in (1.2.6) therefore corresponds to the left- and right-hand sides of Figures 1.1 and 1.4. Note that this auxiliary formulation is "static" and omits analytically cumbersome path-dependency. It is static in the sense that agent is solely concerned about $\widehat{c}_t$ over $[0,T]$.[19] The optimal value for the corresponding trading strategy, π_t, can always be deduced from the specification(s) of $\widehat{c}_t$ and/or c_t. Due to the preceding features of the auxiliary formulation, acquiring optimal solutions to (1.2.6) is quite straightforward. More specifically, the optimal solution is given by: $\widehat{c}_t = I(\eta \widehat{M}_t)$, for all $t \in [0,T]$. Here, $x \mapsto I(x)$ spells out the inverse of marginal utility. To correspondingly procure c_t, we can use that the following holds: $c_t = \widehat{c}_t + e^{(\beta-\alpha)t}h_0 + \beta \int_0^t e^{(\beta-\alpha)[t-s]}\widehat{c}_s \mathrm{d}s$, for all $t \in [0,T]$. Recall that π_t can be obtained from c_t via particular hedging arguments. Clearly, this identity stipulates how the primal variables, c_t and π_t, depend on the dual control, η, and vice versa. Hence, the former equation outlines a valid duality relation. Unlike (1.2.3), we can employ this duality relation to derive the solutions to (1.2.4) in closed-form. To accomplish which, one solely has to obtain the multiplier, η, from $\mathbb{E}[\int_0^T I(\eta \widehat{M}_t)\widehat{M}_t \mathrm{d}t] = \widehat{X}_0$. Hence, using duality techniques, we have been able to solve the mathematically troublesome problem (1.2.4) in complete analytical form. This, again, demonstrates the value of duality and how it may facilitate a retrieval of optimal solutions.[20]

1.2.3 Brief Discussion

The benefits of duality are not limited to those highlighted by the previous set of two examples. In the domain of portfolio optimisation, there are numerous other situations wherein duality proves useful. For example, in markets with

[19] This phenomenon is referred to as time-separability, cf. Detemple and Karatzas (2003).

[20] In Chapter 3, we concentrate on a problem involving multiplicative habit formation. The formulation central to section 1.2.2 is therefore closely related to the topic of the latter chapter. We would like to point out that the transformation of (1.2.4) to (1.2.6) is indirectly analysed by Schroder and Skiadas (2002). The authors relate the two problems to each other via a mathematically defined isomorphism. In doing so, they term the auxiliary environment the "dual" market, which they employ to unravel the exact primal-dual relations. For concrete applications of this isomorphism, we refer to e.g. Munk (2008), van Bilsen and Laeven (2020), and van Bilsen et al. (2020b).

frictions, duality techniques have given rise to the convenient notion of a shadow price. This process simplifies the derivation of optimal solutions and expedites the burdensome procedure of verification, cf. Kallsen and Muhle-Karbe (2010), Choi et al. (2013), and Bichuch and Guasoni (2018). Similar upsides apply to the computation of super-replication prices in comparable markets, cf. Cvitanić and Karatzas (1996), Cvitanić et al. (1999), and Campi and Schachermayer (2006). Furthermore, due to its close relation to utility-maximisation, duality techniques are also of fundamental importance to the literature on partial hedging. The papers by a.o. Pham (2002), Bouchard et al. (2004), and Kim (2012) derive the corresponding dual formulations and exemplify this importance along applied lines. On a more secondary level, duality techniques may bring forth new methods or improve existing ones. The approximating routines developed by e.g. Haugh et al. (2006), Brown et al. (2010), and Weiss (2020) directly hinge on the duality-induced artificial market. In Bick et al. (2013), the duality gap is employed to measure the accuracy of their approximate solutions. This list is by far not exhaustive. Even though we have only supplied anecdotal evidence of duality's use, the gist of its general importance should be clear. In addition to the upsides listed at the beginning of this chapter, we thereby hope to have shed sufficient light on the many possible answers to the question: *Why is duality useful?*

1.3 Duality in this Dissertation

We conclude this introductory chapter with an overview of the three chapters central to this dissertation. Note that these chapters comprise of the primary research output. All three "core" chapters pertain to duality in either a direct or an indirect manner. Interestingly enough, we are able to relate each chapter to a different part of Figure 1.1. Translated into theory, this means that each chapter touches upon a different feature of duality. To make this clear, as pointed out in section 1.1.4, we subsequently indicate where and how we can position the distinct chapters in Figure 1.1. In a similar sense, we are able to categorise the chapters into three connected topics on the subject of utility-maximisation. This categorisation is primarily based on the different preference qualifications and corresponding targets of optimisation. In the sequel, we visit the separate chapters and address their content in more detail. Moreover, we comment on their contributions to the literature on portfolio optimisation and related duality techniques/applications. We finalise this introductory section with a general outline of the remainder of this dissertation.

1.3.1 Chapter 2

In the first chapter of this dissertation, we focus on a constrained utility-maximisation problem similar to the one in (1.2.1). We modify the preceding formulation to additionally account for possibly non-traded labour income and utility over consumption. The ensuing model specification is therefore general enough to cover many different financial/economic situations. Concerning the category of utility-maximisation, we observe that the target of optimisation involves utility over both terminal wealth *and* consumption.[21] From section 1.2.1, we know that constrained problems of this form may be difficult to solve analytically. This is in the first place attributable to the issues emanating from the non-uniqueness of the equivalent martingale measures. To obtain optimal solutions, one can resort to numerical applications, e.g. backward induction techniques or grid-search routines. However, such methods can be computationally demanding, and generate optima that lack the transparency of closed-form expressions. Due to the latter, practical execution/implementation of the optimal policy rules may be cumbersome.

To circumvent these downsides, in Chapter 2, we develop an approximate dual-control method. Its mechanism bears resemblance to the duality-based routines proposed by Haugh et al. (2006), Brown et al. (2010), Brown and Smith (2011), Ma et al. (2017, 2020), Weiss (2020), and Hambel et al. (2021). More specifically, it constitutes a generalisation of the SAMS (Simulation of Artificial Markets Strategies) scheme developed by Bick et al. (2013). Our method primarily relies on the artificial market specification to construct

[21]The utility functions are assumed to be state-dependent. Preferences are accordingly allowed to vary with respect the state of the economy. In addition to this, both specifications involve so-called benchmark processes. These processes can be modelled as person-specific targets. Valid benchmarks would, for instance, be one's labour income or national GDP. This makes the framework amenable to applications in the domain of prospect theory, cf. Kahneman and Tversky (1979). In Chapter 4, we make explicit use of a loss aversion-linked reference-dependent preference qualification.

analytical approximations. Most of this chapter therefore relates to the left- and right-hand sides of Figure 1.1. The method is in principle three-fold and works as follows. First, it addresses the toughest part, and approximates the dual-linked martingale measure. For this purpose, the method restricts the feasible set of martingale measures to a tractable subspace. In accordance with this restriction, it re-optimises the dual. Second, it uses the artificial market to transform the approximate dual control into a primal counterpart. As this counterpart is typically not admissible in the financial environment, the method makes use of a projection operator. Consequently, at this stage, the method is in possession of feasible approximate analytical primal and dual controls. Third, to measure the accuracy of these approximations, it examines the magnitude of the corresponding duality gap. This gap can be quantified in terms of monetary units that admit clear financial interpretations. In the examples that we consider, the method results in relatively small welfare losses. Hence, we conclude that our dual-control approximate method is capable of rendering closed-form near-optimal investment-consumption policies.

1.3.2 Chapter 3

In the second chapter of this dissertation, we study a utility-maximisation problem involving habit formation. Unlike the formulation in (1.2.4), we make use of a *multiplicative* habit model. The agent is correspondingly assumed to derive utility from the ratio of consumption to the habit component. As this ratio is strictly positive, irrespective of one's consumption behaviour, we are able to relax the additive-specific lower bound. That is, consumption is *not* required to exceed the habit level at all times. Due to the omission of this requirement, the multiplicative habit model gains significant relevance from a micro-related perspective. In particular, we are now able to interpret the habit level as a standard of living unique to some household. Adverse changes in the financial circumstances can namely urge a household to scale down consumption below the level to which they have become accustomed. Regarding the category of utility-maximisation, we note that the objective of optimisation exclusively involves utility over consumption. The specification of this model dates back to Abel (1990), and has been economically advocated by

Carroll (2000) and Carroll et al. (2000). In section 2.3 of Rogers (2013), the author recommends the model from a theoretical point of view.

The objective of the multiplicative habit problem is not fully concave and involves irremovable path-dependency. The latter implies that it is not possible to transform the target of optimisation into a time-separable analog. In fact, Schroder and Skiadas (2002)'s isomorphism only manages to transfer the path-dependency from the objective to the constraint qualifications. As a result of these two attributes, the conventional Lagrangian duality techniques fail to generate a dual formulation. In more precise terms, the ordinarily employed Legendre transform solely applies to time-separable problems with concave objective functions. For this reason, to the best of our knowledge, there is no dual problem known for the multiplicative habit model. To fill this gap in the literature, in Chapter 3, we derive a mathematical formulation of the dual corresponding to this multiplicative model. Most of this chapter is consequently related to the upper half of Figure 1.1. In order to derive the dual, we resort to the less well-known notion of Fenchel duality. Contrary to the Legendre transform, Fenchel is able to deal with path-dependent transformations of the relevant control variables.[22] This form of duality renders an analytically defined dual formulation and simultaneously proves that strong duality holds. Complementary to this fundamental duality result, we develop an approximating mechanism similar to the one of Chapter 2. We test the mechanism on the approximation proposed by van Bilsen et al. (2020a). This approximation is proven to be accurate under a wide variety of circumstances. The ensuing welfare losses are confirmed to be small by our novel method.

1.3.3 Chapter 4

In the third and final "core" chapter, we analyse an unconstrained utility-maximisation problem nearly identical to the one in (1.2.1). The main difference consists in the identification of the objective as a hedging criterion. Although it

[22]While Fenchel duality is able to cope with path-dependent transformations, its statement only applies to fully concave optimisation problems. Therefore, we are required to re-define the control variables. Using a simple logarithmic transformation of both consumption and the habit component, we are able to "concavify" the original problem. We consequently arrive at an objective that can be included under the aegis of Fenchel duality.

formally qualifies as a preference function, we namely employ the so-called lower partial moments (LPM) operator as the target of optimisation. This operator plays a central role in the literature on partial hedging, cf. Sekine (2004), Xia (2005), and Choi and Jonsson (2009). In that context, the LPM criterion is used to model situations wherein one is required to hedge a pre-defined claim with insufficient funds. Utility-maximising agents may be confronted with similar situations. To this end, let us postulate that an agent has in mind some goal with respect to his/her terminal wealth. This goal, target or benchmark is ordinarily referred to as the reference level. In practice, due to underfunding problems, it is often the case that the reference level is not attainable without taking risk. This is specifically true for agents in the pension industry. Hence, to optimise the likelihood of ultimately acquiring the reference level, it is sensible to resort to partial hedging criteria. Within the confines of a terminal wealth problem, we correspondingly consider an LPM operator that incorporates a person-specific reference level. It is therefore clear that the value function solely includes utility over terminal wealth.

Chapter 4 analyses the aforementioned problem against the background of a defined contribution (DC) pension scheme. The value for terminal wealth can consequently be aligned with retirement wealth. Likewise, we identify the reference level as a person-specific life annuity. To model the market's return dynamics, we rely on the financial environment proposed by Koijen et al. (2009). This model assumes an affine-term structure for the interest rates and incorporates four distinct risk-drivers. In addition to this, it distinguishes nominal from real returns. To the best of our knowledge, there are no studies available that consider the LPM problem in such an applied setup. The literature on partial hedging is strongly oriented towards the theory, see e.g. Pham (2000, 2002), Jonsson and Sircar (2002), and Nygren and Lakner (2012). We analytically solve the LPM-linked terminal wealth problem by means of the martingale technique. Whereas this technique directly follows from duality machinery, most of this chapter revolves around the lower half of Figure 1.1. Due to the affine nature of the market model, it is not possible to disclose the distributional properties of the unique SPD process. For this reason, we must make use of the Fourier transform, cf. Carr and Madan (1999), to obtain the optimal policy rules. Our numerical findings show that the LPM operator can improve the likelihood of achieving one's pension goals. In spite of this great performance, the outcomes also suggest that the optimal trading strategies may be hard to implement. Moreover, we demonstrate that the results are highly sensitive to the estimates for the market prices of risk.

1.3.4 Outline

To closely follow the academic conventions, we briefly outline this dissertation's roadmap. In Chapters 2, 3 and 4, we present the main research output. For short summaries on the subjects of these chapters, one may re-consult this introductory section or the abstracts provided at the start of each chapter. In a similar way, as a point of reference, it could be useful to re-consider Figure 1.1 or Figure 1.4 prior to reading these core chapters. For this purpose, let us recall that Chapter 2 concerns the left- and right-hand sides of both figures; Chapter 3 touches upon the corresponding upper halves; and Chapter 4 predominantly pertains to the associated lower halves. The research output accordingly covers all duality-linked domains relevant to portfolio optimisation. Finally, in Chapter 5, we conclude this dissertation. We concretely re-examine the conclusions drawn from the core chapters. In addition to this, we provide (i) a more general statement on the overall contribution of this dissertation to the literature, and (ii) a short outlook on future branches of possibly relevant research. As final attachments, *Duality Methods for Stochastic Optimal Control Problems in Finance* includes a bibliography, a summary of this dissertation, a valorisation or impact paragraph, and the author's curriculum vitae.

2

Near-Optimal Asset Allocation in Financial Markets with Trading Constraints

Adapted from: Kamma, T., & Pelsser, A. (2022c). Near-optimal asset allocation in financial markets with trading constraints. *European Journal of Operational Research*, *297*(2), 766–781.

Abstract

We develop a dual-control method for approximating investment strategies in multidimensional financial markets with convex trading constraints. The method relies on a projection of the optimal solution to an (unconstrained) auxiliary problem to obtain a feasible and near-optimal solution to the original problem. We obtain lower and upper bounds on the optimal value function using convex duality methods. The gap between the bounds indicates the precision of the near-optimal solution. We illustrate the effectiveness of our method in a market with different trading constraints such as borrowing, short-sale constraints and non-traded assets. We also show that our method works well for state-dependent utility functions.

2.1 Introduction

In most cases, investment problems do not allow for an expression of the optimal trading and consumption strategies in closed-form. This is especially true for investment problems with trading constraints. Due to the presence of these constraints, there does not exist a unique financially fair valuation of assets, i.e. a risk-neutral pricing measure. In fact, an infinite amount of such measures is available. The investor therefore faces the task of selecting a risk-neutral valuation of the (partially) unhedgeable assets, in addition to the optimisation of his/her utility from trading and consumption. This selection of the optimal risk-neutral valuation is, in general, analytically troublesome, and obstructs an expression of the solutions to constrained investment problems in closed-form.[1] Since analytical solutions (i) clarify the precise roles played by the model parameters, (ii) simplify practical implementations, and (iii) facilitate comparative statistical analyses, see e.g. van Bilsen et al. (2020a), it is beneficial to have closed-form solutions.

The papers by Kim and Omberg (1996), Wachter (2002), Liu (2006), and Battauz et al. (2015b) address and demonstrate the analytical issues involved with solving constrained investment problems. To circumvent these analytical difficulties and still be able to obtain solutions, several approximate methods have been developed. The nature of these methods can roughly be divided along the two following lines: simulation-based ones that render numerical solutions, cf. Cvitanić et al. (2003), Detemple et al. (2003), Brandt et al. (2005), and Keppo et al. (2007)[2]; and those that strive for closed-form outcomes using duality methods, cf. Haugh et al. (2006), Brown et al. (2010), Brown and Smith (2011), Ma et al. (2017, 2020), and Weiss (2020). We concentrate in this chapter on the last branch of routines. Amongst this set of methods, the one proposed by Bick et al. (2013), SAMS (Simulation of Artificial Markets

[1] Section 6.1 in Detemple (2014) demonstrates that the optimal risk-neutral pricing measure is ordinarily characterised by a forward-backward stochastic differential equation (FBSDE). For the roles that FBSDE's play in asset allocation problems, we refer to El Karoui et al. (1997) and Detemple and Rindisbacher (2010).

[2] For practical implementations of the widely applied numerical grid-search methods, see e.g. Longstaff (2001), Haliassos and Michaelides (2003), and Cocco et al. (2005). We refer to Schröder et al. (2013) for a more comprehensive overview of the available simulation-based numerical schemes.

Strategies), stands out in terms of analytical transparency and accuracy. Their method makes use of convex duality techniques.[3] This technology enables one to solve an auxiliary problem, in which one optimises over the feasible risk-neutral pricing measures instead of over the admissible trading and consumption strategies. Each measure implies a pair of closed-form trading and consumption policies via the so-called barrier cone. In fact, by duality principles, the optimal (minimising) pricing measure implies optimal and admissible (maximising) trading and consumption rules. However, the optimisation of the auxiliary problem itself is, in general, analytically difficult.

To bypass the analytical difficulties involved with specifying the optimal pricing measure, and to also retain closed-form solutions, Bick et al. (2013)'s SAMS method uses a twofold approximating scheme: (i) a tractable approximation of the risk-neutral pricing measure, and (ii) a projection of the analytical trading and consumption strategies. The scope of application of the SAMS method concerns setups with quadratic or affine returns, specific definitions of the trading constraints, and CRRA preferences. Guasoni and Wang (2020) propose a variant of the SAMS method for optimal consumption problems, which is additionally able to cope with investment opportunities that depend on Markovian state variables. To enlarge the scope of application even further, in this chapter we develop an approximate method that generalises the method by Bick et al. (2013), such that it is applicable to problems with (i) general return dynamics with non-quadratic and non-affine structures that admit state variables, (ii) general trading and liquidity constraints, and (iii) state-dependent utility functions that are (possibly) specified over the real line (e.g. exponential utility functions) and embed a stochastic benchmark. As for the relevance of the fact that our method works as well for item (iii), we refer to Battauz et al. (2011) and Battauz et al. (2015a), who address the economic appeal of state-dependent utility functions and their complicating effect on the recovery of analytical solutions.

We derive results for a more general class of models, and modify Bick et al. (2013)'s routine. In particular, the first step of our approximate method

[3]The convex duality methodology and the related (unconstrained) artificial market theory are developed by Karatzas et al. (1991a), Cvitanić and Karatzas (1992), and Xu and Shreve (1992).

consists of restricting the space containing all feasible pricing measures to a tractable convex subspace. Then, we approximate the optimal pricing measure by optimising over this subspace. The optimal pricing measure implies a unique closed-form expression for the trading and consumption policies. However, due to sub-optimality of the approximate measure, these controls are inadmissible.[4] Therefore, the second step of our approximate method consists of projecting the trading and consumption policies to be admissible. Correspondingly, we use the inadmissible policies as a guide for obtaining admissible closed-form near-optimal solutions. As a result, we acquire closed-form approximations to the optimal investment and consumption rules. By convex duality principles, the approximation provides a so-called duality gap. The magnitude of this gap decreases with the accuracy of the approximation. Hence, to evaluate the accuracy of the approximated policies, we can use the magnitude of the duality gap. The approximate asset allocation and consumption rules therefore come with a hard guarantee concerning their precision.

To illustrate the effectiveness of our method in a concrete setup, we use a modified version of Cocco et al. (2005)'s financial market model. This model contains a cash account, two assets and labour income. The changes in labour income cannot be fully hedged. Furthermore, we assume that the agent has a state-dependent utility function, where utility is specified in terms of the ratio of terminal wealth (and consumption) relative to a stochastic price index. Additionally, we replace utility from terminal wealth in their setup by Chen et al. (2011)'s SAHARA function, and enlarge its asset mix by adding one asset. The objective of the agent is to maximise expected utility by selecting a consumption and trading strategy over the life-cycle, under trading constraints imposed by the non-traded labour income and/or price inflation. We study three different specifications of these constraints, which represent various economically relevant situations. For all three setups, closed-form solutions are not available. To nevertheless acquire closed-form solutions, we apply our approximate method.[5]

[4]Unless the contracted subspace contains the optimal pricing measure, the approximation implies trading and consumption rules that are inadmissible, see e.g. Proposition 12.1 in Cvitanić and Karatzas (1992).

[5]The SAHARA qualification is specified over the real line, and is therefore illustrative for one of the elements that differentiates our method from Bick et al., 2013's. The three definitions of the constraints involve a requirement to maintain wealth above zero, as well as impediments to trade, short-sale and borrow. Accordingly, these manage to adequately

The outcomes show that the bounds on the annual welfare losses, suffered due to implementation of the approximate rules, vary between 0.000% and 0.051% of the investor's initial amount of wealth. These welfare losses are negligibly small, which demonstrates that the method can be stable, and may be capable of rendering near-optimal strategies.

The remainder of the chapter is structured as follows. Section 2.2 introduces the financial market model. Subsequently, section 2.3 establishes the theoretical background with regard to convex duality and constrained investment problems. Section 2.4 specifies the approximate method, and provides a numerical evaluation thereof. Finally, section 2.5 concludes.

2.2 Model Setup

In this section, we introduce and revise some existing concepts/results that are necessary for the remainder of the chapter. First, we introduce the financial market model. Second, we outline the investment problem. This problem relies on state-dependent utility functions that incorporate benchmark processes. As a result, these functions are more general than the conventional ones. We therefore analyse these functions in detail.

2.2.1 Financial Market Model

Our model forms a combination of the markets in Cuoco (1997) and Detemple and Rindisbacher (2010)'s section 2.1. Let $T > 0$, and fix a probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, on which an $\mathbb{R}^N$-valued standard Brownian motion, $\{W_t\}_{t\in[0,T]}$, is specified. Here, $\mathbb{P}$ represents the physical probability measure, and $\mathbb{F} = \{\mathcal{F}_t\}_{t\in[0,T]}$ denotes the $\mathbb{P}$-augmentation of W_t's canonical filtration, $\{\mathcal{F}_t^W\}_{t\in[0,T]}$. Henceforth, (in)equalities between stochastic processes are understood in a $\mathbb{P}$-a.s. sense.

We outline a market, $\mathcal{M}$, that consists of an investor who continuously trades over $[0, T]$ in a riskless instrument and N risky assets, i.e. stocks. The risk-free

demonstrate the other distinguishing attribute(s).

asset reads

$$\frac{\mathrm{d}B_t}{B_t} = r_t \mathrm{d}t,\ B_0 = 1, \tag{2.2.1}$$

where r_t defines the $\mathbb{R}$-valued and $\mathcal{F}_t$-progressively measurable instantaneous interest rate. We assume here that $r_t \in \mathbb{D}^{1,2}([0,T])$ holds. Here, $\mathbb{D}^{1,2}([0,T])$ represents the so-called Sobolev-Watanabe space. This space contains all $L^2(\Omega \times [0,T])$-processes that are Malliavin differentiable, cf. Karatzas et al. (1991b) and Nualart (2006). The price processes for the N risky assets follow the following stochastic differential equation (SDE):

$$\frac{\mathrm{d}S_{i,t}}{S_{i,t}} = \mu_{i,t}\mathrm{d}t + \sigma_{i,t}^\top \mathrm{d}W_t,\ S_{i,0} = 1, \tag{2.2.2}$$

in which $\mu_{i,t}$ represents the $\mathbb{R}$-valued expected return, and $\sigma_{i,t}$ denotes the $\mathbb{R}^N$-valued corresponding volatility process: both processes are $\mathcal{F}_t$-progressively measurable. We postulate that $\|\mu_t\|_{\mathbb{R}^N}, \mathrm{Tr}(\sigma_t\sigma_t^\top) \in L^1([0,T])^N$, in which $\mu_t \in \mathbb{R}^N$ has entries $\mu_{i,t}$, and $\sigma_t \in \mathbb{R}^{N\times N}$ rows $\sigma_{i,t}$, $i = 1,\dots,N$. In addition to this, we suppose that σ_t fulfills the strong non-degeneracy assumption, $\phi^\top\sigma_t\sigma_t^\top\phi \geq \epsilon\|\phi\|^2_{\mathbb{R}^N}$ for all $\phi \in \mathbb{R}^N$ and some $\epsilon > 0$, which ensures that σ_t is invertible. The space of all p-integrable $\mathbb{R}^N$-valued $\mathcal{F}_t$-progressively measurable processes is given by $L^p([0,T])^N$; the space of all such processes with finite expectations reads $L^p(\Omega\times[0,T])^N$.

In the absence of unhedgeable risk, by Delbaen and Schachermayer (1994), we know that there exists a unique state price density (SPD), $\{M_t\}_{t\in[0,T]}$. Let $\lambda_t := \sigma_t^{-1}(\mu_t - r_t 1_N)$, then M_t follows for all $t \in [0,T]$

$$M_t = \exp\left\{-\int_0^t r_s \mathrm{d}s - \frac{1}{2}\int_0^t \|\lambda_s\|^2_{\mathbb{R}^N}\mathrm{d}s - \int_0^t \lambda_s^\top \mathrm{d}W_s\right\}, \tag{2.2.3}$$

where we assume that $\lambda_t \in \mathbb{D}^{1,2}([0,T])^N$, and in which $\{B_t\}_{t\in[0,T]}$ serves as numéraire. As a result of $\lambda_t \in \mathbb{D}^{1,2}([0,T])$, λ_t satisfies Novikov's condition: $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T \|\lambda_s\|^2_{\mathbb{R}^N}\mathrm{d}s\right)\right] < \infty$, cf. Karatzas and Shreve (2012). Observe that M_t may serve to price traded assets: e.g. $\{M_tS_t\}_{t\in[0,T]}$ is a $\mathbb{P}$-martingale with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$.

The finite-horizon investor in $\mathcal{M}$ receives a continuous stream of non-negative stochastic labour income over $[0, T]$. This random endowment is exogenous and evolves as:

$$\frac{\mathrm{d}Y_t}{Y_t} = \mu_{Y,t}\mathrm{d}t + \sigma_{Y,t}^\top \mathrm{d}W_t, \;\; Y_0 \in \mathbb{R}_+, \tag{2.2.4}$$

where we assume that expected income growth, $\mu_{Y,t}$, and income volatility, $\sigma_{Y,t}$, satisfy $\mu_{Y,t}, \sigma_{i,Y,t} \in \mathbb{D}^{1,2}\left([0,T]\right)$ for all entries, $i = 1, \ldots, N$, of the $\mathbb{R}^N$-valued vector $\sigma_{Y,t}$. As a consequence, the income process fulfills $Y_t \in L^2\left([0,T]\right)$, see El Karoui and Jeanblanc-Picqué (1998).

The agent's wealth process is endogenously determined by consumption, c_t, and the allocation to assets, π_t. We let $\{c_t\}_{t\in[0,T]}$ be the $\mathbb{R}$-valued $\mathcal{F}_t$-progressively measurable consumption process and let $\{\pi_{0,t}, \pi_t\}_{t\in[0,T]}$ be the $\mathbb{R}^{N+1}$-valued $\mathcal{F}_t$-progressively measurable portfolio process. For a fixed initial endowment $X_0 \in \mathbb{R}_+$, the dynamic wealth process of the investor follows:

$$\begin{aligned} \mathrm{d}X_t &= \pi_{0,t}B_t^{-1}\mathrm{d}B_t + \pi_t^\top \mathrm{diag}\left(S_t\right)^{-1}\mathrm{d}S_t - \left(c_t - Y_t\right)\mathrm{d}t \\ &= \pi_{0,t}r_t\mathrm{d}t + \pi_t^\top\left(\mu_t\mathrm{d}t + \sigma_t\mathrm{d}W_t\right) - \left(c_t - Y_t\right)\mathrm{d}t, \;\; X_0 \in \mathbb{R}_+. \end{aligned} \tag{2.2.5}$$

We call a trading-consumption pair $\{c_t, \pi_{0,t}, \pi_t\}_{t\in[0,T]}$ *admissible* if it satisfies: $X_t = \pi_{0,t} + \pi_t^\top 1_N \geq -C$, $\int_0^T \pi_t^\top \sigma_t \sigma_t^\top \pi_t \mathrm{d}t < \infty$, $\int_0^T \left|\pi_t^\top \sigma_t \lambda_t + r_t X_t\right| \mathrm{d}t < \infty$, and $\mathbb{E}[\int_0^T \left|c_t\right|^2 \mathrm{d}t] < \infty$, for some $C \in \mathbb{R}_+$.[6] We denote the class of admissible pairs by $\mathcal{A}_{X_0}$.

Finally, let us introduce a non-empty, closed and convex set $K \subseteq \mathbb{R}^{N+1}$ that comprises of the constraints enforced on $(\pi_{0,t}, \pi_t)$ in a $\mathrm{d}t \otimes \mathbb{P}$-a.e. sense. Accordingly, we define $\widehat{\mathcal{A}}_{X_0}$ as the set of all admissible trading-consumption pairs, $\{c_t, \pi_{0,t}, \pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$, such that $(\pi_{0,t}, \pi_t) \in K$ holds $\mathrm{d}t \otimes \mathbb{P}$-a.e. We assume that $0_{N+1} \in K$ holds for technical purposes. As in Haugh et al. (2006), we note that K is constant, but it can as well be made dependent on time and the values of exogenous (state) variables.

[6] Most setups require $C = 0$, cf. He and Pages (1993), Koo (1998), and Cvitanić et al. (2001). As in e.g. El Karoui and Jeanblanc-Picqué (1998), Owen and Žitković (2009), and Dybvig and Liu (2010), we conversely allow the finite-horizon investor to borrow against future labour income, with the possibility of obtaining terminal debt. Consistent with this, we specify the investor's preferences.

2.2.2 Utility Function and Problem Description

The economic environment $\mathcal{M}$ consists of a finite-horizon investor who is at $t=0$ equipped with a prefixed endowment $X_0 \in \mathbb{R}_+$, and who retires at $t=T$. Over the course of the trading interval, $[0,T]$, this individual aims to maximise expected working life utility from consumption and expected utility from terminal wealth by holding a continuously rebalanced portfolio. We assume that the investor compares consumption and terminal wealth to two different individual-specific stochastic benchmark processes. In particular, the agent derives utility from consumption and terminal wealth, in relation to these benchmark processes. These benchmarks with respect to the endogenous rules are completely exogenous and evolve according to the following SDE's:

$$\mathrm{d}\Pi_{i,t} = \mu_{\Pi,i,t}\mathrm{d}t + \sigma_{\Pi,i,t}^{\top}\mathrm{d}W_t,\ \Pi_0 \in \mathbb{R}. \tag{2.2.6}$$

We assume $(\mu_{\Pi,i,t}, \sigma_{\Pi,i,t}) \in \mathbb{D}^{1,2}([0,T]) \times \mathbb{D}^{1,2}([0,T])^N$, for $i=1,2$. The economic interpretation of these stochastic benchmarks could vary from national GDP to the labour income of one's neighbour, provided that these remain unaffected by the finite-horizon investor's decisions (exogenous). The utility-maximising investor compares $\{c_t\}_{t\in[0,T]}$ to $\{\Pi_{1,t}\}_{t\in[0,T]}$, and X_T to $\Pi_{2,T}$. We stress that our framework allows the latter semi-martingale benchmark processes to attain values on the entire real line.

The utility-maximising investor faces the following dynamic stochastic optimal control problem:

$$\begin{aligned} J_P^{\mathrm{opt}}\left(\bar{X}_0\right) &= \sup_{\{\pi_{0,t},\pi_t,c_t\}_{t\in[0,T]}\in\widehat{\mathcal{A}}_{X_0}} \mathbb{E}\left[\int_0^T u\left(t,c_t,\Pi_{1,t}\right)\mathrm{d}t + U\left(X_T,\Pi_{2,T}\right)\right] \\ \text{s.t.}\quad \mathrm{d}X_t &= \pi_{0,t}r_t\mathrm{d}t + \pi_t^{\top}\left(\mu_t\mathrm{d}t + \sigma_t\mathrm{d}W_t\right) - \left(c_t - Y_t\right)\mathrm{d}t, \end{aligned} \tag{2.2.7}$$

for all $\bar{X}_0 := [X_0, Y_0]^{\top} \in \mathbb{R}_+^2$ and two utility functions $u : [0,T]\times\mathbb{R}_+\times\mathbb{R}_+\times\Omega \to \mathbb{R}$ (utility from consumption) and $U : \mathbb{R}\times\mathbb{R}\times\Omega \to \mathbb{R}_+$ (utility from retirement wealth), both of which are state-dependent and incorporate the benchmark processes, $\{\Pi_{i,t}\}_{t\in[0,T]}$, $i=1,2$.[7] More precisely, to maximise (2.2.7), the

[7]Throughout, we suppress the dependency of all utility-related functionals on $\omega \in \Omega$ for

investor chooses an allocation of funds to the $N+1$ assets, $\{\pi_{0,t}, \pi_t\}_{t\in[0,T]}$, and chooses a particular consumption pattern, $\{c_t\}_{t\in[0,T]}$, throughout the trading interval, $[0,T]$. The selection of controls must be consistent with (2.2.5), such that the admissibility conditions are met.[8] Note here that U is specified over $\mathbb{R}\times\mathbb{R}$, unless explicitly stated otherwise; whereas the domain of u is given by $\mathbb{R}_+\times\mathbb{R}$, for fixed values of $t\in[0,T]$.

Let us consider $u:[0,T]\times\mathbb{R}_+\times\mathbb{R}\times\Omega\to\mathbb{R}$. We assume that u satisfies:

$$\lim_{x\to\infty} u'_X(t,x,y)=0, \quad \lim_{x\to-\infty} u'_X(t,x,y)=\infty, \tag{2.2.8}$$

and

$$\limsup_{x\to\infty} x\frac{u'_X(t,x,y)}{u(t,x,y)}<1 \tag{2.2.9}$$

for all $y\in\mathbb{R}$ and $t\in[0,T]$. Here, $u'_X:[0,T]\times\mathbb{R}_+\times\mathbb{R}\times\Omega\to\mathbb{R}_+$ and $u''_X:[0,T]\times\mathbb{R}_+\times\mathbb{R}\times\Omega\to\mathbb{R}_-$ represent the first and second derivatives of u with respect to x, respectively. The two limits in (2.2.8) are the ordinary Inada conditions. The limit in (2.2.9) is the reasonable asymptotic elasticity requirement, cf. Kramkov and Schachermayer (1999). Jointly, these conditions ensure the validity of convex duality applications. On $U:\mathbb{R}\times\mathbb{R}\times\Omega\to\mathbb{R}$, making use of similar notation for the first and second derivatives (U'_X and U''_X), we impose the same conditions as on u (for fixed $t\in[0,T]$), with the additional requirement that $\limsup_{x\to-\infty} x\frac{U'_X(x,y)}{U(x,y)}>1$ holds.

We define the convex conjugate functions of u and U, for all $x\in\mathbb{R}_+$, $y\in\mathbb{R}$, and $t\in[0,T]$ as follows

$$\begin{aligned} v(t,x,y) &= \sup_{z\in\mathbb{R}_+}\{u(t,z,y)-xz\} \\ &= u(t,\iota(t,x,y),y)-x\iota(t,x,y), \end{aligned} \tag{2.2.10}$$

notational elegance, as all subsequent results hold on a per-state basis.

[8] Combining the arguments in Schachermayer (2001) and Chen and Vellekoop, 2017, dictating that $\|\pi_t\|^2_{\mathbb{R}^N}\le Y\left(1+X_t^2\right)$ holds for some $Y\in L^0(\Omega)$, guarantees that $\{X_t\}_{t\in[0,T]}$ is indeed well-specified, i.e. "martingale-generating", and thus that the problem is well-posed in spite of U allowing for negative arguments. We henceforth assume that $X_t\ge -C$ is in $\mathcal{A}_{X_0}$ and $\widehat{\mathcal{A}}_{X_0}$ replaced by the latter condition.

and

$$V(x,y) = \sup_{z\in\mathbb{R}} \{U(x,y) - xz\} \\ = U(I(x,y),y) - xI(x,y), \tag{2.2.11}$$

in which $\iota : [0,T] \times \mathbb{R}_+ \times \mathbb{R} \times \Omega \to \mathbb{R}_+$ and $I : \mathbb{R}_+ \times \mathbb{R} \times \Omega \to \mathbb{R}$ are the inverses of marginal utility, u'_x and U'_X, respectively such that $u'_X(t, \iota(t,x,y), y) = x$ and $U'_X(I(x,y),y) = x$ hold for all $y \in \mathbb{R}$ and $t \in [0,T]$. Furthermore, we let $u'_Y : [0,T] \times \mathbb{R}_+ \times \mathbb{R} \times \Omega \to \mathbb{R}_-$ and $U'_Y : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}_-$ be the first derivatives of u and U in the y-direction. For all $t \in [0,T]$, we note that u and U ought to be once continuously differentiable in both arguments, $u \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ and $U \in \mathcal{C}(\mathbb{R}, \mathbb{R}; \mathbb{R})$. Likewise, for all $t \in [0,T]$, we assume that u'_x, ι, U'_X and I are at least once piecewise continuously differentiable in both arguments, $u'_X, \iota \in \mathcal{PC}(\mathbb{R}_+, \mathbb{R}; \mathbb{R}_+)$ and $U'_X, I \in \mathcal{PC}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$.

The descriptions of u and U are based on those in Detemple and Zapatero, 1991. We enlarge their generality by relying on Lakner and Nygren (2006) who prove that the following holds true

$$F(G,H) \in \mathbb{D}^{1,2}([0,T]), \quad \text{for} \quad G,H \in \mathbb{D}^{1,2}([0,T]), \tag{2.2.12}$$

for all $F \in \mathcal{PC}(\mathbb{R}, \mathbb{R}; \mathbb{R})$, or all $F \in \mathcal{PC}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ conditional on $G \in \mathbb{D}^{1,2}_+([0,T])$. The latter property implies that one is able to employ utility functions, whose marginal analogues incorporate breakpoints, in standard utility-maximising investment frameworks, without these breakpoints affecting the derivation of the optimal dynamic asset allocation in closed-form. One may consult Ocone and Karatzas (1991) for the link between Malliavin-differentiability and the analytical derivation of optimal portfolio rules. For an example of a utility function that includes breakpoints, see the dual-CRRA preference qualification in Balter et al. (2020).

Moreover, the utility functions expand the conventional specifications via the inclusion of an additional argument (y) within their definitions. These functions consequently admit a widespread variety of preferences that apply to different theoretical frameworks, varying from prospect theory to external habit formation, cf. Abel (1990), Campbell and Cochrane (1999), and Alvarez-Cuadrado et al. (2004). In particular, if we attach the interpretation of a person-specific benchmark to the second arguments, this enables us to characterise the preferences of an individual around this reference level. This results in more target-focused optimal investment strategies, e.g. in the context of pension schemes, cf. Lim and Wong (2010) and Han and Hung (2012), or in setups involving risk-management, see Tepla (2001) and Basak et al. (2007).

2.3 Convex Duality

We continue by analysing the constrained optimal investment problem in (2.2.7). Due to the presence of trading constraints, it is generally not possible to solve the problem for $\{\pi_{0,t}, \pi_t, c_t\}_{t\in[0,T]}$ in closed-form on the basis of its primal specification in (2.2.7). Therefore, we resort to convex duality techniques instead. This allows us to derive closed-form expressions for the optimal controls. We divide this procedure into three distinct steps. First, we derive the dual problem that is associated with the primal in (2.2.7). Second, we demonstrate that this dual problem implies an unconstrained financial market model that embeds a corresponding auxiliary investment problem. Third, we derive the optimal consumption and trading strategies in this unconstrained market, and we show the solution's connection with the primal and dual problems. To make this procedure intuitive, we provide in Figure 2.1 a visual illustration of its underlying mechanism.

2.3.1 Dual Problem Specification

In this section, we rely on convex duality principles, developed by Karatzas et al. (1991a), Cvitanić and Karatzas (1992), and Xu and Shreve (1992), to derive the dual specification of the constrained investment problem (2.2.7). In terms of Figure 2.1, we subsequently derive the technical details corresponding to the depicted upper half. Note here that the lower half concerns the primal problem, which has been introduced in section 2.2.2. Concretely, we give a heuristic derivation of the dual, by using the method of Klein and Rogers (2007) and Rogers (2003, 2013). This approach applies Lagrangian concepts to dynamic optimisation problems of the kind at hand in three steps. Subsequently, we elaborate in a general sense on these three steps.

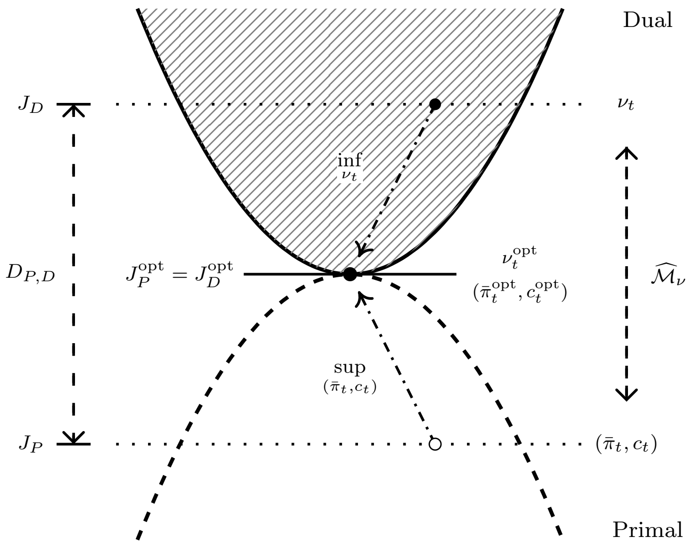

Figure 2.1. Duality mechanism. This figure illustrates the mechanism that underpins convex duality. The upper part represents the dual side of the problem; the lower part the primal side. The upper curve corresponds to the set of feasible dual controls; the lower curve to the set of admissible primal controls. The horizontal lines stress that each set of primal controls ($\bar{\pi}_t = (\pi_{0,t}, \pi_t), c_t$) and each dual control (ν_t) render a primal and dual value function (J_D and J_P), respectively. The vertical arrow on the left stands for the resulting duality gap ($D_{P,D} = J_D - J_P$). The remaining arrows relate to the mechanism itself. Namely, the vertical one on the right demonstrates that the dual controls and primal controls interact with each other via the artificial market ($\widehat{\mathcal{M}}_\nu$). The mechanism tries to select a "point" ν_t that minimises the dual ($\inf_{\nu_t}$), as indicated by the arrow in the upper curve. Via $\widehat{\mathcal{M}}_\nu$, the dual-optimal ν_t^{opt} implies a "point" $(\bar{\pi}_t, c_t)$ that maximises the primal ($\sup_{(\bar{\pi}_t, c_t)}$), as indicated by the arrow in the lower curve. That is, the mechanism attempts to steer the two primal and dual "points", through $\mathcal{M}_\nu$, in the direction of the curves' tangent point, which represents the point of optimality ($(\bar{\pi}_t^{\text{opt}}, c_t^{\text{opt}}), \nu_t^{\text{opt}}$), at which strong duality ($J_P^{\text{opt}} = J_D^{\text{opt}}$, $D_{P,D} = 0$) holds.

First, it introduces a Lagrange multiplier process that enforces the equality upon the dynamic budget constraint in (2.2.7). This Lagrange multiplier process evolves according to: $\mathrm{d}Z_t = Z_t \left[\alpha_t \mathrm{d}t + \theta_t^\top \mathrm{d}W_t\right]$, for some $Z_0 \in \mathbb{R}_+$, where we assume that $\alpha_t \in \mathbb{D}^{1,2}([0,T])$ and $\theta_t \in \mathbb{D}^{1,2}([0,T])^N$ hold. Note that the solution to this SDE is similar to the expression for M_t in (2.2.3). In particular, $Z_t = \exp\{\int_0^t \alpha_s \mathrm{d}s - \frac{1}{2}\int_0^t \|\theta_s\|^2_{\mathbb{R}^N} \mathrm{d}s - \int_0^t \theta_s^\top \mathrm{d}W_s\}$, for all $t \in [0,T]$. Second, it derives corresponding complementary slackness (CS) conditions. These conditions serve to ensure finiteness of the "Lagrangian" corresponding to (2.2.7). For problem (2.2.7), the CS conditions imposed on the drift and diffusion terms in $\{Z_t\}_{t\in[0,T]}$ read $\alpha_t = -\nu_{0,t} - r_t$, $\theta_t = -\lambda_t - \sigma_t^{-1}(\nu_{N,t} - \nu_{0,t}1_N)$ for $\nu_t = (\nu_{0,t}, \nu_{N,t}) \in \mathbb{D}^{1,2}([0,T]) \times \mathbb{D}^{1,2}([0,T])^N$. As a result, Z_t lives by a structure resembling the configuration of an SPD process. The analogous CS conditions imposed on c_t and X_T are:

$$c_t^{\text{opt}} = \iota(t, Z_t, \Pi_{1,t}), \quad \text{and} \quad X_T^{\text{opt}} = I(Z_T, \Pi_{2,T}). \tag{2.3.1}$$

The final CS condition, which is imposed on Z_0, is given by $Z_0^{\text{opt}} = \mathcal{G}^{-1}(X_0)$, where $\mathcal{G}^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ represents the inverse function of $\mathcal{G} : \mathbb{R}_+ \to \mathbb{R}_+$, which is defined in the following manner $\mathcal{G}(Z_0) = \mathbb{E}[X_T^{\text{opt}} Z_0^{-1} Z_T + \int_0^T [c_t^{\text{opt}} - Y_t - \delta(\nu_t)] Z_0^{-1} Z_t \mathrm{d}t]$. Here, $\delta : \mathbb{R}^{N+1} \to \mathbb{R} \cup \{\infty\}$ denotes the support function, $\delta(x) = \sup_{(\pi_{0,t},\pi_t)\in K}(-x_0\pi_{0,t} - x_N^\top \pi_t)$ for all $x = (x_0, x_N) \in \mathbb{R} \times \mathbb{R}^N$. Third, and last, it spells out the (candidate) dual problem itself, see Theorem 2.3.1.

Theorem 2.3.1. *Consider the investment problem in* (2.2.5). *Let* $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ *be the set that contains all* $\{\nu_t\}_{t\in[0,T]}$ *such that* $\nu_t \in \mathbb{D}^{1,2}([0,T])^{N+1}$ *and* $\|\delta(\nu_t)\|^2_{L^2(\Omega\times[0,T])} < \infty$ *are true, and fix* $\widehat{X}_{\nu,0} = X_0 + \mathbb{E}[\int_0^T Z_0^{-1} Z_t [Y_t + \delta(\nu_t)] \mathrm{d}t]$ *for* $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ *and all* $\bar{X}_0 \in \mathbb{R}^2_+$. *Then the dual problem of* (2.2.5) *is given by:*

$$\begin{aligned} J_D^{\text{opt}}(\bar{X}_0) = \inf_{\{\nu_t\}_{t\in[0,T]}\in\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}, Z_0\in\mathbb{R}_+} \mathbb{E}\Bigg[\int_0^T & v(t, Z_t, \Pi_{1,t}) \mathrm{d}t \\ & + V(Z_T, \Pi_{2,T})\Bigg] + \widehat{X}_{\nu,0} Z_0, \end{aligned} \tag{2.3.2}$$

for all $\bar{X}_0 \in \mathbb{R}^2_+$. *The dual value function is therefore characterised as follows:* $J_D(\bar{X}_0, Z_0, \{\nu_t\}_{t\in[0,T]}) = \mathbb{E}[\int_0^T v(t, Z_t, \Pi_{1,t})\mathrm{d}t + V(Z_T, \Pi_{2,T})] + \widehat{X}_{\nu,0} Z_0$. *In addition to this, the following holds holds for all* $\bar{X}_0 \in \mathbb{R}^2_+$*:*

$$J_P^{\text{opt}}(\bar{X}_0) = J_D^{\text{opt}}(\bar{X}_0). \tag{2.3.3}$$

Proof. The proof is given in Appendix A.1. □

Remark 2.3.1. *Suppose that* $C = 0$ *in the admissibility set,* $\widehat{\mathcal{A}}_{X_0}$, *and assume that* $U \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ *holds, i.e. that its first argument is specified over* $\mathbb{R}_+$

rather than over $\mathbb{R}$. In particular, assume that U fulfills the same requirements as u (for all $t \in [0,T]$). Then, the results in Theorem 2.3.1 are still applicable. We emphasise this to be able to show at a later stage that our approximate method also holds for $U \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ and $C = 0$. Moreover, we remark that the specifications of $\nu_t \in \mathbb{D}^{1,2}([0,T])^{N+1}$ and $\delta : \mathbb{R}^{N+1} \to \mathbb{R} \cup \{\infty\}$ differ from those in ordinary setups, e.g. Cvitanić, 1997, Lim and Choi (2009), and Yener (2015). In order to elicit the similarities between these, suppose that $K = \mathbb{R} \times \mathbb{R}^M \times \{0\}^{N-M}$, then $\nu_{0,t} = 0$ and $\nu_{N,t} \in \{0\}^M \times \mathbb{R}^{N-M}$, such that $\frac{Z_t}{Z_0} = B_t^{-1} \frac{\mathrm{d}\mathbb{Q}_\nu}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$ for $\mathbb{Q}_\nu \sim \mathbb{P}$ that only depends on $\nu_{N,t} \in \mathbb{D}^{1,2}([0,T])^N$. Consider Cvitanić and Karatzas (1993), Karatzas and Kou (1996), and Cvitanic (2000) for more examples.

With the exception of the specification of the dual controls, $\nu_t \in \mathbb{D}^{1,2}([0,T])^{N+1}$, and the corresponding support function, $\delta : \mathbb{R}^{N+1} \to \mathbb{R} \cup \{\infty\}$, the dual (2.3.2) coincides with the conventional ones. For these conventional specifications, see e.g. Cvitanić and Karatzas (1992), Cuoco (1997), and Cuoco and Cvitanić (1998). The differences are due to the fact that U's domain is specified over $\mathbb{R} \times \mathbb{R}$ rather than over $\mathbb{R}_+ \times \mathbb{R}$. Notwithstanding these differences, the results in Theorem 2.3.1 are also applicable to utility functions of which the domain lives by $\mathbb{R}_+ \times \mathbb{R}$, provided that $C = 0$ (see Remark 2.3.1 above). In the sequel, if the $C = 0$ and $U \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ conditions imply different results, we will explicitly indicate so. Ultimately, we note that $\{Z_t\}_{t\in[0,T]}$ implies a unique probability measure, for each choice of $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. This identification of Z_t with such a measure enables us to determine an auxiliary unconstrained environment.

2.3.2 Artificial Financial Market

In this section, we specify the auxiliary, artificial, market corresponding to the dual problem in Theorem 2.3.1. Thereby, we establish the fundamental link between between the upper half and lower half of Figure 2.1, i.e. the primal and dual problems, which is represented by the arrow on the right. This auxiliary market, resulting after a "fictitious completion of assets", say $\widehat{\mathcal{M}}_\nu$, involves an investment problem similar to (2.2.5). This auxiliary investment problem excludes trading constraints. Due to the absence of trading constraints in this environment, the recovery of closed-form expressions for the (actual) optimal trading policies is easier. The artificial market, $\widehat{\mathcal{M}}_\nu$, relies on the same probability space, $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, and the same processes as in $\mathcal{M}$ are active, up to the assets. Instead of B_t and S_t, it includes $B_{\nu,t}$ and $S_{\nu,t}$, for $B_{\nu,0} = S_{i,\nu,0} = 1$:

$$\frac{\mathrm{d}B_{\nu,t}}{B_{\nu,t}} = [r_t + \nu_{0,t}]\,\mathrm{d}t, \quad \text{and} \quad \frac{\mathrm{d}S_{i,\nu,t}}{S_{i,\nu,t}} = [\mu_{i,t} + \nu_{i,N,t}]\,\mathrm{d}t + \sigma_{i,t}^\top \mathrm{d}W_t. \tag{2.3.4}$$

As in (2.2.2), we let μ_t and $\nu_{N,t}$ be the $\mathbb{R}^N$-valued vectors with, elements $\mu_{i,t}$ and $\nu_{i,N,t}$, respectively, for $i = 1, \dots, N$. Likewise, σ_t is defined to be the $\mathbb{R}^{N\times N}$-valued matrix with rows $\sigma_{i,t}$ for $i = 1, \dots, N$. Additionally, we invoke the same postulates on the included local drift terms and diffusion coefficients, as on their analogues in (2.2.1) and (2.2.2). In line with Theorem 2.3.1, we assume that $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ is satisfied in $\widehat{\mathcal{M}}_\nu$, for $\nu_t = [\nu_{0,t}, \nu_{N,t}]^\top$. As a consequence of the dependence of $\{B_{\nu,t}\}_{t\in[0,T]}$ and $\{S_{\nu,t}\}_{t\in[0,T]}$ on $\{\nu_t\}_{t\in[0,T]}$, there exists a unique fictitious completion $\widehat{\mathcal{M}}_\nu$, for each $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. We observe that the definitions for the artificial assets, $B_{\nu,t}$ and $S_{\nu,t}$, can be derived on the basis of the one for the dual process, Z_t.

By Delbaen and Schachermayer (1994), we know that the SPD in $\widehat{\mathcal{M}}_\nu$ is:

$$\frac{\mathrm{d}Z_{\nu,t}}{Z_{\nu,t}} = [-r_t - \nu_{0,t}]\,\mathrm{d}t - \left[\lambda_t + \sigma_t^{-1}\left(\nu_{N,t} - \nu_{0,t}1_N\right)\right]^\top \mathrm{d}W_t, \tag{2.3.5}$$

with $Z_{\nu,0} = 1$, as we essentially work with $\left\{Z_0^{-1} Z_t\right\}_{t\in[0,T]}$ rather than with $\{Z_t\}_{t\in[0,T]}$. The strictly positive starting value $Z_0 \in \mathbb{R}_+$ appears in the auxiliary investment problem. We note here that $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ ought to hold, as a consequence of which all assumptions that are active for $\{M_t\}_{t\in[0,T]}$, are likewise active for $\{Z_{\nu,t}\}_{t\in[0,T]}$. The process $\{Z_{\nu,t}\}_{t\in[0,T]}$ therefore establishes a well-defined SPD process. Consistent with Remark 2.3.1, we note that $Z_{\nu,t} = B_{\nu,t}^{-1} \frac{\mathrm{d}\mathbb{Q}_\nu}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$, for the $\widehat{\mathcal{M}}_\nu$-valid pricing measure, $\mathbb{Q}_\nu \sim \mathbb{P}$, and corresponding Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{Q}_\nu}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = \mathbb{E}[\frac{\mathrm{d}\mathbb{Q}_\nu}{\mathrm{d}\mathbb{P}} \mid \mathcal{F}_t]$. Under the pricing measure $\mathbb{Q}_\nu$, the process $\{W_t^{\mathbb{Q}_\nu}\}_{t\in[0,T]}$ with SDE $\mathrm{d}W_t^{\mathbb{Q}_\nu} = \mathrm{d}W_t + \left[\lambda_t + \sigma_t^{-1}\left(\nu_{N,t} - \nu_{0,t}1_N\right)\right]\mathrm{d}t$ is a standard $\mathbb{Q}_\nu$-Brownian motion.

The dynamic wealth process of the finite-horizon investor in $\widehat{\mathcal{M}}_\nu$ is given by:

$$\begin{aligned} \mathrm{d}X_{\nu,t} &= \pi_{0,\nu,t}B_{\nu,t}^{-1}\mathrm{d}B_{\nu,t} + \pi_{\nu,t}^\top \mathrm{diag}\left(S_{\nu,t}\right)^{-1}\mathrm{d}S_{\nu,t} - \left(c_{\nu,t} - Y_{\nu,t}\right)\mathrm{d}t \\ &= \pi_{0,\nu,t}\left[r_t + \nu_{0,t}\right]\mathrm{d}t + \pi_{\nu,t}^\top\left(\left[\mu_t + \nu_{N,t}\right]\mathrm{d}t + \sigma_t\mathrm{d}W_t\right) - \left(c_{\nu,t} - Y_{\nu,t}\right)\mathrm{d}t, \end{aligned} \tag{2.3.6}$$

with $X_{\nu,0} = X_0$. We redefine the labour income stream from $\mathcal{M}$, i.e. $\{Y_t\}_{t\in[0,T]}$, in $\widehat{\mathcal{M}}_\nu$ as: $Y_{\nu,t} = Y_t + \delta\left(\nu_t\right)$ for all $t \in [0,T]$. Now, $\{\pi_{\nu,0,t}, \pi_{\nu,t}\}_{t\in[0,T]}$ represents the $\mathbb{R}^{N+1}$-valued and $\mathcal{F}_t$-progressively measurable portfolio process. To be more precise, $\pi_{\nu,0,t}$ is allocated to $B_{\nu,t}$ and $\pi_{\nu,t}$ to $S_{\nu,t}$ over $t \in [0,T]$. Similar to the specification in $\mathcal{M}$, we define $\{c_{\nu,t}\}_{t\in[0,T]}$ to be the $\mathbb{R}$-valued and $\mathcal{F}_t$-progressively measurable consumption process in $\widehat{\mathcal{M}}_\nu$. The admissiblity set in $\widehat{\mathcal{M}}_\nu$, denoted by $\widehat{\mathcal{A}}_{\nu,X_0}$, contains all artificial trading-consumption pairs $\{\pi_{\nu,0,t}, \pi_{\nu,t}, c_{\nu,t}\}_{t\in[0,T]}$ that satisfy $\int_0^T \pi_{\nu,t}^\top\sigma_t\sigma_t^\top\pi_{\nu,t}\mathrm{d}t < \infty$, $\int_0^T \left|\pi_{\nu,t}^\top\left[\mu_t + \nu_{N,t}\right] + \pi_{\nu,0,t}r_t\right|\mathrm{d}t < \infty$, $\mathbb{E}[\int_0^T |c_t|^2\mathrm{d}t] < \infty$, as well as $X_{\nu,t} = \pi_{\nu,0,t} + \pi_{\nu,t}^\top 1_N \geq -C$, for some $C \in \mathbb{R}_+$ and for all $t \in [0,T]$.

Proposition 2.3.2. *Consider the dual problem in* (2.3.2) *of Theorem 2.3.1. The unconstrained dynamic asset allocation problem in* $\widehat{\mathcal{M}}_\nu$ *is given by*

$$\begin{aligned} \sup_{\{\bar{\pi}_{\nu,t}, c_{\nu,t}\}_{t\in[0,T]}\in\widehat{\mathcal{A}}_{\nu,X_0}} \quad & \mathbb{E}\left[\int_0^T u\left(t, c_{\nu,t}, \Pi_{1,t}\right)\mathrm{d}t + U\left(X_{\nu,T}, \Pi_{2,T}\right)\right] \\ \text{s.t.} \quad & \mathrm{d}X_{\nu,t} = \pi_{0,\nu,t}r_{\nu,t}\mathrm{d}t + \pi_{\nu,t}^\top\left(\mu_{\nu,t}\mathrm{d}t + \sigma_t\mathrm{d}W_t\right) \\ & \qquad - \left(c_{\nu,t} - Y_{\nu,t}\right)\mathrm{d}t, \end{aligned} \tag{2.3.7}$$

for $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$*, in which we let* $\bar{\pi}_{\nu,t} = (\pi_{\nu,0,t}, \pi_{\nu,t})$*,* $r_{\nu,t} = r_t + \nu_{0,t}$ *and* $\mu_{\nu,t} = \mu_t + \nu_{N,t}$ *for notational simplicity. Now, suppose that* $J_{\widehat{\mathcal{M}}_\nu}(\bar{X}_0, \{\nu_t\}_{t\in[0,T]})$ *specifies the optimal value function for* (2.3.7)*. Then, the following equality is true for all* $\bar{X}_0 \in \mathbb{R}_+^2$*:*

$$J_D^{\mathrm{opt}}\left(\bar{X}_0\right) = \inf_{\{\nu_t\}_{t\in[0,T]}\in\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}} J_{\widehat{\mathcal{M}}_\nu}\left(\bar{X}_0, \{\nu_t\}_{t\in[0,T]}\right). \tag{2.3.8}$$

By strong duality, we note that for each choice of $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$*, it holds that* $J_P^{\mathrm{opt}}\left(\bar{X}_0\right) \leq J_{\widehat{\mathcal{M}}_\nu}(\bar{X}_0, \{\nu_t\}_{t\in[0,T]})$*: in the auxiliary market,* $\widehat{\mathcal{M}}_\nu$*, the investor always derives more utility.*

Proof. The proof is given in Appendix A.2. $\square$

Remark 2.3.2. *The artificial market, $\widehat{\mathcal{M}}_\nu$, is uniquely identified by its probability measure, $\mathbb{Q}_\nu \sim \mathbb{P}$, and its perturbation on the interest rate, $\nu_{0,t}$. To observe that the dual is identical to determining the least-favourable artificial market, we rewrite* (2.3.8) *as: $J_D^{\text{opt}}(\bar{X}_0) = \inf_{(\mathbb{Q}_\nu, \nu_{0,t}) \in \mathcal{Q}} J_{\widehat{\mathcal{M}}_\nu}(\bar{X}_0, \mathbb{Q}_\nu, \nu_{0,t})$. Here, $\mathcal{Q}$ contains all measures $\mathbb{Q}_\nu \sim \mathbb{P}$ and local drift terms $\{\nu_{0,t}\}_{t\in[0,T]}$, corresponding to the auxiliary SPD process $\{Z_t\}_{t\in[0,T]}$, that harmonise with all $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. Furthermore, $J_{\widehat{\mathcal{M}}_\nu}(\bar{X}_0, \mathbb{Q}_\nu, \nu_{0,t})$ spells out the optimal value function of* (2.3.7)*, for all $(\mathbb{Q}_\nu, \nu_{0,t}) \in \mathcal{Q}$. This formulation of the dual demonstrates that one must determine the worst-case artificial environment to ensure that the optimal solutions to* (2.3.7) *are admissible and optimal in $\mathcal{M}$.*

The main takeaway from Proposition 2.3.2 is that we may resort to solving an auxiliary problem by means of a fictitious completion of assets, instead of solving the primal problem (2.2.7). This auxiliary problem (2.3.7) excludes trading constraints, which enables us to determine the optimal artificial allocation to assets in closed-form. In line with the objective of the dual in Theorem 2.3.1, we then only ought to determine the shadow price process in a least-favourable fashion to establish optimality as well as admissibility of the preceding controls in the primal environment, cf. (2.3.8). In particular, optimality to the constrained problem is achieved via the auxiliary problem, due to strong duality. The optimisation of the dual namely ensures that the auxiliary-optimal controls are projected into the admissibility region.

2.3.3 Auxiliary Optimality Conditions

We conclude this section on convex duality by deriving and analysing the analytical solutions to the auxiliary unconstrained investment problem in Proposition 2.3.2. Pertaining to Figure 2.1, this section serves to analytically unravel the intimate link between the optimal primal controls, $(\bar{\pi}_t^{\text{opt}}, c_t^{\text{opt}})$, and dual controls, ν_t^{opt}. In particular, it shows how the duality mechanism "steers" ν_t and $(\bar{\pi}_t, c_t)$ via $\widehat{\mathcal{M}}_\nu$ towards the optimum, as indicated by the arrows in both curves. After the derivation of the closed-form solutions, we illustrate the analytical difficulties that may arise when determining the optimal shadow price process, i.e. the least-favourable fictitious completion. For this purpose, we consider a concrete specification of the set of trading constraints, K, cf. Example 2.3.1. This illustration clarifies the potential benefits of an approximating procedure that circumvents these analytical obstacles, and retains the closed-form nature of the optimal solutions.

Examining (2.3.7), we can apply the martingale method, cf. Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991), which enables us to rewrite (2.3.7) as

$$\begin{aligned} \sup_{(X_{\nu,T},c_{\nu,t})\in\widehat{L}^2(\Omega\times[0,T])} \quad & \mathbb{E}\left[\int_0^T u\left(t,c_{\nu,t},\Pi_{1,t}\right)\mathrm{d}t + U\left(X_{\nu,T},\Pi_{2,T}\right)\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\int_0^T \left(c_{\nu,t}-Y_{\nu,t}\right)Z_{\nu,t}\mathrm{d}t + X_{\nu,T}Z_{\nu,T}\right] \leq X_0, \end{aligned} \tag{2.3.9}$$

with $\widehat{L}^2\left(\Omega\times[0,T]\right) := L^2\left(\Omega\right)\times L^2\left(\Omega\times[0,T]\right)$. Concretely, we rewrite the dynamic specification (2.3.7) into its static variational analogue (2.3.9). Rather than selecting a continuous trading strategy and consumption pattern, the investor in this static problem firstly selects terminal wealth in combination with a consumption strategy. The ensuing optimal rules coincide with X_T^{opt} and c_t^{opt} from Theorem 2.3.1. Secondly, the optimal trading strategy, $\{\pi_{\nu,t}^{\text{opt}}\}_{t\in[0,T]}$, can be determined in analytical form by means of standard hedging (martingale) arguments. Proposition 2.3.3 specifies the closed-form optimal dynamic allocation of assets and consumption strategy.

Proposition 2.3.3. *Consider the unconstrained investment problem* (2.3.9)*, as well as its dynamic counterpart* (2.3.7)*. Optimal terminal wealth and optimal consumption are specified as follows:*

$$\begin{aligned} c_{\nu,t}^{\text{opt}} &= \iota\left(t,\eta^{\text{opt}}Z_{\nu,t},\Pi_{1,t}\right), \\ X_{\nu,T}^{\text{opt}} &= I\left(\eta^{\text{opt}}Z_{\nu,T},\Pi_{2,T}\right), \end{aligned} \tag{2.3.10}$$

for all $t\in[0,T]$*, with* $\eta^{\text{opt}} = \mathcal{G}^{-1}\left(X_0\right)$*. The optimal wealth over the trading interval is* $X_{\nu,t}^{\text{opt}} = \mathbb{E}[X_{\nu,T}^{\text{opt}}\frac{Z_{\nu,T}}{Z_{\nu,t}} + \int_t^T \left(c_{\nu,s}^{\text{opt}}-Y_{\nu,s}\right)\frac{Z_{\nu,s}}{Z_{\nu,t}}\mathrm{d}s \mid \mathcal{F}_t]$*, for all* $t\in[0,T]$*.*

Define:

$$\mathcal{R}_{1,c,t}=-\frac{u''_X\left(t,c^{\mathrm{opt}}_{\nu,t},\Pi_{1,t}\right)}{u'_X\left(t,c^{\mathrm{opt}}_{\nu,t},\Pi_{1,t}\right)},\quad \mathcal{R}_{1,X,T}=-\frac{U''_X\left(X^{\mathrm{opt}}_{\nu,T},\Pi_{2,T}\right)}{U'_X\left(X^{\mathrm{opt}}_{\nu,T},\Pi_{2,T}\right)},$$
$$\mathcal{R}_{2,c,t}=-\iota'_Y\left(t,\eta^{\mathrm{opt}}Z_{\nu,t},\Pi_{1,t}\right),\quad \text{and}\quad \mathcal{R}_{2,X,T}=-I'_Y\left(\eta^{\mathrm{opt}}Z_{\nu,T},\Pi_{2,T}\right), \tag{2.3.11}$$

where $\iota'_Y:[0,T]\times\mathbb{R}_+\times\mathbb{R}\to\mathbb{R}$ *is the derivative of* ι *in the* y*-direction; and* $I'_Y:\mathbb{R}_+\times\mathbb{R}\to\mathbb{R}$ *similarly represents the derivative of* I *in its second argument. The terms in* (2.3.11) *identify (proxies to) the absolute risk-aversion (ARA) coefficients. Then, we can express the optimal allocation to assets,* $\{\pi^{\mathrm{opt}}_{\nu,t}\}_{t\in[0,T]}$*, in terms of the following decomposition:* $\pi^{\mathrm{opt}}_{\nu,t}=\pi^m_t+\pi^Z_t+\pi^\Pi_t+\pi^Y_t$*. The first two weights in this decomposition read*

$$\begin{aligned}\pi^m_t&=\mathbb{E}\left[\frac{1}{\mathcal{R}_{1,X,T}}\frac{Z_{\nu,T}}{Z_{\nu,t}}+\int_t^T\frac{1}{\mathcal{R}_{1,c,s}}\frac{Z_{\nu,s}}{Z_{\nu,t}}\mathrm{d}s\,\middle|\,\mathcal{F}_t\right]\sigma_t^{\top^{-1}}\widehat{\lambda}_t\\ \pi^Z_t&=\sigma_t^{\top^{-1}}\mathbb{E}\left[\frac{1}{\widehat{\mathcal{R}}_{1,X,T}}\frac{Z_{\nu,T}}{Z_{\nu,t}}D_{Z,t,T}+\int_t^T\frac{1}{\widehat{\mathcal{R}}_{1,c,s}}\frac{Z_{\nu,s}}{Z_{\nu,t}}D_{Z,t,s}\mathrm{d}s\,\middle|\,\mathcal{F}_t\right],\end{aligned} \tag{2.3.12}$$

for all $t\in[0,T]$*, where* $\widehat{\lambda}_t:=\lambda_t+\sigma_t^{-1}\left(\nu_{N,t}-\nu_{0,t}1_N\right)$*,* $\widehat{\mathcal{R}}^{-1}_{1,X,T}:=X^{\mathrm{opt}}_{\nu,T}-\mathcal{R}^{-1}_{1,X,T}$*,* $\widehat{\mathcal{R}}^{-1}_{1,c,t}:=c^{\mathrm{opt}}_{\nu,t}-Y_{\nu,t}-\mathcal{R}^{-1}_{1,c,t}$*, and* $D_{Z,t,s}=\mathcal{D}^W_t\log\widehat{Z}_{\nu,s}=\mathcal{D}^W_t\log Z_{\nu,s}+\widehat{\lambda}_t$ *for all* $s\geq t$*, such that* $s,t\in[0,T]$*. Additionally,* $\mathcal{D}^W_t:\mathbb{D}^{1,2}([0,T])\to L^2(\Omega\times[0,T])^N$ *defines the Malliavin derivative kernel.*[9] *The remaining two weights in the optimal portfolio process read as follows*

$$\begin{aligned}\pi^Y_t&=-\sigma_t^{\top^{-1}}\mathbb{E}\left[\int_t^T\frac{Z_{\nu,s}}{Z_{\nu,t}}Y_s\left(\mathcal{D}^W_t\log Y_s+\widehat{\sigma}_{Y,t,s}\right)\mathrm{d}s\,\middle|\,\mathcal{F}_t\right],\\ \pi^\Pi_t&=-\sigma_t^{\top^{-1}}\mathbb{E}\left[\frac{Z_{\nu,T}}{Z_{\nu,t}}\mathcal{R}_{2,X,T}\mathcal{D}^W_t\Pi_{2,T}+\int_t^T\frac{Z_{\nu,s}}{Z_{\nu,t}}\mathcal{R}_{2,c,s}\mathcal{D}^W_t\Pi_{1,s}\mathrm{d}s\,\middle|\,\mathcal{F}_t\right],\end{aligned} \tag{2.3.13}$$

for all $t\in[0,T]$*, wherein we define* $\widehat{\sigma}_{Y,t,s}=\sigma_{Y,t}+Y_s^{-1}\mathcal{D}^W_t\delta\left(\nu_s\right)$ *for all*

[9] We characterise the derivative kernel here as a mapping from a uni-dimensional space to an N-dimensional analogue. If we, instead, accompany the aforementioned kernel by an N-dimensional argument, i.e. one that attains values in the $\mathbb{D}^{1,2}([0,T])^N$ space, the relevant mapping alters into $\mathcal{D}^W_t:\mathbb{D}^{1,2}([0,T])^N\to L^2(\Omega\times[0,T])^{N\times N}$. Consider e.g. Nualart (2006) for more technical details underscoring this derivative kernel.

$s \geq t$, such that $s, t \in [0,T]$. Here, $\{X^{\mathrm{opt}}_{\nu,t} - 1_N^\top \pi^{\mathrm{opt}}_{\nu,t}\}_{t\in[0,T]}$ specifies the optimal allocation to the cash account, $\{\pi^{\mathrm{opt}}_{0,\nu,t}\}_{t\in[0,T]}$. Note that the Malliavin derivatives included in (2.3.12) *and* (2.3.13)*, are all well-posed.*[10]

Proof. The proof is given in Appendix A.3. □

For economic intuition corresponding to the solutions in Proposition 2.3.3, see Karatzas and Shreve (1998). For theoretical and economic analyses of the disentangled hedge demands, we refer to Detemple, 2014, and Li et al. (2020), who examine the mathematical decomposition of optimal portfolios in detail. As for the duality principles, we note that Proposition 2.3.3 spells out the optimal controls in $\widehat{\mathcal{M}}_\nu$ in closed-form, for a yet unspecified shadow price process, $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. To make these controls likewise optimal and admissible in the true market, $\mathcal{M}$, we must determine $\{\nu_t\}_{t\in[0,T]}$ as set out by Theorem 2.3.1 and Proposition 2.3.2. That is, we must determine the least-favourable artificial market. The following example illustrates that acquiring the optimal $\{\nu_t\}_{t\in[0,T]}$ in closed-form is generally troublesome.

Example 2.3.1. *(Non-traded Assets) Consider $\mathcal{M}$, and set $\sigma_t = [\bar{\sigma}_{1,t}, \bar{\sigma}_{2,t}]$ for $(\bar{\sigma}_{1,t}, \bar{\sigma}_{2,t}) \in \mathbb{R}^{N\times(N-M)} \times \mathbb{R}^{N\times M}$, wherein $\bar{\sigma}_{1,t} := \left[\sigma_{1,t}, 0_{M\times(N-M)}\right]^\top$ and $\bar{\sigma}_{2,t} := \left[0_{(N-M)\times M}, \sigma_{2,t}\right]^\top$ for $(\sigma_{1,t}, \sigma_{2,t}) \in \mathbb{R}^{(N-M)\times(N-M)} \times \mathbb{R}^{M\times M}$. Consequently, the first $N-M$ elements of S_t are driven exclusively by $W_{N-M,t}$; similarly, the last M elements of S_t are driven solely by $W_{M,t}$; here, $W_t = [W_{N-M,t}, W_{M,t}]^\top \in \mathbb{R}^{N-M} \times \mathbb{R}^M$. Furthermore, fix $K = \mathbb{R} \times \mathbb{R}^{N-M} \times \{0\}^M$: the last M stocks in S_t are non-traded. From K's definition, $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ implies that $\nu_{0,t} = 0$ and $\nu_{N,t} \in \{0\}^{N-M} \times \mathbb{R}^M$. The dual hence reduces to $\inf_{\nu_{N,M,t}\in\mathbb{D}^{1,2}([0,T])^N} J_D(\bar{X}_0, Z_0^{\mathrm{opt}}, \{\nu_{N,M,t}\}_{t\in[0,T]})$, for $Z_0^{\mathrm{opt}} = \mathcal{G}^{-1}(X_0)$ and $\nu_{N,N-M,t} = 0_{N-M}$ in $\nu_{N,t} = [\nu_{N,N-M,t}, \nu_{N,M,t}]^\top$.*

[10] For notational elegance and general clarity, we have chosen to not spell out the Malliavin derivatives at hand. Conversely, we provide the extended versions here: in (2.3.12), $\mathcal{D}_t^W \log \widehat{Z}_{\nu,s} = \int_t^s \left[-\mathcal{D}_t^W r_u - \mathcal{D}_t^W \nu_{0,u}\right] \mathrm{d}u - \int_t^s \left(\mathcal{D}_t^W \widehat{\lambda}_u\right)\widehat{\lambda}_u \mathrm{d}u - \int_t^s \mathcal{D}_t^W \widehat{\lambda}_u \mathrm{d}W_u$, and in (2.3.13), $\mathcal{D}_t^W \log Y_s = \int_t^s \left(\mathcal{D}_t^W \mu_{Y,u}\mathrm{d}s + \mathcal{D}_t^W \sigma_{Y,u}\mathrm{d}W_u^Y\right)$, with $W_t^Y = W_t - \int_0^t \sigma_{Y,s}\mathrm{d}s$, and $\mathcal{D}_t^W \Pi_{i,s} = \int_t^s \mathcal{D}_t^W \mu_{\Pi,i,u}\mathrm{d}u + \int_t^s \mathcal{D}_t^W \sigma_{\Pi,i,u}\mathrm{d}W_u + \sigma_{\Pi,i,t}$, $i = 1,2$, ought to hold true for all $s \geq t$ such that $s,t \in [0,T]$.

In Appendix A.4, we demonstrate that optimisation of this dual results in the following optimality condition:

$$\widehat{\lambda}_{M,t} X^{\text{opt}}_{\nu,t} = -\mathbb{E}\left[\frac{\mathcal{D}^{W_M}_t X^{\text{opt}}_{\nu,T} Z_{\nu,T}}{Z_{\nu,t}} + \int_t^T \frac{\mathcal{D}^{W_M}_t c_{\nu,Y,s} Z_{\nu,s}}{Z_{\nu,t}} \mathrm{d}s \,\middle|\, \mathcal{F}_t \right], \qquad (2.3.14)$$

for all $t \in [0, T]$*, wherein* $\widehat{\lambda}_{M,t} = \lambda_{M,t} + \sigma^{-1}_{2,t} \nu_{N,M,t}$ *for* $\lambda_t = [\lambda_{N-M,t}, \lambda_{M,t}]^\top \in \mathbb{R}^{N-M} \times \mathbb{R}^M$ *such that* $\widehat{\lambda}_t = [\widehat{\lambda}_{N-M,t}, \widehat{\lambda}_{M,t}]^\top \in \mathbb{R}^{N-M} \times \mathbb{R}^M$ *and* $\widehat{\lambda}_{N-M,t} = \lambda_{N-M,t}$*. Concretely, equation* (2.3.14) *spells out the identity from which one is able to derive the dual-optimal* $\{\widehat{\lambda}_{M,t}\}_{t\in[0,T]}$*. In consideration of the characterisation of* $\{\pi^{\text{opt}}_{\nu,t}\}_{t\in[0,T]}$ *in* (A.3.19)*, we observe that the latter specification of* $\{\widehat{\lambda}_{M,t}\}_{t\in[0,T]}$ *ensures that* $\pi_{\nu,M,t} = 0_M$ *holds for all* $t \in [0, T]$*, where* $\pi^{\text{opt}}_{\nu,t} = [\pi^{\text{opt}}_{\nu,N-M,t}, \pi^{\text{opt}}_{\nu,M,t}]^\top \in \mathbb{R}^{N-M} \times \mathbb{R}^M$*. On the grounds of strong duality and the latter satisfaction of trading constraints, it holds that* $\{\pi^{\text{opt}}_t\}_{t\in[0,T]} = \{\pi^{\text{opt}}_{\nu,t}\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$*. That is,* $\{\pi^{\text{opt}}_{\nu,t}\}_{t\in[0,T]}$ *is admissible and optimal in* $\mathcal{M}$*.*

Example 2.3.1 demonstrates the optimal shadow price process' ability to ensure optimality and admissibility of the artificial-optimal trading rules in the true environment ($\mathcal{M}$). However, when inspecting (2.3.14) more closely, we observe that the RHS of this equation depends in general non-linearly on past and future values of the shadow price process. More specifically, (2.3.14) essentially describes a forward-backward SDE (FBSDE), which obstructs a recovery of $\{\widehat{\lambda}_{M,t}\}_{t\in[0,T]}$ in closed-form. Therefore, in order to derive $\{\pi^{\text{opt}}_t\}_{t\in[0,T]}$ analytically, one still must resort to numerical applications so as to solve the FBSDE. This unavailability of an analytically defined shadow price process is not unique to the example ($K = \mathbb{R} \times \mathbb{R}^{N-M} \times \{0\}^M$). Other definitions of K ultimately require numerical procedures for solving the optimal shadow price. For examples of such numerical applications, we refer the reader to e.g. Haugh and Kogan (2007), and Rogers and Zaczkowski (2015).

2.4 Approximate Method

In this section, we explain our approximate method. The method is based on a generalisation of the twofold procedure as outlined in Bick et al. (2013). First, we approximate analytically the optimal shadow price process by restricting the dual space to a tractable subspace. The ensuing approximate shadow price process implies primal controls that are generally inadmissible, cf. Proposition 2.3.3 and Example 2.3.1. Second, therefore, we project these implied primal controls to the admissibility set in the primal financial environment ($\mathcal{M}$). Consequently, we are able to obtain closed-form approximate expressions for the optimal solutions to the constrained problem (2.2.5). We provide a visual illustration of this twofold approximating mechanism in Figure 2.2. In the following part, we elaborate on the relevant technicalities, and we illustrate the method's functional principle in case of a specific economic environment.

2.4.1 Twofold Approximating Procedure

Propositions 2.3.2 and 2.3.3 show that the difficulty in attaining optimal solutions to (2.2.7) in closed-form originates from the absence of tractable expressions for the optimal shadow price processes. These optimal shadow price processes are typically characterised by FBSDE's that require numerical routines to be solved, cf. Example 2.3.1. In an attempt to circumvent these computational burdens, and to be able to acquire "simple" analytical expressions, it therefore seems sensible to seek approximations to the optimal shadow price processes. This brings us to to the first step of our twofold approximate method, which is depicted by (1) and the upper dotted arrow in Figure 2.2. Our method's first step concretely consists of confining the dual space to a convex subset, where we determine $\{\nu_t^*\}_{t\in[0,T]}$ as:

$$\{\nu_t^*\}_{t\in[0,T]} := \underset{\{\nu_t\}_{t\in[0,T]}\in\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}}{\arg\inf} \; J_D\left(\bar{X}_0, Z_0^*, \{\nu_t\}_{t\in[0,T]}\right), \tag{2.4.1}$$

for any $\bar{X}_0 \in \mathbb{R}_+^2$. Here, $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} \subseteq \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ denotes the convex subset of the dual space, $Z_0^* = \mathcal{G}^{-1}(X_0)$ the approximate Lagrange multiplier, and $\{\nu_t^*\}_{t\in[0,T]}$ the approximate shadow price process.[11] We determine $\{\nu_t^*\}_{t\in[0,T]}$ by minimising the dual over the relevant subset, as shown in (2.4.1). Namely, amongst all

[11]We endow the controls with an asterisk to indicate that these are approximations of the optimal policies.

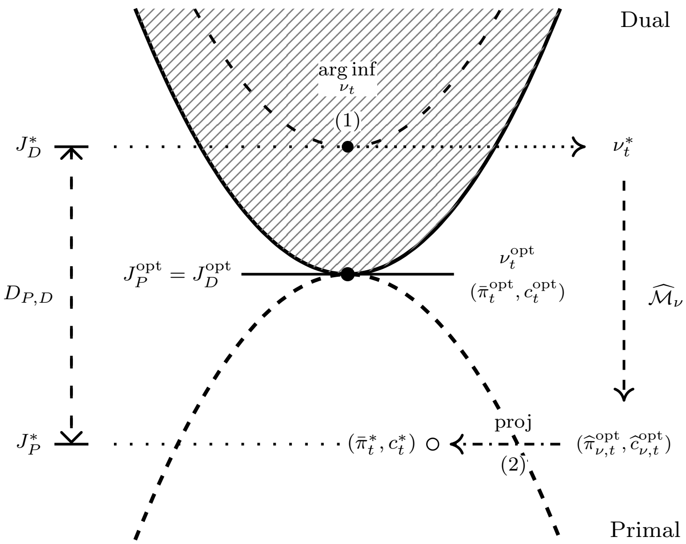

Figure 2.2. Approximate method. This figure illustrates our approximate method. The depicted symbols and shapes represent the same as in Figure 2.1, up to the dashed upper curve and the arrows on the right. Namely, the dashed upper curve corresponds to the tractable subset of feasible dual controls. The arrows relate to the approximate method itself as follows. As indicated by the upper dotted arrow, the first step of our method (1) consists of minimising the dual over all shadow prices in the subset ($\arg\inf_{\nu_t}$). This results in an approximate shadow price (ν_t^*), which is situated in the curve's "minimal" point. The dashed arrow on the right shows that this shadow price implies via the artificial market ($\widehat{\mathcal{M}}_\nu$) an auxiliary pair of primal controls ($\widehat{\pi}_{\nu,t}^{\mathrm{opt}}, \widehat{c}_{\nu,t}^{\mathrm{opt}}$). This pair is typically inadmissible, which is represented by its position outside of the lower curve. To make this pair admissible, the second step of our method (2) projects this pair into the admissibility set (proj), as illustrated by the lower dash-dotted arrow. This results in the approximate primal controls ($\bar{\pi}_t^*, c_t^*$). The horizontal dotted lines stress that ($\bar{\pi}_t^*, c_t^*$) and ν_t^* render a primal and dual value function (J_P^* and J_D^*), respectively. The difference between J_D^* and J_P^* is the duality gap ($D_{P,D} = J_D^* - J_P^*$) and can be used to quantify the approximation's accuracy.

shadow price processes that are contained in $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$, the one that minimises the dual tilts the artificial-optimal controls (in $\widehat{\mathcal{M}}_\nu$) the closest to their actual-optimal analogues (in $\mathcal{M}$). This feature follows from the specific nature of (convex) duality relations. The artificial-optimal controls implied by $\{\nu_t^*\}_{t\in[0,T]}$ consequently provide a proper point of departure in the formulation of approximate primal controls. We note that the tractability of the approximate shadow price processes in (2.4.1) depends on the specific choice for $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$. For instance, $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} = \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}} \cap \mathbb{R}$ guarantees tractability.

This leads us to the second step of our method, which is represented by (2) and the dash-dotted lower arrow in Figure 2.2. Let $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ be equal to $\pi^{\text{opt}}_{\nu,t}$ and $c^{\text{opt}}_{\nu,t}$, wherein $\nu_t = \nu^*_t$ is fixed for all $t \in [0, T]$. This second step then consists of approximating the primal-controls as:

$$\pi^*_t = \text{proj}_{K_2}\left(\widehat{\pi}^{\text{opt}}_{\nu,t}\right)\mathcal{K}_{1,t}, \quad \text{and} \quad c^*_t = \widehat{c}^{\text{opt}}_{\nu,t}\mathcal{K}_{2,t}, \tag{2.4.2}$$

where we let $\text{proj}_{K_2} : \mathbb{R}^N \to K_2$ be the operator that projects an N-dimensional argument to the region of constraints imposed on π_t, i.e. $K_2 \subseteq \mathbb{R}^N$ in $K = K_1 \times K_2 \subseteq \mathbb{R} \times \mathbb{R}^N$, under a case-specific metric. Moreover, $\mathcal{K}_{1,t}$ and $\mathcal{K}_{2,t}$ represent scalar-valued $\mathcal{F}_t$-progressively measurable processes that enforce jointly additional (liquidity) constraints upon X^*_t. Namely, as K_2 only contains constants, the projection operator is not in all cases able to fully account for time-dependent borrowing/liquidity constraints. For instance, if $K_1 = \mathbb{R}, K_2 = \mathbb{R}^N_+$ and $X_t \geq 0$ must hold, then $\text{proj}_{K_2}(x) = (x)^+$, for $x \in \mathbb{R}^N$, ensures that $\text{proj}_{K_2}(\widehat{\pi}^{\text{opt}}_{\nu,t})$ meets the trading constraints. However, this projection alone does not guarantee that $X^*_t \geq 0$ holds, because one may consume or invest more than one possesses, resulting in $X^*_t < 0$. Here, the processes $\mathcal{K}_{i,t}$, $i = 1, 2$, become important. That is, by specifying $\mathcal{K}_{i,t}$ e.g. as $\mathcal{K}_{i,t} = \mathbb{1}_{\{X^*_t > 0\}}$ the approximation ensures that $X^*_t \geq 0$ is satisfied. We consider more examples in sections 2.4.1.1 and 2.4.1.2.

The rationale behind (2.4.2) is as follows. The truly optimal π^{opt}_t and c^{opt}_t are equal to $\pi^{\text{opt}}_{\nu,t}$ and $c^{\text{opt}}_{\nu,t}$, for some $\{\nu_t\}_{t\in[0,T]}$, which ensures that $\{\pi^{\text{opt}}_{0,\nu,t}, \pi^{\text{opt}}_{\nu,t}, c^{\text{opt}}_{\nu,t}\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$ holds. Hence, the artificial rules that are implied by ν^*_t are up to ν^{opt}_t completely identical to the truly optimal policies. Furthermore, under $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} \subseteq \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$, these artificial policies are situated the closest to the optimal ones. Consequently, $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ are plausible starting points in the analytical approximation of π^{opt}_t and c^{opt}_t. It then remains to adjust $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ for remaining inaccuracies that interfere with their admissibility in $\mathcal{M}$. To that end, we project $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ towards the admissibility region, using the transformations in (2.4.2). The resulting approximations, π^*_t and c^*_t, are admissible in $\mathcal{M}$.[12] Approximate wealth

[12]By virtue of $\pi_{0,t} = X_t - \pi_t^\top 1_N$, and X_t being endogenously only affected by $\{\pi_s\}_{s\in[0,t]}$ and $\{c_s\}_{s\in[0,t]}$, our approximate method solely concerns π_t and c_t – and via these

correspondingly evolves for all $t \in [0, T]$ as:

$$X_t^* = X_0 + \int_0^t r_s X_s^* \mathrm{d}s + \int_0^t {\pi_s^*}^\top \left([\mu_s - r_s 1_N] \, \mathrm{d}s + \sigma_s \mathrm{d}W_s \right) - \int_0^t (c_s^* - Y_s) \, \mathrm{d}s. \tag{2.4.3}$$

Our approximating procedure renders analytical expressions for the portfolio process, π_t^*, and the consumption pattern, c_t^*. To quantify the accuracy of these approximations, we use strong duality, cf. Theorem 2.3.1. That is, any deviation from the truly optimal primal and/or dual controls results in a duality gap. Substituting $\{c_t^*\}_{t\in[0,T]}$ and X_t^* evaluated at $t = T$ into the objective in (2.2.7) gives a lower bound on the optimal value function, say $J_P^*(\bar{X}_0, \{\pi_t^*, c_t^*\}_{t\in[0,T]})$. In the same way, $\{\nu_t^*\}_{t\in[0,T]}$ gives an upper bound on the optimal value function, say $J_D^*(\bar{X}_0, Z_0^*, \{\nu_t^*\}_{t\in[0,T]})$. The duality gap is equal to the difference between these two bounds, $D_{P,D} := J_D^* - J_P^*$ (see the most left arrow in Figure 2.2). Note that $D_{P,D} = 0$ holds if and only if $\{\nu_t^{\mathrm{opt}}\}_{t\in[0,T]} \in \mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$ or $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} = \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. In all other cases, $D_{P,D}$ is non-negative and can be used to quantify the accuracy of the approximation. We express the approximation's precision in terms of $\mathcal{V} \in \mathbb{R}_+$ from:

$$J_P^* \left(\bar{X}_0 \left[1 + \mathcal{V}\right], \{\pi_t^*, c_t^*\}_{t\in[0,T]} \right) = J_D^* \left(\bar{X}_0, Z_0^*, \{\nu_t^*\}_{t\in[0,T]} \right). \tag{2.4.4}$$

The compensating variation, $\mathcal{V}$, expresses the size of the duality gap as the fraction of total wealth (X_0 and Y_0). It represents the additional amount of capital needed to "bridge" the duality gap. Hence, $\mathcal{V}$ specifies an upper bound on the "welfare loss", suffered due to implementation of the approximate strategy. Naturally, $\mathcal{V}$ depends on the quality of (i) the approximate duality set's selection, and of the (ii) artificial-optimal controls' projection to the admissibility region. Its annualised equivalent, $(1 + \mathcal{V})^{\frac{1}{T}} - 1$, represents an annual management fee that one pays to some representative investor to be protected against the undiversifiability that arises from K. By paying this fee, the agent is thus assured of being able to maintain his/her optimal consumption pattern, and of acquiring the optimal amount of terminal wealth. We refer

$\pi_{0,t}$. Namely, (2.2.7) boils down to determining the former two controls, of which the combination is capable of ensuring that $\pi_{0,t} \in K_1$.

to de Palma and Prigent (2008) and de Palma and Prigent (2009) for more economic details on $\mathcal{V}$.

2.4.1.1 Utility over the Real Line

In case of an economic setup wherein $C = 0$ is fixed in the admissibility set, $\mathcal{A}_{X_0}$, and in which $U \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}; \mathbb{R})$ holds, our approximate method must be modified slightly. These requirements outline an environment that only admits utility functions whose first arguments are specified over $\mathbb{R}_+$ rather than $\mathbb{R}$. Due to the non-negative domain of these utility functions, borrowing against future labour income must be prohibited ($C = 0$). For the setting at hand, we consequently have to account for a concrete liquidity constraint: $X_t \geq 0$ for all $t \in [0, T]$. In the discussion around (2.4.2), we have argued that $\mathcal{K}_{i,t}$ for $i = 1, 2$ are capable of enforcing the liquidity constraints upon $\{X_t^*\}_{t \in [0,T]}$ via the approximate controls, $\text{proj}_{K_2}(\widehat{\pi}_{\nu,t}^{\text{opt}})$ and $\widehat{c}_{\nu,t}^{\text{opt}}$. Indeed, as Bick et al. (2013), one way of assuring that $X_t^* \geq 0$ holds is by specifying $\mathcal{K}_{i,t}$ for $i = 1, 2$ as:

$$\mathcal{K}_{1,t} = \mathbb{1}_{\{X_t^* > 0\}}, \quad \text{and} \quad \mathcal{K}_{2,t} = \mathbb{1}_{\{X_t^* > 0\}} + k \widehat{c}_{\nu,t}^{\text{opt}^{-1}} Y_t \mathbb{1}_{\{X_t^* \leq 0\}}, \tag{2.4.5}$$

for some $k \in [0, 1]$. That is, the moment that approximate wealth, X_t^*, equates to zero, the agent sells all of his/her investments in the risky assets, and only consumes a fraction, k, of his/her current labour income. By this means, approximate wealth is guaranteed of being non-negative over the trading interval. Observe that our method does not require conceptual changes for the present situation, in which one specifies the asset allocation in terms of proportions of wealth, say ψ_t in $\pi_t = \psi_t X_t$, rather than in terms of explicit monetary units, π_t. Since $X_t, X_{\nu,t} > 0$, a mere replacement in (2.4.2) of $\widehat{\pi}_{\nu,t}^{\text{opt}}$ by $\widehat{\psi}_{\nu,t}^{\text{opt}} := \widehat{\pi}_{\nu,t}^{\text{opt}} \widehat{X}_{\nu,t}^{\text{opt}^{-1}}$, where $\widehat{X}_{\nu,t}^{\text{opt}}$ is identical to optimal wealth in Proposition 2.3.3 for $\nu_t = \nu_t^*$, results in a routine that is appropriate for this setup.

2.4.1.2 Concrete Constraint Specifications

The second step of our twofold approximating procedure, displayed in (2.4.2), can appear to be rather abstract. In order to make this operation more tangible, let us consider a set of concrete specifications of K. For now, we do not impose any liquidity constraints, so as to be able to highlight at a later stage that $\mathcal{K}_{i,t}$ for $i = 1, 2$ suffice to ensure the relevant liquidity of X_t^*. Consider the following characterisations of trading constraints, K: $K = \mathbb{R} \times \mathbb{R}^{N-M} \times \{0\}^M$ (non-traded assets); $K = \mathbb{R} \times \mathbb{R}_+^N$ (short-sale constraints); and $K = \mathbb{R} \times (-\infty, M]^N$ (buying constraints). Then, the following projections manage to ensure that $\left(\pi_{0,t}^*, \pi_t^*\right) \in K$ holds true for all $t \in [0, T]$, under the respective sets of constraints:

$$\mathrm{proj}_{K_2}\left(\widehat{\pi}_{\nu,t}^{\mathrm{opt}}\right) = \left[\widehat{\pi}_{\nu,N-M,t}^{\mathrm{opt}}, 0_M\right]^\top, \ \max\left\{0_N, \widehat{\pi}_{\nu,t}^{\mathrm{opt}}\right\}, \ \min\left\{\widehat{\pi}_{\nu,t}^{\mathrm{opt}}, M1_N\right\}. \tag{2.4.6}$$

The max- and min-operators are functioning in an element-wise manner: e.g. $\max\{0_N, \widehat{\pi}_{\nu,t}^{\mathrm{opt}}\} = (\max\{0, \widehat{\pi}_{\nu,i,t}^{\mathrm{opt}}\})_{i=1}^N$, in which $\widehat{\pi}_{\nu,i,t}^{\mathrm{opt}}$ represents the i^{th} element of $\widehat{\pi}_{\nu,t}^{\mathrm{opt}}$. Another common set of trading constraints is set out by a combination of the short-sale and borrowing restrictions ($K = \mathbb{R} \times [0, M]^N$), in case of which $\mathrm{proj}_{K_2}(\widehat{\pi}_{\nu,t}^{\mathrm{opt}}) = \max\{0_N, \min\{\widehat{\pi}_{\nu,t}^{\mathrm{opt}}, M1_N\}\}$ ensures that $\left(\pi_{0,t}^*, \pi_t^*\right) \in K$ holds. Note that any combination involving non-traded assets, would require a projection that additionally nullifies the relevant elements in $\widehat{\pi}_{\nu,t}^{\mathrm{opt}}$.

Now, let us suppose that $(\pi_{0,t}, \pi_t)$ ought to attain values in the preceding definitions of K. Additionally, assume that both $(\pi_{0,t}, \pi_t)$ and c_t must be chosen, such that $X_t \geq -C$ holds for some $C \in \mathbb{R}_+$. Then,

$$\mathcal{K}_{1,t} = \mathbb{1}_{\left\{X_t^* > -C\right\}}, \quad \text{and} \quad \mathcal{K}_{2,t} = \mathbb{1}_{\left\{X_t^* > -C\right\}} + k\widehat{c}_{\nu,t}^{\mathrm{opt}^{-1}} Y_t \mathbb{1}_{\left\{X_t^* \leq -C\right\}}, \tag{2.4.7}$$

for some $k \in [0, 1]$, would outline two specifications of $\mathcal{K}_{i,t}$ for $i = 1, 2$ that enforce the liquidity constraint upon $\{X_t^*\}_{t \in [0,T]}$. Observe that we do not have to alter the projections in (2.4.6) to ensure that $(\pi_{0,t}, \pi_t)$ attains values in the new definitions of K (these trading restrictions are independent of time). Clearly, the characterisations of $\mathcal{K}_{i,t}$ for $i = 1, 2$ in (2.4.7) alone manage to constrain approximate wealth from below, in line with the liquidity constraint. In case of $C \in \mathbb{R}_-$, one arrives at a portfolio insurance setting, cf. Basak (1995), Basak and Chabakauri (2012), and Browne (2013), and the definitions in (2.4.7) are still valid. Nevertheless, C is not allowed to exceed a pre-specified amount of monetary units, cf. Basak (1995). Ultimately, note that enforcing $X_t^* \leq \mathcal{C}$ for some $\mathcal{C} \in \mathbb{R}_+$ requires only moderate changes in (2.4.7).

2.4.1.3 Step-wise Overview and Remarks

In order to elicit the essential steps that support our twofold approximating routine, and to provide a clear and concise overview of the method as set forth above, we subsequently explicate our method in a step-wise fashion. Herein, we include an optional step, after (2.4.2). In this additional step, the approximate shadow price process, ν_t^*, which is incorporated in π_t^* and c_t^*, is "reset" to an undetermined $\bar{\nu}_t = \nu_t^*$. Hereafter, $\bar{\nu}_t$ is identified by maximising the approximate primal value function over all $\bar{\nu}_t \in \mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$. We exclude it in section 2.4.1, because our numerical evaluations have shown that this step hardly increases the approximation's quality. However, for completeness, and to account for case-specific peculiarities, we include it here:

- Restrict the dual space, $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$, to some convex subspace, $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} \subseteq \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$.

- Determine the approximate shadow price process by minimising the dual: $\{\nu_t^*\}_{t\in[0,T]} := \arg\inf_{\{\nu_t\}_{t\in[0,T]}\in\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}} J_D(\bar{X}_0, Z_0^*, \{\nu_t\}_{t\in[0,T]})$, and calculate $J_D^*(\bar{X}_0, Z_0^*, \{\nu_t^*\}_{t\in[0,T]})$, i.e. the approximate dual value function.

- Insert $\{\nu_t^*\}_{t\in[0,T]}$ into the artificial-optimal controls in $\widehat{\mathcal{M}}_\nu$, $\{\pi_{\nu,t}^{\text{opt}}\}$ and $\{c_{\nu,t}^{\text{opt}}\}$, to obtain $\{\widehat{\pi}_{\nu,t}^{\text{opt}}\}$ and $\{\widehat{c}_{\nu,t}^{\text{opt}}\}$, respectively.

- Approximate the optimal allocation to assets, and the optimal consumption pattern in $\mathcal{M}$ as follows: $\pi_t^* = \text{proj}_{K_2}(\widehat{\pi}_{\nu,t}^{\text{opt}})\mathcal{K}_{1,t}$ and $c_t^* = \widehat{c}_{\nu,t}^{\text{opt}}\mathcal{K}_{2,t}$, and calculate the approximate primal value function, $J_P^*(\bar{X}_0, \{\pi_t^*, c_t^*\}_{t\in[0,T]})$.

- (Optional) Take π_t^* and c_t^* from the latter step and fix $\nu_t^* = \bar{\nu}_t$ in these controls. Subsequently, determine $\bar{\nu}_t$ by optimising the primal: $\{\bar{\nu}_t\}_{t\in[0,T]} = \arg\sup_{\{\bar{\nu}_t\}_{t\in[0,T]}\in\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}} J_P^*(\bar{X}_0, \{\pi_t^*, c_t^*\}_{t\in[0,T]})$, and re-calculate π_t^*, c_t^*, $J_P^*(\bar{X}_0, \{\pi_t^*, c_t^*\}_{t\in[0,T]})$ under $\nu_t^* = \bar{\nu}_t$.

- Quantify the approximation's accuracy by calculating the compensating variation, $\mathcal{V} \in \mathbb{R}_+$, in the following way: $J_P^*(\bar{X}_0[1+\mathcal{V}], \{\pi_t^*, c_t^*\}_{t\in[0,T]}) = J_D^*(\bar{X}_0, Z_0^*, \{\nu_t^*\}_{t\in[0,T]})$.

For technicalities, as well as for employed definitions and notation, we refer to section 2.4.1. Regarding the fifth optional step, we underline that one is not obliged to determine $\bar{\nu}_t$ by maximising the primal value function over $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$. Concretely, one is free to select any other convex set than $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$. Nevertheless, in consideration of the fact that π_t^* and c_t^* are essentially based on $\pi_{\nu,t}^{\text{opt}}$ and $c_{\nu,t}^{\text{opt}}$, and on $\nu_t \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$'s ability to prune the previous controls towards optimality and admissibility in $\mathcal{M}$, a subset of $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ is more likely to render near-optimal solutions than sets adjacent to $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$. Moreover, one must take into account the need for numerical routines in determining $\bar{\nu}_t$, irrespective of whether one optimises over $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$ or over a different set. Similar routines may be necessary in executing the second and the sixth steps.[13]

Remark 2.4.1. *We finalise by addressing the closed-form nature of the approximate controls. Pertaining to the dual controls, the possibility of acquiring $\{\nu_t^*\}_{t\in[0,T]}$ (and Z_0^*) in closed-form depends on the choice for $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$. As to the primal controls, when inspecting* (2.4.2)*, we note that the projection operator leaves the closed-form character of $\widehat{\pi}_{\nu,t}^{\text{opt}}$ intact. Furthermore, the definitions of the processes $\mathcal{K}_{i,t}$, for $i = 1, 2$, are subject to the user's input, and therefore abide by an analytical structure if the user wishes so. Particularly, in section 2.4.1.2, we provide various analytical specifications of these processes, for several regularly imposed liquidity constraints. The qualification of the*

[13] In the second step, identifiable with (2.4.1), it may still be necessary to calculate $\left\{\nu_t^*\right\}_{t\in[0,T]}$ by means of some numerical routine. Looking, for example, at (2.3.14), one can conclude that not all subsets of $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ provide closed-form shadow price processes. The emphasis in the main text on the tractable nature of $\left\{\nu_t^*\right\}_{t\in[0,T]}$ should therefore be understood as the accentuation of the *possibility* to recover π_t^* and c_t^* in analytical form. In the sixth step, identifiable with (2.4.4), due to the X_t^* process, which is unavailable in closed-form, it is clear that the possibility of acquiring $\mathcal{V} \in \mathbb{R}_+$ requires simulations. Moreover, depending on the selected utility functions, obtaining $\mathcal{V}$ may also require numerical root-solving methods. Similar arguments apply to acquiring Z_0^*.

primal controls, π_t^* and c_t^*, as analytical is thus correct, insofar as the user selects closed-form processes $\mathcal{K}_{i,t}$, $i = 1, 2$, and a set $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$, which engenders a tractable shadow price process, $\{\nu_t^*\}_{t\in[0,T]}$. Conversely, if the user selects non-analytical $\mathcal{K}_{i,t}$ and/or a set $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$, which provides a numerical $\{\nu_t^*\}_{t\in[0,T]}$, then π_t^* and c_t^* are not available in closed-form.

2.4.2 Illustration of the Approximation

Our routine's range of application extends Bick et al. (2013)'s to financial market models that (i) contain general return dynamics that admit non-quadratic and non-affine structures, (ii) incorporate general liquidity constraints and convex trading restrictions, and (iii) include state-dependent preference qualifications that are specified over the real-line and explicitly accommodate an exogenous stochastic benchmark. In this section, we provide an illustration of our approximate method, in which we focus on the last two items. To that end, we employ a slightly modified version of Cocco et al. (2005)'s economic environment. In this market model, we introduce a new asset and assume that the agent derives utility relative to a non-negative stochastic benchmark. Moreover, we postulate that the preferences regarding terminal wealth are described by Chen et al. (2011)'s SAHARA (Symmetric Asymptotic Hyperbolic Absolute Risk Aversion) function, rather than by a CRRA (Constant Relative Risk Aversion) qualification.

2.4.2.1 Cocco et al. (2005)'s Market Model

Let us consider the financial market model in section 2.2.1, $\mathcal{M}$, and set $N = 3$. This market model outlines a three-dimensional setup, with three corresponding risk-drivers, $\{W_t\}_{t\in[0,T]}$, where $W_t = [W_{1,t}, W_{2,t}, W_{3,t}]^\top$. In this environment, we assume that all local drift and diffusion terms in the exogenous processes are constant. We accordingly relax the time-dependence indicator in the corresponding subscripts: for example, in equation (2.2.2), $\mu_{i,t}$ and $\sigma_{i,t}$ become μ_i and σ_i for $i = 1, 2, 3$, respectively. We let $\sigma \in \mathbb{R}_+^{3\times 3}$ be a diagonal matrix, with diagonal entries σ_{11}, σ_{22}, σ_{33}. As a consequence, $S_{i,t}$ is only driven by $W_{i,t}$, for $i = 1, 2, 3$. For the labour income process, $\{Y_t\}_{t\in[0,T]}$, we fix $\sigma_Y :=$

$\sigma_Y[\rho_{SY}, \sqrt{1-\rho_{SY}^2}, 0]^\top$, given some $\sigma_Y \in \mathbb{R}_+$ and $\rho_{SY} \in [-1,1]$, defining the instantaneous correlation parameter between S_t and Y_t.

For the artificial environment, $\widehat{\mathcal{M}}_\nu$, we define the set of approximate dual controls, $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$, as follows:

$$\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} = \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}} \cap \left\{\nu_t \mid \nu_t = \nu_a + \nu_b t,\ \nu_a, \nu_b \in \mathbb{R}^4, t \in [0,T]\right\}. \tag{2.4.8}$$

In view of (2.4.1) and (2.4.2), we restrict all shadow price processes to a subset of the duality space that only contains deterministic affine functions that are linear in time. We use this subset for the purpose of analytical and computational tractability in the numerical evaluation of our method. We denote the drift and diffusion terms in $\widehat{\mathcal{M}}_\nu$ as $\nu_t = \nu_a + \nu_b t$ for $\nu_i = [\nu_{i,0}, \nu_{i,1}, \nu_{i,2}, \nu_{i,3}]^\top$, $i = a, b$. Note that the exact definition of $\{\nu_t\}_{t\in[0,T]}$ under (2.4.8) depends on K's characterisation. More precisely, for general $K \subseteq \mathbb{R}^{N+1}$, we are solely able to state that $\nu_t = \nu_a + \nu_b t$ holds. We observe that the specification of $\mathcal{M}$ as provided here and the shadow price process' restriction to the space $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$, given in (2.4.8), are sufficient for acquiring a complete description of the artificial market, $\widehat{\mathcal{M}}_\nu$. Ultimately, observe that all processes in $\mathcal{M}$ and $\widehat{\mathcal{M}}_\nu$ that are not affected by the definitions above remain entirely intact, e.g. all the endogenous ones (c_t, $c_{\nu,t}$ π_t, $\pi_{\nu,t}$).

Remark 2.4.2. *The financial market model above differs from the one in Cocco et al. (2005), as employed in Bick et al. (2013). The difference is attributable to the fact that the model contains a third traded risky asset. Hence, the model as provided here does apart from the latter addition not enlarge the setting of Bick et al. (2013). Nonetheless, to disclose the differentiating elements of our routine, we define several "versions" of the foregoing environment. For that purpose, we specify alternative trading restrictions (K) and liquidity constraints, and introduce the non-trivial state-dependent ratio-SAHARA qualification. Due to the additional asset, the ability to impose liquidity constraints per se, and the SAHARA function being state-dependent and specified over the real line, the environments noticeably differ from Bick et al. (2013)'s. The (comparative) generality of the possibly enforced constraints and of the potentially utilised preferences can therefore be adequately illustrated.*

2.4.2.2 CRRA and SAHARA Utility Functions

We proceed by introducing the agent's preferences. For this reason, let us consider the utility functions in (2.2.7), under the market model as specified in section 2.4.2.1. We fix $\Pi_t := \Pi_{1,t} = \Pi_{2,t}$, i.e. equate the benchmark processes to each other, and set $\mu_{\Pi,i,t} = \mu_\Pi \Pi_t$ as well as $\sigma_{\Pi,i,t} = \sigma_\Pi \Pi_t$, for some $\mu_\Pi \in \mathbb{R}$ and $\sigma_\Pi \in \mathbb{R}^3$. As a result, Π_t is log-normally distributed. In addition, we assume that the investor derives utility from consumption via the so-called ratio-CRRA function: $u(t,x,y) = e^{-\beta t} \frac{(x/y)^{1-\gamma}-1}{1-\gamma}$ for all $x, y \in \mathbb{R}_+$, $\beta \in \mathbb{R}$, and $\gamma \in \mathbb{R}_+ \setminus [0,1]$. Here, β denotes the agent's time preference parameter, and γ represents his/her coefficient of relative risk-aversion (RRA).

Furthermore, the investor derives utility from the ratio of terminal wealth to the benchmark process, X_T/Π_T. His/her preferences are accordingly specified via the following ratio-SAHARA utility specification:

$$U(x,y) = -\frac{1}{\alpha_1^2 - 1} \frac{(x/y-1) + \alpha_1 \sqrt{\alpha_2^2 + (x/y-1)^2}}{\left[(x/y-1) + \sqrt{\alpha_2^2 + (x/y-1)^2}\right]^{\alpha_1}}, \tag{2.4.9}$$

for all $x \in \mathbb{R}$, $y \in \mathbb{R}_+$ and $\alpha_i \in \mathbb{R}_+$, for $i = 1, 2$, such that $\alpha_1 \neq 1$. For $\alpha_1 = 1$, the function in (2.4.9) abides by $U(x,y) = \frac{1}{2}\log\left(\bar{x} + \sqrt{\alpha_2 + \bar{x}^2}\right) + \frac{1}{2}\frac{1}{\alpha_2^2}\bar{x}[\sqrt{\alpha_2^2 + \bar{x}^2} - \bar{x}]$, where $\bar{x} = \frac{x}{y} - 1$, for all $x \in \mathbb{R}$, $y \in \mathbb{R}_+$ and $\alpha_2 \in \mathbb{R}_+$. Due to the fact that one definition of $I : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ corresponding to (2.4.9) holds for all $\alpha_1 \in \mathbb{R}_+$, in deriving optimality conditions to (2.3.7), henceforth we do not make the latter distinction. Concretely, the inverses of marginal utility read: $\iota(t,x,y) = \left(e^{\beta t} xy\right)^{-\frac{1}{\gamma}} y$ and $I(x,y) = \frac{1}{2}[(xy)^{-\frac{1}{\alpha_1}} - \alpha_2^2 (xy)^{\frac{1}{\alpha_1}}]y + y = \alpha_2 \sinh(-\frac{1}{\alpha_1}\log xy - \log \alpha_2) y + y$ for all $x, y \in \mathbb{R}_+$ and $t \in [0,T]$. Furthermore, the levels of ARA, cf. Proposition 2.3.3, for these utility functions abide by: $-\frac{u''_X(t,x,y)}{u'_X(t,x,y)} = \frac{\gamma}{x}$ for all $x, y \in \mathbb{R}_+$ and $t \in [0,T]$, and $-\frac{U''(x,y)}{U'(x,y)} = \frac{\alpha_1}{y\sqrt{\alpha_2^2 + \bar{x}^2}}$ for all $x \in \mathbb{R}$ and $y \in \mathbb{R}_+$. Note that all other, relevant, terms that are addressed in the main text, such as $-\iota'_Y(t,x,y)$ and $-I'_Y(x,y)$, can be derived on the basis of the information above.

Remark 2.4.3. *Our preference qualifications differ from Cocco et al. (2005)'s in that the investor in our model proportionally compares all relevant endogenous* *quantities ($\{c_t\}_{t\in[0,T]}$ and X_T) to a benchmark process, $\{\Pi_t\}_{t\in[0,T]}$. Due to this feature, a state-dependent component affects the utility levels. On top of that, the agent in our environment derives utility from terminal wealth via a SAHARA function,* (2.4.9)*, rather than via a CRRA function. Clearly, this utility specification is defined over the real line, and explicitly models the preferences around the benchmark. Namely, the SAHARA qualification constitutes a generalisation of the exponential utility function (which exhibits constant ARA), because its ARA is wealth-dependent. Concretely, its ARA is given by $\frac{\alpha_1}{y\sqrt{\alpha_2^2+\bar{x}^2}}$ for $\bar{x}=\frac{x}{y}-1$, which shows that for deviations of x away from y, ARA decreases and the agent becomes more willing to engage in risky trades. The agent is therefore inclined to trade in a manner that is concerning X_T strongly target-oriented, cf. Chen et al. (2011).*

2.4.2.3 Approximated Primal and Dual Controls

We conclude with an illustration of the approximate method that focuses on its mathematical aspects. We use the economic environment described in sections 2.4.2.1 and 2.4.2.2, with three distinct characterisations of the included trading and liquidity constraints (K). In our approximate method, the first step, (2.4.1), is entirely dependent on K's definition, whereas the second step, (2.4.2), only depends partially on K. In particular, in (2.4.2), we note that the projection operator and the functions that ensure satisfaction of the liquidity constraint are completely dependent on K. Conversely, in the same equation, $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ are entirely unaffected by K. Therefore, we first focus on the latter step. We revisit Proposition 2.3.3 to obtain $\widehat{\pi}^{\text{opt}}_{\nu,t}$ and $\widehat{c}^{\text{opt}}_{\nu,t}$ in closed-form. Corollary 2.4.1 outlines the optimal solutions to (2.3.7) for the model at hand. For notational purposes and brevity of Corollary 2.4.1, we introduce:

$$f\left(t,T\right)=\int_t^T e^{h\left(t,s,-\frac{1}{\gamma}\right)-\frac{1}{\gamma}\beta(s-t)}\mathrm{d}s, \tag{2.4.10}$$

where h is given by:

$$h\left(t,s,x\right)=\left(1+x\right)\int_t^s\left[-r-\nu_{0,u}+\mu_\Pi-\sigma_\Pi^\top\widehat{\lambda}_u+\frac{1}{2}x\|\sigma_\Pi-\widehat{\lambda}_s\|^2_{\mathbb{R}^3}\right]\mathrm{d}u. \tag{2.4.11}$$

Corollary 2.4.1. *Consider the unconstrained investment problem* (2.3.7), *in the market model defined by sections 2.4.2.1 and 2.4.2.2. Define the following two processes for all* $t \in [0,T]$*:* $\mathcal{X}_{t,T} = \alpha_{2,t,T} \sinh\left(-\frac{1}{\alpha_1} \log\left[\eta^{\text{opt}} Z_{\nu,t} \Pi_t e^{-\alpha_1 h\left(t,T,-\frac{1}{\alpha_1}\right)}\right] - \log(\alpha_{2,t,T})\right)$ *and* $\widehat{\mathcal{X}}_{t,T} = \frac{1}{2}[(\eta^{\text{opt}} Z_{\nu,t}\Pi_t)^{-\frac{1}{\alpha_1}} e^{h\left(t,T,-\frac{1}{\alpha_1}\right)} - \alpha^2_{2,t,T} \frac{\alpha_1+1}{\alpha_2+2} (\eta^{\text{opt}} Z_{\nu,t}\Pi_t)^{\frac{1}{\alpha_1}} e^{-h\left(t,T,-\frac{1}{\alpha_1}\right)}]$*, for* $\alpha_{2,t,T} = \alpha_2 e^{\frac{1}{2}[h(t,T,-\frac{1}{\alpha_1})+h(t,T,\frac{1}{\alpha_1})]}$*. Then, optimal consumption is given by*

$$\widehat{c}^{\text{opt}}_{\nu,t} = e^{-\frac{1}{\gamma}\beta t} \left(\eta^{\text{opt}} Z_{\nu,t}\right)^{-\frac{1}{\gamma}} \Pi_t^{1-\frac{1}{\gamma}}, \tag{2.4.12}$$

for all $t \in [0,T]$*, where* $\eta^{\text{opt}} = \mathcal{G}^{-1}(X_0)$*. Furthermore, the optimal allocation to assets,* $\{\widehat{\pi}^{\text{opt}}_{\nu,t}\}_{t\in[0,T]}$*, reads, for all* $t \in [0,T]$*, as* $\widehat{\pi}^{\text{opt}}_{\nu,t} = \widehat{\pi}^m_t + \widehat{\pi}^Y_t + \widehat{\pi}^\Pi_t$*. The first component of this sum lives by*

$$\widehat{\pi}^m_t = \left(\frac{1}{\alpha_1}\sqrt{\alpha^2_{2,t,T} + \mathcal{X}^2_{t,T}}\,\Pi_t + \frac{1}{\gamma}\widehat{c}^{\text{opt}}_{\nu,t} f(t,T)\right) \sigma^{\top^{-1}}\widehat{\lambda}, \tag{2.4.13}$$

and specifies the ordinary mean-variance hedge demand, corresponding to π^m_t *in* (2.3.12)*. Since the shadow prices of risk are deterministic,* $\mathcal{D}^W_t \log \widehat{Z}_{\nu,s} = 0_3$ *holds true for all* $s \geq t$ *such that* $s,t \in [0,T]$*. As a consequence, the other hedge demand in* (2.3.12)*,* π^Z_t*, is equal to zero. The remaining two portfolio weights in* $\widehat{\pi}^{\text{opt}}_{\nu,t}$*'s decomposition are labour income (*$\widehat{\pi}^Y_t$*) and benchmark (*$\widehat{\pi}^\Pi_t$*) hedge demands corresponding to* π^Y_t *and* π^Π_t *in* (2.3.13)*, respectively. These demands are specified by the following two identities*

$$\begin{aligned}\widehat{\pi}^Y_t &= -\sigma^{\top^{-1}} Y_t \sigma_Y \int_t^T e^{-\int_t^s \left(r+\nu_{0,u}-\mu_Y+\sigma_Y^\top\widehat{\lambda}_u\right)\mathrm{d}u}\mathrm{d}s, \quad \text{and}\\ \widehat{\pi}^\Pi_t &= \sigma^{\top^{-1}}\left(\frac{\alpha_1-1}{\alpha_1}\widehat{\mathcal{X}}_{t,T}\Pi_t + \Pi_t e^{h(0,T,0)} + \frac{\gamma-1}{\gamma}\widehat{c}^{\text{opt}}_{\nu,t} f(t,T)\right)\sigma_\Pi.\end{aligned} \tag{2.4.14}$$

Proof. The proof is given in Appendix A.5. □

Using Corollary 2.4.1, we are able to iterate step-wise through our twofold approximating procedure for concrete specifications of K. The three specifications of K that we will investigate are: (i) $K = \mathbb{R}^3 \times \{0\}$, (ii) $K = \mathbb{R}\times\mathbb{R}^2_+ \times \{0\}$, and (iii) $K = [-C,\infty)\times\mathbb{R}_+\times\{0\}^2$ for some $C \in \mathbb{R}_+$. Item (i) implies a setup, in which $S_{3,t}$ is non-traded and Π_t consequently (partially) unhedgeable;

item (ii)'s is identical to the previous setup, wherein a short-sale constraint is added; item (iii)'s is equal to the latter situation, in which additionally $S_{2,t}$ is non-traded (making Y_t partially undiversifiable) and a borrowing/liquidity constraint is imposed ($\pi_{0,t} \geq -C$ for all $t \in [0,T]$). This third specification of the trading constraints is up to the (partial) non-tradeability of the benchmark process identical to that of Cocco et al. (2005). In Examples 2.4.1 and 2.4.2, we exemplify the evident need for an approximate method in the present setup, and we present the approximate controls for the definitions of K.

Example 2.4.1. *(Optimal Shadow Price) Consider the financial market model defined in sections 2.4.2.1 and 2.4.2.2, and relax the assumption* $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} = \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}} \cap \{\nu_t \mid \nu_t = \nu_a + \nu_b t,\ \nu_a, \nu_b \in \mathbb{R}^4, t \in [0,T]\}$, *cf.* (2.4.8). *Instead, let us assume that* $\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}} = \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ *holds true, which means that* $\mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ *is not contracted to some convex subspace* ($\mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$). *Suppose that* $K = \mathbb{R}^3 \times \{0\}$. *Following Example 2.3.1, we then find that if* $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$, *where* ν_t *decomposes as* $\nu_t = [\nu_{0,t}, \nu_{N,t}]^\top \in \mathbb{R} \times \mathbb{R}^3$ *for* $\nu_{N,t} = [\nu_{1,t}, \nu_{2,t}, \nu_{3,t}]^\top$, *the following must hold:* $\nu_{0,t} = 0$, $\nu_{1,t} = \nu_{2,t} = 0$ *and* $\nu_{3,t} \in \mathbb{D}^{1,2}([0,T])$. *In the present setting, in line with* (2.3.14), *we know that the optimal shadow price,* $\nu_{3,t}$ *can for all* $t \in [0,T]$ *be determined from:*

$$\widehat{\lambda}_{3,t} X^{\text{opt}}_{\nu,t} = \mathbb{E}\left[\frac{\widehat{\mathcal{R}}_{2,X,T}}{Z_{\nu,t}} - \frac{\mathcal{D}^{W_3}_t Z_{\nu,T}}{\widehat{\mathcal{R}}_{1,X,T} Z_{\nu,t}} + \int_t^T \left\{ \frac{\widehat{\mathcal{R}}_{2,c,s}}{Z_{\nu,t}} - \frac{\mathcal{D}^{W_3}_t Z_{\nu,s}}{\widehat{\mathcal{R}}_{1,c,s} Z_{\nu,t}} \right\} \mathrm{d}s \,\middle|\, \mathcal{F}_t \right]. \tag{2.4.15}$$

In this identity, we define $\widehat{\mathcal{R}}_{2,X,T} = \mathcal{R}_{2,X,T} Z_{\nu,T} \Pi_T \sigma_{\Pi,3}$ *and* $\widehat{\mathcal{R}}_{2,c,t} = \mathcal{R}_{2,c,t} Z_{\nu,t} \Pi_t \sigma_{\Pi,3}$ *for all* $t \in [0,T]$. *Additionally,* $\mathcal{D}^{W_3}_t Z_{\nu,s} = Z_{\nu,s}\big(- \int_t^s (\mathcal{D}^{W_3}_t \widehat{\lambda}_u) \widehat{\lambda}_u \mathrm{d}u - \int_t^s \mathcal{D}^{W_3}_t \widehat{\lambda}_u \mathrm{d}W_{3,u} - \widehat{\lambda}_{3,t}\big)$ *holds for all* $s \geq t$ *such that* $s,t \in [0,T]$ *and* $\widehat{\lambda}_{3,t}$ *in* $\widehat{\lambda}_t = [\widehat{\lambda}_{1,t}, \widehat{\lambda}_{2,t}, \widehat{\lambda}_{3,t}]^\top \in \mathbb{R}^3$. *After inserting the definitions from section 2.4.2.2 into* (2.4.15), *and noting that* $Z_{\nu,t}$ *depends on* $\{\widehat{\lambda}_{3,s}\}_{s\in[0,t]}$, *it is clear that the RHS of* (2.4.15) *depends on past and future values of* $\widehat{\lambda}_{3,t}$. *Concretely,* (2.4.15) *implies an FBSDE, like* (2.3.14), *cf. section 6.1 in Detemple (2014). Therefore, we cannot recover* $\{\widehat{\lambda}_{3,t}\}_{t\in[0,T]}$, *and thus* $\{\nu_{3,t}\}_{t\in[0,T]}$, *in closed-form. Due to the inclusion of* (2.4.15) *in all specifications of the dual-optimal* $\{\widehat{\lambda}_t\}_{t\in[0,T]}$, *corresponding to the other descriptions of* K, *the same applies to the other two setups.*

Example 2.4.2. ***(i)*** *(Non-traded Asset) Consider the same model as in Example 2.4.1, with* $K = \mathbb{R}^3 \times \{0\}$*, which means that* $\{S_{3,t}\}_{t\in[0,T]}$ *is not available for trading, and* $\{\Pi_t\}_{t\in[0,T]}$ *is partially unhedgeable. Note that there is no liquidity constraint active. The shadow price process is for* $\{\nu_t\}_{t\in[0,T]} \in \mathcal{P}_{\widehat{\mathcal{A}}_{X_0}}$ *given by* $\nu_{0,t} = 0$*,* $\nu_{1,t} = \nu_{2,t} = 0$ *and* $\nu_{3,t} = \nu_{a,3} + \nu_{b,3}t$*, for some* $\nu_{a,3}, \nu_{b,3} \in \mathbb{R}$*. In light of* (2.4.2)*, let* $\mathcal{K}_{1,t} = \mathcal{K}_{2,t} = 1$ *for all* $t \in [0,T]$*, since no liquidity constraints are active. Furthermore, we employ a projection operator that nullifies any allocation of assets to* $\{S_{3,t}\}_{t\in[0,T]}$*:* $\mathrm{proj}_{K_2}(x) = [x_1, x_2, 0]^\top$ *for all* $x \in \mathbb{R}^3$ *such that* $x = [x_1, x_2, x_3]^\top$*. Ultimately, set* $\widehat{\pi}^{\text{opt}}_{\nu,t} = [\widehat{\pi}^{\text{opt}}_{\nu,1,t}, \widehat{\pi}^{\text{opt}}_{\nu,2,t}, \widehat{\pi}^{\text{opt}}_{\nu,3,t}]^\top$*. Then, utilising our method in* (2.4.1) *and* (2.4.2)*, we determine the approximate dual controls*[14] *in the following manner:* $[\nu_{a,3}, \nu_{b,3}] = \arg\inf_{[\nu_{a,3},\nu_{b,3}]\in\mathbb{R}^2} J_D(\bar{X}_0, Z^*_0, \{\nu_t\}_{t\in[0,T]})$*. Furthermore, the expressions for the admissible approximate primal controls read:*

$$\pi^*_t = \left[\widehat{\pi}^{\text{opt}}_{\nu,1,t}, \widehat{\pi}^{\text{opt}}_{\nu,2,t}, 0\right]^\top \quad \text{and} \quad c^*_t = \widehat{c}^{\text{opt}}_{\nu,t}. \tag{2.4.16}$$

(ii) *(Non-traded Asset and Short-sale Constraints) Consider the market model above, with* $K = \mathbb{R} \times \mathbb{R}^2_+ \times \{0\}$*, which implies that in addition to the preceding non-tradeability of* $\{S_t\}_{t\in[0,T]}$*, short-sales are prohibited (we again exclude a liquidity constraint). As a result, the shadow price process is for* $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ *given by* $\nu_0 = 0$*, and (yet) undetermined* $\nu_{i,1}, \nu_{i,2}, \nu_{i,3} \in \mathbb{R}$*, for* $i = a, b$*, that ought to ensure* $\nu_{1,t}, \nu_{2,t} \geq 0$ *for all* $t \in [0,T]$*. Due to the exclusion of liquidity constraints we define* $\mathcal{K}_{1,t} = \mathcal{K}_{2,t} = 1$ *for all* $t \in [0,T]$*, with an eye on* (2.4.2)*. Moreover, we make use of the projection operator on which* (2.4.16) *relies, which additionally ensures non-negativity of the allocation to assets:* $\mathrm{proj}_{K_2}(x) = \left[(x_1)^+, (x_2)^+, 0\right]^\top$ *for all* $x \in \mathbb{R}$ *such that* $x = [x_1, x_2, x_3]^\top$*, where* $(x_i)^+ = \max\{x_i, 0\}$ *for* $i = 1, 2$*. Then, in keeping with our method, we identify* ν_a *and* ν_b *as in* (2.4.1)*. Moreover, following* (2.4.2)*, the approximate*

[14] The approximate shadow price processes are characterised by analytical identities that require straightforward root-searching algorithms to be solved. In the subsequent evaluation of the method, we therefore accordingly, i.e. numerically, determine the approximate shadow prices (ν_a and ν_b). We stress that this does not interfere with the analytical nature of the primal controls. Since the identities for ν_a and ν_b infer little to nothing about the economic meaning of the processes, we exclude them in this example. Similarly, we exclude expressions for the approximate dual value functions. As the primal value functions have to be calculated by means of simulations, we likewise utilise simulations to obtain the dual value function – to rule out a simulation-bias.

primal controls live by:

$$\pi_t^* = \left[\left(\widehat{\pi}_{\nu,1,t}^{\text{opt}}\right)^+, \left(\widehat{\pi}_{\nu,2,t}^{\text{opt}}\right)^+, 0\right]^\top \quad \text{and} \quad c_t^* = \widehat{c}_{\nu,t}^{\text{opt}} \tag{2.4.17}$$

(iii) *(Non-traded Asset, Short-sale and Liquidity Constraints) Ultimately, consider the market model from the previous setting and let K as $K = [-C, \infty) \times \mathbb{R}_+ \times \{0\}^2$ for some $C \in \mathbb{R}_+$. This means that $\{S_{2,t}\}_{t\in[0,T]}$ and $\{S_{3,t}\}_{t\in[0,T]}$ are non-traded, short-sale constraints are active, and a borrowing/liquidity constraint is imposed ($\pi_{0,t} \geq -C$ for all $t \in [0,T]$). To be precise, we explicitly impose the following liquidity constraint: $X_t^* \geq -C$ for all $t \in [0,T]$. Consequently, the shadow price process is for $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ characterised by $\nu_a, \nu_b \in \mathbb{R}$ such that $\nu_{0,t}, \nu_{1,t} \geq 0$ holds for all $t \in [0,T]$. Unlike in the previous two setups, in the present one the support function, $\delta(\nu)$, is non-zero and equates to $\delta(\nu) = \nu_0 C$. To enforce the borrowing/liquidity constraint upon wealth, in consideration of (2.4.2), we define $\mathcal{K}_{1,t} = \mathbb{1}_{\{X_t^* \geq -C\}}$ and $\mathcal{K}_{2,t} = \mathbb{1}_{\{X_t^* \geq -C\}} + k\widehat{c}_{\nu,t}^{\text{opt}^{-1}} Y_t \mathbb{1}_{\{X_t^* < -C\}}$ for all $t \in [0,T]$ and some $k \in [0,1]$. We utilise a slightly modified version of the projection operator from the previous setup, which additionally nullifies any allocation of assets to $\{S_{2,t}\}_{t\in[0,T]}$ and respects the borrowing constraint: $\text{proj}_{K_2}(x) = \left[\min\{(x_1)^+, X_t^* + C\}, 0, 0\right]^\top$ for all $x \in \mathbb{R}$ such that $x = [x_1, x_2, x_3]^\top$. Then, consistent with our routine, we determine ν_a and ν_b from (2.4.1). Likewise, relying on the identities in (2.4.2), the approximate primal controls are given by:*

$$\pi_t^* = \left[\min\left\{\left(\widehat{\pi}_{\nu,t}^{\text{opt}}\right)^+, X_t^* + C\right\}, 0, 0\right]^\top \mathcal{K}_{1,t} \quad \text{and} \quad c_t^* = \widehat{c}_{\nu,t}^{\text{opt}} \mathcal{K}_{2,t}. \tag{2.4.18}$$

Let us make a few remarks regarding the previous example.[15] First, we note that the choices for the projection operators and "liquidity-ensuring"

[15] In spite of its absence in Example 2.4.2, the Lagrange multiplier, $\eta^{\text{opt}} = \mathcal{G}^{-1}(X_0)$, also constitutes a part of the approximation on the dual-side, corresponding to the first step (2.4.1). However, as the foregoing function depends on the shadow price process, an approximation to the latter implies one to the former (cf. Corollary 2.4.1). We have therefore opted for an exclusion of an according explication of the approximate Lagrange multiplier. Moreover, we observe in Example 2.4.2's last case that the fraction of labour income that the agent consumes once the time-dependent liquidity constraint binds, $k \in [0,1]$, is not specified. Indeed, in the numerical evaluation, we compute k in a manner such that the approximate primal value function is maximised. One may subsume this operation under the fifth step in section 2.4.1.2.

functions are not unique. For instance, for a situation similar to the last one in Example 2.4.2, Bick et al. (2013) employ a specification of $\mathcal{K}_{2,t}$ that differs from ours. In selecting the operators and functions, we aim for an uncomplicated interpretation of the approximations. Second, we address that closed-form expressions for the approximate primal value functions are not available. This absence of analytical solutions is attributable to the integral expression of approximate wealth in (2.4.3). We must employ simulations to acquire the evolution of approximate wealth over $[0, T]$. As a consequence, the primal value function can only be obtained by means of simulations. Third and last, we would like to stress that the controls in Example 2.4.2 provide closed-form expressions for the solutions to a problem that does ordinarily not allow for a characterisation of optimal rules in closed-form at all.

2.4.3 Evaluation of Approximate Method

To demonstrate the accuracy of our approximate method, we conclude this section with a numerical evaluation. We make use of the framework defined in sections 2.4.2.1 and 2.4.2.2, and compute the upper bounds on the welfare losses, $\mathcal{V}$, for the approximate controls in Example 2.4.2. Based on the initialisation of parameters as reported in Brennan and Xia (2002), Kraft and Munk (2011), and Bick et al. (2013), we fix our model parameters as follows: $C = 0$, $\gamma = \alpha_1 = 5$, $\alpha_2 = 0.5$, $X_0/\bar{\Pi}_0 = 1$, $\beta = 0.03$, $r = 0.01$, $\mu_Y = 0.01$, $\sigma_Y = 0.075$, $Y_0 = \Pi_0 = 1$, $\rho_{SY} = 0.5$, $\mu_\Pi = 0.05$, $\sigma_\Pi = 0.01$, $\mu_i = 0.06$, $\sigma_{ii} = 0.2$, for $i = 1, 2, 3$, where $\bar{\Pi}_0 = \mathbb{E}\left[M_T \Pi_T\right]$. All results are based on Monte-Carlo simulations and an Euler scheme, with $10,000$ sample paths and $T * 10$ time-steps: the approximate strategies are adjusted 10 times a year. More frequent adjustments (such as Bick et al. (2013)'s 20 times a year) do not improve our results significantly. In Tables 2.1, 2.2, and 2.3, we report the annual welfare losses, expressed in terms of the percentages of the investor's initial amount of total wealth, for cases (i), (ii), and (iii) of Example 2.4.2, respectively. We present the bounds for different values of γ, α_1, α_2, $X_0/\bar{\Pi}_0$, and T *ceteris paribus*.

First, we analyse the outcomes for item (i) of Example 2.4.2. Item (i), $K = \mathbb{R}^3 \times \{0\}$, outlines a setup, in which the benchmark process, $\{\Pi_t\}_{t\in[0,T]}$, is (partially) unhedgeable. Note that this case constitutes a special version of the

	Coefficient of risk-aversion (γ)				Risk-aversion parameter (α_1)			
	4	6	8	10	4	6	8	10
Time (T)								
5	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002
10	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002
20	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003
40	0.000	0.001	0.001	0.002	0.001	0.001	0.002	0.002
	Scale parameter (α_2)				Funding ratio ($X_0/\bar{\Pi}_0$)			
	0.1	0.4	0.7	1	0.7	0.85	1	1.15
Time (T)								
5	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
10	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
20	0.002	0.000	0.001	0.002	0.000	0.000	0.000	0.000
40	0.002	0.001	0.002	0.003	0.002	0.001	0.001	0.000

Table 2.1. Upper bounds on annual welfare losses for $K = \mathbb{R}^3 \times \{0\}$. For the first case of Example 2.4.2, the table reports the upper bounds on the annual welfare losses that the agent may incur from implementing the approximate strategy in (2.4.16). These annual welfare losses are expressed in terms of the percentages of the investor's initial amount of total wealth (X_0 and Y_0), and calculated on the basis of (2.4.4) as follows: $[(1+\mathcal{V})^{\frac{1}{T}} - 1] \times 100$. The presented numbers ensue from variation over one of the four included parameters (the values of which are shown in the upper rows of the two panels), for different time-horizons (displayed in the most-left column of the two panels), and a predetermined baseline initialisation of the parameters. This baseline initialisation is fixed in the following manner: $C = 0$, $\gamma = \alpha_1 = 5$, $\alpha_2 = 0.5$, $X_0/\bar{\Pi}_0 = 1$, $\beta = 0.03$, $r = 0.01$, $\mu_Y = 0.01$, $\sigma_Y = 0.075$, $Y_0 = \Pi_0 = 1$, $\rho_{SY} = 0.5$, $\mu_\Pi = 0.05$, $\sigma_\Pi = 0.01$, $\mu_i = 0.06$, $\sigma_{ii} = 0.2$, for $i = 1, 2, 3$. The results are based on $10,000$ simulated paths, wherein the agent adjusts his/her portfolio and consumption behaviour 10 times a year, at equidistant points.

problem in Brennan and Xia (2002), involving labour income and intertemporal consumption, where Π_t is the price index. Table 2.1 reveals that the bounds on the annual welfare losses for the present case vary between 0.000% and 0.003% of the investor's total wealth. For a broad range of values for the agent's risk-profile and the trading-horizon, the welfare losses are lower than 0.001%. As a result of the extended exposure to the unhedgeable risk and the interrelated magnification of inconsistencies inherent in the approximation, increases in the trading horizon (T) give rise to small increases in the (maximal) welfare losses. For larger values of γ and α_1, the size of the bounds increases marginally. This can be explained by the positive dependence of the benchmark hedge demand on the γ and α_1 parameters, implied by equation (2.4.14). When we interpret the reported bounds as annual management fees, or as annual extra returns needed to compensate for the sub-optimal approximate strategies, then it is safe to state that the approximate rules for case (i) of Example 2.4.2

	Coefficient of risk-aversion (γ)				Risk-aversion parameter (α_1)			
	4	6	8	10	4	6	8	10
Time (T)								
5	0.000	0.009	0.014	0.023	0.000	0.001	0.003	0.006
10	0.000	0.007	0.007	0.021	0.000	0.002	0.006	0.008
20	0.000	0.011	0.036	0.051	0.000	0.003	0.008	0.014
40	0.001	0.005	0.018	0.022	0.000	0.001	0.001	0.004
	Scale parameter (α_2)				Funding ratio ($X_0/\bar{\Pi}_0$)			
	0.1	0.4	0.7	1	0.7	0.85	1	1.15
Time (T)								
5	0.005	0.001	0.000	0.001	0.002	0.001	0.000	0.000
10	0.015	0.004	0.000	0.001	0.005	0.002	0.001	0.000
20	0.020	0.004	0.001	0.002	0.004	0.002	0.001	0.000
40	0.019	0.002	0.002	0.003	0.002	0.002	0.001	0.001

Table 2.2. Upper bounds on annual welfare losses for $K = \mathbb{R} \times \mathbb{R}^2_+ \times \{0\}$. For the second case of Example 2.4.2, the table reports the upper bounds on the annual welfare losses that the agent may incur from implementing the approximate strategy in (2.4.17). Consider the description of Table 2.1 for further details on the reported values.

are near-optimal.[16]

Second, we focus on item (ii) in Example 2.4.2. Item (ii), $K = \mathbb{R} \times \mathbb{R}^2_+ \times \{0\}$, implies a setup that is identical to item (i)'s, in which short-selling is prohibited. Table 2.2 discloses that the corresponding approximate strategies result in upper bounds that vary between 0.000% and 0.051% of total wealth. For most of the values of the agent's risk-profile, the reported bounds are even lower than 0.010%. Relative to Table 2.1, we observe that these losses are more pronounced. This can be understood from the fact that the approximate rules are forced to account for an additional constraint. This effect partially carries over to the relationship between the time-horizon and the reported bounds. Indeed, we discern a more noticeable positive link between the former two, up to the $T = 20$ horizon. For the $T = 40$ horizon, the losses converge towards the

[16] Here and in the sequel, we draw conclusions as to the near-optimality of the approximate trading-consumption pairs, which are proposed in Example 2.4.2, on the basis of the annual welfare losses, rather than their non-annualised counterparts. We do so, for the purpose of a fair comparison amongst the reported bounds in the dimension of the time-horizon (T). In the interest of of being able to compute the non-annualised welfare losses, and thus to compare the magnitudes of these upper bounds to the ones that are documented in e.g. Bick et al. (2013), we note that $\mathcal{V} \times 100 = \left[\left(\frac{\mathcal{AL}}{100} + 1\right)^T - 1\right] \times 100 \approx \mathcal{AL} \times T$ spawns the non-annualised equivalents of the annual welfare losses (denoted by $\mathcal{AL}$). Comparable to Bick et al. (2013), we observe that the *absolute* bounds reported in Tables 2.1, 2.2 and 2.3 vary between 0.000% and 1.300%.

	Coefficient of risk-aversion (γ)				Risk-aversion parameter (α_1)			
	4	6	8	10	4	6	8	10
Time (T)								
5	0.008	0.007	0.008	0.008	0.006	0.009	0.011	0.014
10	0.012	0.009	0.011	0.013	0.009	0.011	0.013	0.016
20	0.035	0.008	0.008	0.012	0.009	0.012	0.016	0.020
40	0.037	0.017	0.021	0.030	0.016	0.023	0.026	0.032
	Scale parameter (α_2)				Funding ratio ($X_0/\bar{\Pi}_0$)			
	0.1	0.4	0.7	1	0.7	0.85	1	1.15
Time (T)								
5	0.010	0.008	0.006	0.004	0.008	0.008	0.007	0.007
10	0.015	0.011	0.009	0.008	0.011	0.010	0.010	0.010
20	0.022	0.013	0.009	0.007	0.014	0.012	0.011	0.011
40	0.031	0.015	0.014	0.009	0.019	0.016	0.015	0.013

Table 2.3. Upper bounds on annual welfare losses for $K = [-C, \infty) \times \mathbb{R}_+ \times \{0\}^2$. For the third case of Example 2.4.2, the table reports the upper bounds on the annual welfare losses that the agent may incur from implementing the approximate strategy in (2.4.18). Consider the description of Table 2.1 for further details on the reported values.

values in Table 2.1, which indicates that there is virtually no negative demand for the traded assets, $S_{1,t}$ and $S_{2,t}$.[17] Given the relatively small values for σ_Π, it is logical that increases in γ and α_1 put more emphasis on the negative labour income hedge demand, see (2.4.13) and (2.4.14). Consistent with the resulting enhanced inclination to short-sale, Table 2.2 shows a distinct positive relationship between these parameters and the bounds. Due to SAHARA's ARA function, the opposite reasoning applies to the connection between α_2 and $X_0/\bar{\Pi}_0$, respectively, and the welfare losses. In summary, the findings suggest near-optimality of the approximate investment and consumption policies for case (ii) of Example 2.4.2.

Third and last, we concentrate on item (iii) in Example 2.4.2. The setup implied by item (iii), $K = [0, \infty) \times \mathbb{R}_+ \times \{0\}^2$, is equivalent to item (ii)'s, in which labour income is (partially) unhedgeable and borrowing is prohibited. Observe

[17]Whereas changes in α_1 or γ and their impact on the optimal portfolio can be intuitively explained by relating these to the relevant ARA functions and Corollary 2.4.1, the impact of T on the optimal portfolio is less clear. This is specifically true for a situation in which one is not allowed to short-sell and/or borrow. For example, an increase in T does *not* affect the demand for $S_{3,t}$ in the first case of Example 2.4.2, i.e. $\pi_{3,t}^{\text{opt}} = 0$ must hold regardless. But a similar increase in T may very well affect both the (constrained) demands for $S_{1,t}$ *and* $S_{2,t}$ in the second case of Example 4.2. As a result, it is hard to tell how well the selected projection performs in this instance. It is, therefore, not easy to provide clear intuition for the patterns that are visible in the dimension of T.

that this instance is similar to Cocco et al. (2005)'s investment problem. Table 2.3 shows that the bounds on the annual welfare losses are between 0.004% and and 0.037% of total wealth. Most of the reported bounds are lower than 0.015%. In comparison to the previous tables, the welfare losses for this example are larger. We can explain this increase by the inclusion of two additional trading constraints and a borrowing/liquidity constraint. Also in this case, the welfare losses monotonically increase in the trading horizon. Similar patterns are visible in the dimension of all the risk-profile's parameters, with the exception of γ. Due to the large impact on the portfolio (cf. the bounds under γ in Tables 2.1 and 2.2), and the enforced cap and floor on the portfolio process, it is clear that low and high values for γ lead to infeasible allocations that ought to be adjusted most drastically. This phenomenon results in larger welfare losses for the same coefficients of risk-aversion. In conclusion, the approximation has proven to lead to negligible welfare losses, and corresponding near-optimal closed-form trading strategies, as well for item (iii) in Example 2.4.2. Accordingly, the outcomes exemplify the approximation's potential accuracy, and its stable performance, even when the complexity of the trading constraints increases.[18]

2.5 Conclusion

This chapter has developed a dual-control method for finding approximate closed-form solutions to investment problems in multi-dimensional financial markets with convex trading constraints. This method works as follows: (i) it approximates the optimal shadow price process; (ii) it "projects" the auxiliary-optimal analytical controls, which are implied by (i), into the admissibility region; (iii) it evaluates the accuracy by comparing the primal lower and dual

[18] In Tables 2.1, 2.2 and 2.3, we have deliberately decided on an omission of the standard errors, associated with the documented welfare losses. Namely, the magnitudes of these excluded errors are rather insignificant compared to the reported sizes of the welfare losses, and therefore redundant. As for the errors of the approximation itself, here encapsulated by the magnitudes of the bounds, we note that these arise as a result of two reasons: in line with our essentially twofold procedure, from (i) the approximation of the shadow price on one side, and (ii) the projection of the artificial-optimal rules on the other side. The mechanism does not allow one to disentangle the exact attribution of either of these procedures to the totality of the error – only of these together. Nevertheless, in light of the small-sized welfare losses for the examples of interest, the latter assignment of errors to the total amount is irrelevant.

upper bounds on the optimal value function. Our method differs from the literature, in that its range of application covers setups with (i) general return dynamics, (ii) general liquidity constraints and convex trading restrictions, and (iii) state-dependent utility functions that are (possibly) specified over the entirety of the real line and embed an exogenous stochastic benchmark. To highlight these distinguishing features, we have evaluated the quality of our method in Cocco et al. (2005)'s environment, given both CRRA and SAHARA preferences. For three different sets of the trading constraints, the numerical evaluations have shown that the bounds on the annual welfare losses are always lower than 0.051% of the investor's initial amount of total wealth. Hence, our method is capable of rendering near-optimal policy rules.

Appendix A Proofs

A.1 Proof of Theorem 2.3.1

We use the procedure by Klein and Rogers (2007) and Rogers (2003, 2013), to arrive at a candidate for the dual. To that end, we derive

$$\begin{aligned} \mathrm{d}X_t Z_t = {} & \pi_{0,t} r_t Z_t \mathrm{d}t + \pi_t^\top Z_t \left(\mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t\right) \\ & - \left(c_t - Y_t\right) Z_t \mathrm{d}t + X_t Z_t \left[\alpha_t \mathrm{d}t + \theta_t^\top \mathrm{d}W_t\right] + \pi_t^\top \sigma_t \theta_t Z_t \mathrm{d}t, \end{aligned} \tag{A.1.1}$$

which has the following solution:

$$\begin{aligned} X_T Z_T = {} & X_0 Z_0 - \int_0^T \left(c_t - Y_t\right) Z_t \mathrm{d}t + \int_0^T \pi_{0,t} Z_t \left(r_t + \alpha_t\right) \mathrm{d}t \\ & + \int_0^T \pi_t^\top Z_t \left(\mu_t + \alpha_t 1_N + \sigma_t \theta_t\right) \mathrm{d}t + \int_0^T \left(\pi_t^\top \sigma_t + \theta_t^\top X_t\right) Z_t \mathrm{d}W_t. \end{aligned} \tag{A.1.2}$$

To simplify the latter expression, without loss of generality, we write α_t and θ_t for some $\nu_t = [\nu_{0,t}, \nu_{N,t}] \in \mathbb{D}^{1,2}\left([0,T]\right)^{N+1}$ as follows:

$$\begin{aligned} \alpha_t &= -\nu_{0,t} - r_t, \\ \theta_t &= -\lambda_t - \sigma_t^{-1}\left(\nu_{N,t} - \nu_{0,t} 1_N\right). \end{aligned} \tag{A.1.3}$$

Then, from (A.1.2), we are able to deduce:

$$\begin{aligned}\mathbb{E}\Bigg[-X_T Z_T - \int_0^T \left(c_t - Y_t\right) Z_t \mathrm{d}t \\ - \int_0^T \left(\pi_{0,t}\nu_{0,t} + \pi_t^\top \nu_{N,t}\right) Z_t \mathrm{d}t \Bigg] + X_0 Z_0 = 0.\end{aligned} \tag{A.1.4}$$

To assemble a Lagrangian functional for the problem in (2.2.7), we simply include the latter identity into the value function. Let $\mathcal{L}$ denote the Lagrangian functional. Then, it can be shown that $\mathcal{L}$ equates to:

$$\begin{aligned}\mathcal{L} = \mathbb{E}\Bigg[\int_0^T u\left(t, c_t, \Pi_{1,t}\right) \mathrm{d}t + U\left(X_T, \Pi_{2,T}\right) - \int_0^T c_t Z_t \mathrm{d}t \\ - X_T Z_T - \int_0^T \left(\pi_{0,t}\nu_{0,t} + \pi_t^\top \nu_{N,t}\right) Z_t \mathrm{d}t + \int_0^T Y_t Z_t \mathrm{d}t \Bigg] + X_0 Z_0.\end{aligned} \tag{A.1.5}$$

Optimising the Lagrangian, $\mathcal{L}$, over all c_t and X_T results in the following two first order conditions (FOCs):

$$\begin{aligned} u'_X\left(t, c_t, \Pi_{1,t}\right) - Z_t = 0, \\ U'_X\left(X_T, \Pi_{2,T}\right) - Z_T = 0.\end{aligned} \tag{A.1.6}$$

These equations are solved by c_t^{opt} and X_T^{opt}. Consequently:

$$\begin{aligned}\mathcal{L} = \mathbb{E}\Bigg[\int_0^T v\left(t, Z_t, \Pi_{1,t}\right) \mathrm{d}t + V\left(Z_T, \Pi_{2,T}\right) \\ - \int_0^T \left(\pi_{0,t}\nu_{0,t} + \pi_t^\top \nu_{N,t}\right) Z_t \mathrm{d}t + \int_0^T Y_t Z_t \mathrm{d}t \Bigg] + X_0 Z_0.\end{aligned} \tag{A.1.7}$$

Then, we optimise $\mathcal{L}$ over all $\{\pi_{0,t}, \pi_t\}_{t\in[0,T]} \in K$. This results in $\mathcal{L} = \infty$, unless the CS condition holds true:

$$\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}. \tag{A.1.8}$$

As a result, for $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$:

$$\sup_{\{\pi_{0,t},\pi_t,c_t\}_{t\in[0,T]}\in\widehat{\mathcal{A}}_{X_0}} \mathcal{L} = J_D\left(\bar{X}_0, Z_0, \{\nu_t\}_{t\in[0,T]}\right). \tag{A.1.9}$$

Minimising the latter expression over all $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ and $Z_0 \in \mathbb{R}_+$ yields (2.3.2). Therefore, it is easy to show that optimal $Z_0 \in \mathbb{R}_+$ reads:

$$Z_0^{\mathrm{opt}} = \mathcal{G}^{-1}\left(X_0\right) \tag{A.1.10}$$

For proofs that the candidate in (2.3.3) is the legitimate dual of the primal problem (2.2.5), we refer the reader to, for example, Karatzas, Žitković, et al. (2003), and Hugonnier, Kramkov, et al. (2004). The derivations in these papers can easily be adapted to our mildly adjusted environment.

A.2 Proof of Proposition 2.3.2

It is clear that

$$\inf_{Z_0\in\mathbb{R}_+} J_D\left(\bar{X}_0, Z_0, \{\nu_t\}_{t\in[0,T]}\right) = J_{\widehat{\mathcal{M}}_\nu}\left(\bar{X}_0, \{\nu_t\}_{t\in[0,T]}\right), \tag{A.2.11}$$

holds for all $\{\nu_t\}_{t\in[0,T]} \in \mathcal{H}_{\widehat{\mathcal{A}}_{X_0}}$ and $\bar{X}_0 \in \mathbb{R}_+^2$. Namely, the latter minimisation procedure results in $Z_0^{\mathrm{opt}} = \mathcal{G}^{-1}\left(X_0\right)$. The previous definition eliminates the budget constraint from the specification of J_D, which is given by:

$$-\mathbb{E}\left[\int_0^T \left(c_{\nu,t} - Y_{\nu,t}\right) Z_{\nu,t}\mathrm{d}t + X_{\nu,T}Z_{\nu,T}\right] + X_0. \tag{A.2.12}$$

An alternative approach is to note that

$$\begin{aligned} \sup_{(X_{\nu,T},c_{\nu,t})\in\widehat{L}^2(\Omega\times[0,T])} \quad & \mathbb{E}\left[\int_0^T u\left(t, c_{\nu,t}, \Pi_{1,t}\right)\mathrm{d}t + U\left(X_{\nu,T}, \Pi_{2,T}\right)\right] \\ \text{s.t.} \quad & \mathbb{E}\left[\int_0^T \left(c_{\nu,t} - Y_{\nu,t}\right) Z_{\nu,t}\mathrm{d}t + X_{\nu,T}Z_{\nu,T}\right] \leq X_0, \end{aligned} \tag{A.2.13}$$

generates a minimax formulation inherent in its Lagrangian functional that is identical to the dual formulation central to this proof.

A.3 Proof of Proposition 2.3.3

We first solve (2.3.9) by means of standard Lagrangian methods. To this end, introduce $\eta \in \mathbb{R}_+$ as the Lagrange multiplier. Fix $\mathcal{L} := \mathcal{L}(X_{\nu,T}, c_{\nu,t}, \eta)$. Then, the Lagrangian, $\mathcal{L} : \widehat{L}^2(\Omega \times [0,T]) \times \mathbb{R}_+^2 \to \mathbb{R}$, reads as:

$$\begin{aligned}\mathcal{L} = \mathbb{E}\Bigg[&\int_0^T u\left(t, c_{\nu,t}, \Pi_{1,t}\right) \mathrm{d}t + U\left(X_{\nu,T}, \Pi_{2,T}\right) \\ -\eta &\left(\int_0^T c_{\nu,t} Z_{\nu,t} \mathrm{d}t + X_{\nu,T} Z_{\nu,T} - X_0 - \int_0^T Y_{\nu,t} Z_{\nu,t} \mathrm{d}t \right) \Bigg].\end{aligned} \tag{A.3.14}$$

We can compute the following two Fréchet derivatives, cf. Definition 5 in Battauz et al. (2015a):

$$\begin{aligned} D_{X_{\nu,T}} \mathcal{L} \xi_1 &= \langle U_X' \left(X_{\nu,T}, \Pi_{2,T}\right) - \eta Z_{\nu,T}, \xi_1 \rangle_{L^2(\Omega)} \\ &= \mathbb{E}\left[\left(U_X' \left(X_{\nu,T}, \Pi_{2,T}\right) - \eta Z_{\nu,T} \right) \xi_1 \right] = 0, \end{aligned} \tag{A.3.15}$$

and

$$\begin{aligned} D_{c_{\nu,t}} \mathcal{L} \xi_2 &= \langle u_X' \left(t, c_{\nu,t}, \Pi_{1,t}\right) - \eta Z_{\nu,t}, \xi_2 \rangle_{L^2(\Omega \times [0,T])} \\ &= \mathbb{E}\left[\int_0^T \left(u_X' \left(t, c_{\nu,t}, \Pi_{1,t}\right) - \eta Z_{\nu,t} \right) \xi_{2,t} \mathrm{d}t \right] = 0, \end{aligned} \tag{A.3.16}$$

for all $\xi_1 \in L^2(\Omega)$ and $\xi_2 \in L^2(\Omega \times [0,T])$, in which $D_{X_\nu} \mathcal{L} : L^2(\Omega) \to \mathbb{R}$ and $D_{c_{\nu,t}} \mathcal{L} : L^2(\Omega \times [0,T]) \to \mathbb{R}$ specify the Fréchet derivatives in the $X_{\nu,T}$-direction and in the $c_{\nu,t}$-direction, respectively.

Solving (A.3.15) and (A.3.16) for $X_{\nu,T}$ and $c_{\nu,t}$, yields us the following

$$\begin{aligned} c_{\nu,t} &= \iota\left(t, \eta Z_{\nu,t}, \Pi_{1,t}\right), \\ X_{\nu,T} &= I\left(\eta Z_{\nu,T}, \Pi_{2,T}\right). \end{aligned} \tag{A.3.17}$$

These are the stationary points of $\mathcal{L}$. Here, the corresponding value for optimal η, i.e. $\eta^{\text{opt}} = \mathcal{G}^{-1}(X_0)$, follows from

$$\begin{aligned} D_\eta \mathcal{L} \xi_3 = -\Bigg\langle \int_0^T c_{\nu,t} Z_{\nu,t} \mathrm{d}t + X_{\nu,T} Z_{\nu,T} - X_0 \\ - \int_0^T Y_{\nu,t} Z_{\nu,t} \mathrm{d}t, \xi_3 \Bigg\rangle_{L^2(\Omega)} = -\mathcal{G}(\eta) + X_0 = 0, \end{aligned} \tag{A.3.18}$$

for all $\xi_3 \in \mathbb{R}_+$, where $D_\eta \mathcal{L} : \mathbb{R}_+ \to \mathbb{R}$ is $\mathcal{L}$'s Fréchet derivative in the η-direction. Note that these identities hold for all $\omega \in \Omega$ and generate (2.3.10).

Now, using the Clark-Ocone formula, we derive:

$$\begin{aligned} \pi^{\text{opt}}_{\nu,t} = \sigma_t^{\top^{-1}} \Bigg\{ \mathbb{E} \Bigg[\mathcal{D}^W_t \Bigg(X^{\text{opt}}_{\nu,T} \frac{Z_{\nu,T}}{Z_{\nu,t}} \\ + \int_t^T \left(c^{\text{opt}}_{\nu,s} - Y_{\nu,s} \right) \frac{Z_{\nu,s}}{Z_{\nu,t}} \mathrm{d}s \Bigg) \Bigg| \mathcal{F}_t \Bigg] + \widehat{\lambda}_t X^{\text{opt}}_{\nu,t} \Bigg\}, \end{aligned} \tag{A.3.19}$$

for all $t \in [0,T]$. The general result underscoring this identity can be found in the study by Ocone and Karatzas (1991). The four hedge demands follow from a straightforward expansion of the Malliavin derivative in the same equation (A.3.19) for $\pi^{\text{opt}}_{\nu,t}$, i.e. $\mathcal{D}^W_t \{X^{\text{opt}}_{\nu,T} Z_{\nu,T} + \int_0^T (c^{\text{opt}}_{\nu,s} - Y_{\nu,s}) Z_{\nu,s} \mathrm{d}s\}$.

A.4 Derivation of (2.3.14)

Minimisation of J_D over $\nu_{N,M,t} \in \mathbb{D}^{1,2}([0,T])^N$ yields

$$D_{\{\nu_{N,M,t}\}} J_D \{\epsilon_t\} = \mathbb{E} \left[X^{\text{opt}}_{\nu,T} Z_T \mathcal{P}_{\epsilon,T} + \int_0^T \left(c^{\text{opt}}_{\nu,t} - Y_{\nu,t} \right) Z_t \mathcal{P}_{\epsilon,t} \mathrm{d}t \right], \tag{A.4.20}$$

which describes for all $\epsilon_t \in \mathbb{D}^{1,2}([0,T])^M$ the Fréchet derivative of $J_D : \mathbb{R}^2_+ \times \mathbb{R}_+ \times \mathbb{D}^{1,2}([0,T])^N \to \mathbb{R}$ in the $\nu_{N,M,t}$-direction, such that $D_{\{\nu_{N,M,t}\}} J_D : \mathbb{D}^{1,2}([0,T])^M \to \mathbb{R}$. Here, we define for all $t \in [0,T]$:

$$\mathcal{P}_{\epsilon,t} := \int_0^t \epsilon_s^\top \sigma_{2,s}^{-1^\top} \widehat{\lambda}_{M,s} \mathrm{d}s + \int_0^t \epsilon_s^\top \sigma_{2,s}^{-1^\top} \mathrm{d}W_{M,s}. \tag{A.4.21}$$

By means of Skorokhod's duality result, cf. Chapter 4.41 of Rogers and Williams, 2000, we find that:

$$\mathbb{E}\left[X^{\mathrm{opt}}_{\nu,T} Z_T \zeta\left(\sigma_2^{-1}\epsilon\right)\right] = \mathbb{E}\left[\int_0^T \epsilon_t^\top \sigma_{2,t}^{-1\top}\, \mathbb{E}\left[\mathcal{D}_t^{W_M} X^{\mathrm{opt}}_{\nu,T} Z_T \,\middle|\, \mathcal{F}_t\right] \mathrm{d}t\right], \quad \text{(A.4.22)}$$

where $\zeta(\cdot)$ specifies the divergence operator or Skorokhod integral. Note that the Skorokhod integral is defined as:

$$\zeta\left(\sigma_2^{-1}\epsilon\right) = \int_0^T \epsilon_s^\top \sigma_{2,s}^{-1\top}\, \mathrm{d}W_{M,s}. \quad \text{(A.4.23)}$$

Likewise, after changing the order of integration,

$$\mathbb{E}\left[\int_0^T c^{\mathrm{opt}}_{\nu,Y,t} Z_t \zeta\left(\widehat{\epsilon}\right) \mathrm{d}t\right] = \mathbb{E}\left[\int_0^T \int_0^t \widehat{\epsilon}_s^\top\, \mathbb{E}\left[\mathcal{D}_s^{W_M} c^{\mathrm{opt}}_{\nu,Y,t} Z_t \,\middle|\, \mathcal{F}_s\right] \mathrm{d}s\mathrm{d}t\right], \quad \text{(A.4.24)}$$

for all $\epsilon_t \in \mathbb{D}^{1,2}\left([0,T]\right)^M$ in which $c^{\mathrm{opt}}_{\nu,Y,t} = c^{\mathrm{opt}}_{\nu,t} - Y_t$ and $\widehat{\epsilon} = \sigma_2^{-1}\epsilon$.

Combining results and solving

$$D_{\{\nu_{N,M,t}\}} J_D\left\{\epsilon_t\right\} = 0, \quad \text{(A.4.25)}$$

for all $\epsilon_t \in \mathbb{D}^{1,2}\left([0,T]\right)^M$, we find the first-order condition in (2.3.14) for all $t \in [0,T]$. Note that, as for $\mathcal{L}$ in Proposition 2.3.3, J_D lays out stationary points at which $D_{\{\nu_{N,M,t}\}} J_D\left\{\epsilon_t\right\} = 0$ holds for all $\epsilon_t \in \mathbb{D}^{1,2}\left([0,T]\right)^M$: convexity arguments suffice to conclude on the optimality of $\{\widehat{\lambda}_{M,t}\}_{t\in[0,T]}$.

A.5 Proof of Corollary 2.4.1

Noting that $\iota\left(t,x,y\right) = \left(e^{\beta t}xy\right)^{-\frac{1}{\gamma}} y$, (2.4.12) follows from (2.3.10). Regarding $\widehat{\pi}^{\mathrm{opt}}_{\nu,t}$, consider (2.3.12), and observe that the following holds:

$$\begin{aligned}
\frac{1}{\mathcal{R}_{1,c,t}} &= \frac{1}{\gamma}\widehat{c}^{\mathrm{opt}}_{\nu,t} = \frac{1}{\gamma} e^{-\frac{1}{\gamma}\beta t}\left(\eta^{\mathrm{opt}} Z_{\nu,t}\right)^{-\frac{1}{\gamma}} \Pi_t^{1-\frac{1}{\gamma}}, \\
\frac{1}{\mathcal{R}_{1,X,T}} &= \frac{\alpha_2}{\alpha_1}\cosh\left(\frac{1}{\alpha_1}\log\left(\eta^{\mathrm{opt}} Z_T \Pi_T\right) + \log\left(\alpha_2\right)\right)\Pi_T.
\end{aligned} \quad \text{(A.5.26)}$$

It is then straightforward to show that

$$\begin{aligned}\mathbb{E}\left[\int_t^T \frac{1}{\mathcal{R}_{1,c,s}}\frac{Z_{\nu,s}}{Z_{\nu,t}}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right] &= \widehat{c}_{\nu,t}^{\mathrm{opt}} f(t,T), \\ \mathbb{E}\left[\frac{1}{\mathcal{R}_{1,X,T}}\frac{Z_{\nu,T}}{Z_{\nu,t}} \,\middle|\, \mathcal{F}_t\right] &= \frac{1}{\alpha_1}\left[\alpha_{2,t,T}^2 + \mathcal{X}_{t,T}^2\right]^{\frac{1}{2}} \Pi_t.\end{aligned} \tag{A.5.27}$$

Now, consider (2.3.13), and note that the following holds: $\mathcal{R}_{2,c,t} = -(1 - \frac{1}{\gamma})\widehat{c}_{\nu,t}^{\mathrm{opt}}\Pi_t^{-1}$. In addition to this, we are able to derive that

$$\begin{aligned}\mathcal{R}_{2,X,t} = -\frac{1}{2}\Bigg[&\left(1-\frac{1}{\alpha_1}\right)\left(\eta^{\mathrm{opt}} Z_T \Pi_T\right)^{-\frac{1}{\alpha_1}} \\ &-\alpha_2^2\left(1+\frac{1}{\alpha_1}\right)\left(\eta^{\mathrm{opt}} Z_T \Pi_T\right)^{\frac{1}{\alpha_1}}\Bigg] - 1,\end{aligned} \tag{A.5.28}$$

and $\mathcal{D}_t^W \Pi_{1,s} = \mathcal{D}_t^W \Pi_{2,s} = \mathcal{D}_t^W \Pi_s = \sigma_\Pi \Pi_s$ for all $s \geq t$. Then,

$$\begin{aligned}\mathbb{E}\left[\int_t^T \frac{Z_{\nu,s}}{Z_{\nu,t}}\mathcal{R}_{2,c,s}\mathcal{D}_t^W \Pi_{1,s}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right] &= -\left[1-\frac{1}{\gamma}\right]\widehat{c}_{\nu,t}^{\mathrm{opt}} f(t,T)\,\sigma_\Pi, \\ \mathbb{E}\left[\frac{Z_{\nu,T}}{Z_{\nu,t}}\mathcal{R}_{2,X,T}\mathcal{D}_t^W \Pi_{2,T} \,\middle|\, \mathcal{F}_t\right] &= \left(\left[1-\frac{1}{\alpha_1}\right]\widehat{\mathcal{X}}_{t,T}\Pi_t + \Pi_t e^{h(0,T,0)}\right)\sigma_\Pi.\end{aligned} \tag{A.5.29}$$

hold, yielding $\widehat{\pi}_t^{\Pi}$. Finally, one is able to show that $\mathcal{D}_t^W \mu_{Y,u} = 0_3$, $\mathcal{D}_t^W \sigma_{Y,u} = 0_{3\times 3}$ and $\mathbb{E}[\int_t^T \frac{Z_{\nu,s}}{Z_{\nu,t}} Y_s \mathrm{d}s \mid \mathcal{F}_t] = Y_t g(t,T)$ hold true. These results render the demand corresponding to the agent's labour income, $\widehat{\pi}_t^Y$. We stress that

$$\begin{aligned}\mathcal{D}_t^W \log \widehat{Z}_{\nu,s} &= \mathcal{D}_t^W \log Z_{\nu,s} + \widehat{\lambda}_t \\ &= \int_t^s \left[-\mathcal{D}_t^W r_u - \mathcal{D}_t^W \nu_{0,u}\right]\mathrm{d}u \\ &\quad - \int_t^s \mathcal{D}_t^W \widehat{\lambda}_u \mathrm{d}\left(W_u + \widehat{\lambda}_u\right) = 0_3,\end{aligned} \tag{A.5.30}$$

such that

$$\pi_t^Z = {\sigma_t^\top}^{-1}\mathbb{E}\left[\frac{1}{\widehat{\mathcal{R}}_{1,X,T}}\frac{Z_{\nu,T}}{Z_{\nu,t}}D_{Z,t,T} + \int_t^T \frac{1}{\widehat{\mathcal{R}}_{1,c,s}}\frac{Z_{\nu,s}}{Z_{\nu,t}}D_{Z,t,s}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right] = 0_3. \tag{A.5.31}$$

3

Dual Formulation of the Optimal Consumption Problem with Multiplicative Habit Formation

Adapted from: Kamma, T., & Pelsser, A. (2022a). Dual formulation of the optimal consumption problem with multiplicative habit formation. *Working paper.*

Abstract

This chapter provides a dual formulation of the optimal consumption problem with internal multiplicative habit formation. In this problem, the agent derives utility from the ratio of consumption to the internal habit component. Due to this multiplicative specification of the habit model, the optimal consumption problem is not strictly concave and incorporates irremovable path-dependency. As a consequence, standard Lagrangian techniques fail to supply a candidate for the corresponding dual formulation. Using Fenchel's Duality Theorem, we manage to identify a candidate formulation and prove that it satisfies strong duality. On the basis of this strong duality result, we develop an evaluation mechanism to measure the accuracy of analytical or numerical approximations to the optimal solutions.

3.1 Introduction

Habit formation describes the phenomenon of an individual growing accustomed to a certain standard of living. In a financial context, this standard of living is dependent on a person's past decisions with regard to saving and consumption. Consuming more or less than a person-specific living standard may impact the utility levels of an individual, cf. Kahneman and Tversky (1979). It is therefore plausible that habit formation affects the current consumption behaviour of a person. To model and analyse the impact of habit-forming tendencies on this behaviour, a wide variety of studies have investigated optimal consumption problems that incorporate a habit level, representing the agent's living standard. These studies can be distinguished into two categories: (i) those that focus on additive habits, and (ii) those that concentrate on multiplicative habits.

We start by discussing the additive habits. In optimal consumption problems with additive habits, the utility-maximising individual draws utility from the *difference* between consumption and a habit level. The literature on these habits is pioneered by Constantinides (1990) and has been studied by e.g. Detemple and Zapatero (1991), Campbell and Cochrane (1999), Munk (2008), Muraviev (2011), and Yu (2015). Additive habit models typically employ arithmetic habit levels, which monotonically increase over time, cf. Detemple and Karatzas (2003), Bodie et al. (2004), and Polkovnichenko (2007). Furthermore, as most standard utility functions only admit strictly positive arguments, additive habit specifications force the agent to maintain consumption above the habit level. For this reason, the habit component is sometimes interpreted as a subsistence level, see e.g. Yogo (2008). This interpretation is sensible for exogenous habits. However, if we assume that habits are endogenous, the habit level depends on the individual's past decisions and becomes person-specific. Consequently, for endogenous habits, it is hard not to consider the habit component as a standard of living that increases over time.

Although individuals have a natural incentive to maintain consumption at least above their living standard, it is clear that additive habit models are too restrictive to be realistic. We attribute this restrictiveness to two main reasons. First of all, in practice, adverse changes in the financial circumstances can urge people to scale down consumption below the level to which they have become accustomed. Second, because of the latter phenomenon, an individual's standard of living may decrease over the course of a lifetime. To arrive at a more realistic model setup that manages to deal with the preceding two situations, the following two modifications can be made. As for the possibility of a declining standard of living, one can employ a geometric specification of the habit level, cf. Kozicki and Tinsley (2002), Corrado and Holly (2011), and van Bilsen et al. (2020a). Unlike the arithmetic habit levels, this geometric specification relies on the logarithmic transformation of consumption, and can therefore decrease over time. As for the possibility of scaling down consumption below the habit level, one can make use of multiplcative habit models.

We now continue with a discussion of the multiplicative habits. Optimal consumption problems with multiplicative habit formation assume that the utility-maximising individual derives utility from the *ratio* of consumption to a habit level. The specification of these habits dates back to Abel (1990), and has been economically advocated by Carroll (2000) and Carroll et al. (2000). Contrary to the additive case, consumption is in this multiplicative setup not constrained to achieve values above the habit level. Namely, since the ratio of consumption to the habit level is always strictly positive, it can be included as an argument in all standard utility functions. The multiplicative habit model consequently allows the agent to reduce consumption levels below the habit component. Furthermore, the multiplicative habit model endows the utility-maximising agent with a strong incentive to fix consumption near/above the habit level. This incentive is due to the fact that the utility function of the agent increases with the magnitude of the ratio.

When the habit level is endogenously determined (internal), standard solution techniques generally fail to solve optimal consumption problems with multiplicative habit formation in closed-form. Because of its dependence on past consumption decisions, the habit component gives rise to path-dependency in the objective function. This path-dependency is irremovable and cannot be handled in an analytical manner.[1] Due to the structure of multiplicative habits, the optimal consumption problem is not strictly concave. In general,

[1] This analytical intractability is unique to problems involving multiplicative habits. In case of additive habits, the path-dependency can be eliminated from the problem, cf. Schroder and Skiadas (2002).

non-concave optimisation problems are more difficult to solve than concave ones, see e.g. Chen et al. (2019). To be able to analyse the corresponding optimal solutions, the general approach is to fall back on (i) numerical routines, (ii) approximations or (iii) duality techniques. In a discrete-time setup, Fuhrer (2000) and Gomes and Michaelides (2003) employ numerical methods to analyse the internal multiplicative habit model. More recently, in a continuous-time setup, van Bilsen et al. (2020a) and Li et al. (2021) have made use of an approximation and numerical routines, respectively.[2] Although these studies provide valuable insights into the (optimal) solutions, they ignore potential benefits and insights from duality approaches. In fact, to the best of our knowledge, a dual formulation for the multiplicative habit model is not known.

In this chapter, we provide a dual formulation of the optimal consumption problem with internal multiplicative habit formation. We derive this formulation in a continuous-time setup with power utility and a finite trading-horizon. The habit level of the utility-maximising individual is assumed to live by a geometric form. The conventional Lagrangian methods for obtaining dual formulations, e.g. those in Klein and Rogers (2007) and Rogers (2003, 2013), are unable to supply a dual for this multiplicative habit problem. Namely, due to the fact that the problem is non-concave and involves path-dependency, the ordinary Legendre transform fails to establish the necessary conjugacy properties. Therefore, we resort to Fenchel's Duality Theorem and a change of variables to derive a dual formulation and prove that strong duality holds. Inspired by Bick et al. (2013) and Kamma and Pelsser (2022c), we make use of this strong duality result to develop an evaluation mechanism, suitable for quantifying the accuracy of analytical or numerical approximations to the optimal solutions. This evaluation mechanism spawns a hard upper bound on the welfare losses associated with the approximations, and requires little to no numerical effort. For the approximation developed by van Bilsen et al. (2020a), we employ this mechanism to show that the corresponding welfare losses can

[2]We exclusively mention studies that focus on the consumption problem with internal multiplicative habits. Problems involving external habit formation, see e.g. Carroll et al. (1997), Chan and Kogan (2002) and Gómez et al. (2009), do not pose issues when it comes to deriving optimal (duality) results. Martingale duality techniques, developed in the seminal contributions by Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991), suffice to analytically solve these consumption problems.

be very small.

The remainder of the chapter is organised as follows. Section 3.2 introduces the model setup and the optimal consumption problem. Section 3.3 presents our main result: the dual formulation. We divide this section into three parts. In the first part, we provide the dual and a rough sketch of its proof. In the second part, we address some technical features of the dual. In the third part, we comment on particular implications of the strong duality result. Subsequently, section 3.4 outlines the evaluation mechanism and provides some numerical results. Appendix A contains the proof of our main duality result.

3.2 Model Setup

In this section, we introduce the model setup. First, we lay out the financial market model. Second, we define the agent's wealth process. Third, we specify the agent's habit level. Fourth, we outline the optimal consumption problem.

3.2.1 Financial Market Model

Our financial market model is N-dimensional, defined in continuous-time, and based on the economic environments provided in Detemple and Rindisbacher (2010), van Bilsen et al. (2020a) and Kamma and Pelsser (2022c). We define $T > 0$ as the finite trading or planning horizon, and $[0, T]$ as the corresponding trading interval. Moreover, we introduce the complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. The components of this space live by their typical definitions, and its randomness is generated by an $\mathbb{R}^N$-valued standard Brownian motion, $\{W_t\}_{t\in[0,T]}$. As of now, all (in)equalities between random variables and stochastic processes are understood in a $\mathbb{P}$-a.s. or a $\mathrm{d}t \otimes \mathbb{P}$-a.e. sense.

The financial market, $\mathcal{M}$, contains a money market account and N risky assets that are represented by N semi-martingale processes. The money market account submits to the following ordinary differential equation (ODE):

$$\frac{\mathrm{d}B_t}{B_t} = r_t \mathrm{d}t, \ B_0 = 1. \tag{3.2.1}$$

Here, r_t represents the $\mathbb{R}$-valued instantaneous interest rate. We assume that r_t is $\mathcal{F}_t$-progressively measurable and fulfills $\int_0^T |r_t| \, \mathrm{d}t < \infty$. The price processes for the N risky assets (stocks) evolve according to the following stochastic differential equation (SDE) for all $i = 1, \ldots, N$:

$$\frac{\mathrm{d}S_{i,t}}{S_{i,t}} = \mu_{i,t}\mathrm{d}t + \sigma_{i,t}^\top \mathrm{d}W_t, \ S_{i,0} = 1, \tag{3.2.2}$$

where $\mu_{i,t}$ denotes the $\mathbb{R}$-valued instantaneous expected return on stock i and $\sigma_{i,t}$ the $\mathbb{R}^N$-valued vector containing the volatility processes for stock i, both of which are $\mathcal{F}_t$-progressively measurable. We postulate that $\int_0^T \|\mu_t\|_{\mathbb{R}^N} \, \mathrm{d}t < \infty$ and $\int_0^T \mathrm{Tr}(\sigma_t \sigma_t^\top)\mathrm{d}t < \infty$, in which $\mu_t \in \mathbb{R}^N$ has entries $\mu_{i,t}$, and $\sigma_t \in \mathbb{R}^{N \times N}$ rows $\sigma_{i,t}$, $i = 1, \ldots, N$. Observe here that $\|\cdot\|_{\mathbb{R}^N}$ denotes the N-dimensional Euclidean norm and that $\mathrm{Tr}(\cdot)$ represents the trace operator. To ensure invertibility of σ_t, we assume that σ_t is non-singular.

Due to the absence of trading restrictions, this financial market is complete, i.e. all traded risks are hedgeable. Hence, by the fundamental theorem of asset pricing, as formulated by Delbaen and Schachermayer (1994), there must exist a unique equivalent martingale measure. Correspondingly, there must exist a unique state price density (SPD), $\{M_t\}_{t \in [0,T]}$. Define $\lambda_t := \sigma_t^{-1}(\mu_t - r_t 1_N)$ as the market price of risk, then M_t reads:

$$\frac{\mathrm{d}M_t}{M_t} = -r_t \mathrm{d}t - \lambda_t^\top \mathrm{d}W_t, \ M_0 = 1. \tag{3.2.3}$$

Note that $\{B_t\}_{t \in [0,T]}$ is selected as the numéraire quantity. We assume that $\{\lambda_t\}_{t \in [0,T]}$ satisfies $\mathbb{E}[\exp(\frac{1}{2}\int_0^T \|\lambda_s\|_{\mathbb{R}^N}^2 \mathrm{d}s)] < \infty$, cf. Karatzas and Shreve (2012). Moreover, we postulate that $\{\lambda_t\}_{t \in [0,T]}$ is such that M_t and $\log M_t$ attain values in $L^2(\Omega \times [0,T])$.[3] The latter assumption is necessary to assure well-posedness of the dual formulation. In order to evaluate financial instruments in a risk-neutral fashion, one can make use of M_t. For example, $M_t B_t$ and $M_t S_t$ are both $\mathbb{P}$-martingales with respect to $\{\mathcal{F}_t\}_{t \in [0,T]}$.

[3] We define $L^p(\Omega \times [0,T]; \mathbb{R}^n)$ as the standard Lebesgue space of all $\mathcal{F}_t$-progressively measurable functions, $f : \Omega \times [0,T] \to \mathbb{R}^n$, satisfying $\left(\int_{\Omega \times [0,T]} \|f_t\|_{\mathbb{R}^n}^p \, \mathbb{P}(\mathrm{d}t)\right)^{1/p} = \left(\mathbb{E}\left[\int_0^T \|f_t\|_{\mathbb{R}^n}^p \, \mathrm{d}t\right]\right)^{1/p} < \infty$. If $n = 1$, we drop the "$\mathbb{R}$"-notation from the definition of the L^p space.

3.2.2 Dynamic Wealth Process

In this environment, the agent is free to continuously select an investment and a consumption strategy over $[0, T]$. Specifically, the agent's wealth process, $\{X_t\}_{t\in[0,T]}$, is affected by two endogenous terms: (i) a process for the proportion of wealth that is allocated to the stock, $\{\pi_t\}_{t\in[0,T]}$, and (ii) a consumption process, $\{c_t\}_{t\in[0,T]}$. We assume that both preceding endogenous processes are $\mathcal{F}_t$-progressively measurable. Let us fix a deterministic initial endowment, $X_0 \in \mathbb{R}_+$. Then, the agent's wealth process is defined by:

$$\mathrm{d}X_t = X_t \left[\left(r_t + \pi_t^\top \sigma_t \lambda_t \right) \mathrm{d}t + \pi_t^\top \sigma_t \mathrm{d}W_t \right] - c_t \mathrm{d}t, \tag{3.2.4}$$

Clearly, $\{c_t\}_{t\in[0,T]}$ is $\mathbb{R}_+$-valued and $\{\pi_t\}_{t\in[0,T]}$ is $\mathbb{R}^N$-valued. A trading-consumption pair, $\{c_t, \pi_t\}_{t\in[0,T]}$, is said to be *admissible* if it satisfies the following set of conditions: $X_t \geq 0$, $\int_0^T \pi_t^\top \sigma_t \sigma_t^\top \pi_t \mathrm{d}t < \infty$, $\int_0^T \left| \pi_t^\top \sigma_t \lambda_t + r_t X_t \right| \mathrm{d}t < \infty$, and $\log c_t \in L^2 (\Omega \times [0, T])$. The set containing all admissible trading-consumption pairs is denoted by $\mathcal{A}_{X_0}$. Observe that the proportion of wealth that is allocated to the cash account can be recovered from $1 - \pi_t^\top 1_N$, where 1_N is an $\mathbb{R}^N$-valued vector containing only 1's. This specific proportion only plays a role through $\{\pi_t\}_{t\in[0,T]}$, due to which it can be excluded from $\{X_t\}_{t\in[0,T]}$. See e.g. Cuoco (1997) for a situation in which this is not the case.

3.2.3 Habit Level

The economic environment $\mathcal{M}$ consists of a utility-maximising agent who is internally habit-forming. As a consequence, the individual is in possession of a habit level, h_t at time $t \in [0, T]$. This habit level represents the level of consumption to which the agent has become accustomed. Naturally, h_t depends on the agent's preferences and his/her corresponding past consumption behaviour. Due to this dependence on past consumption decisions, the habit level constitutes an endogenous (internal) component. If h_t is exogenously determined ($\beta = 0$ below), the agent is externally habit-forming. By analogy with van Bilsen et al. (2020a) and references therein, we suppose that:

$$\mathrm{d} \log h_t = \left(\beta \log c_t - \alpha \log h_t \right) \mathrm{d}t, \;\; \log h_0 = 0. \tag{3.2.5}$$

The parameter $\beta \in \mathbb{R}_+$ expresses the relative importance of past consumption decisions in the specification of $\log h_t$. For large values of β, more weight is attached to these past consumption choices. For small values of β, the converse is true. The parameter $\alpha \in \mathbb{R}_+$ stands for the habit level's rate of depreciation. For small values of α, the habit level depends on past consumption decisions over a large time-horizon. For large values of α, the converse is true. We assume that $\alpha \geq \beta$ holds, for concavity purposes related to the optimal consumption problem. The limiting case $\alpha = \beta = 0$ results in $h_t = 1$ for all $t \in [0, T]$. Setting $\alpha = \beta = 0$ consequently recovers a model without habit formation.

We note that the solution to the ODE in (3.2.5) reads for all $t \in [0, T]$ as:

$$\log h_t = \beta \int_0^t e^{-\alpha(t-s)} \log c_s \mathrm{d}s. \tag{3.2.6}$$

Hence, the habit level lives by a geometric form. That is, $h_t = \exp\{\beta \int_0^t e^{-\alpha(t-s)} \log c_s \mathrm{d}s\}$ holds for all $t \in [0, T]$. In contrast to arithmetic habits, cf. Constantinides (1990) and van Bilsen et al. (2020b), this specification of h_t is not strictly increasing in time. As the geometric form consequently allows for decreases in h_t over $t \in [0, T]$, the interpretation of this habit component as a standard of living is more sensible. Hence, in our setup, h_t is not necessarily a subsistence level. Ultimately, we observe that $\log h_t$ in (3.2.6) can be represented as follows: $\log h_t = \alpha \int_0^t e^{-\alpha(t-s)} \log c_s^{\beta/\alpha} \mathrm{d}s$, for all $t \in [0, T]$. This representation indicates that h_t can be interpreted as the geometric weighted moving average (GWMA) of transformed past consumption decisions, $\{c_s^{\beta/\alpha}\}_{s \in [0,t]}$. Clearly, if $\alpha = \beta \neq 0$, $\frac{c_t}{h_t}$ becomes a dimensionless quantity, and h_t reduces to the ordinary GWMA of (non-transformed) past consumption decisions, $\{c_s\}_{s \in [0,t]}$.

3.2.4 Optimal Consumption Problem

The habit-forming agent in $\mathcal{M}$ is at $t = 0$ in possession of a predetermined amount of cash, $X_0 \in \mathbb{R}_+$, and lives until $t = T$. Throughout the trading interval, $[0, T]$, this agent seeks to maximise expected lifetime utility from the ratio of consumption to the habit process by continuously selecting his/her consumption levels and corresponding portfolio weights. The habit-forming agent must determine these controls in agreement with the dynamic budget constraint in (3.2.4), such that the admissibility conditions are met. We assume that the preferences of the individual are characterised by the Von-Neumann-Morgenstern index: $\mathbb{E}[\int_0^T e^{-\delta t} U(c_t/h_t)\, \mathrm{d}t]$, cf. Detemple and Zapatero (1991). Consistent with this description, the agent faces the following problem:

$$
\begin{aligned}
\sup_{\{c_t,\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} \quad & \mathbb{E}\left[\int_0^T e^{-\delta t} U\left(\frac{c_t}{h_t}\right) \mathrm{d}t\right] \\
\text{s.t.} \quad & \mathrm{d}X_t = X_t\left[\left(r_t + \pi_t^\top \sigma_t \lambda_t\right)\mathrm{d}t + \pi_t^\top \sigma_t \mathrm{d}W_t\right] - c_t \mathrm{d}t, \\
& \mathrm{d}\log h_t = \left(\beta \log c_t - \alpha \log h_t\right)\mathrm{d}t, \quad h_0 = 1,\ X_0 \in \mathbb{R}_+ .
\end{aligned}
\tag{3.2.7}
$$

In this problem, $\delta \in \mathbb{R}_+$ represents the agent's time-preference parameter and $U : \mathbb{R}_+ \to \mathbb{R}$ denotes the agent's utility function. For simplicity, we assume that utility is given by an ordinary power (CRRA) function, $U(x) = \frac{x^{1-\gamma}}{1-\gamma}$ for all $x \in \mathbb{R}_+$. Here, γ defines the coefficient of relative risk-aversion. For purposes related to concavity of the optimisation problem, we fix $\gamma > 1$. We denote the first and second derivatives of U by U' and U'', respectively. The first derivative of U, U', is also known as marginal utility. By $I : \mathbb{R}_+ \to \mathbb{R}_+$, we denote the inverse of marginal utility ($I = (U')^{-1}$). The preceding optimisation problem is not strictly concave and does not permit a derivation of the optimal controls in closed-form, because of the path-dependency induced by h_t. Namely, utility at time t depends, apart from on c_t alone, through h_t as well on $\{c_s\}_{s\in[0,t]}$.

3.3 Dual Formulation

In this section, we provide the main result of this chapter (Theorem 3.3.1): the dual formulation of the optimal consumption problem in (3.2.7). We divide this section into three parts. First, we present the dual formulation and formalise that it satisfies strong duality, in Theorem 3.3.1. In the same part, we provide a rough sketch of the proof that culminates in the preceding strong duality result. Second, we present a set of three technical remarks related to the dual formulation. Third, we discuss certain implications of the strong duality result concerning the analytical structure of the optimal controls. These analytical implications primarily pertain to the duality relations.

3.3.1 Main Result: Strong Duality

Theorem 3.3.1 contains the main result central to this chapter. Its statement formalises the fact that the optimal (dual) control problem in (3.3.1) and the optimal consumption problem in (3.2.7) are dual to each other. The latter concretely means that these problems satisfy strong duality. First, we provide the theorem itself. Second, we comment on its corresponding proof.

Theorem 3.3.1. *Consider the optimal consumption problem in* (3.2.7) *and define the primal objective function:* $J\left(X_0, \{c_t, \pi_t\}\right) = \mathbb{E}[\int_0^T e^{-\delta t} U(\frac{c_t}{h_t}) \mathrm{d}t]$. *Furthermore, introduce the following concave conjugate:* $V\left(x\right) = \inf_{z \in \mathbb{R}} \left\{xz - (-e^{-z})\right\} = x - x \log x$, *for all* $x \in \mathbb{R}_+$. *Then, the dual formulation of the optimal consumption problem in* (3.2.7) *is given by:*

$$\inf_{\psi_t \in L^2(\Omega \times [0,T]), \eta \in \mathbb{R}_+} \mathbb{E}\left[\int_0^T \left\{ e^{-\delta t} \frac{1}{1-\gamma} V\left(e^{\delta t} \psi_t\right) - \eta M_t V\left(\frac{\psi_t - \beta \mathbb{E}\left[\int_t^T e^{-\alpha(s-t)} \psi_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right]}{\eta M_t}\right)\right\} \mathrm{d}t\right] + \eta X_0. \tag{3.3.1}$$

That is, suppose that $\mathcal{V}\left(X_0, \psi_t, \eta\right)$ *represents the dual objective function of* (3.3.1). *Then, the problems in* (3.2.7) *and* (3.3.1) *satisfy strong duality:*

$$\sup_{\{c_t, \pi_t\}_{t \in [0,T]} \in \mathcal{A}_{X_0}} J\left(X_0, \{c_t, \pi_t\}\right) = \inf_{\psi_t \in L^2(\Omega \times [0,T]), \eta \in \mathbb{R}_+} \mathcal{V}\left(X_0, \psi_t, \eta\right), \tag{3.3.2}$$

for all $X_0 \in \mathbb{R}_+$

Proof. The proof is given in Appendix A. □

Typically, the Legendre transform alone suffices to establish a strong duality result. However, due to the non-concavity and path-dependency of the objective of (3.2.7), the Legendre transform cannot be used to derive strong duality.

Therefore, to prove Theorem 3.3.1, we apply a change of variables and employ Fenchel Duality, cf. Proposition A.1. This form of duality can be regarded as a generalisation of the Legendre result to problems involving path-dependent linear transformations of one of the control variables. On the basis of Fenchel Duality, deriving strong duality for problems (3.2.7) and (3.3.1) is straightforward. First, we re-express the primal problem (3.2.7) in terms of its static equivalent and $\log c_t$. Second, we use Fenchel Duality to demonstrate that strong duality holds for the static problem and $\inf_{\psi_t \in L^2(\Omega\times[0,T])} \mathcal{V}\left(X_0, \psi_t, \eta\right)$. Third and last, we resort to a technical argument (Lemma A.2) in order to extend this strong duality result to (3.2.7) and (3.3.1).

3.3.2 Technical Remarks

In this section, we address three technical aspects of the dual formulation in Theorem 3.3.1. The first two aspects touch upon the dual control variable, $\{\psi_t\}_{t\in[0,T]}$. As for the first aspect, we know that this control variable attains values in the "unconstrained" set $L^2\left(\Omega\times[0,T]\right)$. However, the function V is defined over $\mathbb{R}_+$. Consequently, the dual forces the following two constraints upon $\{\psi_t\}_{t\in[0,T]}$: $\psi_t > 0$ and $\psi_t - \beta\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\psi_s\mathrm{d}s \mid \mathcal{F}_t] > 0$ for all $t\in[0,T]$. The second aspect is closely related to the latter and concerns an alternative representation of the dual in (3.3.1). Suppose that we define a process $p_t = \psi_t - \beta\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\psi_s\mathrm{d}s \mid \mathcal{F}_t]$ for all $t\in[0,T]$. Then, the latter identity can be regarded as a Volterra equation for ψ_t with the following solution: $\psi_t = p_t + \beta\mathbb{E}[\int_t^T e^{-[\alpha-\beta](s-t)}p_s\mathrm{d}s \mid \mathcal{F}_t]$ for all $t\in[0,T]$. Therefore, the dual formulation in (3.3.1) is identical to the following one:

$$\begin{aligned}\inf_{p_t\in L^2(\Omega\times[0,T]),\eta\in\mathbb{R}_+} \mathbb{E}\Bigg[\int_0^T \Bigg\{ & -\eta M_t V\left(\frac{p_t}{\eta M_t}\right) \\ & e^{-\delta t}\frac{1}{1-\gamma}V\left(e^{\delta t}\left\{p_t+\beta\mathbb{E}\left[\int_t^T e^{-[\alpha-\beta](s-t)}p_s\mathrm{d}s\,\middle|\,\mathcal{F}_t\right]\right\}\right)\Bigg\}\mathrm{d}t\Bigg] + \eta X_0.\end{aligned} \tag{3.3.3}$$

As in case of $\{\psi_t\}_{t\in[0,T]}$, it is clear that the the preceding dual formulation forces the following constraints upon $\{p_t\}_{t\in[0,T]}$: $p_t > 0$ and $p_t + \beta\mathbb{E}[\int_t^T e^{-[\alpha-\beta](s-t)}p_s\mathrm{d}s \mid \mathcal{F}_t] > 0$ for all $t\in[0,T]$. Although this re-definition of the dual control variable does not affect the dual optimality conditions, it implies slightly different duality relations. In the subsequent section 3.3.3, we address the nature of these relations in more detail.

The third and final aspect concerns the dual formulation for the model setup without habit formation. To recover this no-habit case, it suffices to fix $\alpha = \beta = 0$. Setting $\alpha = \beta = 0$ in the dual of Theorem 3.3.1 provides us with the following dual formulation:

$$\inf_{\psi_t \in L^2(\Omega \times [0,T]), \eta \in \mathbb{R}_+} \mathbb{E}\left[\int_0^T \left\{ e^{-\delta t} \frac{V\left(e^{\delta t}\psi_t\right)}{1-\gamma} - \eta M_t V\left(\frac{\psi_t}{\eta M_t}\right)\right\} \mathrm{d}t\right] + \eta X_0. \tag{3.3.4}$$

Here, the dual forces $\{\psi_t\}_{t\in[0,T]}$ to satisfy $\psi_t > 0$ for all $t \in [0,T]$. In line with the exclusion of h_t in the primal, the no-habit dual does not contain the $\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\psi_s \mathrm{d}s \mid \mathcal{F}_t]$ term. The dual in (3.3.4) differs from the conventional one in e.g. Cvitanić and Karatzas (1992). Note that this is not troublesome, as the dual formulations for convex optimisation problems are not unique, cf. Rockafellar (2015). In fact, after inserting the optimal dual control, say ψ_t^{opt} (cf. Example 3.3.1), into (3.3.4), we find the conventional formulation. The aforementioned difference is attributable to the fact that this dual ensues from an application of Fenchel Duality. Namely, this notion of duality involves two convex conjugates instead of one. Moreover, it requires one to re-express the primal control as follows: $c_t = e^{-(-\log c_t)}$ for all $t \in [0,T]$. Due to these two features, the dual accommodates two functions that coincide with the concave conjugate of the exponential utility function ($x \mapsto -e^{-x}$).

3.3.3 Duality Relations

For convex optimisation problems, duality theory can be employed to disclose the relationship between the primal and dual controls, i.e. the duality relation. This duality relation infers how the primal controls analytically depend on the dual controls, and vice versa. The key characteristic of this relation is that it yields the optimal primal (dual) controls after insertion of the optimal dual (primal) controls (respectively). Therefore, the duality relation contains important information about the analytical structure of the optimal primal and dual variables. In addition to this, it provides an alternative view on the mechanisms that are involved with optimising the primal and dual problems. As the dual in (3.3.1) follows from Fenchel Duality rather than from the Legendre transform, its implied duality relations differ from the conventional ones. In fact, the duality relations[4] for the problems in (3.2.7) and (3.3.1) are twofold and for all $t \in [0, T]$ given by the following identities:

$$c_t^* = \frac{\psi_t - \beta\mathbb{E}\left[\int_t^T e^{-\alpha(s-t)}\psi_s \mathrm{d}s \;\middle|\; \mathcal{F}_t\right]}{\eta M_t} \quad \text{and} \quad h_t^* = c_t^* \left(e^{\delta t}\psi_t\right)^{\frac{1}{\gamma-1}}. \tag{3.3.5}$$

In a technical sense, the duality relation for consumption in (3.3.5), c_t^*, can be regarded as a specification of optimal consumption in some auxiliary (artificial) market. To ensure that consumption defined by this relation is admissible and optimal in the true market, the dual problem in (3.3.1) aims to characterise this identity for c_t^* in such a manner that it generates the habit level in (3.3.5), h_t^*. In an economic sense, we note that dual-implied consumption c_t^* is endowed with a "penalty term". Concretely, selecting high values for ψ_t at future dates, requires one to increase ψ_t today so as to arrive at similar utility levels. This mechanism inversely reflects the agent's viewpoint in the primal problem. Namely, if this agent selects high values for c_t today, via h_t, he/she is required to increase c_t even further to maintain similar utility levels. To obtain some insights into the role that ψ_t plays in minimising the dual value function, $\mathcal{V}$, we now conclude with Example 3.3.1.

Example 3.3.1. *Suppose that $\alpha = \beta = 0$. Then, we have the following:*

$$\psi_t^{\mathrm{opt}} = \left(\eta^{\mathrm{opt}} M_t\right)\left[e^{\delta t}\eta^{\mathrm{opt}} M_t\right]^{-\frac{1}{\gamma}}, \tag{3.3.6}$$

for all $t \in [0, T]$. Here, ψ_t^{opt} denotes the optimal dual process, and η^{opt} represents the corresponding dual-optimal constant. In particular, η^{opt} can

[4]These duality relations follow from the fact that the primal and dual objectives, in (3.2.7) and (3.3.1), are conjugate to each other. This concretely means that these expressions bind in the unique "point" outlined by (3.3.5), conditional on $\eta X_0 = \eta\mathbb{E}\left[\int_0^T c_t M_t \mathrm{d}t\right]$ being true. Note that the duality relation in (3.3.5) corresponds to the dual in (3.3.1). For the alternative, howbeit identical, representation in (3.3.3), the duality relations read: $c_t^* = \frac{p_t}{\eta M_t}$ and $h_t^* = c_t^*\left(e^{\delta t}\left\{p_t + \beta\mathbb{E}\left[\int_t^T e^{-[\alpha-\beta](s-t)} p_s \mathrm{d}s \;\middle|\; \mathcal{F}_t\right]\right\}\right)^{\frac{1}{\gamma-1}}$ for all $t \in [0, T]$.

be obtained from solving $\mathbb{E}[\int_0^T \frac{\psi_t^{\text{opt}}}{\eta^{\text{opt}} M_t} \mathrm{d}t] = X_0$ *for* η^{opt}*. From the duality relations provided in* (3.3.5)*, we know that* ψ_t^{opt} *should generate* c_t^{opt} *via* $c_t^* = \frac{\psi_t - \beta \mathbb{E}\left[\int_t^T e^{-\alpha(s-t)} \psi_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right]}{\eta M_t}$*. Using that* $c_t^* = \frac{\psi_t}{\eta M_t}$ *for* $\alpha = \beta = 0$*, we therefore find that optimal consumption is given by:*

$$c_t^{\text{opt}} = \frac{\psi_t^{\text{opt}}}{\eta^{\text{opt}} M_t} = \left(e^{\delta t} \eta^{\text{opt}} M_t\right)^{-\frac{1}{\gamma}}, \tag{3.3.7}$$

for all $t \in [0, T]$*. Employing the definition of* $I : \mathbb{R}_+ \to \mathbb{R}_+$*, we accordingly have that* $c_t^{\text{opt}} = I\left(e^{\delta t} \eta^{\text{opt}} M_t\right)$ *holds. Moreover, in the optimum characterised by* c_t^{opt} *and* ψ_t^{opt}*, the value for* $\eta^{\text{opt}} \in \mathbb{R}_+$ *is determined such that* $\mathbb{E}[\int_0^T c_t^{\text{opt}} M_t \mathrm{d}t] = X_0$ *holds. Hence, it is clear that* c_t^{opt} *in* (3.3.7) *coincides with optimal consumption in the no-habit case* ($\alpha = \beta = 0$)*. See for instance Merton (1971) for a similar representation of* c_t^{opt} *and the corresponding optimal portfolio.*

3.4 Approximations

In this section, we develop a duality-based mechanism for evaluating the accuracy of approximations to the optimal solutions of (3.2.7). We break this section down into three parts. First, we provide the general evaluation mechanism and comment on related technicalities. Second, we present the approximate solution proposed by van Bilsen et al. (2020a). In addition to this, we rely on the duality relation in (3.3.5) to develop a corresponding dual approximation. Third, we make use of the evaluation mechanism to numerically study the precision of these analytical approximations.

3.4.1 Evaluation Mechanism

To quantify the accuracy of approximations to the optimal solutions of (3.2.7), we develop an evaluation mechanism. This mechanism is predicated on the evaluation techniques proposed in Bick et al. (2013) and Kamma and Pelsser (2022c). These techniques make use of strong duality to note that any departure from the optimal primal and/or dual controls results in a duality gap. Concretely, the primal value function, $J\left(X_0, \{c_t, \pi_t\}\right)$, delivers a lower bound on the optimal dual value function, for each admissible pair $\{c_t, \pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$. Likewise, the dual value function, $\mathcal{V}(X_0, \psi_t, \eta)$, spawns an upper bound on the optimal primal value function, for each feasible pair $(\eta, \psi_t) \in \mathbb{R}_+ \times \Psi$, where $\Psi := \{\psi_t \in L^2(\Omega \times [0,T]) \mid \psi_t > 0,\ \psi_t > \beta\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\psi_s \mathrm{d}s \mid \mathcal{F}_t],\ \forall\, t \in [0,T]\}$. To be more precise, for all $X_0 \in \mathbb{R}_+$, $\{c_t, \pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$ and $(\eta, \psi_t) \in \mathbb{R}_+ \times \Psi$, the following holds true:

$$J(X_0, \{c_t, \pi_t\}) \leq \mathcal{V}(X_0, \psi_t, \eta). \tag{3.4.1}$$

Theorem 3.3.1 infers that the inequality in (3.4.1) binds if and only if $(c_t, \pi_t) = \left(c_t^{\mathrm{opt}}, \pi_t^{\mathrm{opt}}\right)$ and $(\eta, \psi_t) = \left(\eta^{\mathrm{opt}}, \psi_t^{\mathrm{opt}}\right)$ for all $t \in [0,T]$. Here, we employ the superscript "opt" to indicate that these concern the optimal primal/dual control variables. For the former reason, the difference between J and $\mathcal{V}$ grows, the farther $\{c_t, \pi_t\}_{t\in[0,T]}$ and/or (η, ψ_t) are situated from the optima. We can employ this observation to gauge the accuracy of particular approximations as follows. Suppose that $\{c'_t, \pi'_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$ represents an arbitrary admissible trading-consumption pair and that $(\eta', \psi'_t) \in \mathbb{R}_+ \times \Psi$ specifies a feasible pair of dual controls. Then, $D(X_0) = \mathcal{V}(X_0, \psi'_t, \eta') - J(X_0, \{c'_t, \pi'_t\})$ characterises for all $X_0 \in \mathbb{R}_+$ the corresponding duality gap. As it is difficult to interpret the quantity $D(X_0) \in \mathbb{R}_+$, we compute the so-called compensating variation (CV), denoted by $\mathcal{C} \in \mathbb{R}_+$. The CV can be calculated from:

$$J(X_0, \{c'_t, \pi'_t\}) = \mathcal{V}(X_0[1 - \mathcal{C}], \psi'_t, \eta'). \tag{3.4.2}$$

Here, $\mathcal{C}$ can be interpreted as a fee that one pays to gain access to the optimal trading-consumption pair. In Bick et al. (2013) and Kamma and Pelsser (2022c), a similar interpretation is used. Note that $\mathcal{C}$ grows with the magnitude of $D(X_0)$, and thus with the difference(s) between $\{c'_t, \pi'_t\}_{t\in[0,T]}$ and $\left\{c_t^{\mathrm{opt}}, \pi_t^{\mathrm{opt}}\right\}_{t\in[0,T]}$, as well as between (η', ψ'_t) and $\left(\eta^{\mathrm{opt}}, \psi_t^{\mathrm{opt}}\right)$. Note that this way of calculating the approximation error is numerically not demanding, as one does not require the optimal controls to obtain the error.

Suppose that we are in possession of an approximate trading-consumption pair, $\{c'_t, \pi'_t\}_{t\in[0,T]}$. Calculating the corresponding lower bounds (J) is straightforward. To see how we may obtain matching upper bounds ($\mathcal{V}$), we note that

$$c'_t = \frac{\psi'_t - \beta\mathbb{E}\left[\int_t^T e^{-\alpha(s-t)}\psi'_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right]}{\eta' M_t} \quad \text{and} \quad \widehat{c}'_t = \left(e^{\delta t}\psi'_t\right)^{\frac{1}{1-\gamma}}, \tag{3.4.3}$$

follows from the duality relations in (3.3.5). Here, we define $\widehat{c}'_t := \frac{c'_t}{h'_t}$ as approximate ratio consumption, where h'_t represents the corresponding approximate habit level. The first identity in (3.4.3) indicates that each admissible consumption strategy corresponds to dual control variable, ψ'_t. Similarly, the second identity shown in (3.4.3) demonstrates that each ratio consumption strategy implies an (analytical) expression for ψ'_t. One can then obtain a matching approximation to the dual constant, η', from: $\mathbb{E}[\int_0^T \frac{1}{\eta'}(\psi'_t - \beta\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\psi'_s \mathrm{d}s \mid \mathcal{F}_t])\mathrm{d}t] = X_0$ (conditional on $\psi'_t \in \Psi$). Since it could be the case that $\psi'_t \notin \Psi$, it may be necessary to project $\{\psi'_t\}_{t\in[0,T]}$ into Ψ, to ensure dual-feasibility of (η', ψ'_t). For this purpose, we introduce a projection operator $\mathrm{proj}_\Psi : L^2(\Omega \times [0,T]) \to \Psi$. Using this operator, we re-define the primal-implied dual controls in the following way:

$$\widehat{\psi}'_t = \mathrm{proj}_\Psi(\psi'_t) \quad \text{and} \quad \widehat{\eta}' = \mathcal{R}^{-1}(X_0). \tag{3.4.4}$$

We define $\mathcal{R}^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ as the inverse of the following function: $\mathcal{R}(\eta) = \mathbb{E}[\int_0^T \frac{1}{\eta'}(\widehat{\psi}'_t - \beta\mathbb{E}[\int_t^T e^{-\alpha(s-t)}\widehat{\psi}'_s \mathrm{d}s \mid \mathcal{F}_t])\mathrm{d}t]$. Clearly, $(\widehat{\eta}', \widehat{\psi}'_t)$ is feasible and generates an upper bound, $\mathcal{V}$, on the optimal value function. This enables us to determine $\mathcal{C}$ as in equation (3.4.2).

We would like to make three remarks. First, we observe that one can distil (feasible) dual controls from either c'_t or $\widehat{c}'_t$ in (3.4.3). The dual controls implied by c'_t differ from those implied by $\widehat{c}'_t$, except when $c'_t = c^{\mathrm{opt}}_t$ and $\widehat{c}'_t = \widehat{c}^{\mathrm{opt}}_t$. Second, we note that it is possible to obtain an analytical expression for ψ'_t from the duality relation for c'_t in (3.4.3). Similar to the alternative representation of the dual in (3.3.3), it suffices to identify the appropriate Volterra equation. Third and last, we stress that our evaluation mechanism does not need to be conceptually modified for the alternative dual in (3.3.3). In fact, all steps relevant to the mechanism remain the same. The mere adjustment that has to be made is the specification of the duality relations in (3.4.3).

3.4.2 Approximate Controls

Subsequently, we present the analytical approximation to optimal (ratio) consumption, $\widehat{c}_t^{\text{opt}}$, proposed by van Bilsen et al. (2020a). On the grounds of our evaluation mechanism, and the relations in (3.4.3), we develop a corresponding novel dual approximation. In the sequel, we do not pay attention to the trading strategy that hedges approximate consumption. The trading strategy does not play a role in the specification of the approximate value function, and can always be obtained from admissible consumption processes. The approximation of van Bilsen et al. (2020a) is based on a first-order Taylor expansion of the budget constraint in the static representation of (3.2.7), i.e. $\mathbb{E}[\int_0^T M_t\widehat{c}_t h_t \mathrm{d}t] \leq X_0$, around $\{\widehat{c}_t\}_{t\in[0,T]} = 1$. The motivation for such an expansion is that the habit level closely tracks optimal consumption. According to Theorem 3.1 in van Bilsen et al. (2020a), this approximation is for all $t \in [0,T]$ given by:

$$\widehat{c}_t' = \left(\eta' e^{\delta t} M_t \left\{1 + \beta\mathbb{E}\left[\int_t^T e^{-[\alpha-\beta](s-t)}\frac{M_s}{M_t}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right]\right\}\right)^{-\frac{1}{\gamma}}. \tag{3.4.5}$$

Here, $\eta' \in \mathbb{R}_+$ is determined such that the budget constraint holds: $\mathbb{E}[\int_0^T M_t\widehat{c}_t' h_t' \mathrm{d}t] = X_0$. Moreover, h_t' is the approximate habit level generated by $\{\widehat{c}_s'\}_{s\in[0,T]}$, i.e. $h_t' = e^{\beta\int_0^t e^{-[\alpha-\beta](t-s)}\log\widehat{c}_s'\mathrm{d}s}$.

Consistent with the outline of our evaluation mechanism, we must determine a corresponding dual approximation in order to measure the accuracy of $\widehat{c}_t'$ in (3.4.5). For this purpose, we must utilise either of the two duality relations shown in (3.4.3). Although the identity for $\widehat{c}_t'$ allows for an easy recovery of ψ_t', it may be the case that $\psi_t' \notin \Psi$. As we wish to avoid rigorous modifications enforced upon ψ_t' by a projection operator, we resort to the identity for c_t' instead. Note that $c_t' = \widehat{c}_t' h_t'$ for all $t \in [0,T]$. In the identity for c_t' in (3.4.3), we can recognize a clear Volterra equation. Its solution for ψ_t' can accordingly be formulated as: $\psi_t' = \eta M_t c_t' + \beta\mathbb{E}[\int_t^T e^{-[\alpha-\beta](s-t)}\eta M_s c_s'\mathrm{d}s \mid \mathcal{F}_t]$ for all $t \in [0,T]$. Hence, after re-arranging some terms, we are able to derive that

$$\psi_t' = \eta' M_t c_t' \left(1 + \beta\mathbb{E}\left[\int_t^T e^{-[\alpha-\beta](s-t)}\frac{M_s c_s'}{M_t c_t'}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right]\right), \tag{3.4.6}$$

	Coefficient of risk-aversion (γ)					Initial endowment (X_0)				
	6	8	10	12	14	8	9	10	11	12
$\mathcal{C}$ (%)	0.103	0.061	0.042	0.031	0.025	0.093	0.047	0.042	0.067	0.114
RSS	0.062	0.037	0.026	0.020	0.016	0.028	0.011	0.026	0.050	0.075
	Speed of mean-reversion ($\alpha = \beta$)					Time-preference (δ)				
	0.01	0.05	0.1	0.15	0.2	0.01	0.02	0.03	0.04	0.05
$\mathcal{C}$ (%)	0.000	0.010	0.042	0.095	0.167	0.048	0.045	0.042	0.039	0.036
RSS	0.003	0.014	0.026	0.037	0.048	0.027	0.026	0.026	0.026	0.026

Table 3.1. Upper bounds on welfare losses ($\mathcal{C}$). In the row denoted by $\mathcal{C}$, the table reports the upper bounds on the welfare losses corresponding to the approximate solution provided in equation (3.4.5). Additionally, in the row denoted by RSS, the table reports the "root" sum of squares (RSS) corresponding to this approximate solution. The welfare losses are calculated by solving (3.4.2) for $\mathcal{C}$, and are expressed in terms of percentages (%). The RSS is calculated as follows: $\text{RSS} = \mathbb{E}\left[\sum_{i=1}^{M}(c'_{t_i} - c''_{t_i})^2\right]^{1/2}$, where c'_{t_i} is approximate consumption for (3.4.5), and c''_{t_i} is approximate consumption implied by ψ'_t in (3.4.6) via the second duality relation in (3.4.3). Moreover, $t_1 < t_2 < \ldots < t_M$ represent M time-steps in the Euler scheme. The table reports $\mathcal{C}$ and RSS for different values of the four displayed parameters (γ, X_0, $\alpha = \beta$ and δ), under a baseline initialisation of the parameters. This baseline set is fixed as follows: $X_0 = 10$, $T = 10$, $\gamma = 10$, $\delta = 0.03$, $\alpha = \beta = 0.1$, $\mu = 0.05$, $r = 0.01$ and $\sigma = 0.2$. The results are based on $10,000$ simulations and an Euler scheme with 20 equidistant time-steps.

characterises the approximation to ψ_t^{opt}, implied by the duality relation for c'_t in (3.4.3). Note that $\eta M_t c'_t > 0$ holds for all $t \in [0, T]$. As a consequence, we have that $\psi'_t > 0$ is true for all $t \in [0, T]$. Moreover, by construction, it is the case that $\psi'_t - \beta \mathbb{E}\left[\int_t^T e^{-\alpha(s-t)} \psi'_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right] = \eta' M_t c'_t > 0$ for all $t \in [0, T]$. Therefore, $\psi'_t \in \Psi$, implying that the approximation in (3.4.6) is dual-feasible. It should be noted that both approximations are truly optimal in the no-habit case ($\alpha = \beta = 0$). For the approximation to optimal consumption in (3.4.5), fixing $\alpha = \beta = 0$, yields that $\widehat{c}'_t = c'_t = \left(e^{\delta t} \eta' M_t\right)^{-\frac{1}{\gamma}}$, which coincides with c_t^{opt} in (3.3.7). Similarly, letting $\alpha = \beta = 0$, the approximation to the optimal dual control in (3.4.6) becomes $\psi'_t = \eta' M_t c'_t = \eta' M_t \left[e^{\delta t} \eta' M_t\right]^{-\frac{1}{\gamma}}$, which coincides with (3.3.6). Observe that we have not distinguished between the primal and dual values η'. These values are, in fact, identical.

3.4.3 Numerical Results

We evaluate the accuracy of the approximation shown in equation (3.4.5), using the evaluation mechanism of section 3.4.1. To this end, we set $N = 1$

in the market model, $\mathcal{M}$, and fix $r_t = r$, $\sigma_t = \sigma$, $\mu_t = \mu$, where r, σ and μ are constants. Based upon the parameter initialisation in van Bilsen et al. (2020a), we define: $X_0 = 10$, $T = 10$, $\gamma = 10$, $\delta = 0.03$, $\alpha = \beta = 0.1$, $\mu = 0.05$, $r = 0.01$ and $\sigma = 0.2$. In Table 3.1, we present the upper bounds on the welfare losses ($\mathcal{C}$) associated with the approximation, for different values of γ, X_0, $\alpha = \beta$ and δ. We compute the welfare losses from the equality displayed in (3.4.2), in which the value functions, J and $\mathcal{V}$, follow directly from the primal and dual approximations in (3.4.5) and (3.4.6), respectively. In addition to this, the table displays the "root" sum of squares (RSS) corresponding to the approximations. The RSS is calculated as follows: $\text{RSS} = \mathbb{E}[\sum_{i=1}^{M}(c'_{t_i} - c''_{t_i})^2]^{\frac{1}{2}}$, where c'_{t_i} is approximate consumption for (3.4.5), and c''_{t_i} is approximate consumption implied by ψ'_t in (3.4.6) via the second duality relation in (3.4.3). Here, $t_1 < t_2 < \ldots < t_M$ represent M time-steps in the Euler scheme. These results ensue from $10,000$ Monte-Carlo simulations and an Euler scheme with 20 equidistant time-steps.

Table 3.1 shows that the maximal welfare losses generated by the approximation in (3.4.5) vary between 0.000% and 0.167% of X_0. Bearing in mind that these numbers constitute upper bounds on the true errors, we can conclude that the approximation is near-optimal. This finding coincides with the results reported in van Bilsen et al. (2020a) and is confirmed by the magnitude of the values reported for RSS. In view of the fact that $c'_t \to c_t^{\text{opt}}$ as $\alpha = \beta \to 0$, it is clear why the table displays a positive relation between $\alpha = \beta$ and the magnitude of both $\mathcal{C}$ and the RSS. As for the same relations involving γ and δ, we can observe from (3.4.5) that increases in γ and δ result in $\widehat{c}_t$ attaining values closer to 1. That is, due to increases in γ and δ, the habit level tracks consumption more closely. Consequently, as shown in Table 3.1, approximate consumption in (3.4.5) becomes more accurate for increases in γ and δ, as c'_t is based on a Taylor expansion around $\{\widehat{c}_t\}_{t\in[0,T]} = 1$. To explain the relation between X_0 and both $\mathcal{C}$ and RSS, we note that $X_0 = \frac{1}{r}\left(1 - e^{-rT}\right) \approx T$ roughly implies that $\widehat{c}_t \approx 1$. The latter is a consequence of the budget constraint, $\mathbb{E}[\int_0^T M_t \widehat{c}_t h_t \mathrm{d}t] = X_0$. Hence, the performance of c'_t should decrease for values of X_0 and T that deviate from $X_0 = T$. This is shown in Table 3.1.[5]

[5] We note that the running time for calculating $\mathcal{C}$ and RSS in Table 3.1 is effectively equal to zero. The mere required computational effort stems from the simulations. This is a

Appendix A Proof of Theorem 3.3.1

To prove Theorem 3.3.1, we make use of Fenchel Duality as formalised in Theorem 4.3.3 of the textbook by Borwein and Zhu (2004). As this theorem lies at the heart of our proof, we provide its statement in the following proposition.

Proposition A.1. *Let $f : X \to \mathbb{R} \cup \{\infty\}$ and $g : Y \to \mathbb{R} \cup \{\infty\}$ be two continuous and convex functions. Additionally, introduce the bounded linear map A, defined as $A : X \to Y$. Here, X and Y outline two Banach spaces. Then, the Fenchel problems, denoted by p^* and d^*, are given by:*

$$\begin{aligned} p^* &= \inf_{x \in X} \{f(x) + g(Ax)\}, \\ d^* &= \sup_{y^* \in Y} \{-f^*(A^* y^*) - g^*(-y^*)\}, \end{aligned} \tag{A.1}$$

and satisfy weak duality, $d^* \leq p^*$. Here, f^* and g^* represent the convex conjugates of f and g, respectively, i.e. $f^*(x) = \sup_{z \in X} \{\langle x, z \rangle - f(z)\}$ and $g^*(y) = \sup_{z \in Y} \{\langle y, z \rangle - g(z)\}$, for all $x \in X^*$ and $y \in Y^*$. Note that X^* and Y^* are the dual spaces of X and Y, respectively. Moreover, A^*, defined as $A^* : Y \to X$, is the adjoint of A. That is, A and A^* must satisfy: $\langle Ax, y \rangle_Y = \langle y, A^* y \rangle_X$, for all $x \in X$ and $y \in Y$. Strong duality, $p^* = d^*$, holds if either of the following conditions is fulfilled:

- *(i)* $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$ *and f and g are both lower semi-continuous. Here,* core *stands for the algebraic interior, and* dom h *is given by* $\operatorname{dom} h = \{z \mid h(z) < \infty\}$ *for any function h;*

- *(ii)* $A \operatorname{dom} f \cap \operatorname{cont} g \neq \phi$, *where* cont *are the points where the function is continuous.*

Moreover, if $|d^*| < \infty$ holds, then the supremum in (A.1) *is attained.*

Proof. See page 136 of Borwein and Zhu (2004). □

significant advantage, as most evaluation procedures employ computationally demanding grid-search routines.

We continue by aligning the notation of our primal and dual problems with the notation of Proposition A.1. To this end, we start by deriving an alternative representation of (3.2.7). This alternative representation is based on the static formulation of optimal investment-consumption problems, due to Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991). We provide this formulation in the subsequent lemma.

Lemma A.2. *Define the following function:*

$$\begin{aligned}\mathcal{J}\left(X_0, -\log c_t, \eta\right) = \mathbb{E}\left[\int_0^T e^{-\delta t}\frac{e^{[1-\gamma](\log c_t - \log h_t)}}{1-\gamma})\mathrm{d}t\right] \\ - \eta\mathbb{E}\left[\int_0^T e^{\log c_t} M_t \mathrm{d}t\right] + \eta X_0.\end{aligned} \tag{A.2}$$

Then, for all $X_0 \in \mathbb{R}_+$, the following optimisation problems are identical:

$$\sup_{\{c_t,\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} J\left(X_0, \{c_t, \pi_t\}\right) = \inf_{\eta\in\mathbb{R}_+} \sup_{-\log c_t\in L^2(\Omega\times[0,T])} \mathcal{J}\left(X_0, -\log c_t, \eta\right). \tag{A.3}$$

Proof. By arguments similar to those that yield Lemma 2.2 in Cox and Huang (1989) and Proposition 7.3 in Cvitanić and Karatzas (1992), we know that $\{c_t,\pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$ if and only if $\{c_t\}_{t\in[0,T]}$ satisfies $\mathbb{E}\left[\int_0^T c_t M_t \mathrm{d}t\right] \leq X_0$ and $\log c_t \in L^2\left(\Omega \times [0,T]\right)$. Therefore, maximisation of $J\left(X_0, \{c_t, \pi_t\}\right)$ over $\{c_t,\pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$ is the same as maximisation of $J\left(X_0, \{c_t, \pi_t\}\right)$ over all $\log c_t \in L^2\left(\Omega \times [0,T]\right)$ such that $\mathbb{E}\left[\int_0^T c_t M_t \mathrm{d}t\right] \leq X_0$ holds. As a result of the preceding equivalence, we are able to derive the following equations:

$$\begin{aligned}&\sup_{\{c_t,\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} \mathbb{E}\left[\int_0^T e^{-\delta t} U\left(\frac{c_t}{h_t}\right)\mathrm{d}t\right] \\ &= \sup_{\log c_t\in L^2 \text{ s.t. } \mathbb{E}\left[\int_0^T c_t M_t \mathrm{d}t\right]\leq X_0} \mathbb{E}\left[\int_0^T e^{-\delta t} U\left(\frac{c_t}{h_t}\right)\mathrm{d}t\right] \\ &= \inf_{\eta\in\mathbb{R}_+}\left(\sup_{-\log c_t\in L^2}\left\{\mathbb{E}\left[\int_0^T e^{-\delta t} U\left(\frac{c_t}{h_t}\right)\mathrm{d}t\right] - \eta\mathbb{E}\left[\int_0^T c_t M_t \mathrm{d}t\right] + \eta X_0\right\}\right).\end{aligned} \tag{A.4}$$

Here, we set $L^2 := L^2(\Omega \times [0,T])$.

The last equality is a result of the following ingredients. First, we know that $c_t = X_0\epsilon(\mathbb{E}[\int_0^T M_t \mathrm{d}t])^{-1}$ for $\epsilon \in (0,1)$ is a strictly feasible solution to the static formulation of the consumption problem. Second, we have that $h_t > 0$ and $c_t > 0$. Hence, $c_t = e^{\log c_t}$ and $h_t = e^{\log h_t}$. Using this, we derive that $\mathbb{E}[\int_0^T e^{-\delta t} U(\frac{c_t}{h_t})\mathrm{d}t]$ is strictly concave in $-\log c_t \in L^2(\Omega \times [0,T])$. Similarly, we have that $\eta\mathbb{E}[\int_0^T c_t M_t \mathrm{d}t]$ is strictly convex in $-\log c_t \in L^2(\Omega \times [0,T])$. Third, by concavity of U, and the fact that $\log c_t, \log h_t \in L^2(\Omega \times [0,T])$, it holds that $\mathbb{E}[\int_0^T e^{-\delta t} U(\frac{c_t}{h_t})\mathrm{d}t] < \infty$. These properties validate the last equality, cf. Theorem 1 on page 217 of Luenberger (1997). The step from (A.4) to (A.3) is trivial using the definition of U, and that $c_t = e^{\log c_t}$ and $h_t = e^{\log h_t}$. $\square$

To align our notation with the one of Proposition A.1, we should have:

$$d^* = \sup_{-\log c_t \in L^2(\Omega \times [0,T])} \mathcal{J}(X_0, -\log c_t, \eta). \tag{A.5}$$

Accordingly, in the nomenclature of the aforementioned proposition, we have that $y^* = -\log c_t$ and $Y = L^2(\Omega \times [0,T])$. Note here that $Y = L^2(\Omega \times [0,T])$ outlines a Banach space. In addition to this, in terms of the functions f^* and g^*, and the mapping A, we must have the following:

$$\begin{aligned} -f^*(A^* y^*) &= \mathbb{E}\left[\int_0^T e^{-\delta t} \frac{e^{-[1-\gamma]A^*(-\log c_t)}}{1-\gamma} \mathrm{d}t\right], \\ -g^*(-y^*) &= -\eta\mathbb{E}\left[\int_0^T e^{\log c_t} M_t \mathrm{d}t\right] + \eta X_0, \end{aligned} \tag{A.6}$$

where the linear map A^* is given by:

$$A^*(-\log c_t) = -\log c_t + \beta \int_0^t e^{-\alpha(t-s)} \log c_s \mathrm{d}s. \tag{A.7}$$

Clearly, $A^* : L^2(\Omega \times [0,T]) \to L^2(\Omega \times [0,T])$. Therefore, by adjointess arguments, we must have that $A : L^2(\Omega \times [0,T]) \to L^2(\Omega \times [0,T])$, too.

We note the following: $-f(z_1) = \inf_{x \in X^*} \{f^*(x) - \langle x, z_1 \rangle\}$ and $-g(z_2) =$

$\inf_{y \in Y^*} \{g^*(y) - \langle y, z_2 \rangle\}$, for all $z_1 \in X$ and $z_2 \in Y$. It is easy to show that:

$$\begin{aligned} f(x) &= \mathbb{E}\left[\int_0^T e^{-\delta t} \frac{1}{1-\gamma} V\left(e^{\delta t} x_t\right) \mathrm{d}t\right], \\ g(x) &= -\mathbb{E}\left[\int_0^T \eta M_t V\left(\frac{x_t}{\eta M_t}\right) \mathrm{d}t\right] + \eta X_0. \end{aligned} \tag{A.8}$$

We observe that $X = Y = L^2(\Omega \times [0,T])$, and that the preceding definitions of $f : X \to \mathbb{R} \cup \{\infty\}$ and $g : Y \to \mathbb{R} \cup \{\infty\}$ constitute two continuous and convex functions. Furthermore, we find that A is given by:

$$Ax_t = x_t - \beta \mathbb{E}\left[\int_t^T e^{-\alpha(s-t)} x_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right]. \tag{A.9}$$

Note here that:

$$\begin{aligned} \|Ax_t\|_{L^2} &\leq \|x_t\|_{L^2} + \beta \left\| \mathbb{E}\left[\int_t^T e^{-\alpha(s-t)} x_s \mathrm{d}s \,\middle|\, \mathcal{F}_t\right] \right\|_{L^2} \\ &\leq \|x_t\|_{L^2} + \beta \mathbb{E}\left[\int_0^T \mathbb{E}\left[\int_t^T e^{-2\alpha(s-t)} x_s^2 \mathrm{d}s \,\middle|\, \mathcal{F}_t\right] \mathrm{d}t\right]^{\frac{1}{2}} \\ &= \|x_t\|_{L^2} + \frac{1}{2}\frac{\beta}{\alpha} \left\| x_t \left(1 - e^{-2\alpha t}\right)^{\frac{1}{2}} \right\|_{L^2} \leq \frac{3}{2} \|x_t\|_{L^2}. \end{aligned} \tag{A.10}$$

Again, we use that $L^2 := L^2(\Omega \times [0,T])$. The first inequality is due to the triangle inequality; the second inequality is a result of Hölder's inequality; the final inequality is trivial ($1 - e^{-2\alpha t} < 1$ for all $t \in [0,T]$). As a consequence of (A.10), we know that $A : X \to Y$ is a bounded linear map.

Considering Proposition A.1, we note that A dom $f \cap$ cont $g = \left(L^2(\Omega \times [0,T]) \cap \mathbb{R}\right) \cap \left(L^2(\Omega \times [0,T]) \cap \mathbb{R}_+\right) \neq \phi$. Hence, by Proposition A.1, we have strong duality, which finalises – via Lemma A.2 – the proof:

$$\begin{aligned} p^* &= \inf_{\psi_t \in L^2(\Omega \times [0,T])} \mathcal{V}(X_0, \psi_t, \eta) \\ &= \sup_{-\log c_t \in L^2(\Omega \times [0,T])} \mathcal{J}(X_0, -\log c_t, \eta) = d^*. \end{aligned} \tag{A.11}$$

4

Investing Towards an Exogenous Reference Level Using a Lower Partial Moments Criterion

Adapted from: Kamma, T., & Pelsser, A. (2022b). Investing towards an exogenous reference level using a lower partial moments criterion. *Working Paper*.

Abstract

This chapter analyses an optimal investment problem, in which the agent aims to minimise a lower partial moments (LPM) criterion that depends on an exogenous reference level. The problem concerns terminal wealth alone and is specified in an affine-term structure model with four risk-drivers. We derive and present closed-form expressions for the optimal portfolio rules and the optimal wealth process. Moreover, we analytically disentangle the distributional features of optimal terminal wealth. In the numerical illustrations, we examine the problem in the context of a defined contribution (DC) pension scheme. The reference level is accordingly identified as a life annuity. Our findings suggest that LPM-based investment policies can improve a pension fund's recovery potential. Despite their potentially outstanding performance, we illustrate that these policies may be difficult to implement. Furthermore, we show that the optima strongly depend on the estimates for the market prices of risk.

4.1 Introduction

Reference levels, otherwise known as benchmarks or targets, play a non-negligible role in the specification of an individual's preferences.[1] The critical nature of a reference level can be attributed to (i) its impact on preference qualifications and (ii) the fact that it constitutes a crucial part of an agent's risk profile. The impact on preferences is well-documented in studies on utility-based frameworks, see e.g. Browne (1999), Wagner (2002), Gómez and Zapatero (2003), and Berkelaar et al. (2004). The fact that it forms a pivotal element of an agent's risk profile is confirmed by e.g. Thaler (1980), Tversky and Kahneman (1991), Bateman et al. (1997), Munro and Sugden (2003), and Marzilli Ericson and Fuster (2011).[2] Due to their typical dependence on preference qualifications, the importance of a reference level carries over to optimal investment problems.

To illustrate this importance, let us visit the following stylised pension-related example.[3] Suppose that an individual receives a lump-sum of 100,000 monetary units upon retirement. We postulate that this amount is large enough to avail the individual of all resources necessary for an appropriate continuation of his/her pre-retirement life. In addition to this, assume that the individual's neighbours obtain a similar lump-sum, however, of 1,000,000 monetary units. Even though the former amount is in principle sufficient for the individual, in comparison to the neighbours' lump-sum, it appears rather small (only one tenth of the 1,000,000 monetary units). Depending on the value that this comparative magnitude has for the individual of interest, his/her experienced levels of happiness or utility will differ. In case of a strong reference-oriented individual, the 100,000 monetary units are clearly disappointing. The converse would be true for an individual who is indifferent with respect to the neighbours' financial circumstances. Note that this illustration extends to setups beyond the pension context, wherein the neighbours' retirement wealth can be identified

[1]Throughout the remainder of this chapter, we use "reference level", "benchmark", "goal" and "target" interchangeably.

[2]This short outline of empirical studies on reference-dependent preferences is by far not exhaustive. We refer to O'Donoghue and Sprenger (2018) for a more comprehensive overview. In the study by Zank (2010), one can find a closely related overview.

[3]This illustration builds on the "catching/keeping up with the Joneses" idea, formulated and analysed in a.o. Abel (1990), Gali (1994), and Gómez (2007).

as e.g. a life annuity, one's last earned wage or the stock index.

This example corroborates the claims above. In particular, it intuitively demonstrates that a reference level cannot be ignored in specifying an agent's preferences and, therefore, in optimising his/her utility from wealth. This intuition is scientifically supported by a great body of empirical findings dating back to a.o. Markowitz (1952), Edwards (1955), and Hershey and Schoemaker (1985).[4] These papers concretely argue that individuals tend to measure accumulated amounts of cash in relative terms, i.e. compared to a reference level. This feature gave rise to a myriad of studies on portfolio problems that explicitly incorporate such a level. We refer to section 2 of van Bilsen et al. (2020b) for an overview of such studies. In the contributions by a.o. Bernard and Ghossoub (2010), Balter et al. (2020), and van Bilsen and Laeven (2020), the inclusion of a reference level in investment-linked frameworks is shown to have a non-negligible impact on the optimal decision variables. By virtue of the aforementioned empirical evidence and the latter non-negligible impact, research on reference-oriented or goal-based investment routines is highly relevant. Within the confines of portfolio optimisation, the literature on reference levels can roughly be distinguished into three categories: those that concentrate on (i) loss aversion, (ii) risk aversion, and (iii) preference-independent hedging criteria.

Loss aversion outlines a key concept in the domain of prospect theory, cf. Kahneman and Tversky (1979). It refers to the empirically confirmed tendency of individuals to value losses greater than equivalent gains. Losses and gains are defined with respect to a person-specific reference level. To model this phenomenon, loss aversion models ordinarily rely on S-shaped utility functions, see e.g. He and Zhou (2011). Risk aversion setups do not explicitly distinguish between losses and gains. These setups hinge on Inada-type preferences that characterise an agent's attitude towards risk. The agent's risk profile is accordingly defined across the *absolute* accumulation of capital. To model preferences around a benchmark, risk aversion frameworks redefine this absolute accumulation in relative terms. This redefinition is usually carried out by incorporating a reference level into the conventional definitions of preferences, see e.g. Detemple

[4] We refer to Zank (2010) for an extensive overview of these empirical studies.

and Zapatero (1991), Van Binsbergen et al. (2008), and Kamma and Pelsser (2022a). Preference-independent hedging criteria ignore an individual's attitude towards risk and solely concentrate on acquiring/replicating a reference level. For these criteria, the reference level primarily serves as a goal or a target. Due to the independence from an individual's risk profile, hedging criteria involve a strong target-orientation. Models of this type are defined on a broad spectrum that includes e.g. super-replication and expected shortfall hedging, cf. Cvitanić et al. (1999) and Cvitanic (2000).

In this chapter, we focus on an investment problem, in which the agent aims to minimise a lower partial moments (LPM) criterion. We assume that this criterion depends on an exogenous reference level. In terms of the three categories above, the LPM framework combines a preference-independent hedging criterion with the concept of risk aversion. Although not explicitly, as pointed out by Jarrow and Zhao (2006), aspects unique to loss aversion return in the LPM operator.[5] We stress that the LPM operator enters into the problem via the agent's objective function. Unlike most studies, e.g. Harlow and Rao (1989), Leitner (2008), and Gao et al. (2017), this implies that it does not define a downside risk constraint. Similar problems have been analysed by e.g. Föllmer and Leukert (2000), Jarrow and Zhao (2006), and Krabichler and Wunsch (2021). We emphasise that Nawrocki and Viole (2014) advocate the use of such partial moments for the study of utility and portfolio theory from the perspective of behavioural economics.

Mathematically speaking, the LPM operator specifies a smoothed variant of the expected shortfall criterion. The latter identifies a valid hedging criterion that is unaffected by an agent's risk profile. As a result of this smoothing procedure, the LPM operator can be interpreted as a utility function that is closely related to the Inada-family. The agent's preference qualification or attitude towards risk is correspondingly specified as follows. If wealth is equal to the benchmark, the agent becomes infinitely risk averse in an attempt to "lock in" wealth at the current level. On the contrary, if wealth falls below

[5] In Jarrow and Zhao (2006), the authors demonstrate that the LPM problem can be subsumed under the umbrella of prospect theory. To this end, they employ a definition of (downside) loss aversion that relaxes the convexity requirement with regard to losses, cf. equation (1) and footnote 4 of their study.

the benchmark, the agent becomes considerably less risk averse. The previous decline in the agent's level of risk aversion comes close to the "gamble for resurrection" behaviour specific to loss aversion frameworks. Adapted to setups involving risk aversion, Yang et al. (2021) refer to this phenomenon as "risk-taking for resurrection". These attributes of the LPM criterion imply that the agent is strongly target-oriented.

We analyse the problem in continuous-time for terminal wealth alone. Moreover, we place the problem in the financial market model proposed by Koijen et al. (2009). This model assumes an affine-term structure for the (real) interest rate and involves four risk-drivers. The market distinguishes nominal from real returns and accommodates market prices of risk that depend on a mean-reverting state variable. For the exogenous reference level, we postulate a general log-normal process. To the best of our knowledge, there are no studies available that couple the LPM objective to such a complex model specification.[6] Due to the state-dependency of the market prices of risk, it is highly nontrivial to derive analytical solutions to the LPM problem. Nevertheless, using the Fourier transform similar to Carr and Madan (1999), we are able to derive closed-form expressions for the optimal portfolio rules and the optimal wealth process. In addition to this, we manage to disentangle an analytical formulation for the distributional properties of optimal terminal wealth. These analytical expressions form our first main contribution to the literature

In the numerical experiments, we analyse the problem in the context of a defined contribution (DC) pension scheme. For this reason, we identify the reference level as a life annuity that generates annual payments until the agent's fixed date of death. We numerically analyse the distribution of optimal retirement wealth and execute a sensitivity analysis of the optimal portfolio rules. The findings suggest that LPM-based investment policies can increase

[6]In Föllmer and Leukert (2000), the LPM problem was first introduced. The authors consider the problem in the context of partial hedging. Our chapter can therefore be regarded as a contribution to the literature on *applied* partial hedging. Although there are more studies available on partial hedging than on LPM problems, these papers typically concentrate on the relevant *mathematical* aspects. For such theoretical treatments, consider e.g. Pham (2000, 2002), Sekine (2004), Xia (2005), and Choi and Jonsson (2009). In related technical studies by Jonsson and Sircar (2002), Bouchard et al. (2004), and Nygren and Lakner (2012), the utility-dimension of a.o. the LPM operator is highlighted. For an applied deep learning approach to partial hedging, cf. Hou et al. (2022).

the likelihood of achieving one's pension goals, i.e. improve the pension fund's recovery potential. In spite of their possibly outstanding performance, we exemplify that these policies may be difficult to implement in reality. Due to the extraordinarily large and/or highly leveraged positions implied by the optimal trading rules, we recommend to include trading constraints. In addition to this, we demonstrate that both the recovery potential and the portfolio rules grandly depend on the estimates for the market prices of risk. This finding supports the use of policies that account for parameter/model uncertainty, cf. Balter (2016) and references therein. These three economic takeaways constitute our second main contribution to the literature.

The remainder of the chapter is structured as follows. Section 4.2 introduces the model setup. Section 4.3 presents the optimality conditions corresponding to the LPM problem. Section 4.4 contains the numerical analysis. Finally, section 4.5 concludes. We collect all mathematical proofs in the appendix.

4.2 Model Setup

In this section, we spell out the model setup. First, we introduce the financial market model. Second, we provide the agent's dynamic budget constraint. Third, we specify the dynamics of a benchmark process, i.e. the exogenous reference level, which models the agent's "goal" with regard to retirement. Fourth, we outline the optimal terminal wealth problem.

4.2.1 Financial Market Model

We make use of the financial market introduced in the paper by Koijen et al. (2009). Their model is defined in continuous-time, contains four independent risk-drivers and involves two distinct state variables. This setup is similar to Brennan and Xia (2002)'s. To specify this environment, we introduce a finite-valued trading horizon, $T > 0$. Correspondingly, the trading interval reads $[0, T]$. The uncertainty is described by a complete filtered probability space, $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$. Note that the separate components of this space live by their conventional definitions. On this space, the four factors are defined by an $\mathbb{R}^4$-valued standard Brownian motion process, $\{W_t\}_{t\in[0,T]}$, such that $W_t = [W_{1,t}, W_{2,t}, W_{3,t}, W_{4,t}]^\top$. Henceforth, we assume that all (in)equalities between processes hold in either a $\mathbb{P}$-a.s. or $\mathrm{d}t \otimes \mathbb{P}$-a.e. sense.

In line with the market specification in Koijen et al. (2009), we introduce an $\mathbb{R}^2$-dimensional state variable process, $Z_t = [Z_{1,t}, Z_{2,t}]^\top$. These state variables are assumed to evolve according to an Ornstein-Uhlenbeck process as follows:

$$\mathrm{d}Z_t = -K_Z Z_t \mathrm{d}t + \Sigma_Z \mathrm{d}W_t,\ Z_0 = 0_2, \tag{4.2.1}$$

where 0_2 is defined as follows: $0_2 = [0,0]^\top$ Here, the drift and diffusion terms, K_Z and Σ_Z, naturally achieve values in $\mathbb{R}^{2\times 2}$ and $\mathbb{R}^{4\times 4}$, respectively. We postulate that K_Z constitutes a lower-triangular matrix. Moreover, we set $\Sigma_Z = [I_{2\times 2}, 0_{2\times 2}]$, where $I_{2\times 2}$ characterises a two-dimensional identity matrix and $0_{2\times 2}$ identifies an $\mathbb{R}^{2\times 2}$-valued matrix containing zeros. Note here that Z_t spells out a two-dimensional mean-reverting process.

Similar to the models in Brennan and Xia (2002) and Sangvinatsos and Wachter (2005), this market makes a distinction between nominal and real return dynamics. To model this distinction, we introduce three processes for (i) the nominal interest rate, (ii) the commodity price index (CPI), and (iii) the inflation rate. The first is specified as:

$$r_t = \delta_{0,r} + \delta_{1,r}^\top Z_t, \tag{4.2.2}$$

where $\delta_{0,r} \in \mathbb{R}_+$ and $\delta_{1,r} \in \mathbb{R}^2$. That is, the process for the instantaneous interest rate is affine in both state variables, $Z_{1,t}$ and $Z_{2,t}$. For the CPI, we assume a geometric form, such that it evolves according to the following stochastic differential equation (SDE):

$$\frac{\mathrm{d}\Pi_t}{\Pi_t} = \pi_t \mathrm{d}t + \sigma_\Pi^\top \mathrm{d}W_t,\ \Pi_0 = 1, \tag{4.2.3}$$

where π_t spells out the instantaneous expected inflation process, and σ_Π outlines an $\mathbb{R}^4$-valued vector. As r_t, the process for π_t is assumed to be affine in Z_t:

$$\pi_t = \delta_{0,\pi} + \delta_{1,\pi}^\top Z_t, \tag{4.2.4}$$

where $\delta_{0,\pi} \in \mathbb{R}_+$ and $\delta_{1,\pi} \in \mathbb{R}^2$.

In this financial market, $\mathcal{M}$, the agent is allowed to continuously trade in five financial instruments: a money market account, a stock, two nominal bonds, and an inflation-linked bond. We first introduce the money market account and the stocks. After determining the term structure of interest rates, we provide the dynamics of the two bonds. The money market account lives by the following ordinary differential equation (ODE):

$$\frac{\mathrm{d}B_t}{B_t} = r_t \mathrm{d}t, \; B_0 = 1. \tag{4.2.5}$$

Note that B_t is specified in nominal terms. The return dynamics of the non-dividend paying (nominal) stock are characterised by the following SDE:

$$\frac{\mathrm{d}S_{1,t}}{S_{1,t}} = (r_t + \eta_S)\,\mathrm{d}t + \sigma_S^\top \mathrm{d}W_t, \; S_{1,0} = 1, \tag{4.2.6}$$

Here, σ_S defines the $\mathbb{R}^4$-valued vector for the volatility of the stock. Observe that σ_S consists of scalars alone. Additionally, η_S characterises the time-independent scalar-valued equity risk premium of the stock.

We assume that the financial market excludes frictions, i.e. (proportional) transaction costs. Moreover, we postulate that all risk factors are traded, due to which the market is complete. Therefore, by the fundamental theorem of asset pricing, cf. Delbaen and Schachermayer (1994), we know that there exists a unique equivalent martingale measure. In view of its dependency on this measure, we consequently know that there exists a unique nominal state price density process, $\{M_t\}_{t\in[0,T]}$. Under the assumption that $\{B_t\}_{t\in[0,T]}$ serves as the numéraire quantity, this process is given by the subsequent SDE:

$$\frac{\mathrm{d}M_t}{M_t} = -r_t \mathrm{d}t - \Lambda_t^\top \mathrm{d}W_t, \; M_0 = 1, \tag{4.2.7}$$

where Λ_t denotes the $\mathbb{R}^4$-valued process for the nominal market prices risk. Following the reasoning around equations (1)-(4) in Detemple et al. (2003), we are able to write the state price density process as follows: $M_t = \exp\{-\int_0^t r_s \mathrm{d}s - \frac{1}{2}\int_0^t \Lambda_s^\top \Lambda_s \mathrm{d}s - \int_0^t \Lambda_s^\top \mathrm{d}W_s\}$, for all $t \in [0,T]$. This representation plays an important role in the LPM problem.

To arrive at an affine model for the term structure of interest rates, Koijen et al. (2009) assume that these nominal market prices of risk are affine in both state variables, $Z_{1,t}$ and $Z_{2,t}$. As a result, Λ_t is specified in conformity with r_t and π_t. Concretely, we suppose that the process for Λ_t is given by:

$$\Lambda_t = \Lambda_0 + \Lambda_1 Z_t, \tag{4.2.8}$$

where $\Lambda_0 \in \mathbb{R}^4$ and $\Lambda_1 \in \mathbb{R}^{4\times 2}$. It should be noted that the specification of $\{S_t\}_{t\in[0,T]}$ in (4.2.6) forces restrictions upon the preceding Λ_0 and Λ_1 parameters. Namely, by martingale arguments, it must hold that $\sigma_S^\top \Lambda_t = \eta_S$ for all $t \in [0, T]$. Suppose that $\Lambda_{0(i)}$ and $\Lambda_{1(i,j)}$ denote the i^{th} entry of Λ_0 and the $(i,j)^{\text{th}}$ element of Λ_1, respectively. Throughout the remainder of this chapter, we adopt this notation for all vectors and matrices. In Koijen et al. (2009), the former condition is handled by assuming that $\Lambda_{0(4)}$, $\Lambda_{1(4,1)}$, and $\Lambda_{1(4,2)}$ are defined such that $\sigma_S^\top \Lambda_0 = \eta_S$ and $\sigma_S^\top \Lambda_1 = 0_2^\top$. We assume the same:

$$\begin{aligned}\Lambda_{0(4)} &= \frac{\eta_S}{\sigma_{S(4)}} - \frac{1}{\sigma_{S(4)}} \sum_{i=1}^{3} \sigma_{S(i)} \Lambda_{0(i)}, \\ \Lambda_{1(4,j)} &= -\frac{1}{\sigma_{S(4)}} \sum_{i=1}^{3} \sigma_{S(i)} \Lambda_{1(i,j)}, \quad j = 1, 2.\end{aligned} \tag{4.2.9}$$

We complete the asset mix by introducing two nominal bonds and an inflation-linked bond. For a given time to maturity, $t + \tau_i$,

$$\frac{\mathrm{d}P_{t,t+\tau_i}}{P_{t,t+\tau_i}} = r_t \mathrm{d}t + B(\tau_i)^\top \Sigma_Z (\mathrm{d}W_t + \Lambda_t \mathrm{d}t), \; P_{0,\tau_i} = e^{A(\tau_i)}, \tag{4.2.10}$$

characterises the dynamics of the two nominal bonds. Here, we set $i = 1, 2$, such that $\tau_1 \neq \tau_2$. Suppose that $\widetilde{\Lambda}_0 = \Sigma_Z \Lambda_0$ and $\widetilde{\Lambda}_1 = \Sigma_Z \Lambda_1$. Then, the deterministic functions $A(x)$ and $B(x)$ are for all $x \in \mathbb{R}_+$ given by:

$$\begin{aligned}A(x) &= -\int_0^x \left[B(s)^\top \widetilde{\Lambda}_0 - \frac{1}{2} B(s)^\top B(s) + \delta_{0,r} \right] \mathrm{d}s, \\ B(x) &= \left(K_Z^\top + \widetilde{\Lambda}_1^\top \right)^{-1} \left[\exp\left\{ -\left[K_Z^\top + \widetilde{\Lambda}_1^\top \right] x \right\} - I_{2\times 2} \right] \delta_{1,r}.\end{aligned} \tag{4.2.11}$$

Note that the return dynamics of the nominal bonds in (4.2.10) are based on the following identity: $P_{t,T} = \mathbb{E}\left[\frac{M_T}{M_t} \,\middle|\, \mathcal{F}_t\right] = e^{A(T-t)+B(T-t)^\top Z_t}$ for all $t \in [0,T]$. In a similar sense, we are able to derive the nominal return dynamics of an inflation-linked bond with a given time to maturity, $t+\tau$, say $P^R_{t,t+\tau}$. That is, the identity for the inflation-linked bond is predicated on: $P^R_{t,T} = \mathbb{E}\left[\frac{M_T \Pi_T}{M_t} \,\middle|\, \mathcal{F}_t\right] = e^{A^R(T-t)+B^R(T-t)^\top Z_t}\Pi_t$ for all $t \in [0,T]$. As a result, for a given time to maturity, $t+\tau$, the SDE of $P^R_{t,t+\tau}$ evolves according to:

$$\frac{\mathrm{d}P^R_{t,t+\tau}}{P^R_{t,t+\tau}} = r_t \mathrm{d}t + \left(B^R(\tau)^\top \Sigma_Z + \sigma_\Pi^\top\right)(\mathrm{d}W_t + \Lambda_t \mathrm{d}t), \; P^R_{0,t} = e^{A^R(\tau)}. \tag{4.2.12}$$

To be able to define the deterministic functions, $\tau \mapsto A^R(\tau)$ and $\tau \mapsto B^R(\tau)$, we introduce the following convenient notation: $\widehat{\delta}_{0,r} = \delta_{0,r} - \delta_{0,\pi} + \sigma_\Pi^\top \Lambda_0$, $\widehat{\delta}_{1,r} = \delta_{1,r} - \delta_{1,\pi} + \Lambda_1^\top \sigma_\Pi$, and $\widehat{\Lambda}_0 = \widetilde{\Lambda}_0 - \widetilde{\sigma}_\Pi$, for $\widetilde{\sigma}_\Pi = \Sigma_Z \sigma_\Pi$. Then, similar to (4.2.11), it can be shown that $\tau \mapsto A^R(\tau)$ and $\tau \mapsto B^R(\tau)$ are for all $\tau \in \mathbb{R}_+$ given by the subsequent specifications:

$$\begin{aligned} A^R(\tau) &= -\int_0^\tau \left[B^R(s)^\top \widehat{\Lambda}_0 - \frac{1}{2} B^R(s)^\top B^R(s) + \widehat{\delta}_{0,r}\right] \mathrm{d}s, \\ B^R(\tau) &= \left(K_Z^\top + \widetilde{\Lambda}_1^\top\right)^{-1} \left[\exp\left\{-\left[K_Z^\top + \widetilde{\Lambda}_1^\top\right]\tau\right\} - I_{2\times 2}\right] \widehat{\delta}_{1,r}. \end{aligned} \tag{4.2.13}$$

We conclude the introduction of the two bonds with three remarks. First, following Pelsser (2019), we note that not all EU countries issue inflation-linked bonds (e.g. the Netherlands). However, as most other countries in the euro-area do issue such contracts (e.g. France), it is realistic to include a single inflation-linked bond in the asset mix. Second, in outlining both $P_{t,t+\tau_i}$ and $P^R_{t,t+\tau}$ in (4.2.10) and (4.2.12), respectively, we have employed the notation of Chen et al. (2020). To avoid confusion, we stress that $\widetilde{\Lambda}_0$, $\widetilde{\Lambda}_1$ and $\widetilde{\sigma}_\Pi$ are defined as: $\widetilde{\Lambda}_0 = \left[\Lambda_{0(1)}, \Lambda_{0(2)}\right]^\top$, $\widetilde{\Lambda}_1 = \left[\Lambda_{1(1,1:2)}, \Lambda_{1(2,1:2)}\right]^\top$, and $\widetilde{\sigma}_\Pi = \left[\sigma_{\Pi(1)}, \sigma_{\Pi(2)}\right]^\top$. Third and last, we observe that the three bonds are assumed to have continuously adjusted times to maturity, $t+\tau_i$ and $t+\tau$. We are obliged to make this simplifying assumption in order to ensure that the three bonds are traded throughout the entire trading interval.[7]

[7]The yields or the continuously compounded zero-coupon rates of the nominal bonds

Finally, suppose that S_t contains all traded financial instruments: $S_t = \left[S_{1,t}, P_{t,t+\tau_1}, P_{t,t+\tau_2}, P^R_{t,t+\tau}\right]^\top$. Then, the SDE of S_t is given by:

$$\mathrm{d}S_t = \mathrm{diag}\left(S_t\right)\left[\left(r_t + \sigma\Lambda_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t\right], \tag{4.2.14}$$

with starting value $S_0 = \left[1, e^{A(\tau_1)}, e^{A(\tau_2)}, e^{A^R(\tau)}\right]^\top$. In the latter SDE, $\mathrm{diag}\left(S_t\right)$ stands for the $\mathbb{R}^{4\times 4}$-valued diagonal matrix, which contains the four distinct entries of S_t on its diagonal. To be more precise: $\mathrm{diag}\left(S_t\right) = \left(S_t 1_4^\top\right) \odot I_{4\times 4}$, where 1_4 spells out an $\mathbb{R}^4$-valued vector of 1's, $I_{4\times 4}$ characterises a four-dimensional identity matrix, and "$\odot$" denotes the Hadamard product. The matrix σ outlines the volatility process of S_t and attains values in $\mathbb{R}^{4\times 4}$. Here, σ consists of scalars alone and is defined as follows: $\sigma = \left[\sigma_S, B\left(\tau_1\right)^\top \Sigma_Z, B\left(\tau_2\right)^\top \Sigma_Z, B^R\left(\tau\right)^\top \Sigma_Z + \sigma_\Pi^\top\right]^\top$. We cast the four different risky financial instruments into the vector-format provided by S_t in order to both facilitate derivations and simplify notation. We conclude this section by pointing out that $\left\{S_t\right\}_{t\in[0,T]}$ is indeed driven by the four Brownian motions, $\left\{W_t\right\}_{t\in[0,T]}$. As a consequence, trading in S_t alone suffices to hedge all the risk present in this financial environment.

4.2.2 Dynamic Budget Constraint

We proceed with the introduction of the agent's dynamic budget constraint. In the utility maximisation problem of interest, this agent is solely concerned about terminal wealth. As a result, consumption does not play a role and can be excluded from the specification of the budget constraint. Taking into account that trading takes place in continuous-time over the financial instruments from section 4.2.1, it suffices to introduce a single endogenous process, $\left\{\pi_t\right\}_{t\in[0,T]}$, as the control variable. This process is $\mathbb{R}^4$-valued, $\mathcal{F}_t$-progressively measurable, and records the agent's decisions with regard to investment. More precisely, $\left\{\pi_t\right\}_{t\in[0,T]}$ denotes the amount of monetary units that the agent allocates to the four traded risky assets, $\left\{S_t\right\}_{t\in[0,T]}$. Let us assume that the agent is at the start

and the index-linked bond, say y_{t,τ_i} and $y^R_{t,\tau}$, are as follows: $y_{t,\tau_i} = -\frac{\log P_{t,t+\tau_i}}{\tau_i} = -\frac{A(\tau_i)}{\tau_i} - \frac{B(\tau_i)^\top}{\tau_i}X_t$ and $y^R_{t,\tau} = -\frac{\log P^R_{t,t+\tau}}{\tau} = -\frac{A^R(\tau)}{\tau} - \frac{B^R(\tau)^\top}{\tau}X_t - \frac{1}{\tau}\int_0^t \pi_s \mathrm{d}s + \frac{1}{\tau}\frac{1}{2}\sigma_\Pi^\top\sigma_\Pi - \frac{1}{\tau}\sigma_\Pi^\top W_t$.

of the trading interval, $t = 0$, in possession of a fixed initial endowment equal to $X_0 \in \mathbb{R}_+$. Then, given $X_0 \in \mathbb{R}_+$, the agent's dynamic budget constraint evolves according to the following SDE:

$$\begin{aligned} \mathrm{d}X_t &= \left(X_t - \sum_{i=1}^{4} \pi_{i,t} \right) \frac{\mathrm{d}B_t}{B_t} + \sum_{i=1}^{4} \pi_{i,t} \frac{\mathrm{d}S_{i,t}}{S_{i,t}} \\ &= \left(X_t r_t + \pi_t^\top \sigma \Lambda_t \right) \mathrm{d}t + \pi_t^\top \sigma \mathrm{d}W_t. \end{aligned} \tag{4.2.15}$$

For the purpose of clarity, we note that $S_{i,t}$ represents the i^{th} element of the $\mathbb{R}^4$-valued vector, S_t, in (4.2.14). Moreover, we observe that $X_t - \sum_{i=1}^{4} \pi_{i,t}$ stands for the amount of monetary units that the agent leaves in the money market account, B_t, at time $t \in [0, T]$. Note that the allocation of assets to B_t is, therefore, entirely dependent on the allocation to S_t. To prevent the agent from implementing ill-posed investment decisions, e.g. doubling strategies, we introduce an admissibility set, $\mathcal{A}_{X_0}$. This set contains all *admissible* or well-posed investment strategies: $\mathcal{A}_{X_0}$ consists of all $\{\pi_t\}_{t \in [0,T]}$, such that $X_t \geq 0$, $\int_0^T \pi_t^\top \sigma_t \sigma_t^\top \pi_t \mathrm{d}t < \infty$, and $\int_0^T \left| \pi_t^\top \sigma_t \Lambda_t \right| \mathrm{d}t < \infty$ hold for all $t \in [0, T]$. Naturally, we restrict the agent's investment decisions to those that are contained in the admissibility set, $\mathcal{A}_{X_0}$. See for instance Karatzas and Shreve (1998), van Bilsen et al. (2020a), or Kamma and Pelsser (2022c), for similar definitions of admissible trading strategies.

4.2.3 Specification of Benchmark

In this financial environment, the agent is assumed to retire at $t = T$. Accordingly, the agent's retirement wealth is equal to X_T. In this context, X_T can be regarded as (i) a lump-sum that is paid out to the agent at $t = T$, or (ii) as a specific amount of monetary units that is converted into an annuity, which renders annual payments until the agent's date of death. Within the confines of a pension scheme, it is reasonable to assume that participants have particular expectations with regard to their retirement wealth. That is, the agent in our model setup may have in mind a certain benchmark or reference level that he/she ideally acquires at retirement. Note that such a reference level may also be relevant to agents outside of the pension industry, cf. the introduction.

To model this benchmark, we make use of a log-normally distributed random variable.[8] Suppose that Y_T represents the agent's person-specific (real) benchmark. Henceforth, with a slight abuse of notation, we set Y_t equal to $Y_T|_{T=t}$ for all $t \in [0,T]$. Let $Y_0 \in \mathbb{R}_+$ be the benchmark's starting value. Moreover, define α_t and β_t as two deterministic processes of time ($t \in [0,T]$) alone. Here, $t \mapsto \alpha_t$ achieves values in $\mathbb{R}$; $t \mapsto \beta_t$ attains values in $\mathbb{R}^4$. One could generalise α_t to be affine in Z_t. However, due to the log-normal structure of Y_T, the corresponding randomness generated by α_t can easily be handled by β_t alone. With this notation at hand, we postulate that the benchmark constitutes the solution to $\frac{\mathrm{d}Y_t}{Y_t} = \alpha_t \mathrm{d}t + \beta_t^\top \mathrm{d}W_t$ at time $t = T$. In other words, the value of the retirement-linked reference level is specified as follows:

$$Y_T = Y_0 \exp\left\{\int_0^T \alpha_s \mathrm{d}s - \frac{1}{2}\int_0^T \beta_s^\top \beta_s \mathrm{d}s + \int_0^T \beta_s^\top \mathrm{d}W_s\right\}. \tag{4.2.16}$$

We conclude by touching upon two features relevant to the benchmark in (4.2.16): (i) possible interpretations of Y_T, and (ii) the funding or coverage ratio corresponding to Y_T. Concerning item (i), as mentioned in Kamma and Pelsser (2022c), there is a wide variety of interpretations available for Y_T. We should stress here that these interpretations should respect the exogeneity of Y_T, meaning that Y_T must be left unaffected by the agent's own decisions. For example, we could interpret Y_T as one's labour income, the net worth of the agent's neighbour, or a fraction of a nation's GDP. As for item (ii), we note that the funding or coverage ratio corresponding to Y_T is defined as: $F_0 = \frac{X_0}{\mathbb{E}[Y_T M_T \Pi_T]}$. If $F_0 < 1$, the agent is not in possession of enough funds (assets) at $t = 0$ to risk-neutrally cover his/her benchmark (liabilities). The converse holds true if $F_0 \geq 1$. Taking into account that F_0 is linear in Y_0, it is clear that we can alter Y_0 to modify the coverage ratio. In the sequel, for technical reasons, we are limited to the $F_0 < 1$ case

[8] We assume log-normality of this benchmark process for two main reasons. First, due to its distributional elegance, the log-normal benchmark does not interfere with a closed-form characterisation of optimal solutions. Second, because it outlines a strictly positive process that can be identified with a.o. B_t and S_t, the log-normal specification allows for a fairly broad selection of interpretations. In the numerical illustration related to the investment problem of interest, we define the benchmark as the analytical approximation to a life annuity. In this definition of the benchmark, its log-normality plays a central role via the so-called Fenton-Wilkinson approximate method, cf. Fenton (1960).

4.2.4 Optimal Investment Problem

The subsequent investment problem is predicated on the LPM formulation that is addressed in section 5.2 of Föllmer and Leukert (2000). To supply this problem, we cast the agent's situation into the following context. The financial market, $\mathcal{M}$, occupies an agent, who is in possession of an initial endowment $X_0 \in \mathbb{R}_+$. The agent uses this entire prefixed amount of monetary units for investment over the trading interval, $[0, T]$. Over the course of this interval, the agent is allowed to continuously modify the weights of his/her portfolio. These portfolio weights, π_t, are defined as the amounts of monetary units that the agent allocates to the four risky assets, S_t. By modifying these weights, the agent aims to minimise the so-called lower partial moment of the difference between *real* retirement wealth, $\frac{X_T}{\Pi_T}$, and the (real) benchmark Y_T. In doing so, the agent tries to acquire an amount of real retirement wealth that is as close as possible to his/her goal: the benchmark. In mathematical terms, the agent faces the following investment problem:

$$\begin{aligned} \sup_{\{\pi_t\}_{t\in[0,T]}\in\mathcal{A}_{X_0}} \quad & \mathbb{E}\left[-\frac{1}{p}\left[\left(Y_T - \frac{X_T}{\Pi_T}\right)^+\right]^p\right] \\ \text{s.t.} \quad & \mathrm{d}X_t = \left(X_t r_t + \pi_t^\top \sigma \Lambda_t\right)\mathrm{d}t + \pi_t^\top \sigma \mathrm{d}W_t, \\ & X_0 < \sup_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}^{\mathbb{Q}}\left[\frac{Y_T \Pi_T}{B_T}\right] = \mathbb{E}\left[Y_T M_T \Pi_T\right]. \end{aligned} \tag{4.2.17}$$

Here, $p > 1$ holds and $\mathcal{Q}$ describes the set of all equivalent martingale measures under the numéraire $\{B_t\}_{t\in[0,T]}$. Notation-wise, we let $\mathbb{E}^{\mathbb{Q}}[\cdot]$ represent the expectation operator under the $\mathbb{Q}$ measure, and $(\cdot)^+$ stands for the max-operator: $(y)^+ = \max\{0, y\}$ for all $y \in \mathbb{R}$. If $p = 1$, the formulation in (4.2.17) reduces to the expected shortfall minimisation problem, cf. section 4 in Föllmer and Leukert (2000). However, $p = 1$ cannot be regarded as a special case of the optimal solution to (4.2.17), as $x \mapsto -\frac{[(x)^+]^p}{p}$ is only continuously differentiable for $p > 1$. Although the formulation in (4.2.17) is specified in the canonical form of a maximisation problem, it should be noted that it is equivalent to minimising $\mathbb{E}[\frac{1}{p}[(Y_T - \frac{X_T}{\Pi_T})^+]^p]$. The latter criterion characterises the p^{th} lower partial moment of $\frac{X_T}{\Pi_T} - Y_T$. The agent aims to minimise this criterion over all admissible trading strategies, $\{\pi_t\}_{t\in[0,T]} \in \mathcal{A}_{X_0}$, given that $X_0 < \mathbb{E}\left[Y_T M_T \Pi_T\right]$ holds true. The preceding condition states that X_0 ought to be smaller than the super-replication price of Y_T.[9] Since the market $\mathcal{M}$ is complete, this super-replication price coincides with the risk-neutral value of Y_T, given by $\mathbb{E}\left[Y_T M_T \Pi_T\right]$. Hence, since $X_0 < \mathbb{E}\left[Y_T M_T \Pi_T\right]$ is true, we have that $F_0 < 1$ holds. Observe that if $F_0 \geq 1$, X_0 exceeds the super-replication price, which enables $\frac{X_T}{\Pi_T}$ to exceed Y_T in all states of the world. As this makes problem (4.2.17) superfluous, we solely analyse the $F_0 < 1$ case.

4.3 Optimality Conditions

In this section, we analytically spell out and analyse the optimality conditions corresponding to the investment problem in (4.2.17). First, we provide a general formulation of the optimality conditions for the LPM problem in the spirit of Föllmer and Leukert (2000). For this reason, we employ the martingale method. Second, we introduce the optimal solutions to the dynamic optimisation problem (4.2.17) in closed-form. In line with the martingale method, these solutions are two-fold: (i) one part of the solution clearly concerns the optimal trading strategy, and (ii) the remaining part pertains to the agent's optimal wealth process. In view of the fact that item (i) follows directly from item (ii), we first focus on the latter and conclude with the former. All derivations relevant to this section can be found in the appendix.

4.3.1 General Optimal Solutions

In deriving the optimality conditions to the formulation in (4.2.17), we make use of the martingale method. This method is developed in the seminal contributions by Pliska (1986), Karatzas et al. (1987), Cox and Huang (1989, 1991), and revolves around a static alternative to the dynamic problem in (4.2.17). This approach slightly differs from the one that Föllmer and Leukert (2000) employ to solve (4.2.17). Whereas their approach treats the LPM operator as a hedging criterion, the martingale method emphasises its specification as

[9] For the definition of a super-replication price, cf. El Karoui and Quenez (1995), Kramkov (1996), Föllmer and Kabanov (1997), and Example 2 of Rogers (2003).

a utility function. To identify the LPM operator as a utility function, let us introduce the mapping $U : \mathbb{R}_+^2 \to \mathbb{R}_-$ defined by: $U(x, y) = -\frac{1}{p}\left[(y-x)^+\right]^p$, for all $x, y \in \mathbb{R}_+$. Clearly, $U : \mathbb{R}_+^2 \to \mathbb{R}_-$ qualifies as a utility function and $\mathbb{E}\left[U\left(X_T, \Pi_T\right)\right]$ spells out the objective of the problem in (4.2.17). We observe that the inverse of marginal utility is characterised by: $I(z, y) = y - z^{\frac{1}{p-1}} \wedge y$. In Theorem 4.3.1, we introduce the general optimality conditions for the investment problem in (4.2.17).

Theorem 4.3.1. *Consider the optimal dynamic LPM investment problem in* (4.2.17)*. The corresponding optimal wealth process, denoted by* X_t^{opt}*, is for all* $t \in [0, T]$ *characterised by:*

$$\begin{aligned} X_t^{\text{opt}} = & \frac{1}{M_t}\mathbb{E}\left[Y_T M_T^R \mathbb{1}_{\{\mathcal{A}_T\}} \mid \mathcal{F}_t\right] \\ & - \frac{1}{M_t}\mathbb{E}\left[\left(\mathcal{H}^{-1}\left(X_0\right)^{\frac{1}{p}} M_T^R\right)^{\frac{p}{p-1}} \mathbb{1}_{\{\mathcal{A}_T\}} \,\middle|\, \mathcal{F}_t\right]. \end{aligned} \tag{4.3.1}$$

Here, $\mathcal{H}^{-1} : \mathbb{R}_+ \to \mathbb{R}_+$ *outlines the inverse function of* $\mathcal{H} : \mathbb{R}_+ \to \mathbb{R}_+$*, which is specified as:* $\mathcal{H}(\eta) = \mathbb{E}\left[I\left(\eta M_T \Pi_T, Y_T\right)\Pi_T M_T\right] = X_0$*. The event* $\mathcal{A}_T$ *is given by:* $\mathcal{A}_T = \{Y_T \geq (\mathcal{H}^{-1}(X_0)\, M_T^R)^{\frac{1}{p-1}}\}$*. We define* M_t^R *as the real pricing kernel:* $M_t^R = M_t \Pi_t$*, for all* $t \in [0, T]$*. Consistent with* X_t^{opt}*, there exists an* $L^2([0, T])$*-valued process,* $\{\psi_t\}_{t\in[0,T]}$*, such that the optimal investment strategy,* π_t^{opt}*, reads for all* $t \in [0, T]$ *as:*

$$\pi_t^{\text{opt}} = \sigma^{\top^{-1}}\frac{\psi_t}{M_t} + \sigma^{\top^{-1}}\Lambda_t X_t^{\text{opt}}. \tag{4.3.2}$$

Proof. The proof is given in Appendix A. □

The specifications of X_t^{opt} and π_t^{opt} are valid for general Y_t and general return dynamics. More specifically, the presented optima in Theorem 4.3.1 still hold true for non-log-normal characterisations of Y_t, e.g. a semi-martingale definition. The same applies to the specification of S_t, i.e. the return dynamics. Nevertheless, for concrete setups, it is in general possible to spell out X_t^{opt} and π_t^{opt} more explicitly. In case of $\mathcal{M}$, one can indeed derive closed-form expressions for both processes. Due to the distributional features of M_t, the evaluation of the conditional expectations in (4.3.1) demands special care. We elaborate on the details in section 4.3.2. As the retrieval of ψ_t in (4.3.2) entirely depends on the (analytical) definition of X_t^{opt}, the distributional properties of M_t also play a role in the identification of π_t^{opt}. We discuss this at greater length in section 4.3.3. Note that ψ_t can be derived from X_t in closed-form by means of Itô's Lemma or Malliavin calculus. For details on the latter less well-known approach, we refer to Nualart (2006) or Appendix C.

4.3.2 Optimal Wealth Process

We continue with the introduction of the optimal wealth process corresponding to the dynamic investment problem in (4.2.17): X_t^{opt} for all $t \in [0, T]$. As addressed in the discussion of the general optimality conditions in section 4.3.1, the derivation of X_t^{opt} comes down to an evaluation of the conditional expectations in (4.3.1). To analytically evaluate these expectations, we are required to employ Fourier transforms. Namely, the stochastic processes included in this identity for X_t^{opt} are, for non-zero Λ_1, not log-normally distributed. In fact, the distributional features of these processes are not known, which encumbers a closed-form recovery of X_t^{opt} by means of standard machinery. Therefore, the Fourier transform comes in handy, because it enables us to evaluate the conditional expectations in (4.3.1), without requiring the exact distributional properties of the relevant processes. Details concerning the Fourier transform in application to (4.3.1) can be found in Appendix B.1. In Proposition 4.3.2, we formally introduce the analytical expression for X_t^{opt}.

Proposition 4.3.2. *Consider the optimal dynamic investment problem in* (4.2.17)*. The corresponding value for the optimal wealth process,* X_t^{opt}*, at time* $t \in [0, T]$ *is given in* (4.3.1)*. After analytical evaluation of the conditional expectations in this identity, the following holds:*

$$\begin{aligned} X_t^{\mathrm{opt}} &= Y_t \Pi_t P_1(t, Z_t) \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\kappa^*(T, \omega)\, \phi_{1,T-t}(-\omega - i\kappa, h)\, \mathrm{d}\omega \\ &\quad - \left(\eta^{\mathrm{opt}} M_t \Pi_t\right)^{\frac{1}{p-1}} \Pi_t P_2(t, Z_t) \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\kappa^*(T, \omega)\, \phi_{2,T-t}(-\omega - i\kappa, h)\, \mathrm{d}\omega, \end{aligned} \tag{4.3.3}$$

for all $t \in [0, T]$*. Here,* $\eta^{\text{opt}} = \mathcal{H}^{-1}(X_0) \in \mathbb{R}_+$ *characterises the optimal Lagrange multiplier, which can be retrieved from solving* $X_0^{\text{opt}} = X_0$ *for* η^{opt}*. Moreover,* $P_1(t, Z_t)$ *and* $P_2(t, Z_t)$ *are for all* $t \in [0, T]$ *specified as:*

$$\begin{aligned} P_1(t, Z_t) &= \exp\left\{\tilde{A}(t) + \tilde{B}(t)^\top Z_t\right\}, \text{ and} \\ P_2(t, Z_t) &= \exp\left\{\widehat{A}(t) + \widehat{B}(t)^\top Z_t + Z_t^\top \widehat{C}(t) Z_t\right\}. \end{aligned} \tag{4.3.4}$$

Here, $t \mapsto \tilde{A}(t)$ *and* $t \mapsto \tilde{B}(t)$ *outline two deterministic functions that are given in* (B.1.13)*. Likewise,* $t \mapsto \widehat{A}(t)$*,* $t \mapsto \widehat{B}(t)$ *and* $t \mapsto \widehat{C}(t)$ *spell out three deterministic functions that jointly solve the system of ODE's in* (B.1.14)*. The deterministic function* $f_\kappa^*(T, \omega)$ *is for all* $\omega \in \mathbb{R}$ *and some* $\kappa \in \mathbb{R}_+$ *given by:* $f_\kappa^*(T, \omega) - \frac{e^{(i\omega - \kappa)\frac{1}{p-1}\log \mathcal{H}^{-1}(X_0)}}{i\omega - \kappa}$*. Ultimately, let* $Q_j(t, Z_t, \omega) = e^{\bar{A}_j(t,\omega) + \bar{B}_j(t,\omega)^\top Z_t + Z_t^\top \bar{C}_j(t,\omega) Z_t}$ *for* $j = 1, 2$*, all* $\omega \in \mathbb{R}$ *and* $t \in [0, T]$*. The functions,* $(t, \omega) \mapsto \bar{A}_j(t, \omega)$*,* $(t, \omega) \mapsto \bar{B}_j(t, \omega)$ *and* $(t, \omega) \mapsto \bar{C}_j(t, \omega)$*, jointly solve the system of ODE's in* (B.1.29)*, for* $j = 1, 2$*. Then, for* $j = 1, 2$*, all* $\omega \in \mathbb{R}$ *and* $t \in [0, T]$*, the characteristic function is given by:*

$$\phi_{j,T-t}(\omega, h) = Q_j(t, Z_t, \omega)\left(M_t^{R^{-\frac{1}{p-1}}} Y_t\right)^{i\omega}. \tag{4.3.5}$$

Proof. The proof is given in Appendix B.1. □

The expression for optimal wealth over the trading interval, X_t^{opt} in (4.3.3), can be analysed along the following technical lines. The $Y_t \Pi_t P_1(t, Z_t)$ term coincides with the risk-neutral value of $Y_T \Pi_T$ at time $t \in [0, T]$. Similarly, the $(\mathcal{H}^{-1}(X_0) M_t \Pi_t)^{\frac{1}{p-1}} \Pi_t P_2(t, Z_t)$ term identifies the risk-neutral value of $\left(\mathcal{H}^{-1}(X_0) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T$ at time $t \in [0, T]$. Moreover, the two integral expressions, $\frac{1}{2\pi}\int_{-\infty}^{\infty} f_\kappa^*(T, \omega)\, \phi_{j,T-t}(-\omega - i\kappa, h)\, \mathrm{d}\omega$ for $j = 1, 2$, specify the Fourier transforms of two distinct conditional probabilities. These conditional probabilities concern the $\mathcal{A}_T$ event under different (nearly-identical) probability measures that are equivalent to $\mathbb{P}$, i.e. $\mathbb{X}_1$ and $\mathbb{X}_2$ in Appendix B.1. It is clear that these conditional probabilities determine how much weight is attached to the risk-neutral value of $Y_T \Pi_T - \left(\mathcal{H}^{-1}(X_0) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T$, at time $t \in [0, T]$. That is, the better the state of the economy, the closer these probabilities will be to 1;

for the converse case, these probabilities will attain values near 0. Hence, X_t^{opt} can be regarded as the risk-neutral value of $Y_T\Pi_T - (\mathcal{H}^{-1}(X_0) M_T^R)^{\frac{1}{p-1}} M_T$, weighted in accordance with the state of the economy. In Corollary 4.3.3, we introduce X_t^{opt} for the case where Λ_1 is equal to 0.

Corollary 4.3.3. *Consider the optimal dynamic investment problem in* (4.2.17)*. The corresponding optimal wealth process,* X_t^{opt}*, is given in* (4.3.3) *of Proposition 4.3.2. Suppose that* $\Lambda_1 = 0_{4\times 2}$*, where* $0_{4\times 2} = [0_4, 0_4]$ *and* 0_4 *is an* $\mathbb{R}^4$*-valued vector of zeros. Then,*

$$\begin{aligned} X_t^{\text{opt}} &= Y_t \Pi_t e^{\tilde{A}(t) + \tilde{B}(t)^\top Z_t} \Phi\left(d_{1,t,T}\right) \\ &\quad - \left(\mathcal{H}^{-1}\left(X_0\right) M_t \Pi_t\right)^{\frac{1}{p-1}} \Pi_t e^{\widehat{A}(t) + \widehat{B}(t)^\top Z_t} \Phi\left(d_{2,t,T}\right), \end{aligned} \tag{4.3.6}$$

characterises for all $t \in [0,T]$ *the optimal wealth process. Here,* $t \mapsto \tilde{A}(t)$ *and* $t \mapsto \tilde{B}(t)$ *outline two deterministic functions that are given in* (B.2.32)*. Likewise,* $t \mapsto \widehat{A}(t)$ *and* $t \mapsto \widehat{B}(t)$ *spell out two deterministic functions that are given in* (B.2.33)*. Additionally,* $\Phi(\cdot)$ *represents the cumulative distribution function (CDF) of a random variable that is standard normally distributed. The two arguments inside this function, i.e. the processes* $d_{1,t,T}$ *and* $d_{2,t,T}$*, are for* $j = 1, 2$ *and all* $t \in [0,T]$ *given by the following identity:*

$$d_{j,t,T} = \frac{-\log\left(\frac{\mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}}{M_t^{R^{-\frac{1}{p-1}}} Y_t}\right) + \mathbb{E}^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t\right]}{\sqrt{\mathbb{V}ar^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t\right]}}. \tag{4.3.7}$$

Here, $\mathbb{V}ar^{\mathbb{X}_j}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T / M_t^{R^{-\frac{1}{p-1}}} Y_t \,\middle|\, \mathcal{F}_t\right]$ *and* $\mathbb{E}^{\mathbb{X}_j}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T / M_t^{R^{-\frac{1}{p-1}}} Y_t \,\middle|\, \mathcal{F}_t\right]$ *are given in* (B.2.40) *and* (B.2.41)*, respectively, for* $j = 1, 2$ *and all* $t \in [0,T]$.

Proof. The proof is given in Appendix B.2. □

As addressed in the discussion preceding Proposition 4.3.2, an analytical evaluation of the conditional expectations in (4.3.1) is complicated by the distributional features of the relevant processes. In particular, for non-zero Λ_1, these distributional properties are unknown. To arrive at an analytical expression for X_t^{opt}, we therefore employed the Fourier transform, cf. Proposition 4.3.2. However, if Λ_1 is equal to zero, the distributional properties of both $Y_T M_T^R$ and $(\mathcal{H}^{-1}(X_0)^{\frac{1}{p}} M_T^R)^{\frac{p}{p-1}}$ are given. In precise terms, both of the latter processes are log-normally distributed. This result can be attributed to the fact that Λ_t and $\Lambda_t^R = \Lambda_t - \sigma_\Pi$ spell out non-affine constants, as a consequence of $\Lambda_1 = 0_{4\times 2}$. On the grounds of this log-normality, the closed-form evaluation of the conditional expectations in (4.3.1) is significantly facilitated. Furthermore, the expression for X_t^{opt} is more tractable, as it neither depends on Fourier transforms nor on analytically unsolvable systems of ODE's. Note that the analysis regarding the weighted risk-neutral value of $Y_T\Pi_T - \left(\mathcal{H}^{-1}(X_0) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T$ from (4.3.3) also applies to (4.3.6). The sole difference consists in the known distributional features of the probability weights.

Remark 4.3.1. *In spite of its analytical transparency, the expression for optimal wealth (X_t^{opt}) in* (4.3.3) *can be computationally challenging. Namely, in addition to $P_2(t, Z_t)$'s dependence on an analytically unsolvable matrix Riccati differential equation for $t \mapsto \widehat{C}(t)$, both $\phi_{1,T-t}$ and $\phi_{2,T-t}$ depend on similar Riccati equations. As $\phi_{1,T-t}$ and $\phi_{2,T-t}$ are parts of the distinct integrals in* (4.3.3)*, their dependence on a.o. two separate Riccati equations may complicate computational analyses. Therefore, to facilitate numerical evaluations of the integral(s) characterising X_t^{opt}, we provide a computationally friendlier expression for X_t^{opt} (at the cost of analytical clarity). This expression is based on a direct application of the Fourier transform to X_t^{opt} in* (4.3.1). *Define $\widehat{f}_\kappa^*(T,\omega) = \frac{e^{(i\omega-\kappa+1)\frac{1}{p-1}\log \mathcal{H}^{-1}(X_0)}}{(i\omega-\kappa+1)[i\omega-\kappa]}$, for some $\kappa > 1$, and $\widehat{\phi}_{T-t}(\omega, g, j) = \widehat{Q}(t, Z_t, \omega)\, M_t^{R\,\frac{p-i\omega}{p-1}} Y_t^{i\omega}$, cf.* (B.3.48) *and* (B.3.49) *for the definition of $\widehat{Q}$. Then*[10]*, for all $t \in [0,T]$, X_t^{opt} is specified as follows:*

$$X_t^{\text{opt}} = \frac{1}{2\pi}\frac{1}{M_t}\int_{-\infty}^{\infty} \widehat{f}_\kappa^*(T,\omega)\,\widehat{\phi}_{T-t}(-\omega - i\kappa, g, j)\,\mathrm{d}\omega. \tag{4.3.8}$$

[10] For the full derivation of X_t^{opt} in (4.3.8), see Appendix B.3. We stress here that the expressions in (4.3.3) and (4.3.8) naturally result in the same optimal wealth processes (X_t^{opt}).

4.3.3 Optimal Trading Strategy

In this section, we present the optimal trading strategy that solves the dynamic investment problem in (4.2.17): π_t^{opt} for all $t \in [0, T]$. According to the analysis in section 4.3.1, π_t^{opt} can be retrieved on the basis of the analytical specification for X_t^{opt}. To emphasise the link between π_t^{opt} and X_t^{opt}, let us turn to equation (4.3.2) for π_t^{opt}, and observe that we are already in possession of X_t^{opt}, cf. Proposition 4.3.2 and Corollary 4.3.3. As a result, the mere unknown in this equation is the process $\{\psi_t\}_{t\in[0,T]}$. Note here that $\{\psi_t\}_{t\in[0,T]}$ outlines the integrand in $X_t^{\text{opt}} M_t$'s martingale representation and consequently depends on X_t^{opt}. Therefore, in order to identify π_t^{opt}, we utilise X_t^{opt} to derive the expression for this integrand by means of Itô's Lemma or Malliavin calculus. Details concerning the applications of the latter two concepts to X_t^{opt} in (4.3.3) and (4.3.6) are provided in Appendices C.1 and C.2, respectively. In Proposition 4.3.4, we present the ensuing final expression for π_t^{opt}.

Proposition 4.3.4. *Consider the optimal dynamic investment problem in* (4.2.17)*. The optimal trading strategy, i.e. the solution of* (4.2.17)*, is given in* (4.3.2)*. In this identity, the expression for* X_t^{opt} *can be found in* (4.3.3)*. After determining* $\{\psi_t\}_{t\in[0,T]}$*,* π_t^{opt} *can be decomposed in the following way:* $\pi_t^{\text{opt}} = \pi_t^M + \pi_t^\Pi + \pi_t^Y + \pi_t^{FT} + \pi_t^R$*, for all* $t \in [0, T]$*.*

The first two weights in this portfolio decomposition, π_t^M *and* π_t^R*, read:*

$$\begin{aligned} \pi_t^M &= \frac{\sigma^{\top^{-1}} \Lambda_t}{p-1} \left(\mathcal{H}^{-1} \left(X_0 \right)^{\frac{1}{p}} M_t^R \right)^{\frac{p}{p-1}} \frac{P_2\left(t, Z_t\right)}{M_t} R_{2,t}, \\ \pi_t^\Pi &= \sigma^{\top^{-1}} \sigma_\Pi X_t^{\text{opt}} - \frac{\sigma^{\top^{-1}} \sigma_\Pi}{p-1} \left(\mathcal{H}^{-1} \left(X_0 \right)^{\frac{1}{p}} M_t^R \right)^{\frac{p}{p-1}} \frac{P_2\left(t, Z_t\right)}{M_t} R_{2,t}, \end{aligned} \tag{4.3.9}$$

for all $t \in [0, T]$*. Here,* $R_{j,t}$ *is defined as:* $R_{j,t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\kappa^*\left(T, \omega\right) \phi_{j,T-t}\left(-\omega - i\kappa, h\right) \mathrm{d}\omega$*, for* $j = 1, 2$*, and all* $t \in [0, T]$*. The deterministic function* $f_\kappa^*\left(T, \omega\right)$ *is spelled out in Proposition 4.3.2. Likewise, the characteristic functions,* $\phi_{j,T-t}\left(\omega, h\right)$ *for* $j = 1, 2$*, are specified in* (4.3.5)*. Now, let* $\widehat{R}_{j,t} = \frac{1}{2\pi} P_j\left(t, Z_t\right) \int_{-\infty}^{\infty} \left(f_\kappa^*\left(T, \omega\right) \phi_{1,T-t}\left(-\omega - i\kappa, h\right) [\Sigma_Z^\top \bar{B}_j\left(t, -\omega - i\kappa\right) + 2\Sigma_Z^\top \bar{C}_j\left(t, -\omega - i\kappa\right) Z_t + i\left(-\omega - i\kappa\right)\left(\beta_t + \frac{1}{p-1}\Lambda_t^R\right)]\right)\mathrm{d}\omega$*, for* $j = 1, 2$*, and* *all* $t \in [0,T]$*, where* $(t,\omega) \mapsto \bar{B}_j(t,\omega)$ *and* $(t,\omega) \mapsto \bar{C}_j(t,\omega)$ *solve the multidimensional system of non-linear ODE's in* (B.1.29)*.*

Then, the remaining weights, π_t^Y*,* π_t^{FT} *and* π_t^R*, in* π_t^{opt}*'s decomposition read:*

$$\begin{aligned}
\pi_t^Y &= \left(\sigma^{\top^{-1}}\beta_t\right) Y_t \Pi_t P_1(t, Z_t) R_{1,t}, \\
\pi_t^{FT} &= \sigma_t^{\top^{-1}}\left[Y_t\Pi_t\widehat{R}_{1,t} - \left(\mathcal{H}^{-1}(X_0)^{\frac{1}{p}} M_t^R\right)^{\frac{p}{p-1}} \frac{1}{M_t}\widehat{R}_{2,t}\right], \\
\pi_t^R &= \sigma^{\top^{-1}}\Sigma_Z^\top\left(\tilde{B}(t) X_t^{\text{opt}} - \tilde{D}(t)\left(\mathcal{H}^{-1}(X_0)^{\frac{1}{p}} M_t^R\right)^{\frac{p}{p-1}} P_2(t,Z_t) R_{2,t}\right),
\end{aligned} \tag{4.3.10}$$

for all $t \in [0,T]$*, where the mapping* $t \mapsto \tilde{B}(t)$ *is given in* (B.1.13)*. In addition to this, the mapping* $t \mapsto \tilde{D}(t)$ *is defined as follows:* $\tilde{D}(t) = \widehat{B}(t) + 2\widehat{C}(t) Z_t + \tilde{B}(t)$*, for all* $t \in [0,T]$*. In the latter identity,* $t \mapsto \widehat{B}(t)$ *and* $t \mapsto \widehat{C}(t)$ *jointly solve the multidimensional system of non-linear ODE's in* (B.1.14)*.*

Proof. The proof is given in Appendix C.1. □

In Proposition 4.3.4, we decompose the optimal trading strategy (π_t^{opt}) into five distinct hedge demands. In disentangling these demands, we primarily adhere to the decomposition principles proposed by Detemple and Rindisbacher (2010) and Li et al. (2020). These papers concentrate on utility functions of the conventional Inada-family. Although the LPM function in the objective of (4.2.15) cannot be entirely subsumed under this family, the corresponding optimal portfolio weights live by a structure similar to theirs. We identify the demands in (4.3.9), π_t^M and π_t^Π, as a variant of the optimal mean-variance portfolio and a CPI hedge, respectively. As for π_t^M, it is in this regard noteworthy that $\frac{\sigma^{\top^{-1}}\Lambda_t}{\gamma}$ corresponds to the Merton portfolio for $\gamma = p-1 > 0$, cf. Merton (1969, 1971). In a similar sense, π_t^Π incorporates Merton-like weights that depend on σ_Π instead of Λ_t. The remaining demands in (4.3.10), π_t^Y, π_t^{FT} and π_t^R, can be identified as a benchmark hedge, a probability hedge, and a real interest rate hedge, respectively. Due to their dependence on β_t and the parameters in $R_t = r_t - \pi_t + \sigma_\Pi^\top\Lambda_t$, the identification of π_t^Y and π_t^R is straightforward. The π_t^{FT} weight is referred to as a probability hedge, as its components follow from the Fourier transforms in (4.3.3) that characterise the relevant conditional probabilities. In Corollary 4.3.5, we introduce the optimal trading strategy, π_t^{opt}, given that $\Lambda_1 = 0_{4\times 2}$ holds.

Corollary 4.3.5. *Consider the optimal dynamic investment problem in* (4.2.17)*. The optimal trading strategy, i.e. the solution of* (4.2.17)*, is given in* (4.3.2)*. Suppose that* $\Lambda_1 = 0_{4\times 2}$*, and* $\widehat{M}_t^R = \mathcal{H}^{-1}(X_0)^{\frac{1}{p}} M_t^R$*. Then,* $\pi_t^{\text{opt}} = \pi_t^M + \pi_t^\Pi + \pi_t^Y + \pi_t^R$ *holds, for all* $t \in [0,T]$*. Here,* π_t^M *and* π_t^Π *read:*

$$\begin{aligned} \pi_t^M &= \frac{\sigma^{\top^{-1}}\Lambda_0}{p-1}\left(\widehat{M}_t^R\right)^{\frac{p}{p-1}} \frac{e^{\widehat{A}(t)+\widehat{B}(t)^\top Z_t}}{M_t}\Phi\left(d_{2,t,T}\right), \\ \pi_t^\Pi &= \sigma^{\top^{-1}}\sigma_\Pi X_t^{\text{opt}} - \frac{\sigma^{\top^{-1}}\sigma_\Pi}{p-1}\left(\widehat{M}_t^R\right)^{\frac{p}{p-1}} \frac{e^{\widehat{A}(t)+\widehat{B}(t)^\top Z_t}}{M_t}\Phi\left(d_{2,t,T}\right), \end{aligned} \tag{4.3.11}$$

for all $t \in [0,T]$*. The mappings* $t \mapsto \tilde{A}(t)$ *and* $t \mapsto \tilde{B}(t)$ *correspond to the deterministic functions in Corollary 4.3.3 and are given in* (B.2.32)*. Similarly, the mappings* $t \mapsto \widehat{A}(t)$ *and* $t \mapsto \widehat{B}(t)$ *follow from Corollary 4.3.3 and are presented in* (B.2.33)*. Moreover, the function* $\Phi(\cdot)$ *denotes the CDF of a standard normally distributed random variable. The arguments of this function,* $d_{1,t,T}$ *and* $d_{2,t,T}$*, are defined in equation* (4.3.7)*. Suppose that* $\tilde{D}(t) = \widehat{B}(t) + \tilde{B}(t)$*, for all* $t \in [0,T]$*. Then, the remaining weights,* π_t^Y *and* π_t^R*, in the decomposition of* π_t^{opt} *are specified, for all* $t \in [0,T]$*, as follows:*

$$\begin{aligned} \pi_t^Y &= \left(\sigma^{\top^{-1}}\beta_t\right) Y_t \Pi_t e^{\tilde{A}(t)+\tilde{B}(t)^\top Z_t}\Phi\left(d_{1,t,T}\right), \\ \pi_t^R &= \sigma^{\top^{-1}}\Sigma_Z^\top\left(\tilde{B}(t)\, X_t^{\text{opt}} - \tilde{D}(t)\left(\widehat{M}_t^R\right)^{\frac{p}{p-1}} e^{\widehat{A}(t)+\widehat{B}(t)^\top Z_t}\Phi\left(d_{2,t,T}\right)\right). \end{aligned} \tag{4.3.12}$$

Proof. The proof is given in Appendix C.2. □

Corollary 4.3.5 spells out the optimal trading strategy (π_t^{opt}) corresponding to the optimal wealth process (X_t^{opt}) in Corollary 4.3.5. That is, the optimal portfolio weights in (4.3.11) and (4.3.12) precisely replicate the optimal wealth process in (4.3.6). The analytical tractability of this wealth process correspondingly carries over to the optimal trading strategy. Concretely, unlike the portfolio in Proposition 4.3.4, the optimal investment rules for $\Lambda_1 = 0_{4\times 2}$

neither depend on Fourier transforms nor on analytically troublesome multi-dimensional systems of (non-)linear ODE's. Due to the interdependent links between X_t^{opt} and π_t^{opt}, it is self-explanatory that this tractability is attributable to the advantageous distributional features of X_T^{opt}. As for the economic interpretations of the separate demands, we underline that the weights in (4.3.11) and (4.3.12) represent the same hedges as those in Proposition 4.3.4. Note that the hedge demand for the probability weights, π_t^{FT} in (4.3.10), is equal to 0_4 for $\Lambda_1 = 0_{4\times 2}$. We are able to derive this result through an application of Malliavin calculus.[11] However, because of the analytical structure of X_T^{opt} in (A.3), a similar application for the general $\Lambda_1 \in \mathbb{R}^{4\times 2}$ case does not infer that $\pi_t^{FT} = 0_4$ must hold.[12] Therefore, Proposition 4.3.4 includes this hedge demand as a possibly non-zero one.

Remark 4.3.2. *In the spirit of Remark 4.3.1, we observe that the expression for the optimal trading strategy (π_t^{opt}) in Proposition 4.3.4 may pose computational difficulties. Namely, the dependencies in X_t^{opt} on multiple matrix Riccati differential equations are also present in the latter specification of π_t^{opt}. Hence, consistent with* (4.3.8)*, in order to facilitate numerical evaluations of the integral(s) outlining π_t^{opt}, we provide a computationally friendlier expression for this process. We derive this expression on the grounds of the identity for X_t^{opt} in equation* (4.3.8)*. For this purpose, we employ the definitions of $\widehat{f}_\kappa^*(T,\omega)$ and $\widehat{\phi}_{T-t}(\omega,g,j)$ given in Remark 4.3.1. In addition to this, we fix $\widehat{D}_{\widehat{Q}}(t,\omega) = \widehat{B}_{\widehat{Q}}(t,\omega) + 2\widehat{C}_{\widehat{Q}}(t,\omega)\,Z_t$, for all $t \in [0,T]$ and $\omega \in \mathbb{R}$. Here, the definitions of $(t,\omega) \mapsto \widehat{B}_{\widehat{Q}}(t,\omega)$ and $(t,\omega) \mapsto \widehat{C}_{\widehat{Q}}(t,\omega)$ can be found in* (B.3.49)*. Moreover, we set $\widehat{\Lambda}_t^R(p,\omega) = i\omega\beta_t - \frac{p-i\omega}{p-1}\Lambda_t^R$. Then*[13]*, the optimal*

[11] This application strongly depends on the result reported in Lakner and Nygren (2006), concerning Mallivain-differentiability of piecewise continuously differentiable functions.

[12] Such an application ultimately requires one to evaluate a conditional expectation of the following form: $\mathbb{E}\left[N_T \mathbb{1}_{\{\mathcal{A}_T\}} \int_t^T \chi_s \mathrm{d}W_s \;\middle|\; \mathcal{F}_t\right]$, where $N_T \in L_+^0(\Omega)$ and $\{\chi_t\}_{t\in[0,T]}$ represents a deterministic process satisfying $\chi_t \in L^2(0,T)$. Although this conditional expectation can be evaluated by employing the Fourier transform, it does not enable us to determine whether the ensuing expression contributes to the π_t^{FT} weight or not. Hence, by means of the Malliavin-based approach, we are unable to state that $\pi_t^{FT} = 0_4$ holds for general values of $\Lambda_1 \in \mathbb{R}^{4\times 2}$. Note that this condition, $\pi_t^{FT} = 0_4$, neither follows from the expression for π_t^{FT} in Proposition 4.3.4 itself.

[13] For the full derivation of π_t^{opt} in (4.3.13), see Appendix C.3. As in Remark 4.3.1, we stress here that the expressions in Proposition 4.3.4 and (4.3.13) naturally result in the same optimal portfolio (π_t^{opt}).

trading strategy, π_t^{opt}, is for all $t \in [0, T]$ given by:

$$\begin{aligned}\pi_t^{\text{opt}} = \frac{1}{2\pi}\frac{1}{M_t}\sigma^{\top^{-1}} \int_{-\infty}^{\infty} \Big(\widehat{f}_\kappa^* \left(T, \omega\right) \widehat{\phi}_{T-t} \left(-\omega - i\kappa, g, j\right) \\ \times \left[\Sigma_Z^\top \widehat{D}_{\widehat{Q}} \left(t, -\omega - i\kappa\right) + \widehat{\Lambda}_t^R \left(p, -\omega - i\kappa\right)\right]\Big) \, \mathrm{d}\omega + \sigma^{\top^{-1}} \Lambda_t X_t^{\text{opt}}.\end{aligned} \tag{4.3.13}$$

4.4 Numerical Analysis

In this section, we provide a numerical analysis of the optimal solutions to the investment problem in (4.2.15). A numerical examination of the closed-form solutions (cf. section 4.3) can aid and deepen our understanding of a.o. their distributional properties and parameter sensitivity. Subsequently, we first elaborate on the specification of the benchmark process, $\{Y_t\}_{t\in[0,T]}$. Thereby, we aim to cast the matter into the confines of the accumulation phase of a defined contribution (DC) pension scheme. Second, we numerically investigate the probability distribution of the optimal terminal wealth process. In particular, we are interested in the likelihood of $\frac{X_T}{\Pi_T}$ attaining Y_T. Third and last, we study the behaviour of the optimal trading strategy with respect to changes in the model parameters. We particularly concentrate on the strategy's sensitivity with respect to the so-called replacement ratio.

4.4.1 Benchmark and Life Annuity

Henceforth, we assume that the investment problem in (4.2.17) corresponds to a participant in a DC scheme. Accordingly, we choose to model the benchmark, $\{Y_t\}_{t\in[0,T]}$, as a life annuity. As a consequence, the agent or pension fund in (4.2.17) aims to invest in such a manner that real retirement wealth ($\frac{X_T}{\Pi_T}$) completely covers the life annuity (Y_T). Note here that the agent departs from an underfunding situation ($F_0 < 1$), similar to most pension funds. The setup is therefore highly relevant. Now, we postulate that the participant lives τ_A fixed years after his/her predetermined retirement date, T. This concretely implies that the trading interval, $[0, T]$, coincides with the participant's accumulation phase. Correspondingly, the interval that aligns with the participant's decumulation phase is given by $(T, T+\tau_A]$. In the sequel, we suppose that a_T represents the *real* value of the life annuity at time $t = T$. The formal equation for the (real) value of this life annuity, a_T, reads:

$$a_T = C \sum_{i=1}^{\tau_A} \frac{P^R_{T,T+i}}{\Pi_T} = C \sum_{i=1}^{\tau_A} \exp\left\{ A^R(i) + B^R(i)^\top Z_T \right\}. \tag{4.4.1}$$

The specifications of $A^R(\cdot)$ and $B^R(\cdot)$ are provided in (4.2.13). Without loss of generality[14], we assume here that the life annuity pays $C \in \mathbb{R}_+$ monetary units per annum in real terms. We mainly employ C to adjust the coverage or funding ratio (F_0). In Donnelly et al. (2022), the authors make use of a similar definition for the value of a life annuity. Noting that Z_T outlines a normally distributed random variable, it is clear that the expression for a_T identifies a sum of τ_A log-normally distributed processes. Therefore, a_T in (4.4.1) is not log-normally distributed, and cannot be directly identified with Y_T in (4.2.16), cf. Dufresne (2008). Nevertheless, in order to make this identification possible, we assemble an approximation to a_T. For this purpose, we resort to a modified application of the Fenton-Wilkinson (FW) method developed by Fenton (1960). This method is predicated on the observation that sums of log-normally distributed random variables are *approximately* log-normal. Applied to a_T in equation (4.4.1), this method proceeds as follows.

Now, let $\widehat{a}_T$ denote the log-normal approximation to a_T. Then,

$$\widehat{a}_T = C \exp\left\{ \bar{\alpha} + \bar{\beta}^\top Z_T \right\} = C \exp\left\{ \bar{\alpha} + \bar{\beta}^\top \int_0^T e^{-K_Z(T-s)} \Sigma_Z \mathrm{d}W_s \right\}, \tag{4.4.2}$$

must hold according to the FW method. In this definition of $\widehat{a}_T$, the scalar $\bar{\alpha} \in \mathbb{R}$ and the vector $\bar{\beta} \in \mathbb{R}^2$ are to be determined. Ordinarily, the FW method characterises both $\bar{\alpha}$ and $\bar{\beta}$ by matching the first and second moments of $\widehat{a}_T$ and a_T. However, as $\bar{\beta}$ is two-dimensional, this procedure is not appropriate.

[14] Deterministic variable payments can be incorporated into a_T's definition by replacing $A^R(i)$ with $\widehat{A}^R(i) = A^R(i) + \log \frac{C_i}{C}$, given $C_i \in \mathbb{R}_+$, for all $i = 1, \dots, \tau_A$. In a similar manner, stochastic payments can be included in a_T's specification, as long as these respect the log-normality of $\exp\left\{ A^R(i) + B^R(i)^\top Z_T \right\}$. For instance, $C = \alpha S_T$, with $\alpha \in \mathbb{R}_+$, would be an appropriate candidate.

Therefore, to be able to specify the preceding unknowns, we slightly modify the FW approach. In particular, we still match the first moments. Yet, instead of the second moments, we also match the following two expectations: $\mathbb{E}\left[a_T Z_T\right]$ and $\mathbb{E}\left[\widehat{a}_T Z_T\right]$. The idea underscoring the latter operation is that it is identical to fixing: $\frac{\partial}{\partial \beta}\mathbb{E}\left[\widehat{a}_T\right] = \sum_{i=1}^{\tau_A} \frac{\partial}{\partial B^R(i)}\mathbb{E}\left[\exp\left\{A^R\left(i\right) + B^R\left(i\right)^\top Z_T\right\}\right]$. That is, we match in expectation the variations of a_T and $\widehat{a}_T$ with respect to the loadings of Z_T. Note here that both $\mathbb{E}\left[a_T\right]$ and $\mathbb{E}\left[a_T Z_T\right]$ are completely available in closed-form and can be found in equation (D.1.5) and equation (D.1.8), respectively. Based on this mildly adjusted version of the conventional FW method, we are able to determine $\bar{\alpha}$ and $\bar{\beta}$ in closed-form. For the complete derivation of both $\bar{\alpha}$ and $\bar{\beta}$ in (4.4.3), we refer the reader to Appendix D.1. In precise terms, we find that $\bar{\alpha}$ and $\bar{\beta}$ are characterised by the following identities:

$$\begin{aligned}
\bar{\alpha} &= \log\frac{\mathbb{E}\left[a_T\right]}{C} - \frac{1}{2}\int_0^T \bar{\beta}^\top e^{-K_Z(T-s)}\left[e^{-K_Z(T-s)}\right]^\top \bar{\beta}\mathrm{d}s, \\
\bar{\beta} &= \left(\int_0^T e^{-K_Z(T-s)}\left[e^{K_Z(T-s)}\right]^\top \mathrm{d}s\right)^{-1}\frac{\mathbb{E}\left[a_T Z_T\right]}{\mathbb{E}\left[a_T\right]}.
\end{aligned} \tag{4.4.3}$$

To illustrate the accuracy of this approximation, we examine the magnitude of the $L^2\left(\Omega\right)$-distance between $\widehat{a}_T$ and a_T: $\mathbb{E}\left[\left(a_T - \widehat{a}_T\right)^2\right]^{\frac{1}{2}}$. Additionally, we assess the size and sign of the correlation between $\widehat{a}_T$ and a_T: $\mathbb{C}\mathrm{orr}\left(\widehat{a}_T, a_T\right)$. We compute these quantities by means of simulations, to avoid presenting additional complicated expressions. For this reason, we make use of the baseline parameter initialisation provided in Table 4.1. The subsequent numbers are based on 10,000 simulated paths and an Euler scheme with 10 yearly equidistant time-points. Regarding the $L^2\left(\Omega\right)$-distance, we find: $\left\|a_T - \widehat{a}_T\right\|_{L^2(\Omega)} = 0.0923$.[15] For the correlation, we have: $\mathbb{C}\mathrm{orr}\left(a_T, \widehat{a}_T\right) = 0.9994$. Considering that this coefficient is nearly equal to 1, and that the corresponding $L^2\left(\Omega\right)$-distance is negligibly small, we conclude that the approximation to a_T is rather accurate. To finalise this section, we identify $\widehat{a}_T$ with Y_T in (4.2.16). The following unique specifications ensure that $Y_T = \widehat{a}_T$ holds: $Y_0 = Ce^{\bar{\alpha}}$, $\alpha_t = \frac{1}{2}\beta_t^\top\beta_t$ and

[15] We observe that $\inf_{\bar{\alpha},\bar{\beta}}\left\|a_T - \widehat{a}_T\right\|_{L^2(\Omega)} = 0.0897$ holds. As this value is very close to the one in the main text (0.0923), we can conclude that our approximation behaves similar to the $L^2\left(\Omega\right)$-optimal variant.

$\beta_t^\top = \bar{\beta}^\top e^{-K_Z(T-t)}\Sigma_Z$, for all $t \in [0, T]$.

4.4.2 Distribution of Retirement Wealth

In this section, we derive the distributional properties of optimal retirement wealth, X_T^{opt}. From section 4.2.4, we know that the agent in the LPM problem is strongly target-oriented. This focus on a person-specific target or benchmark can be summarised as follows.[16] As long as $\frac{X_T}{\Pi_T} < Y_T$ holds, the agent is able to draw additional non-negative utility by increasing the magnitude of $\frac{X_T}{\Pi_T}$. However, if $\frac{X_T}{\Pi_T} = Y_T$ is true, no additional utility can be derived by enlarging $\frac{X_T}{\Pi_T}$. That is, the agent has reached a maximal level of utility once the desired target has been obtained. These preference-related features translate into the following behaviour. Provided that wealth drifts away from the benchmark, the agent is willing to engage in riskier trades so as to increase the odds of ultimately securing the target. On the contrary, if wealth covers the benchmark, the agent becomes more prudent with respect to risky investments in an attempt to "lock in" his/her wealth at the current target level. Since the agent's preferences are explicitly modelled around this target, we confine ourselves to an analysis of $\frac{X_T}{\Pi_T}$ relative to Y_T. With this end in view, we introduce Proposition 4.4.1

Proposition 4.4.1. *Consider optimal terminal wealth, X_T^{opt}, provided in Proposition 4.3.2 or more explicitly in* (A.3)*. The corresponding benchmark, Y_T, is presented in* (4.2.16)*. Suppose that $x \mapsto F_{X/Y}(x)$ represents the CDF of $\frac{X_T}{\Pi_T}\frac{1}{Y_T}$, i.e. $F_{X/Y}(x) = \mathbb{P}\left(\frac{X_T}{\Pi_T}\frac{1}{Y_T} \leq x\right)$ for all $x \in \mathbb{R}$. Then,*

$$F_{X/Y}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}_\kappa^*(T, x, \omega)\, \phi_T(-\omega - i\kappa, h)\, \mathrm{d}\omega, \tag{4.4.4}$$

holds for all $x \in [0, 1)$.[17] *Moreover, $F_{X/Y}(x) = 0$ for all $x \in (-\infty, 0)$, and*

[16] The subsequent analysis is based on U's coefficient of absolute risk aversion (ARA). ARA for U is equal to: $-\frac{U_x''(x,y)}{U_x'(x,y)} = \frac{p-1}{(y-x)^*}$ for all $x, y \in \mathbb{R}_+$.

[17] Note here that $F_{X/Y}(x) = 1 - \mathbb{P}\Big(M_T^{R\,\frac{1}{p-1}}\, Y_T{}^{-1} \leq (1-x)\,\eta^{\mathrm{opt}\,-\frac{1}{p-1}}\Big)$ holds for all $x \in \mathbb{R}$, cf. (D.3.18). Whereas we were able to disentangle more explicit terms for the optimal control processes in Propositions 4.3.2 and 4.3.4, this is not possible for the case at hand. This is entirely attributable to the analytical structure of the indicator function, $x \mapsto \mathbb{1}_{\{x \in \mathcal{A}\}}$, which is not multiplicative in its sole argument. Alternative, indirect

$F_{X/Y}(x) = 1$ for all $x \in [1, \infty)$. The deterministic function $\tilde{f}^*_\kappa(T, x, \omega)$ is for all $\omega \in \mathbb{R}$, all $x \in [0, 1)$ and some $\kappa \in \mathbb{R}_-$ given by: $\tilde{f}^*_\kappa(T, x, \omega) = \frac{1}{i\omega-\kappa} e^{(i\omega-\kappa)\left[\frac{1}{p-1}\log \mathcal{H}^{-1}(X_0) - \log(1-x)\right]}$. Additionally, the characteristic function, $\phi_T(\omega, h)$, is specified as follows: $\phi_T(\omega, h) = e^{\tilde{A}(0,\omega)} Y_0^{i\omega}$ for all $\omega \in \mathbb{R}$. Here, $(t, \omega) \mapsto \tilde{A}(t, \omega)$ represents the deterministic function provided in (D.2.15), *through the system of ODE's in* (D.2.12). *Now, let $x \mapsto f_{X/Y}(x)$ denote the density of $\frac{X_T}{\Pi_T}\frac{1}{Y_T}$ on the domain $(0, 1)$, i.e. $f_{X/Y}(x) = \frac{\partial}{\partial x} F_{X/Y}(x)$ for all $x \in (0, 1)$. Then, for all $x \in (0, 1)$:*

$$f_{X/Y}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{i\omega - \kappa}{1 - x} \tilde{f}^*_\kappa(T, x, \omega)\, \phi_T(-\omega - i\kappa, h)\, \mathrm{d}\omega. \tag{4.4.5}$$

Proof. The proof is given in Appendix D.2. □

Proposition 4.4.1 analytically characterises the distribution of $X_T^{Y,\Pi} = \frac{X_T}{\Pi_T}\frac{1}{Y_T}$ using the following two functions: (i) the CDF of $X_T^{Y,\Pi}$ in (4.4.4), and (ii) the density of $X_T^{Y,\Pi}$ on the domain $(0, 1)$ in (4.4.4). Note that the CDF, $x \mapsto F_{X/Y}(x)$, delivers the likelihood that $X_T^{Y,\Pi}$ attains values in an interval $(-\infty, x]$. The density, $x \mapsto f_{X/Y}(x)$, can be interpreted as a function that constitutes the continuous analogue of a histogram for $X_T^{Y,\Pi}$. We choose to analyse the distributional features of $X_T^{Y,\Pi}$, as this random variable immediately infers how well $\frac{X_T}{\Pi_T}$ performs relative to Y_T. In fact, $X_T^{Y,\Pi}$ can be identified with the so-called replacement ratio, see e.g. Balter et al. (2020). Concretely, $X_T^{Y,\Pi}$ takes on values in the half-open unit interval, $[0, 1)$, and measures the degree up to which real retirement wealth is able to cover or replicate the benchmark. For example, if $X_T^{Y,\Pi}$ achieves a value of 0.75, real retirement wealth is able to cover 75% of the benchmark's value. Great performance is accordingly associated with $X_T^{Y,\Pi} \approx 1$; poor performance with $X_T^{Y,\Pi} \approx 0$. Therefore, the shape of $X_T^{Y,\Pi}$'s distribution explicitly indicates whether and up to what extent the pension fund is able to meet the financial expectations of the participant. Evidently, in particular from the participant's point of view,

applications of the Fourier transform (used for Propositions 4.3.2 and 4.3.4) would consequently not contribute to the analytical transparency of the ensuing identities. Therefore, $x \mapsto F_{X/Y}(x)$ in (4.4.4) is the most analytical expression that we can engender for the CDF. On account of the immediate derivation of the density from the CDF, it is self-explanatory that the same applies to $x \mapsto f_{X/Y}(x)$.

a highly left-skewed distribution is preferred. In Corollary 4.4.2, we provide $x \mapsto F_{X/Y}(x)$ and $x \mapsto f_{X/Y}(x)$ for $\Lambda_1 = 0_{4\times 2}$.

Corollary 4.4.2. *Consider optimal terminal wealth, X_T^{opt}, provided in* (A.3)*. The corresponding benchmark, Y_T, is presented in* (4.2.16)*. Suppose that $\Lambda_1 = 0_{4\times 2}$. Then, the CDF of $\frac{X_T}{\Pi_T}\frac{1}{Y_T}$, i.e. $F_{X/Y}(x) = \mathbb{P}\left(\frac{X_T}{\Pi_T}\frac{1}{Y_T} \leq x\right)$ for all $x \in \mathbb{R}$, is for all $x \in [0,1)$ given by:*

$$F_{X/Y}(x) = \Phi\left(\frac{\log \frac{\mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}}{1-x} - \mathbb{E}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}{\sqrt{\mathbb{V}ar\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}}\right). \tag{4.4.6}$$

Furthermore, $F_{X/Y}(x) = 0$ for all $x \in (-\infty, 0)$, and $F_{X/Y}(x) = 1$ for all $x \in [1, \infty)$. Here, $\mathbb{E}[\log M_T^{R^{-\frac{1}{p-1}}} Y_T]$ and $\mathbb{V}ar[\log M_T^{R^{-\frac{1}{p-1}}} Y_T]$ are given in (D.3.20)*. In addition to this, $\Phi(\cdot)$ denotes the CDF of a standard normally distributed random variable. Now, define $x \mapsto d_{0,T}(x)$ as the argument of the CDF in* (4.4.6)*. Let $x \mapsto f_{X/Y}(x)$ denote the density of $\frac{X_T}{\Pi_T}\frac{1}{Y_T}$ on the domain $(0,1)$, i.e. $f_{X/Y}(x) = \frac{\partial}{\partial x}F_{X/Y}(x)$ for all $x \in (0,1)$. Moreover, let $\phi(\cdot)$ denote the PDF of a random variable that is standard normally distributed: $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ for all $x \in \mathbb{R}$. Then, for all $x \in (0,1)$:*

$$f_{X/Y}(x) = \frac{\phi(d_{0,T}(x))}{(1-x)\sqrt{\mathbb{V}ar\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}}. \tag{4.4.7}$$

Proof. The proof is given in Appendix D.3.[18] □

Following sections 4.3.2 and 4.3.3, we distinguish the $\Lambda_1 = 0_{4\times 2}$ case from the general $\Lambda_1 \in \mathbb{R}^{4\times 2}$ case. As $X_T^{Y,\Pi}$'s distributional features are unknown for $\Lambda_1 \neq 0_{4\times 2}$, we relied on Fourier transforms to derive the results in Proposition

[18] To avoid confusion, we emphasise that both Proposition 4.4.1 and Corollary 4.4.2 are valid for the general specification of the benchmark process, Y_T, provided in (4.2.16). As a consequence, the special case implied by $Y_T = \widehat{a}_T$ in section 4.4.1 is covered by the presented results.

4.4.1. However, provided that $\Lambda_1 = 0_{4\times 2}$ holds, $X_T^{Y,\Pi}$ solely depends on log-normal random variables. As a consequence, the expression for $X_T^{Y,\Pi}$ allows for more explicit specifications of both $x \mapsto F_{X/Y}(x)$ and $x \mapsto f_{X/Y}(x)$. These specifications are presented in (4.4.6) and (4.4.7), respectively. Indeed, neither of the two aforementioned identities depends on Fourier transforms or analytically burdensome systems of ODE's. Compared to the definition of $F_{X/Y}$, the characterisation of $f_{X/Y}$ is specifically tractable, due to the involvement of $x \mapsto \phi(x)$ rather than $x \mapsto \Phi(x)$. Namely, whereas the latter function concerns an integral expression[19], the former is completely available in closed-form. Considering the fact that $f_{X/Y}$ can be regarded as the continuous variant of a histogram, it directly tells something about the frequency with which X_T/Π_T comes near Y_T. In the context of our DC scheme, the analytical nature of $f_{X/Y}$ is therefore highly advantageous. Notwithstanding, we must emphasise that $f_{X/Y}$ does not coincide with the formal PDF of $X_T^{Y,\Pi}$, cf. Remark 4.4.1. In spite of technicality, the addressed relation between $X_T^{Y,\Pi}$'s distributional features and $x \mapsto f_{X/Y}(x)$ in (4.4.7) stands.

Remark 4.4.1. *The Borel probability measure corresponding to the CDF of* $X_T^{Y,\Pi}$ *admits atoms. That is, positive mass is assigned to the singleton* $\{X_T^{Y,\Pi} = 0\}$*. In particular, we have:* $F_{X/Y}(0) - F_{X/Y}(0^-) = F_{X/Y}(0)$*, which is strictly positive. As a consequence,* $F_{X/Y}$ *is not absolutely continuous and there does not exist a corresponding probability density function (PDF).*[20] *Nevertheless, although* $x \mapsto f_{X/Y}(x)$ *in both* (4.4.5) *and* (4.4.7) *is no formal PDF, as part of* $X_T^{Y,\Pi}$*'s mixed density function, it does enable us to (numerically) examine the shape of* $X_T^{Y,\Pi}$*'s distribution on the practically relevant domain,* $(0,1)$*. To*

[19] In exact terms, $\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{1}{\sqrt{2}}x\right)\right]$, with $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,\mathrm{d}t$, for all $x \in \mathbb{R}$.

[20] Suppose that $\mu : \mathcal{B}(\mathbb{R}) \to \mathbb{R}_+$ represents a Borel measure, such that $F_{X/Y}(x) = \mu((-\infty, x])$ holds for all $x \in \mathbb{R}$. Clearly, μ identifies the probability measure associated with $F_{X/Y}$. Then, following the Lebesgue Decomposition Theorem: $\mu(\mathcal{A}) = \int_{\mathcal{A}} f_{X/Y}(t)\,\mathrm{d}t + F_{X/Y}(0)\,\mathbb{1}_{\{0\in\mathcal{A}\}}$, where we extend $f_{X/Y}$'s definition to ensure that $f_{X/Y}(x) = 0$ holds for all $x \in (-\infty, 0] \cup [1, \infty)$. Using this expression, it is tempting to argue that the PDF reads: $f_{X/Y}(x) + F_{X/Y}(0)\,\delta(x)$, in which $x \mapsto \delta(x)$ outlines the Dirac delta function. Note that we discard the absolute non-continuity of $x \mapsto \delta(x)$ here. However, $F_{X/Y}(0) - F_{X/Y}\left(0^-\right) = F_{X/Y}(0) > 0$, which contradicts the fact that the PDF should coincide with the derivative of $F_{X/Y}$. Hence, the former expression for the PDF is invalid. Aside from this technical invalidity, at $x = 0$ it results in ∞, rendering the expression meaningless for practical purposes.

get an idea of what happens at the extreme value, $x = 0$*, we can in turn apply numerical machinery to approximate* $\lim_{x \downarrow 0} f_{X/Y}(x)$*. Furthermore, the exact probability for this extreme event,* $\{X^{Y,\Pi} = 0\}$*, can be directly obtained from* (4.4.4) *and* (4.4.6) *as follows:* $\mathbb{P}(X^{Y,\Pi} = 0) = F_{X/Y}(0)$*. By virtue of these reasons, we provide the analytical expressions in* (4.4.5) *and* (4.4.7) *for the densities of* $X_T^{Y,\Pi}$ *on the subdomain* $(0, 1)$*.*

4.4.3 Analysis of Retirement Wealth

We continue with a numerical evaluation of $X_T^{Y,\Pi}$'s distributional features presented in Proposition 4.4.1 and Corollary 4.4.2. For this reason, we employ the parameter estimates reported in Table 4.1. The values for these estimates are based on calibrations to recent data. For details on this calibration procedure, we refer to Pelsser (2019). As we wish to emphasise the salient dependence of the optimal solutions on Λ_t, we vary the parameter estimates over Λ_0 and Λ_1. To this end, we introduce two additional sets of parameter estimates defined by the following values for Λ_0 and Λ_1: (i) $\Lambda_{0(i)} = 0.1$ and $\Lambda_{1(i,j)} = 0.075$, and (ii) $\Lambda_{0(i)} = 0.1$ and $\Lambda_1 = 0_{4\times 2}$, for all $i = 1, \ldots, 4$ and $j = 1, 2$. These values are comparable to the estimates for the market prices of risk provided in table 1 of Brennan and Xia (2002). Note that we modify the value for η_S in conformity with the adjustments in Λ_0 and Λ_1. We label the baseline input in Table 4.1 as "P^0"; the aforementioned two inputs as "P^1" and "P^2", respectively.

Although all relevant details are given in the table, we note that the subsequent results correspond to an accumulation phase of $T = 40$ years. We assume that the related decumulation phase lasts for $\tau_A = 20$ years. The agent endows the pension fund at $t = 0$ with $X_0 = 10$ monetary units, faces a funding ratio of $F_0 = 80\%$, and has a risk profile characterised by $p = 2$. We solely consider the $p = 2$ case. Our primary findings on the subject of robustness do not change for higher values of p. Observe that $p = 2$ comes close to the situation for the celebrated expected shortfall criterion ($p = 1$). In Figures 4.1 and 4.2, we present the CDF of $X_T^{Y,\Pi}$ and the density function of $X_T^{Y,\Pi}$, respectively. In Figure 4.3, we display the success probability of $X_T^{Y,\Pi}$ for different values of Λ_0 and Λ_1. The success probability is defined as the likelihood that $X_T^{Y,\Pi}$ exceeds

Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
$S_{1,t}$		Λ_t		Π_t, Z_t		r_t, π_t	
η_S	0.0451	$\Lambda_{0(1)}$	0.6420	$\sigma_{\Pi(1)}$	−0.0010	$\delta_{0,r}$	0.0097
$\sigma_{S(1)}$	−0.0483	$\Lambda_{0(2)}$	−0.0240	$\sigma_{\Pi(2)}$	0.0013	$\delta_{1,r(1)}$	−0.0094
$\sigma_{S(2)}$	0.0078	$\Lambda_{1(1,1)}$	0.1710	$\sigma_{\Pi(3)}$	0.0055	$\delta_{1,r(2)}$	−0.0024
$\sigma_{S(3)}$	0.0010	$\Lambda_{1(1,2)}$	0.3980	$K_{Z(1,1)}$	0.0479	$\delta_{0,\pi}$	0.0158
$\sigma_{S(4)}$	0.1335	$\Lambda_{1(2,1)}$	−0.5140	$K_{Z(2,1)}$	1.2085	$\delta_{1,\pi(1)}$	−0.0028
		$\Lambda_{1(2,2)}$	−1.1470	$K_{Z(2,2)}$	0.5440	$\delta_{1,\pi(2)}$	−0.0014

Table 4.1. Baseline parameter input. This table contains the baseline parameter input on which we rely to compute the numerical results in section 4.4. The displayed values are derived from the second column of Table 1 in Pelsser (2019). Note that the estimates documented in the latter paper are based on calibrations to recent data. In fact, these estimates are employed by the Dutch Central Bank (DNB). As in both Koijen et al. (2009) and Pelsser (2019), we define $\Lambda_{0(3)} = \Lambda_{1(3,1)} = \Lambda_{1(3,2)} = \sigma_{\Pi(4)} = 0$. In addition to this, we set the trading horizon (retirement date) equal to 40, i.e. $T = 40$. Moreover, we assume that the agent lives 20 fixed years after his/her retirement, i.e. $\tau_A = 20$. The funding or coverage ratio is held fixed at 100%, i.e. $F_0 = 1$, unless stated differently. The times to maturity of the three bonds are: $\tau_1 = 5$, $\tau_2 = 20$ and $\tau = 20$. Last, we set the agent's initial endowment equal to: $X_0 = 10$. Observe that C in the definition of the benchmark process, Y_T, cf. section 4.4.1, follows from the parameter input. That is, $C = \frac{X_0}{F_0} e^{-\bar{\alpha} - \tilde{A}(0)}$, where $\bar{\alpha}$ is shown in (4.4.3), and $t \mapsto \tilde{A}(t)$ is given in (B.1.13).

95%. Given that we depart from a funding ratio of 80%, this definition of "success" seems acceptable.

4.4.3.1 Technical Discussion: Figures 4.1 and 4.2

We proceed with a technical discussion of Figures 4.1 and 4.2. From both figures, one can infer that $X_T^{Y,\Pi}$ is equal to approximately 100% with a probability that approaches 1. This result implies that retirement wealth is in almost all states of the world able to *completely* cover the agent's life annuity. Considering that $F_0 = 80\%$, this outcome seems highly unrealistic. Nevertheless, we should take into account that: (i) the market prices of risk (Λ_t) are fairly large, and (ii) the diffusion coefficient of Y_T (β_t) is small compared to Λ_t. The market prices of risk, Λ_t, play a dominant role in the specifications of both $x \mapsto F_{X/Y}(x)$ and $x \mapsto f_{X/Y}(x)$. This is particularly visible in the definition of $\mathcal{H}^{-1}(X_0)$. The identity for X_t^{opt} in (4.3.3) depends on exponentially compounded quadratic versions of Λ_0 and Λ_1. If $|\Lambda_0|$, $|\Lambda_1|$ and/or T grow, this will (exponentially) increase the value of the second term in (4.3.3). Note that $\mathcal{H}^{-1}(X_0)$ solves $X_0^{\text{opt}} = X_0$. Consequently, large values for $|\Lambda_0|$, $|\Lambda_1|$ and/or T will considerably

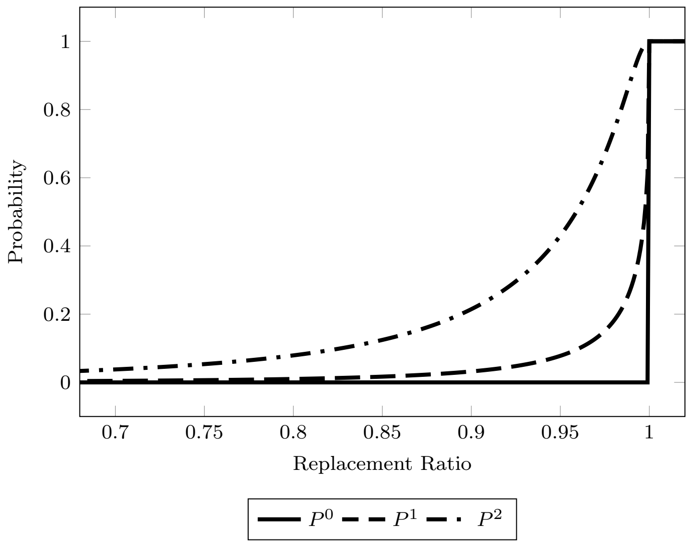

Figure 4.1. CDF of replacement ratio. This figure depicts the CDF of the replacement ratio ($X_T^{Y,\Pi}$) for three different inputs of parameter estimates. The black line corresponds to the P^0 input; the dashed line to the P^1 input; and the dash-dotted line to the P^2 input. For the baseline input of estimates (P^0), we refer to Table 4.1. The remaining inputs are equal to P^0's, where (i) $\Lambda_{0(i)} = 0.1$ and $\Lambda_{1(i,j)} = 0.075$ (P^1), and (ii) $\Lambda_{0(i)} = 0.1$ and $\Lambda_1 = 0_{4\times 2}$ (P^2), for all $i = 1, \ldots, 4$ and $j = 1, 2$. The horizontal axis represents the value of the replacement ratio. The vertical axis represents the value of the probability rendered by the relevant CDF. For this graph, we relied on $X_T^{Y,\Pi}$'s CDF provided in Proposition 4.4.1. Moreover, we fixed $p = 2$ and $F_0 = 80\%$. Although we solely plot the three trajectories on the practically relevant subdomain $[0.7, 1)$, we note that the behaviour of the CDFs on the remaining domains follows from what is presented here. In fact, for $X_T^{Y,\Pi} \geq 1$, the value of the CDF is equal to 1; for $X_T^{Y,\Pi} < 0.7$, all three trajectories tend towards 0 and stay there once $X_T^{Y,\Pi}$ has passed 0.

drive down the magnitude of $\mathcal{H}^{-1}(X_0)$. From (4.4.4) and (4.4.6), we know that declines in $\mathcal{H}^{-1}(X_0)$ result in smaller values for the CDF. Hence, increases in $|\Lambda_0|$ and/or $|\Lambda_1|$ negatively influence the CDF and, by extension, the density function. Note that the opposite holds for increases in β_t. As the substantial impact of Λ_t on both $F_{X/Y}$ and $f_{X/Y}$ is not offset by an equal impact of β_t, we find the shapes in Figures 4.1 and 4.2.[21] In section 4.4.3.3, we comment on

[21] To get an idea of the magnitudes, consider (4.4.6) and let $\Lambda_1 = 0_{4\times 2}$ in P^0. Then, $\mathcal{H}^{-1}(X_0) = 2.8668 \times 10^{-06}$. Moreover, $\mathbb{E}\Big[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\Big] = 18.9333$ and

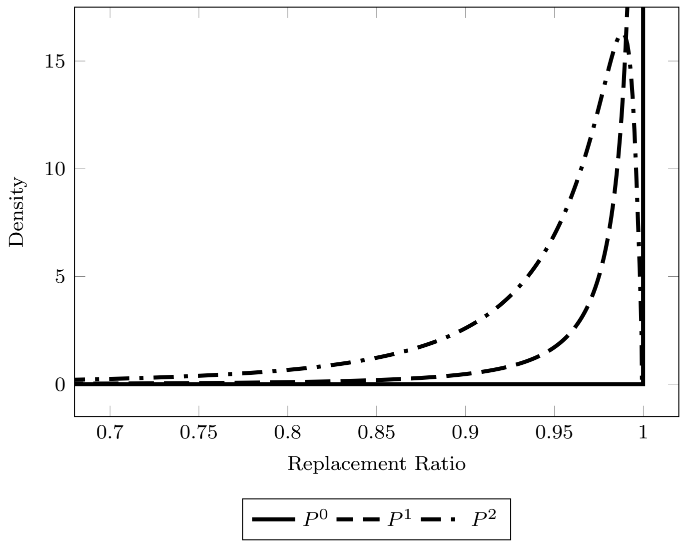

Figure 4.2. Density function of replacement ratio. This figure depicts the density function of the replacement ratio ($X_T^{Y,\Pi}$) for three different inputs of parameter estimates. Similar to Figure 4.1, the black line corresponds to the P^0 input; the dashed line to the P^1 input; and the dash-dotted line to the P^2 input. For the definitions of P^0, P^1 and P^2, we refer the reader to the main text of section 4.4.3 and the description of Figure 4.1. The horizontal axis represents the value of the replacement ratio. The vertical axis represents the value of the density function. For this graph, we relied on $X_T^{Y,\Pi}$'s density function provided in Proposition 4.4.1. Moreover, we fixed $p = 2$ and $F_0 = 80\%$. Although we solely plot the three trajectories on the practically relevant subdomain $[0.7, 1)$, we note that the corresponding behaviour on $(0, 0.7)$ follows from what is presented here. In fact, for $X_T^{Y,\Pi} < 0.7$, all three trajectories tend towards 0. Observe that the density function is only defined on the open unit interval $(0, 1)$. Values for x outside of $(0, 1)$ are consequently irrelevant.

the implications of this finding.

These claims are verified by the trajectories for the P^1 and P^2 cases in Figures 4.1 and 4.2. Due to the smaller values for $|\Lambda_0|$ and $|\Lambda_1|$ in P^1, the CDF of $X_T^{Y,\Pi}$ is indeed driven away from zero. The density function correspondingly lives by a more natural shape. The latter still displays that $X_T^{Y,\Pi}$ is likely to achieve values in a region near 95%. This is partially attributable to the comparatively small values for β_t. However, the agent now also faces notable

$\mathbb{V}\mathrm{ar}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right] = 27.1862$. For all $x \in [0, 1)$, this will clearly steer $F_{X/Y}(x)$ towards values near 0.

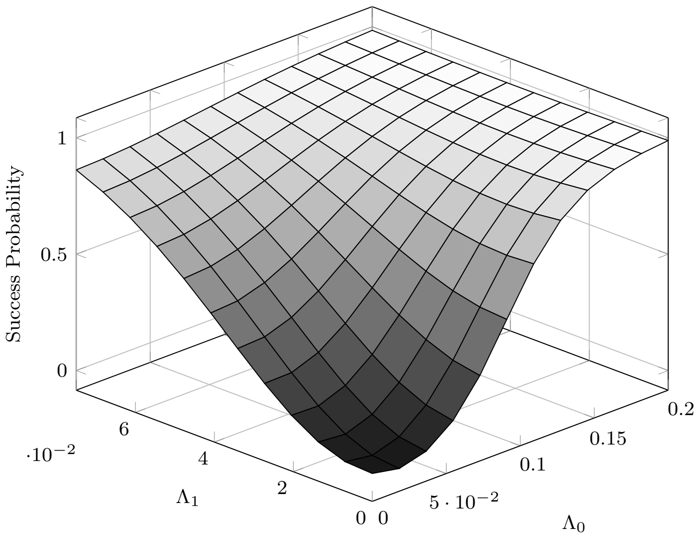

Figure 4.3. Success probability for replacement ratio. This figure depicts the success probability for the replacement ratio ($X_T^{Y,\Pi}$) varied with respect to Λ_0 and Λ_1. The success probability is defined as the likelihood that $X_T^{Y,\Pi}$ exceeds 95%: $\mathbb{P}\left(X_T^{Y,\Pi} \geq 0.95\right)$. Moreover, the elements in Λ_0 and Λ_1 are assumed to be identical: $\Lambda_0 = \Lambda_{0(i)}$ and $\Lambda_1 = \Lambda_{0(i,j)}$, for all $i = 1, \dots 4$ and $j = 1, 2$. The output corresponds to the baseline collection of parameter estimates (P^0) reported in Table 4.1. The vertical axis represents the success probability. The axes labelled Λ_0 and Λ_1 represent the values for Λ_0 and Λ_1, respectively. For this graph, we relied on $X_T^{Y,\Pi}$'s CDF provided in Proposition 4.4.1. In addition to this, we fixed $p = 2$ and $F_0 = 80\%$. Similar to Figures 4.1 and 4.2, we solely plot the three-dimensional trajectories on two distinct subdomains for Λ_0 and Λ_1: $[0, 0.2]$ and $[0, 0.075]$, respectively. The graph's dynamics on the remainder of $\mathbb{R}^2$ namely follow from what is presented here. In fact, for larger values of Λ_0 and Λ_1, the success probability converges to 1; for negative values, the graph behaves as displayed.

positive odds of obtaining $X_T^{Y,\Pi}$ below the level of his/her starting position $(F_0 = 80\%)$. For the P^2 case, the graphical results appear even more plausible. The density function indicates that $X_T^{Y,\Pi}$ is likely to achieve values in a much wider region around 95%. Additionally, the agent faces significantly higher odds of obtaining $X_T^{Y,\Pi}$ below 80%. For a funding ratio of 80%, such outcomes seem reasonable. In our discussion preceding this analysis of P^1 and P^2, we did not distinguish between the impact of Λ_0 and Λ_1 on $F_{X/Y}$ and $f_{X/Y}$. The technical reason for this indifference is that $|\Lambda_0|$ and $|\Lambda_1|$ play similar roles in the characterisation of $\mathcal{H}^{-1}(X_0)$, cf. Proposition 4.3.2. As shown in the two plots for P^1 and P^2, the impact is indeed comparable in terms of its sign.

4.4.3.2 Technical Discussion: Figure 4.3

Figure 4.3 confirms the previous claim. For a fixed value of Λ_0, the success probability positively depends on Λ_1. The same holds with respect to Λ_0, for a fixed value of Λ_1. Note here that the elements in Λ_0 and Λ_1 are assumed to be identical: $\Lambda_0 = \Lambda_{0(i)}$ and $\Lambda_1 = \Lambda_{0(i,j)}$, for all $i = 1, \dots 4$ and $j = 1, 2$ (with a slight abuse of notation). Despite the former similarities, the impact of Λ_1 on the success probability is greater than the impact of Λ_0. This difference can be explained by the fact that Λ_t depends on Λ_1 through Z_t. The variance of Z_t grows in time. Therefore, upon retirement, small non-zero values for Λ_1 are capable of inflating the market prices of risk. We stress that $\Lambda_1 Z_t$ can be regarded as a time-dependent analogue of Λ_0. As a result, small values for Λ_1 have an impact on the success probability that resembles the impact of comparatively large values for Λ_0. We underline that Figure 4.3 substantiates our claims regarding the outcomes' sensitivity to Λ_0 and Λ_1. In fact, for values of these parameters in rather small subdomains, the success probability varies between 0 and 1. In spite of this sensitivity, we observe that the replacement ratio is potentially very likely to exceed a level of 95%. Given that $\Lambda_0 > 0.125$ and/or $\Lambda_1 > 0.050$ hold true, the success probability tends towards 100%.

4.4.3.3 Economic Takeaways

The set of economic takeaways corresponding to the output in Figures 4.1, 4.2 and 4.3 is twofold. First, we find that LPM-based investment strategies can increase the likelihood of achieving one's pension goals. This finding is robust to *some* uncertainty around Λ_0 and Λ_1. Second, the displayed outcomes are highly dependent on the estimates for Λ_0 and Λ_1. This result supports the use of approaches that account for model/parameter uncertainty. We elaborate on these takeaways in the following two summaries:

Takeaway 1. The first takeaway is a straightforward consequence of a.o. the displayed trajectories for $F_{X/Y}$ and $f_{X/Y}$. Based on the plots for $f_{X/Y}$, we can conclude that real retirement wealth is very likely to achieve values in the neighborhood of the life annuity. These high odds are coupled to fairly low odds of the replacement ratio falling below the agent's funding ratio. Note that the pension fund departs from a funding ratio of 80%. In terms of hedging, this means that the fund is able to arrive at a replacement ratio of likewise 80% with a probability of 1. Without sacrificing too much of this certainty, the LPM-framework is able to significantly improve this recovery potential of 80%. Note that these qualitative results hold true for different values of Λ_0 and Λ_1, despite their relatively large impact on the shapes of $F_{X/Y}$ and $f_{X/Y}$. Figure 4.3 indeed demonstrates that the replacement ratio exceeds a level of 95% with a probability near 1, provided that $\Lambda_0 > 0.125$ and/or $\Lambda_1 > 0.050$ hold true The conclusions that we draw here are consequently quite robust to *some* uncertainty around Λ_0 and Λ_1.

Takeaway 2. The second takeaway also follows from the graphs for $F_{X/Y}$ and $f_{X/Y}$. Nevertheless, the sensitivity of the outcomes to the estimates for Λ_0 and Λ_1 is most distinct in Figure 4.3. The pronounced impact of Λ_t on these graphs highlights the strong dependence of the optimality conditions on the market prices of risk. As pointed out in section 4.4.3.1, particular values for Λ_0 and Λ_1 may even lead to nonsensical outcomes. This dependence has an important implication for the way in which one ordinarily treats parameter estimates. Small estimation errors can namely generate meaningless outcomes and/or imply enormous policy changes.[22] In order to avoid such adverse events, it is wise to be careful concerning the parameter estimates and account for parameter/model uncertainty. The latter refers to doubts that one may have about the veracity of particular parameter estimates or model specifications. An agent can account for this by cautiously preparing him- or herself for a worst-case scenario. Accordingly, one becomes less sensitive to estimation-related errors, i.e. robust. This robustness naturally returns in the optima and the corresponding policy rules. For applications of parameter/model uncertainty in similar investment-based frameworks, we refer to Balter (2016) and references therein.

[22] It is noteworthy that the standard errors corresponding to the Λ_t-related estimates reported in Pelsser (2019) are fairly large compared to those for the other estimates.

4.4.4 Analysis of Portfolio Rules

In this section, we present a numerical analysis of the optimal portfolio rules provided in Proposition 4.3.4 and Corollary 4.3.5. We mainly aim to inspect the implications of the LPM-mechanism for the optimal trading behaviour. For this analysis, we make use of the parameter estimates defined by P^2. This collection of estimates is spelled out in section 4.4.3 and the description of Figure 4.1. Due to the nature of the analysis, it is not necessary to employ the P^0 and P^1 collections. Note that all subsequent results correspond to a situation wherein $T = 40$, $\tau_A = 20$, $X_0 = 10$, $F_0 = 80\%$ and $p = 2$. The pension-related interpretation of this initialisation is given at the beginning of section 4.4.3. Observe that the asset mix consists of a stock, two τ_i-year nominal bonds and a τ-year inflation-linked bond. Henceforth, we assume that $\tau_1 = 5$, $\tau_2 = 20$ and $\tau = 20$.

The optimal portfolio rules can be expressed in terms of observable quantities. To this end, consider e.g. (4.3.13) and note that the right-hand side depends on the following processes: M_t, Π_t, Z_t, Y_t and X_t. The values for Y_t, Π_t and X_t are directly observed. The value for M_t uniquely depends on the ones for Z_t, Y_t, Π_t and X_t, cf. (4.3.8). Moreover, Z_t can be expressed in terms of r_t and π_t, both of which are observed. Hence, given the values of r_t, π_t, Y_t, Π_t and X_t, the agent knows precisely how he/she should optimally invest in the four risky assets. This is an outstanding advantage entirely attributable to the closed-form nature of the optimality conditions.

To emphasise this advantage and to examine the implications of the LPM-mechanism for the optimal exposures, we present Figures 4.4 and 4.5. These figures depict the optimal allocation of assets (π_t^{opt}) varied with respect to time ($t \in [0, T]$). In Figure 4.4, the risk-neutral value of the replacement ratio progressively *grows* to a level of 100% at retirement. Figure 4.5 displays the same, but for a risk-neutral value of the replacement ratio that progressively *declines* to a level of 60% at retirement. The risk-neutral value of the replacement ratio lives by: $F_t = \frac{X_t}{\mathbb{E}[Y_T \Pi_T M_T / M_t \mid \mathcal{F}_t]}$, for all $t \in [0, T]$. For Figures 4.4 and 4.5, we fix: $r_t = \delta_{0,r}$, $\pi_t = \delta_{0,\pi}$, $Y_t = Y_0 e^{\delta_{0,r} t}$, $\Pi_t = e^{\delta_{0,\pi} t}$ and $X_t = X_0 e^{(\delta_{0,r} + \delta_{0,\pi} + \frac{1}{T} \log F_T \times Y_0 / X_0) t}$, for all $t \in [0, T]$. It is clear that Π_t is initialised at its approximate expected value. We predicate the specification of Y_t on the assumption that it moves together with the cash account. The definition of X_t therefore ensures that $X_T^{Y,\Pi} = F_T$ holds. Figures 4.4 and 4.5 accordingly set $F_T = 100\%$ and $F_T = 60\%$, respectively. Note that we hold r_t and π_t fixed at their expected values.

4.4.4.1 Technical Discussion: Figure 4.4

We continue with a technical discussion of Figure 4.4. The depicted allocations correspond to a participant who enters the pension scheme with a funding ratio of 80%. Over the course of the accumulation phase, his/her wealth moves closer to the reference level. Upon retirement, the replacement ratio is equal to 100%. For this situation, the optimal portfolio exhibits a clear life-cycle pattern. That is, the percentages of wealth allocated to the risky assets decrease as the individual ages. This trading behaviour can be attributed to the LPM-mechanism. From section 4.4.2, we recall that an LPM-agent's level of risk aversion positively depends on the risk-neutral value of the replacement ratio (F_t). Due to the inverse relation between levels of risk aversion and optimal portfolios, progressive growth of F_t generates life-cycle strategies as shown in the graph. We emphasise that the LPM-agent is willing to bear significant risk as long as the target is not achieved. For instance, at $t = 0$ when $F_0 = 80\%$, the fund is required to invest nearly 15 times their accumulated wealth in the 5-year nominal bond. This extreme behaviour comes close to the "gamble for resurrection" phenomenon. The opposite can be observed when the target is more or less reached. Indeed, for F_t close to 100% (at $t = T$), the fund "de-risks" their portfolio so as to "lock in" wealth at the current level.

4.4.4.2 Technical Discussion: Figure 4.5

Let us now turn to Figure 4.5. As in the previous case, the participant enters the pension scheme with a funding ratio of 80%. However, during the accumulation phase, his/her wealth gradually moves away from the reference level. Upon retirement, the replacement ratio now equals 60%. Consequently, in contrast to Figure 4.4, the risk-neutral value of the replacement ratio decreases over time. This leads to optimal allocations that behave as reversed life-cycle strategies.

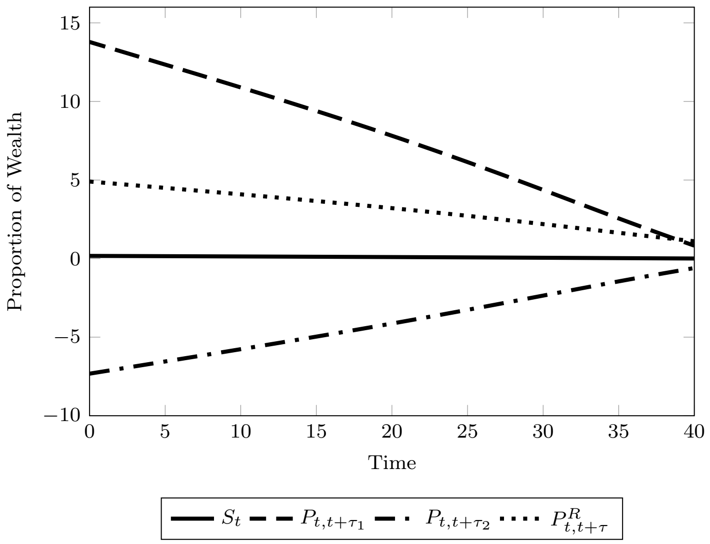

Figure 4.4. Optimal portfolio for progressively improving replacement ratio. This figure depicts the optimal allocation of assets varied with respect to time ($t \in [0, T]$). The risk-neutral value of the replacement ratio progressively grows to a level of 100% at retirement ($T = 40$). The black line corresponds to the demand for the stock (S_t); the dashed line to the demand for the 5-year nominal bond ($P_{t,t+\tau_1}$); the dash-dotted line to the demand for the 20-year nominal bond ($P_{t,t+\tau_2}$); and the dotted line to the demand for the 20-year inflation-linked bond ($P^R_{t,t+\tau}$). The output is based on the set of parameter estimates labelled as P^2, cf. section 4.4.3. The horizontal axis represents the time-dimension. The vertical axis represents the proportion of wealth. For this graph, we relied on the analytical expression for π_t^{opt} in (4.3.13). Moreover, we set $p = 2$ and $F_0 = 80\%$. Throughout the time-dimension, we hold the observable quantities fixed as follows: $r_t = \delta_{0,r}$, $\pi_t = \delta_{0,\pi}$, $Y_t = Y_0 e^{\delta_{0,r} t}$, $\Pi_t = e^{\delta_{0,\pi} t}$ and $X_t = X_0 e^{(\delta_{0,r} + \delta_{0,\pi} + \frac{1}{T} \log Y_0 / X_0) t}$, for all $t \in [0, T]$. As the demands are calculated on a yearly basis, we used cubic-spline interpolation to compute and accordingly smoothen the demands over the entirety of $[0, T]$.

Strategies of this form are in stark contrast with conventional wisdom, cf. Cocco et al. (2005). Nevertheless, using the same arguments as before, one is able to explain the progressively increasing exposure to risk. As a result of the declining value of F_t, the agent namely becomes increasingly less risk averse. This directly leads to trading patterns involving more risk as time passes. In fact, near $t = T$ when F_t approaches 60%, the agent is so worried about achieving the target that the pension fund is required to invest almost 40 times their wealth in the 5-year nominal bond. Note that the overall exposure to risk

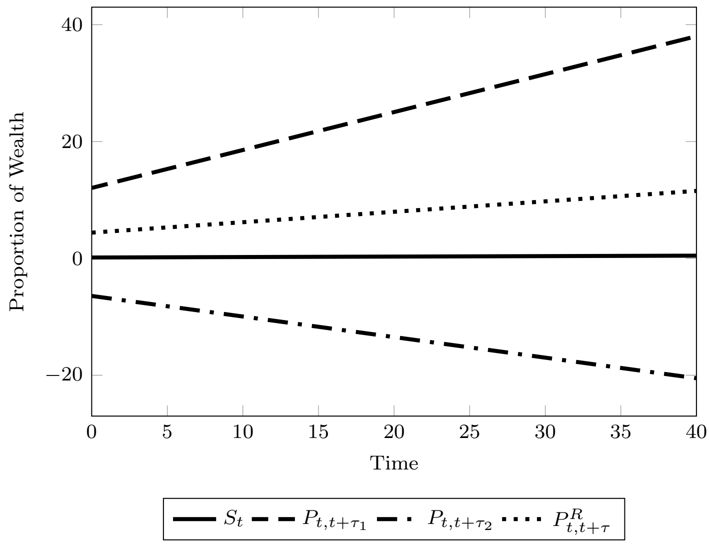

Figure 4.5. Optimal portfolio for progressively worsening replacement ratio. This figure depicts the optimal allocation of assets varied with respect to time ($t \in [0, T]$). The risk-neutral value of the replacement ratio progressively declines to a level of 60% at retirement ($T = 40$). In line with Figure 4.4, the black line corresponds to the demand for the stock (S_t); the dashed line to the demand for the 5-year nominal bond ($P_{t,t+\tau_1}$); the dash-dotted line to the demand for the 20-year nominal bond ($P_{t,t+\tau_2}$); and the dotted line to the demand for the 20-year inflation-linked bond ($P^R_{t,t+\tau}$). Note that these line patterns coincide with those shown in Figure 4.4. The displayed trajectories are predicated on the set of parameter estimates labelled as P^2, cf. section 4.4.3. The horizontal axis represents the time-dimension. The vertical axis represents the proportion of wealth. As this graph presents the same as Figure 4.4 but for $X_t = X_0 e^{(\delta_{0,r} + \delta_{0,\pi} + \frac{1}{T} \log 0.6 \times Y_0 / X_0)t}$, we refer to the description of Figure 4.4 for further details on (i) LPM-specific parameters, (ii) the observable quantities and (iii) the interpolation technique.

was already relatively large at $t = 0$. This behaviour confirms that the LPM operator accommodates features specific to loss aversion setups. Although these strategies are highly impractical, we point out that they are potentially capable of rendering great outcomes for $X_T^{Y,\Pi}$, cf. section 4.4.3.

4.4.4.3 Economic Takeaways

As in section 4.4.3.3, the set of economic takeaways corresponding to Figures 4.4 and 4.5 is twofold. First, we find that LPM-based investment strategies strongly depend on the risk neutral value of the replacement ratio. This dependency highlights the core mechanism of the LPM operator. Second, we show that the optimal portfolio rules can be difficult to implement in practice. For this purpose, it is advisable to account for trading/solvency constraints. We expand on these takeaways in the following two summaries:

Takeaway 1. In section 4.4.2, we addressed the dependence of the agent's attitude towards risk on the difference between the reference level and accumulated wealth. This dependence gives rise to a positive relation between risk aversion and the risk-neutral value of the replacement ratio (F_t). The latter relation constitutes the central feature of the LPM mechanism. As a consequence of this relation, the optimal LPM-based investment policies negatively depend on F_t. In particular, for values of F_t near 100%, the LPM-agent implements a notably prudent investment strategy. In this way, he/she aims to "lock in" wealth at the desired reference level. On the contrary, if F_t decreases in value, the LPM-agent allocates strikingly larger proportions of wealth to the risky assets. Thereby, he/she tries to improve the likelihood of ultimately achieving the target. This phenomenon approaches the "gamble for resurrection" behaviour. Although the agent does not become risk-loving for $F_t < 1$, Figures 4.4 and 4.5 show indeed that he/she is considerably more willing to bear risk. The graphs accordingly support the study by Jarrow and Zhao (2006). In this paper, the authors demonstrate that the LPM operator can be nested under the aegis of prospect theory.

Takeaway 2. The second takeaway ties in with the previous one. In line with the LPM mechanism, declines in F_t namely force the agent to increase his/her exposure to risk. Particular values of F_t may even result in unrealistically leveraged and/or large positions in the financial instruments. This is saliently visible in Figure 4.5. For values of F_t between 60% and 80%, the agent may be required to invest nearly 40 times his/her wealth in the 5-year nominal bond. In reality, pension funds must deal with solvency requirements. These requirements involve specific restrictions concerning a.o. borrowing and short-selling. The LPM-based investment policies are correspondingly difficult to adopt by most pension funds. Due to the very extreme nature of the optimal allocations, this extends to a broader set of agents. Therefore, in order to arrive at more practical portfolio rules, it is recommendable to account for trading/solvency constraints. We refer to Cvitanić and Karatzas (1992) and Basak and Shapiro (2001) for analyses involving such constraints. By the same token, it can be advantageous to take parameter/model uncertainty into account, cf. section 4.4.3.3. The worst-case preparation naturally leads to more conservative investment policies.

4.5 Conclusion

This chapter has studied an optimal terminal wealth problem, in which the agent aims to minimise an LPM criterion. This criterion incorporates a log-normal exogenous reference level. We have placed the problem in the complex market model proposed by Koijen et al. (2009). In this continuous-time framework, the market prices of risk depend on a mean-reverting state variable. As a result, it is highly nontrivial to derive closed-form solutions to the LPM problem. Nevertheless, using Fourier machinery, we have been able to deduce analytical expressions for the optimal portfolio rules and the optimal wealth process. Moreover, we have managed to disentangle a closed-form specification for the distributional features of optimal terminal wealth. In the numerical illustrations, we have cast the LPM problem into the context of a DC pension scheme. The ensuing results have demonstrated that LPM-based investment policies can improve a pension fund's recovery potential. In spite of their possibly outstanding performance, we have exemplified that these policies may be difficult to implement in reality. Furthermore, we have shown that these outcomes strongly depend on the estimates for the market prices of risk.

Appendix A Proof of Theorem 4.3.1

The dynamic problem in (4.2.17) is equivalent to:

$$\sup_{X_T \in L^0_+(\Omega) \text{ s.t. } \mathbb{E}[X_T M_T] \le X_0} \mathbb{E}\left[-\frac{1}{p}\left[\left(Y_T - \frac{X_T}{\Pi_T}\right)^+\right]^p\right]. \tag{A.1}$$

The Lagrangian for (A.1), $\mathcal{L} : L^0_+(\Omega) \times \mathbb{R}_+ \to \mathbb{R}$, is given by:

$$\begin{aligned}\mathcal{L}(X_T, \eta) &= \mathbb{E}\left[U(X_T, Y_T) - \eta X_T M_T\right] + \eta X_0 \\ &= \mathbb{E}\left[-\frac{1}{p}\left[\left(Y_T - \frac{X_T}{\Pi_T}\right)^+\right]^p - \eta X_T M_T\right] + \eta X_0,\end{aligned} \tag{A.2}$$

where $\eta \in \mathbb{R}_+$ represents the Lagrange multiplier. Appropriate optimisation of this Lagrangian results in:

$$X_T^{\text{opt}} = I\left(\mathcal{H}^{-1}(X_0) M_T \Pi_T, Y_T\right) \Pi_T. \tag{A.3}$$

Due to the fact that $\mathbb{E}\left[I\left(\mathcal{H}^{-1}(X_0) M_T \Pi_T, Y_T\right) \Pi_T M_T\right] = X_0$ holds, $\left\{X_t^{\text{opt}} M_t\right\}_{t\in[0,T]}$ spells out a $\mathbb{P}$-martingale process with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$. This results in the identity for X_t^{opt} in (4.3.1).

Moreover, by the martingale representation theorem: $X_t^{\text{opt}} M_t = X_0 + \int_0^T \psi_s^\top \mathrm{d}W_s$, for all $t \in [0,T]$ and some $L^2([0,T])$-valued process $\{\psi_t\}_{t\in[0,T]}$. Then, (4.3.2) follows from:

$$\begin{aligned}X_t^{\text{opt}} M_t &= X_0 + \int_0^t \left(\pi_s^{\text{opt}\top} \sigma - \Lambda_s^\top X_s^{\text{opt}}\right) M_s \mathrm{d}W_s \\ &= X_0 + \int_0^T \psi_s^\top \mathrm{d}W_s, \quad \forall\, t \in [0,T].\end{aligned} \tag{A.4}$$

Appendix B Proofs I

B.1 Proof of Proposition 4.3.2

Define the following two processes for all $t \in [0,T]$:

$$\begin{aligned}\left.\frac{\mathrm{d}\mathbb{X}_1}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} &= \frac{\mathbb{E}\left[Y_T M_T^R \mid \mathcal{F}_t\right]}{\mathbb{E}\left[Y_T M_T^R\right]} = e^{-\frac{1}{2}\int_0^t \lambda_{1,s}^\top \lambda_{1,s} \mathrm{d}s + \int_0^t \lambda_{1,s}^\top \mathrm{d}W_s}, \\ \left.\frac{\mathrm{d}\mathbb{X}_2}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} &= \frac{\mathbb{E}\left[\left(M_T^R\right)^{\frac{p}{p-1}} \,\middle|\, \mathcal{F}_t\right]}{\mathbb{E}\left[\left(M_T^R\right)^{\frac{p}{p-1}}\right]} = e^{-\frac{1}{2}\int_0^t \lambda_{2,s}^\top \lambda_{2,s} \mathrm{d}s + \int_0^t \lambda_{2,s}^\top \mathrm{d}W_s}.\end{aligned} \tag{B.1.1}$$

Both $\frac{\mathrm{d}\mathbb{X}_1}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$ and $\frac{\mathrm{d}\mathbb{X}_2}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t}$ qualify as valid Radon-Nikodym[23] derivatives:

$$\left.\frac{\mathrm{d}\mathbb{X}_i}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = \mathbb{E}\left[\left.\frac{\mathrm{d}\mathbb{X}_j}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_s} \,\middle|\, \mathcal{F}_t\right] > 0,\ \forall s \geq t,\ i = 1, 2. \tag{B.1.2}$$

Note that $\mathbb{X}_i \sim \mathbb{P}$, for $i = 1, 2$. Here, the two processes $\{\lambda_{1,t}\}_{t\in[0,T]}$ and $\{\lambda_{2,t}\}_{t\in[0,T]}$ are to be determined. Under $\mathbb{X}_i$, the following two processes are standard Brownian motions, for $i = 1, 2$ and all $t \in [0, T]$:

$$W_t^{\mathbb{X}_i} = W_t - \int_0^t \lambda_{i,s}\mathrm{d}s. \tag{B.1.3}$$

Using the changes of measure implied by the Radon-Nikodym derivatives in (B.1.1), we are able to rewrite the conditional expectation in (4.3.1) as:

$$\begin{aligned} X_t^{\text{opt}} &= \frac{1}{M_t}\mathbb{E}\left[Y_T M_T^R \mathbb{1}_{\{\mathcal{A}_T\}} \mid \mathcal{F}_t\right] \\ &\quad - \frac{1}{M_t}\mathbb{E}\left[\left(\mathcal{H}^{-1}(X_0)^{\frac{1}{p}} M_T^R\right)^{\frac{p}{p-1}} \mathbb{1}_{\{\mathcal{A}_T\}} \,\middle|\, \mathcal{F}_t\right] \\ &= Y_t \Pi_t \mathbb{E}\left[\frac{Y_T M_T^R}{Y_t M_t^R} \,\middle|\, \mathcal{F}_t\right] \mathbb{X}_1(\mathcal{A}_T \mid \mathcal{F}_t) \\ &\quad - \left(\mathcal{H}^{-1}(X_0) M_t^R\right)^{\frac{1}{p-1}} \Pi_t \mathbb{E}\left[\left(\frac{M_T^R}{M_t^R}\right)^{\frac{p}{p-1}} \,\middle|\, \mathcal{F}_t\right] \mathbb{X}_2(\mathcal{A}_T \mid \mathcal{F}_t). \end{aligned} \tag{B.1.4}$$

In the last line, we use for $i = 1, 2$ that:

$$\mathbb{X}_i(\cdot \mid \mathcal{F}_t) = \mathbb{E}\left[\left.\frac{\mathrm{d}\mathbb{X}_i}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_T} \left.\frac{\mathrm{d}\mathbb{X}_i}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t}^{-1} \mathbb{1}_{\{\cdot\}} \,\middle|\, \mathcal{F}_t\right]. \tag{B.1.5}$$

We start by evaluating the expectations in (B.1.4). Relying on the results in Duffie and Kan (1996), we note that the following holds for all $t \in [0, T]$:

$$P_1(t, Z_t) = \mathbb{E}\left[\frac{Y_T M_T^R}{Y_t M_t^R} \,\middle|\, \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}_1}\left[\exp\left\{-\int_t^T F(s, Z_s)\,\mathrm{d}s\right\} \,\middle|\, Z_t\right], \tag{B.1.6}$$

[23]See e.g. Karatzas and Shreve (1998, 2012).

and

$$P_2\left(t,Z_t\right)=\mathbb{E}\left[\left(\frac{M_T^R}{M_t^R}\right)^{\frac{p}{p-1}}\,\middle|\,\mathcal{F}_t\right]=\mathbb{E}^{\mathbb{Q}_2}\left[\exp\left\{-\int_t^T G\left(s,Z_s\right)\mathrm{d}s\right\}\,\middle|\,Z_t\right]. \tag{B.1.7}$$

Here, the measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ are induced by:

$$\left.\frac{\mathrm{d}\mathbb{Q}_i}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t}=e^{-\frac{1}{2}\int_0^t\widehat{\lambda}_{i,s}^\top\widehat{\lambda}_{i,s}\mathrm{d}s+\int_0^t\widehat{\lambda}_{i,s}^\top\mathrm{d}W_s}, \tag{B.1.8}$$

for $i=1,2$, where $\widehat{\lambda}_{1,t}=-\Lambda_t^R+\beta_t$ and $\widehat{\lambda}_{2,t}=-\frac{p}{p-1}\Lambda_t^R$, for all $t\in[0,T]$. Note that $R_t=r_t+\pi_t-\sigma_\Pi^\top\Lambda_t$ and $\Lambda_t^R=\Lambda_t-\sigma_\Pi$, for all $t\in[0,T]$. Furthermore, $F:[0,T]\times\mathbb{R}\to\mathbb{R}$ and $G:[0,T]\times\mathbb{R}\to\mathbb{R}$, read:

$$\begin{aligned}F\left(t,Z_t\right)&=R_t-\alpha_t+\Lambda_t^{R^\top}\beta_t=a_t+b_t^\top Z_t,\\G\left(t,Z_t\right)&=\frac{p}{p-1}R_t-\frac{1}{2}\frac{p}{\left(p-1\right)^2}\Lambda_t^{R^\top}\Lambda_t^R\\&=\tilde{a}+\tilde{b}^\top Z_t+Z_t^\top\tilde{c}Z_t,\end{aligned} \tag{B.1.9}$$

where $a_t=\widehat{\delta}_{0,r}-\alpha_t+\beta_t^\top\left(\Lambda_0-\sigma_\Pi\right)$, and $b_t=\widehat{\delta}_{1,r}+\Lambda_1^\top\beta_t$; as well as $\tilde{a}=\frac{p}{p-1}\widehat{\delta}_{0,r}-\frac{1}{2}\frac{p}{(p-1)^2}\left(\Lambda_0-\sigma_\Pi\right)^\top\left(\Lambda_0-\sigma_\Pi\right)$, $\tilde{b}=\frac{p}{p-1}\widehat{\delta}_{1,r}-\frac{p}{(p-1)^2}\Lambda_1^\top\left(\Lambda_0-\sigma_\Pi\right)$, and $\tilde{c}=-\frac{1}{2}\frac{p}{(p-1)^2}\Lambda_1^\top\Lambda_1$. Hence, the function F is affine in Z_t, and the function G is affine-quadratic in Z_t.

Now, let us note that the SDE's of $Y_tM_t^R$ and $Y_tM_t^{R^{\frac{p}{p-1}}}$ are given by:

$$\begin{aligned}\frac{\mathrm{d}Y_tM_t^R}{Y_tM_t^R}&=-\left(R_t-\alpha_t+\beta_t^\top\Lambda_t^R\right)\mathrm{d}t-\left(\Lambda_t^R-\beta_t\right)^\top\mathrm{d}W_t,\\\frac{\mathrm{d}M_t^{R^{\frac{p}{p-1}}}}{M_t^{R^{\frac{p}{p-1}}}}&=-\left(\frac{p}{p-1}R_t-\frac{1}{2}\frac{p}{\left(p-1\right)^2}\Lambda_t^{R^\top}\Lambda_t^R\right)\mathrm{d}t-\frac{p}{p-1}\Lambda_t^{R^\top}\mathrm{d}W_t.\end{aligned} \tag{B.1.10}$$

As a result, the following must be true:

$$\begin{aligned}P_{1,t}-P_{1,Z}^\top K_ZZ+\frac{1}{2}\mathrm{tr}\left(P_{1,ZZ}\right)-F\left(t,Z\right)P_1-P_{1,Z}^\top\Sigma_Z\left(\Lambda^R-\beta\right)&=0,\\P_{2,t}-P_{2,Z}^\top K_ZZ+\frac{1}{2}\mathrm{tr}\left(P_{2,ZZ}\right)-G\left(t,Z\right)P_2-\frac{p}{p-1}P_{2,Z}^\top\Sigma_Z\Lambda^R&=0.\end{aligned} \tag{B.1.11}$$

Observe here that $\mathrm{tr}(\cdot)$ spells out the trace operator.

Inspired by Duffie and Kan (1996), Sangvinatsos and Wachter (2005), and Koijen et al. (2009), we define the subsequent *ansatz* functions for P_1 and P_2:

$$\begin{aligned} P_1(t, Z_t) &= \exp\left\{\tilde{A}(t) + \tilde{B}(t)^\top Z_t\right\}, \text{ and} \\ P_2(t, Z_t) &= \exp\left\{\widehat{A}(t) + \widehat{B}(t)^\top Z_t + Z_t^\top \widehat{C}(t) Z_t\right\}. \end{aligned} \tag{B.1.12}$$

Here, we postulate that $\tilde{A} : [0,T] \to \mathbb{R}$, $\tilde{B} : [0,T] \to \mathbb{R}^2$, $\widehat{A} : [0,T] \to \mathbb{R}$, $\widehat{B} : [0,T] \to \mathbb{R}^2$, and $\widehat{C} : [0,T] \to \mathbb{R}^{2\times 2}$, are deterministic functions of time, $t \in [0,T]$, alone. Now, we proceed in the spirit of Dai and Singleton (2002) and Koijen et al. (2009), and insert the (ansatz) definitions for P_1 and P_2 into (B.1.11). Let $c_t = \Lambda_0 - \sigma_\Pi - \beta_t$. For P_1, we then find:

$$\begin{aligned} \tilde{A}(t) &= -\int_t^T \left[\tilde{B}(s)^\top \Sigma_Z c_s - \frac{1}{2}\tilde{B}(s)^\top \tilde{B}(s) + a_s\right] \mathrm{d}s, \\ \tilde{B}(t) &= -\int_t^T \exp\left\{-\left[K_Z^\top + \Lambda_1^\top \Sigma_Z^\top\right](s-t)\right\} b_s \mathrm{d}s. \end{aligned} \tag{B.1.13}$$

For P_2, we derive the subsequent system of ODE's:

$$\begin{aligned} \widehat{A}'(t) &= \frac{p}{p-1}\widehat{B}(t)^\top \Sigma_Z [\Lambda_0 - \sigma_\Pi] - \frac{1}{2}\widehat{B}(t)^\top \widehat{B}(t) + \tilde{a}_t, \\ \widehat{B}'(t) &= \left(\frac{p}{p-1}\Lambda_1^\top \Sigma_Z^\top + K_Z^\top - 2\widehat{C}(t)^\top\right)\widehat{B}(t) + \tilde{b}_t, \\ \widehat{C}'(t) &= 2\left(K_Z^\top + \frac{p}{p-1}\Lambda_1^\top \Sigma_Z^\top\right)\widehat{C}(t) - 2\widehat{C}(t)^\top \widehat{C}(t) + \tilde{c}, \end{aligned} \tag{B.1.14}$$

where we define $\tilde{a}_t = -\mathrm{tr}\left(\widehat{C}(t)\right) + \tilde{a}$ and $\tilde{b}_t = 2\frac{p}{p-1}\widehat{C}(t)^\top \Sigma_Z [\Lambda_0 - \sigma_\Pi] + \tilde{b}$, for all $t \in [0,T]$. Note that: $\widehat{A}(T) = 0$, $\widehat{B}(T) = 0_2$, and $\widehat{C}(T) = 0_{2\times 2}$, where $0_2 = [0,0]^\top$ and $0_{2\times 2} = [0_2, 0_2]^\top$. This system of ODE's involves a matrix Riccati differential equation, for which no closed-form solution is available. Likewise, we cannot analytically solve the second ODE in (B.1.14). We are only able to state that the following holds for all $t \in [0,T]$:

$$\widehat{A}(t) = -\int_t^T \left[\frac{p}{p-1}\widehat{B}(s)^\top \Sigma_Z [\Lambda_0 - \sigma_\Pi] - \frac{1}{2}\widehat{B}(s)^\top \widehat{B}(s) + \tilde{a}_s\right] \mathrm{d}s. \tag{B.1.15}$$

Let us return to (B.1.1) and derive the following SDE's:

$$\begin{aligned}
\mathrm{d}\frac{\mathrm{d}\mathbb{X}_1}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_t} &= -\left(\Lambda_t^{R^\top} - \beta_t^\top - \tilde{B}(t)^\top \Sigma_Z\right)\mathrm{d}W_t,\\
\mathrm{d}\frac{\mathrm{d}\mathbb{X}_2}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_t} &= -\left(\frac{p}{p-1}\Lambda_t^{R^\top} - \left[\widehat{B}(t)^\top + 2Z_t^\top \widehat{C}(t)^\top\right]\Sigma_Z\right)\mathrm{d}W_t,
\end{aligned} \tag{B.1.16}$$

where we use that $\frac{\mathrm{d}\mathbb{X}_1}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = C_1 P_1(t, Z_t)\, Y_t M_t^R$ and $\frac{\mathrm{d}\mathbb{X}_2}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_t} = C_2 P_2(t, Z_t)\, M_t^{R^{\frac{p}{p-1}}}$, for $C_1 = \mathbb{E}\left[Y_T M_T^R\right]^{-1}$ and $C_2 = \mathbb{E}[(M_T^R)^{\frac{p}{p-1}}]$. As a consequence, for all $t \in [0, T]$

$$\begin{aligned}
W_t^{\mathbb{X}_1} &= W_t + \int_0^t \left(\Lambda_t^R - \beta_t - \Sigma_Z^\top \tilde{B}(t)\right)\mathrm{d}s,\\
W_t^{\mathbb{X}_2} &= W_t + \int_0^t \left(\frac{p}{p-1}\Lambda_t^R - \Sigma_Z^\top\left[\widehat{B}(t) + 2\widehat{C}(t)\, Z_t\right]\right)\mathrm{d}s,
\end{aligned} \tag{B.1.17}$$

outline the standard Brownian motions under $\mathbb{X}_1$ and $\mathbb{X}_2$, respectively.

To facilitate the application of the Fourier transform to the two conditional probabilities of interest, note that: $\mathcal{A}_T = \{(M_T^R)^{-\frac{1}{p-1}} Y_T \geq \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}\}$. Therefore, $\mathbb{1}_{\{\mathcal{A}_T\}} = f(T, H_T)$, where we define $H_T = \log[(M_T^R)^{-\frac{1}{p-1}} Y_T]$. This implies that: $f_j(t,h) = \mathbb{X}_j(\mathcal{A}_T \mid \mathcal{F}_t) = \mathbb{E}^{\mathbb{X}_j}[f(T, H_T) \mid H_t = h]$, for $j = 1, 2$ and all $t \in [0, T]$. Via the Fourier transform, for $j = 1, 2$ and all $t \in [0, T]$:

$$\begin{aligned}
f_j(t,h) &= \mathbb{E}^{\mathbb{X}_j}[f(T, H_T) \mid H_t = h]\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} f_\kappa^*(T,\omega)\,\phi_{j,T-t}(-\omega - i\kappa, h)\,\mathrm{d}\omega,
\end{aligned} \tag{B.1.18}$$

where $f_\kappa^*(T,\omega)$ is for some $\kappa > 0$ and all $\omega \in \mathbb{R}$ given by the following:

$$\begin{aligned}
f_\kappa^*(T,\omega) &= \int_{-\infty}^{\infty} e^{(i\omega-\kappa)H} f(T,H)\,\mathrm{d}H\\
&= \int_{-\infty}^{\infty} e^{(i\omega-\kappa)H}\mathbb{1}_{\left\{\left(M_T^R\right)^{-\frac{1}{p-1}} Y_T \geq \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}\right\}}\mathrm{d}H\\
&= \int_{-\infty}^{\infty} e^{(i\omega-\kappa)H}\mathbb{1}_{\left\{H \geq \frac{1}{p-1}\log\mathcal{H}^{-1}(X_0)\right\}}\mathrm{d}H = -\frac{e^{(i\omega-\kappa)\frac{1}{p-1}\log\mathcal{H}^{-1}(X_0)}}{i\omega - \kappa}.
\end{aligned} \tag{B.1.19}$$

Observe here that we have:

$$\phi_{j,T-t}(\omega,h) = \int_{-\infty}^{\infty} e^{i\omega H}\phi_j(H,h)\,\mathrm{d}H$$
$$= \mathbb{E}^{\mathbb{X}_j}\left[e^{i\omega H_T}\;\middle|\; H_t = h\right] = \mathbb{E}^{\mathbb{X}_j}\left[\left(M_T^{R^{-\frac{1}{p-1}}} Y_T\right)^{i\omega}\;\middle|\;\mathcal{F}_t\right], \tag{B.1.20}$$

for $j = 1, 2$ and all $t \in [0, T]$. Note that $\phi_j(H, h)$ characterises the conditional density function (under the $\mathbb{X}_j$ measure) corresponding to H_T.

To determine $\phi_{j,T-t}(\omega, h)$ in (B.1.18), we derive:

$$\frac{\mathrm{d}M_t^{R^{-\frac{i\omega}{p-1}}} Y_t^{i\omega}}{M_t^{R^{-\frac{i\omega}{p-1}}} Y_t^{i\omega}} = \Bigg(i\omega\alpha_t + \frac{1}{2}i\omega\,[i\omega-1]\,\beta_t^\top\beta_t + \frac{i\omega}{p-1}R_t$$
$$-\frac{1}{2}\frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1}-1\right]\Lambda_t^{R^\top}\Lambda_t^R + \frac{(i\omega)^2}{p-1}\beta_t^\top\Lambda_t^R$$
$$+i\omega\left[\beta_t^\top + \frac{1}{p-1}\Lambda_t^{R^\top}\right]\lambda_{j,t}\Bigg)\mathrm{d}t + i\omega\left[\beta_t^\top + \frac{1}{p-1}\Lambda_t^{R^\top}\right]\mathrm{d}W_t^{\mathbb{X}_j}, \tag{B.1.21}$$

for $j = 1, 2$, where we define the processes $\lambda_{1,t}$ and $\lambda_{2,t}$ as follows: $\lambda_{1,t} = -(\Lambda_t^R - \beta_t - \Sigma_Z^\top \tilde{B}(t))$ and $\lambda_{2,t} = -(\frac{p}{p-1}\Lambda_t^R - \Sigma_Z^\top[\widehat{B}(t) + 2\widehat{C}(t)\,Z_t])$ for all $t \in [0, T]$. Then, in the sense of Duffie and Kan (1996), we note that the following holds for all $t \in [0, T]$:

$$Q_j(t, Z_t, \omega) = \mathbb{E}^{\mathbb{X}_j}\left[\frac{M_T^{R^{-\frac{i\omega}{p-1}}} Y_T^{i\omega}}{M_t^{R^{-\frac{i\omega}{p-1}}} Y_t^{i\omega}}\;\middle|\;\mathcal{F}_t\right]$$
$$= \mathbb{E}^{\mathbb{C}_j}\left[\exp\left\{-\int_t^T R_j(s, Z_s)\,\mathrm{d}s\right\}\;\middle|\; Z_t\right]. \tag{B.1.22}$$

The measures $\mathbb{C}_j$ are induced by the following Radon-Nikodym derivatives: $\frac{\mathrm{d}\mathbb{C}_j}{\mathrm{d}\mathbb{X}_j}|_{\mathcal{F}_t} = e^{-\frac{1}{2}\int_0^t \bar{\lambda}_s^\top\bar{\lambda}_s\mathrm{d}s + \int_0^t \bar{\lambda}_s^\top \mathrm{d}W_s^{\mathbb{X}_j}}$, where $\bar{\lambda}_t = i\omega[\beta_t + \frac{1}{p-1}\Lambda_t^R]$. Furthermore, $R_1 : [0, T] \times \mathbb{R} \to \mathbb{R}$ and $R_2 : [0, T] \times \mathbb{R} \to \mathbb{R}$ read:

$$-R_1(t, Z_t) = a_{1,t}(\omega) + b_{1,t}(\omega)^\top Z_t + Z_t^\top c_{1,t}(\omega)\,Z_t,$$
$$-R_2(t, Z_t) = a_{2,t}(\omega) + b_{2,t}(\omega)^\top Z_t + Z_t^\top c_{2,t}(\omega)\,Z_t. \tag{B.1.23}$$

The deterministic function $a_{j,t}(\omega)$ is given by:

$$\begin{aligned} a_{j,t}(\omega) = {} & i\omega\alpha_t + \frac{1}{2} i\omega\left[i\omega - 1\right]\beta_t^\top \beta_t + \frac{i\omega}{p-1}\widehat{\delta}_{0,r} \\ & - \frac{1}{2}\frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1} - 1\right]\left(\Lambda_0 - \sigma_\Pi\right)^\top\left(\Lambda_0 - \sigma_\Pi\right) \\ & + \frac{\left(i\omega\right)^2}{p-1}\beta_t^\top\left(\Lambda_0 - \sigma_\Pi\right) + i\omega\left[\beta_t^\top + \frac{1}{p-1}\left(\Lambda_0 - \sigma_\Pi\right)^\top\right]\lambda_{j,0,t}, \end{aligned} \tag{B.1.24}$$

for $j = 1, 2$, all $t \in [0, T]$ and $\omega \in \mathbb{R}$, where we define

$$\begin{aligned} \lambda_{1,0,t} &= -\left(\left(\Lambda_0 - \sigma_\Pi\right) - \beta_t - \Sigma_Z^\top \tilde{B}(t)\right), \\ \lambda_{2,0,t} &= -\left(\frac{p}{p-1}\left[\Lambda_0 - \sigma_\Pi\right] - \Sigma_Z^\top \widehat{B}(t)\right), \end{aligned} \tag{B.1.25}$$

for all $t \in [0, T]$. Note that $a_t(\omega) \in \mathbb{R}$ holds. Similarly, $b_t(\omega) \in \mathbb{R}^2$ and $c_t(\omega) \in \mathbb{R}^{2\times 2}$ hold. Now, define: $\lambda_{1,1,t} = -\Lambda_1$ and $\lambda_{2,1,t} = -\left(\frac{p}{p-1}\Lambda_1 - 2\Sigma_Z^\top \widehat{C}(t)\right)$, for all $t \in [0, T]$. Then, $\lambda_{j,t} = \lambda_{j,0,t} + \lambda_{j,1,t} Z_t$ for all $t \in [0, T]$. The definitions of $b_{j,t}(\omega)$ and $c_{j,t}(\omega)$ are for $j = 1, 2$, all $t \in [0, T]$ and $\omega \in \mathbb{R}$ given by:

$$\begin{aligned} b_{j,t}(\omega) = {} & \frac{i\omega}{p-1}\widehat{\delta}_{1,r} - \frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1} - 1\right]\Lambda_1^\top\left(\Lambda_0 - \sigma_\Pi\right) + \frac{\left(i\omega\right)^2}{p-1}\Lambda_1^\top \beta_t \\ & + i\omega\lambda_{j,1,t}^\top\left(\beta_t + \frac{1}{p-1}\left[\Lambda_0 - \sigma_\Pi\right]\right) + i\omega\frac{1}{p-1}\Lambda_1^\top \lambda_{j,0,t}, \\ c_{j,t}(\omega) = {} & -\frac{1}{2}\frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1} - 1\right]\Lambda_1^\top \Lambda_1 + i\omega\frac{1}{p-1}\Lambda_1^\top \lambda_{j,1,t}. \end{aligned} \tag{B.1.26}$$

As in (B.1.12), we postulate the following ansatz for $Q_j(t, Z_t, \omega)$:

$$Q_j(t, Z_t, \omega) = \exp\left\{\bar{A}_j(t,\omega) + \bar{B}_j(t,\omega)^\top Z_t + Z_t^\top \bar{C}_j(t,\omega) Z_t\right\}. \tag{B.1.27}$$

Here, $\bar{A}_j : [0, T] \times \mathbb{R} \to \mathbb{R}$, $\bar{B}_j : [0, T] \times \mathbb{R} \to \mathbb{R}^2$, and $\bar{C}_j : [0, T] \times \mathbb{R} \to \mathbb{R}^{2\times 2}$, for $j = 1, 2$, are deterministic functions. We know that:

$$\begin{aligned} & Q_{j,t} - Q_{j,Z}^\top K_Z Z + \frac{1}{2}\mathrm{tr}\left(Q_{j,ZZ}\right) \\ & - R_j(t, Z)\, Q_j + i\omega Q_{j,Z}^\top \Sigma_Z\left[\beta + \frac{1}{p-1}\Lambda^R\right] = 0. \end{aligned} \tag{B.1.28}$$

Then, we derive the following system of ODE's:

$$\begin{aligned}\bar{A}_j'(t) &= -i\omega \bar{B}_j(t)^\top \Sigma_Z \bar{\Lambda}_{0,t}^R - \frac{1}{2}\bar{B}_j(t)^\top \bar{B}_j(t) + \bar{a}_{j,t}(\omega),\\ \bar{B}_j'(t) &= \left(-\frac{i\omega}{p-1}\Lambda_1^\top \Sigma_Z^\top + K_Z^\top - 2\bar{C}_j(t)^\top\right)\bar{B}_j(t) + \bar{b}_{j,t}(\omega),\\ \bar{C}_j'(t) &= 2\left(K_Z^\top - \frac{i\omega}{p-1}\Lambda_1^\top \Sigma_Z^\top\right)\bar{C}_j(t) - 2\bar{C}_j(t)^\top \bar{C}_j(t) - c_{j,t}(\omega),\end{aligned} \tag{B.1.29}$$

for $j = 1, 2$, where we define $\bar{a}_{j,t}(\omega) = -a_{j,t}(\omega) - \operatorname{tr}\left(\bar{C}_j(t)\right)$, $\bar{b}_{j,t}(\omega) = -b_{j,t}(\omega) - 2i\omega\bar{C}_j(t)^\top \Sigma_Z \bar{\Lambda}_{0,t}^R$, and $\bar{\Lambda}_{0,t}^R = \frac{1}{p-1}\left[\Lambda_0 - \sigma_\Pi\right] + \beta_t$, for all $t \in [0, T]$ and $\omega \in \mathbb{R}$. Observe that: $\bar{A}_j(T, \omega) = 0$, $\bar{B}_j(T, \omega) = 0_2$ and $\bar{C}_j(T, \omega) = 0_{2\times 2}$, for $j = 1, 2$ and all $\omega \in \mathbb{R}$. As for the system of ODE's in (B.1.14), $\bar{B}_j(t)$ and $\bar{C}_j(t)$ cannot be solved in closed-form. Hence, we are solely able to state that the following[24] holds, for $j = 1, 2$:

$$\bar{A}_j(t) = -\int_t^T \left[-i\omega\bar{B}_j(s)^\top \Sigma_Z \bar{\Lambda}_{0,s}^R - \frac{1}{2}\bar{B}_j(s)^\top \bar{B}_j(s) + \bar{a}_{j,s}(\omega)\right] \mathrm{d}s. \tag{B.1.30}$$

Then, we conclude by observing that, for all $t \in [0, T]$, $\omega \in \mathbb{R}$, and $j = 1, 2$:

$$\begin{aligned}\phi_{j,T-t}(\omega, h) &= \int_{-\infty}^{\infty} e^{i\omega H}\phi_j(H, h)\,\mathrm{d}H\\ &= \mathbb{E}^{\mathbb{X}_j}\left[\left(M_T^{R^{-\frac{1}{p-1}}} Y_T\right)^{i\omega} \,\middle|\, \mathcal{F}_t\right] = Q_j(t, Z_t, \omega)\left(M_t^{R^{-\frac{1}{p-1}}} Y_t\right)^{i\omega}.\end{aligned} \tag{B.1.31}$$

B.2 Proof of Corollary 4.3.3

For the $\Lambda_1 = 0_{4\times 2}$ case, we have that $\mathbb{E}\left[\frac{Y_T M_T^R}{Y_t M_t^R} \,\middle|\, \mathcal{F}_t\right] = e^{\tilde{A}(t) + \tilde{B}(t)^\top Z_t}$, where

$$\begin{aligned}\tilde{A}(t) &= -\int_t^T \left[\tilde{B}(s)^\top \Sigma_Z c_s - \frac{1}{2}\tilde{B}(s)^\top \tilde{B}(s) + a_s\right] \mathrm{d}s,\\ \tilde{B}(t) &= K_Z^{\top^{-1}}\left[\exp\left\{-K_Z^\top (T-t)\right\} - I_{2\times 2}\right]\widehat{\delta}_{1,r},\end{aligned} \tag{B.2.32}$$

[24] Note that the three deterministic functions, $\bar{A}_j$, $\bar{B}_j$, and $\bar{C}_j$, in a.o. (B.1.29), depend for $j = 1, 2$ on $\omega \in \mathbb{R}$. We have suppressed this dependency for notational simplicity.

for all $t \in [0,T]$, in which $a_t = \widehat{\delta}_{0,r} - \alpha_t + \beta_t^\top (\Lambda_0 - \sigma_\Pi)$ and $c_t = \Lambda_0 - \sigma_\Pi - \beta_t$. Likewise, we have that the following holds true: $\mathbb{E}[(\frac{M_T^R}{M_t^R})^{\frac{p}{p-1}} \mid \mathcal{F}_t] = e^{\widehat{A}(t)+\widehat{B}(t)^\top Z_t}$. Suppose that $\tilde{a} = \frac{p}{p-1}\widehat{\delta}_{0,r} - \frac{1}{2}\frac{p}{(p-1)^2}(\Lambda_0 - \sigma_\Pi)^\top (\Lambda_0 - \sigma_\Pi)$. Then,$\widehat{A}(t)$ and $\widehat{B}(t)$ read for all $t \in [0,T]$ as follows:

$$\begin{aligned} \widehat{A}(t) &= -\int_t^T \left[\frac{p}{p-1}\widehat{B}(s)^\top \Sigma_Z \Lambda_0^R - \frac{1}{2}\widehat{B}(s)^\top \widehat{B}(s) + \tilde{a}\right] \mathrm{d}s, \\ \widehat{B}(t) &= \frac{p}{p-1} K_Z^{\top^{-1}} \left[\exp\left\{-K_Z^\top (T-t)\right\} - I_{2\times 2}\right] \widehat{\delta}_{1,r}. \end{aligned} \tag{B.2.33}$$

Now, from (B.1.21), let us note that the following holds:

$$\begin{aligned} \frac{\mathrm{d}M_t^{R^{-\frac{1}{p-1}}} Y_t}{M_t^{R^{-\frac{1}{p-1}}} Y_t} = &\left(\alpha_t + \frac{\widehat{\delta}_{0,r}}{p-1} + \frac{1}{2}\frac{p}{(p-1)^2}\Lambda_0^{R^\top}\Lambda_0^R \right.\\ &+ \frac{1}{p-1}\beta_t^\top \Lambda_0^R + \left[\beta_t^\top + \frac{1}{p-1}\Lambda_0^{R^\top}\right]\lambda_{j,t} \\ &\left. + \frac{1}{p-1}\widehat{\delta}_{1,r}^\top Z_t\right)\mathrm{d}t + \left[\beta_t^\top + \frac{1}{p-1}\Lambda_0^{R^\top}\right]\mathrm{d}W_t^{\mathbb{X}_j}, \end{aligned} \tag{B.2.34}$$

where $\Lambda_0^R = \Lambda_0 - \sigma_\Pi$, and in which $\lambda_{1,t} = -(\Lambda_0^R - \beta_t - \Sigma_Z^\top \tilde{B}(t))$ and $\lambda_{2,t} = -(\frac{p}{p-1}\Lambda_0^R - \Sigma_Z^\top \widehat{B}(t))$ for all $t \in [0,T]$. Suppose that we write the instantaneous drift term of the preceding SDE in (B.2.34) as follows: $\nu_{j,t} + \frac{1}{p-1}\widehat{\delta}_{1,r}^\top Z_t$, where $\nu_{j,t}$ is defined for $j = 1,2$ and all $t \in [0,T]$ as:

$$\nu_{j,t} = \alpha_t + \frac{\widehat{\delta}_{0,r}}{p-1} + \frac{1}{2}\frac{p}{(p-1)^2}\Lambda_0^{R^\top}\Lambda_0^R + \frac{1}{p-1}\beta_t^\top \Lambda_0^R + \left[\beta_t^\top + \frac{1}{p-1}\Lambda_0^{R^\top}\right]\lambda_{j,t}, \tag{B.2.35}$$

Then, let us derive that the following holds true for all $t \in [0,T]$:

$$\begin{aligned} \int_t^T \frac{1}{p-1}\widehat{\delta}_{1,r}^\top Z_s \mathrm{d}s &= \int_t^T \int_0^s \frac{1}{p-1}\widehat{\delta}_{1,r}^\top e^{-K_Z(s-u)}\Sigma_Z \mathrm{d}W_u \mathrm{d}s \\ &= \frac{1}{p-1}\widehat{\delta}_{1,r}^\top K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-t]}\right) Z_t \\ &+ \frac{1}{p-1}\widehat{\delta}_{1,r}^\top \int_t^T K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-s]}\right)\Sigma_Z \mathrm{d}W_s. \end{aligned} \tag{B.2.36}$$

As $t \mapsto K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-t]}\right)\Sigma_Z$ characterises a deterministic function of time, the latter integral is normally distributed – conditional on $\mathcal{F}_t$. Hence:

$$\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} = \int_t^T \left(-\frac{1}{2} \left[\beta_s + \frac{1}{p-1}\Lambda_0^R \right]^\top \left[\beta_s + \frac{1}{p-1}\Lambda_0^R \right] \right.$$
$$\left. + \nu_{j,s} + \frac{1}{p-1}\widehat{\delta}_{1,r}^\top Z_s \right) \mathrm{d}s + \int_t^T \left(\beta_s^\top + \frac{1}{p-1}\Lambda_0^{R^\top} \right) \mathrm{d}W_s^{\mathbb{X}_j}, \tag{B.2.37}$$

holds for all $t \in [0,T]$. As for the stochastic process in (B.2.36), we note that $t \mapsto \beta_t^\top + \frac{\Lambda_0^{R^\top}}{p-1} + \frac{\widehat{\delta}_{1,r}^\top}{p-1} K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-t]}\right)\Sigma_Z$ is a deterministic function of time, to conclude that the process in (B.2.37) is normally distributed conditional on $\mathcal{F}_t$. This concretely means that the $M_T^{R^{-\frac{1}{p-1}}} Y_T / M_t^{R^{-\frac{1}{p-1}}} Y_t$ process is log-normally distributed, conditional on $\mathcal{F}_t$, for all $t \in [0,T]$. Let us recall that:

$$\mathcal{A}_T = \left\{ \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \geq \frac{\mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \right\} \tag{B.2.38}$$

Hence, combining the preceding arguments/results, we can evaluate $\mathbb{X}_1(\mathcal{A}_T \mid \mathcal{F}_t)$ and $\mathbb{X}_2(\mathcal{A}_T \mid \mathcal{F}_t)$ explicitly. That is:

$$\mathbb{X}_j(\mathcal{A}_T \mid \mathcal{F}_t) = \Phi\left(\frac{-\log\left(\frac{\mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \right) + \mathbb{E}^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t \right]}{\sqrt{\mathbb{V}\mathrm{ar}^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t \right]}} \right). \tag{B.2.39}$$

In this identity, the variance term equates for $j = 1,2$ and all $t \in [0,T]$ to:

$$\mathbb{V}\mathrm{ar}^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t \right] = \int_t^T \left\| \beta_s^\top + \frac{1}{p-1}\Lambda_0^{R^\top} \right.$$
$$\left. + \frac{\widehat{\delta}_{1,r}^\top}{p-1} K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-s]}\right)\Sigma_Z \right\|_{\mathbb{R}^4}^2 \mathrm{d}s. \tag{B.2.40}$$

Here, $\|\cdot\|_{\mathbb{R}^n}$ is the n-dimensional Euclidean norm. Moreover, the expectation on the right-hand side of (B.2.39) reads for $j = 1, 2$ and all $t \in [0, T]$ as:

$$\begin{aligned}\mathbb{E}^{\mathbb{X}_j}\left[\log \frac{M_T^{R^{-\frac{1}{p-1}}} Y_T}{M_t^{R^{-\frac{1}{p-1}}} Y_t} \,\middle|\, \mathcal{F}_t\right] &= \frac{1}{p-1}\widehat{\delta}_{1,r}^{\top} K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-t]}\right) Z_t \\ &+ \int_t^T \left(\nu_{j,s} - \frac{1}{2}\left\|\beta_s + \frac{1}{p-1}\Lambda_0^R\right\|_{\mathbb{R}^4}^2\right)\mathrm{d}s \\ &+ \frac{\widehat{\delta}_{1,r}^{\top}}{p-1}\int_t^T K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-s]}\right)\Sigma_Z \lambda_{j,s}\mathrm{d}s.\end{aligned} \tag{B.2.41}$$

B.3 Derivation of (4.3.8)

Let us define $\widehat{f}(T, G_T, J_T) = M_T^{R^{\frac{p}{p-1}}}(M_T^{R^{-\frac{1}{p-1}}} Y_T - \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}})\mathbb{1}_{\{\mathcal{A}_T\}}$, where $G_T = \log M_T^{R^{-\frac{1}{p-1}}} Y_T$ and $J_T = \log M_T^{R^{\frac{p}{p-1}}}$. As in Appendix B.1: $X_t^{\text{opt}} M_t = \widehat{f}(t, g, j) = \mathbb{E}\left[\widehat{f}(T, G_T, J_T) \,\middle|\, G_t = g,\ J_t = j\right]$. Then:

$$\begin{aligned}\widehat{f}(t, g, j) &= \mathbb{E}\left[\widehat{f}(T, G_T, J_T) \,\middle|\, G_t = g,\ J_t = j\right] \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}_\kappa^*(T, \omega)\,\widehat{\phi}_{T-t}(-\omega - i\kappa, g, j)\,\mathrm{d}\omega,\end{aligned} \tag{B.3.42}$$

for all $t \in [0, T]$, which is a direct result of the Fourier transform. In the latter identity, $\widehat{f}_\kappa^*(T, \omega)$ is for some $\kappa > 1$ and all $\omega \in \mathbb{R}$ given by:

$$\begin{aligned}\widehat{f}_\kappa^*(T, \omega) &= \int_{-\infty}^{\infty} e^{(i\omega - \kappa)G}\widehat{f}(T, G, J)\,\mathrm{d}G \\ &= \int_{-\infty}^{\infty} e^{(i\omega-\kappa)G}\left(e^G - \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}\right)\mathbb{1}_{\left\{G \geq \frac{1}{p-1}\log \mathcal{H}^{-1}(X_0)\right\}}\mathrm{d}G \\ &= -\frac{e^{(i\omega-\kappa+1)\frac{1}{p-1}\log \mathcal{H}^{-1}(X_0)}}{i\omega - \kappa + 1} + \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}\frac{e^{(i\omega-\kappa)\frac{1}{p-1}\log\mathcal{H}^{-1}(X_0)}}{i\omega - \kappa}.\end{aligned} \tag{B.3.43}$$

Here, we have:

$$\widehat{\phi}_{T-t}(\omega, g, j) = \mathbb{E}\left[M_T^{R^{\frac{p-i\omega}{p-1}}} Y_T^{i\omega} \,\middle|\, \mathcal{F}_t\right], \tag{B.3.44}$$

for all $t \in [0,T]$. Note that $\phi(G,J,g,j)$ characterises the conditional joint density corresponding to G_T and J_T.

As the characterisation of $\widehat{\phi}_{T-t}(\omega,g,j)$ is similar to the one of $\phi_{T-t}(\omega,g,j)$ in Appendix B.1, we omit an elaborate derivation. Define the deterministic function $(t,\omega) \mapsto \widehat{a}_t(\omega)$:

$$\begin{aligned} \widehat{a}_t(\omega) &= i\omega\alpha_t + \frac{1}{2} i\omega \left[i\omega - 1\right] \beta_t^\top \beta_t - \frac{p - i\omega}{p-1} \widehat{\delta}_{0,r} \\ &+ \frac{1}{2} \frac{p-i\omega}{p-1} \left[\frac{p-i\omega}{p-1} - 1\right] \Lambda_0^{R\top} \Lambda_0^R - \frac{(p-i\omega)\, i\omega}{p-1} \beta_t^\top \Lambda_0^R. \end{aligned} \tag{B.3.45}$$

Note that $\widehat{a}_t(\omega) \in \mathbb{R}$. In addition to this, introduce the following two functions, $(t,\omega) \mapsto \widehat{b}_t(\omega)$ and $(t,\omega) \mapsto \widehat{c}_t(\omega)$, with $\widehat{b}_t(\omega) \in \mathbb{R}^2$ and $\widehat{c}_t(\omega) \in \mathbb{R}^{2\times 2}$:

$$\widehat{b}_t(\omega) = \frac{p-i\omega}{p-1} \left(-\widehat{\delta}_{1,r} + \left[\frac{p-i\omega}{p-1} - 1\right] \Lambda_1^\top \Lambda_0^R - i\omega \Lambda_1^\top \beta_t \right), \tag{B.3.46}$$

and

$$\widehat{c}_t(\omega) = \frac{1}{2} \frac{p-i\omega}{p-1} \left[\frac{p-i\omega}{p-1} - 1\right] \Lambda_1^\top \Lambda_1. \tag{B.3.47}$$

Then, postulate the next ansatz for $\widehat{Q}(t,Z_t,\omega) = \widehat{\phi}_{T-t}(\omega,g,j)$:

$$\widehat{Q}(t,Z_t,\omega) = \exp\left\{ \widehat{A}_{\widehat{Q}}(t,\omega) + \widehat{B}_{\widehat{Q}}(t,\omega)^\top Z_t + Z_t^\top \widehat{C}_{\widehat{Q}}(t,\omega) Z_t \right\}, \tag{B.3.48}$$

for all $t \in [0,T]$ and $\omega \in \mathbb{R}$. We assume that $\widehat{A}_{\widehat{Q}} : [0,T] \times \mathbb{R} \to \mathbb{R}$, $\widehat{B}_{\widehat{Q}} : [0,T] \times \mathbb{R} \to \mathbb{R}^2$, and $\widehat{C}_{\widehat{Q}} : [0,T] \times \mathbb{R} \to \mathbb{R}^{2\times 2}$ are deterministic functions of time $t \in [0,T]$ and $\omega \in \mathbb{R}$ alone. As $\widehat{Q}(t,Z_t,\omega) M_t^{R\,\frac{p-i\omega}{p-1}} Y_t^{i\omega}$ is a $\mathbb{P}$-martingale, we derive the following system of ODE's:

$$\begin{aligned} \widehat{A}'_{\widehat{Q}}(t) &= -\widehat{B}_{\widehat{Q}}(t)^\top \Sigma_Z \widehat{\Lambda}^R_{\widehat{Q},t}(\omega) - \frac{1}{2} \widehat{B}_{\widehat{Q}}(t)^\top \widehat{B}_{\widehat{Q}}(t) + \bar{a}_{\widehat{Q},t}(\omega), \\ \widehat{B}'_{\widehat{Q}}(t) &= \left(\frac{p-i\omega}{p-1} \Lambda_1^\top \Sigma_Z^\top + K_Z^\top - 2\widehat{C}_{\widehat{Q}}(t)^\top \right) \widehat{B}_{\widehat{Q}}(t) + \bar{b}_{\widehat{Q},t}(\omega), \\ \widehat{C}'_{\widehat{Q}}(t) &= 2\left(K_Z^\top + \frac{p-i\omega}{p-1} \Lambda_1^\top \Sigma_Z^\top \right) \widehat{C}_{\widehat{Q}}(t) - 2\widehat{C}_{\widehat{Q}}(t)^\top \widehat{C}_{\widehat{Q}}(t) - \widehat{c}_t(\omega), \end{aligned} \tag{B.3.49}$$

where we define:

$$\begin{aligned}
\bar{a}_{\widehat{Q},t}(\omega) &= -\widehat{a}_t(\omega) - \operatorname{tr}\left(\widehat{C}_{\widehat{Q}}(t)\right), \\
\bar{b}_{\widehat{Q},t}(\omega) &= -\widehat{b}_t(\omega) - 2\widehat{C}_{\widehat{Q}}(t)^\top \Sigma_Z \widehat{\Lambda}^R_{\widehat{Q},t}(\omega), \\
\widehat{\Lambda}^R_{\widehat{Q},t}(\omega) &= -\frac{p - i\omega}{p-1}\left[\Lambda_0 - \sigma_\Pi\right] + i\omega\beta_t,
\end{aligned} \tag{B.3.50}$$

for all $t \in [0,T]$ and $\omega \in \mathbb{R}$. Note here that we suppress the dependencies of $\widehat{A}_{\widehat{Q}}$, $\widehat{B}_{\widehat{Q}}$, and $\widehat{C}_{\widehat{Q}}$ on $\omega \in \mathbb{R}$. Moreover, we have that: $\widehat{A}_{\widehat{Q}}(T,\omega) = 0$, $\widehat{B}_{\widehat{Q}}(T,\omega) = 0_2$, and $\widehat{C}_{\widehat{Q}}(T,\omega) = 0_{2\times 2}$, for all $\omega \in \mathbb{R}$. Note the similarity between this system and the system in (B.1.29). Therefore, we only know that:

$$\widehat{A}_{\widehat{Q}}(t) = -\int_t^T \left[-\widehat{B}_{\widehat{Q}}(s)^\top \Sigma_Z \widehat{\Lambda}^R_{\widehat{Q},s}(\omega) - \frac{1}{2}\widehat{B}_{\widehat{Q}}(s)^\top \widehat{B}_{\widehat{Q}}(s) + \bar{a}_{\widehat{Q},s}(\omega)\right] \mathrm{d}s. \tag{B.3.51}$$

Then, we are able to conclude by deriving that, for all $t \in [0,T]$ and $\omega \in \mathbb{R}$:

$$\begin{aligned}
\widehat{\phi}_{T-t}(\omega, g, j) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^J e^{i\omega G}\phi(G, J, g, j)\,\mathrm{d}G\mathrm{d}J \\
&= \mathbb{E}\left[M_T^{R\,\frac{p-i\omega}{p-1}} Y_T^{i\omega} \,\middle|\, \mathcal{F}_t\right] = \widehat{Q}(t, Z_t, \omega)\, M_t^{R\,\frac{p-i\omega}{p-1}} Y_t^{i\omega}.
\end{aligned} \tag{B.3.52}$$

Appendix C Proofs II

C.1 Proof of Proposition 4.3.4

The proof of this proposition primarily consists in an application of Itô's Lemma to $X_t^{\text{opt}} M_t$. This enables us to analytically identify ψ_t. Therefore, let us start by noting that $X_t^{\text{opt}} M_t$ is for all $t \in [0,T]$ given by:

$$\begin{aligned}
X_t^{\text{opt}} M_t = {}& Y_t M_t^R P_1(t, Z_t)\frac{1}{2\pi}\int_{-\infty}^{\infty} f_\kappa^*(T,\omega)\,\phi_{1,T-t}(-\omega - i\kappa, h)\,\mathrm{d}\omega \\
&- \left(\eta^{\text{opt}\,\frac{1}{p}} M_t^R\right)^{\frac{p}{p-1}} P_2(t, Z_t)\frac{1}{2\pi}\int_{-\infty}^{\infty} f_\kappa^*(T,\omega)\,\phi_{2,T-t}(-\omega - i\kappa, h)\,\mathrm{d}\omega,
\end{aligned} \tag{C.1.1}$$

where $\eta^{\text{opt}} = \mathcal{H}^{-1}(X_0)$. To identify ψ_t in (4.3.2), we must find the diffusion coefficients for the latter process $X_t^{\text{opt}} M_t$. Making use of a simple application of Itô's Lemma, we find the subsequent SDE's:

$$\begin{aligned} \frac{\mathrm{d}Y_t M_t^R P_1}{Y_t M_t^R P_1} &= \left(- \left[\Lambda_t^R - \beta_t \right]^\top + \tilde{B}(t)^\top \Sigma_Z \right) \mathrm{d}W_t, \\ \frac{\mathrm{d}M_t^{R^{\frac{p}{p-1}}} P_2}{M_t^{R^{\frac{p}{p-1}}} P_2} &= \left(- \frac{p}{p-1} \Lambda_t^{R^\top} + \widehat{D}(t)^\top \Sigma_Z \right) \mathrm{d}W_t. \end{aligned} \tag{C.1.2}$$

Now, define for $j = 1, 2$, and all $t \in [0, T]$:

$$R_j(t, G_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_\kappa^*(T, \omega)\, \phi_{j,T-t}(-\omega - i\kappa, h)\, \mathrm{d}\omega. \tag{C.1.3}$$

Using an application of the multidimensional version of Itô's Lemma, we are able to derive that the SDE's of these processes are given by:

$$\begin{aligned} \mathrm{d}R_j &= \frac{\partial R_j}{\partial t} + (\nabla_G R_j)^\top \mathrm{d}G_t + \frac{1}{2} (\mathrm{d}G_t)^\top (H_G R_j)\, \mathrm{d}G_t \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Big(f_\kappa^*(T, \omega)\, \phi_{j,T-t}(-\omega - i\kappa, h) \Big[\bar{B}_j(t)^\top \Sigma_Z \\ &+2 Z_t^\top \bar{C}_j(t)^\top \Sigma_Z + i(-\omega - i\kappa) \left(\beta_t^\top + \frac{1}{p-1} \Lambda_t^{R^\top} \right) \Big] \Big) \mathrm{d}\omega \mathrm{d}W_t. \end{aligned} \tag{C.1.4}$$

Note that $t \mapsto \bar{B}_j(t)$ and $t \mapsto \bar{C}_j(t)$ are assumed to incorporate the $-\omega - i\kappa$ argument. For notational purposes, we denote the diffusion coefficient of the preceding SDE as $\widehat{R}_j(t, \phi, Z_t)$. More concretely, we fix: $\mathrm{d}R_j = \widehat{R}_j(t, \phi, Z_t)\, \mathrm{d}W_t$. Combining terms, we consequently find that the following holds:

$$\begin{aligned} \mathrm{d}Y_t M_t^R P_1 R_1 = Y_t M_t^R P_1 \Big[\Big(&- \left[\Lambda_t^R - \beta_t \right]^\top \\ &+ \tilde{B}(t)^\top \Sigma_Z \Big) R_1 + \widehat{R}_j(t, \phi, Z_t) \Big] \mathrm{d}W_t. \end{aligned} \tag{C.1.5}$$

Likewise, we are able to derive the subsequent SDE:

$$\begin{aligned} \mathrm{d}M_t^{R^{\frac{p}{p-1}}} P_2 R_2 = M_t^{R^{\frac{p}{p-1}}} P_2 \Big[\Big(&- \frac{p}{p-1} \Lambda_t^{R^\top} + \Big[\widehat{B}(t)^\top \\ &+ 2 Z_t^\top \widehat{C}(t)^\top \Big] \Sigma_Z \Big) R_2 + \widehat{R}_j(t, \phi, Z_t) \Big] \mathrm{d}W_t. \end{aligned} \tag{C.1.6}$$

The sum of the SDE's in (C.1.5) and (C.1.6) is identical to the SDE of $X_t^{\mathrm{opt}} M_t$. To conclude the proof of Corollary 4.3.5, we observe that ψ_t in (4.3.2) is identical to the diffusion coefficient of the SDE for $X_t^{\mathrm{opt}} M_t$:

$$\begin{aligned} \psi_t^\top \mathrm{d}W_t &= \mathrm{d}Y_t M_t^R P_1(t, Z_t) R_1 \left(t, Z_t, M_t^{R\frac{p}{p-1}} Y_t\right) \\ &\quad - \mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}} \mathrm{d}M_t^{R\frac{p}{p-1}} P_2(t, Z_t) R_2 \left(t, Z_t, M_t^{R\frac{p}{p-1}} Y_t\right). \end{aligned} \tag{C.1.7}$$

C.2 Proof of Corollary 4.3.5

According to the Clark-Ocone formula, we have:

$$X_T^{\mathrm{opt}} M_T = X_0 + \int_0^T \mathbb{E}\left[\mathcal{D}_t^W X_T^{\mathrm{opt}} M_T \mid \mathcal{F}_t\right]^\top \mathrm{d}W_t, \tag{C.2.8}$$

cf. Karatzas et al. (1991b) and Ocone and Karatzas (1991). Here, $\mathcal{D}_t^W : \mathbb{D}^{1,2}([0,T]) \to L^2(\Omega \times [0,T])^4$ represents the Mallivain derivative kernel, where $\mathbb{D}^{1,2}([0,T])$ stands for the Sobolev-Watanabe space of all $L^2(\Omega \times [0,T])$-valued Mallivain differentiable processes. Hence, from (C.2.8), for all $t \in [0,T]$:

$$\psi_t = \mathbb{E}\left[\mathcal{D}_t^W X_T^{\mathrm{opt}} M_T \mid \mathcal{F}_t\right]. \tag{C.2.9}$$

Then, we note that:

$$\begin{aligned} \mathcal{D}_t^W X_T^{\mathrm{opt}} M_T &= M_T \mathcal{D}_t^W \left(Y_T \Pi_T - \left(\mathcal{H}^{-1}(X_0) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T\right) \mathbb{1}_{\{\mathcal{A}_T\}} \\ &\quad + \left(Y_T \Pi_T - \left(\mathcal{H}^{-1}(X_0) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T\right) \mathbb{1}_{\{\mathcal{A}_T\}} \mathcal{D}_t^W M_T. \end{aligned} \tag{C.2.10}$$

Using the fact that $\mathcal{D}_t^W H_T = H_T \mathcal{D}_t^W \log H_T$, we are able to derive that the following holds true for all $t \in [0,T]$:

$$\begin{aligned} \frac{\mathcal{D}_t^W M_T}{M_T} &= -\left[\delta_{1,r}^\top K_Z^{-1}\left[I_{2\times 2} - \exp\{-K_Z(T-t)\}\right]\Sigma_Z\right]^\top - \Lambda_0, \\ \frac{\mathcal{D}_t^W Y_T \Pi_T}{Y_T \Pi_T} &= \left[\delta_{1,\pi}^\top K_Z^{-1}\left[I_{2\times 2} - \exp\{-K_Z(T-t)\}\right]\Sigma_Z\right]^\top + \sigma_\Pi + \beta_t. \end{aligned} \tag{C.2.11}$$

Ultimately, we combine all preceding arguments to derive the relevant Malliavin derivative of $\left(\mathcal{H}^{-1}\left(X_0\right) M_T \Pi_T\right)^{\frac{1}{p-1}} \Pi_T$. After re-arranging terms, we find the weights in Corollary 4.3.5. The former reads for all $t \in [0, T]$:

$$\frac{\mathcal{D}_t^W M_T^{R^{\frac{1}{p-1}}} \Pi_T}{M_T^{R^{\frac{1}{p-1}}} \Pi_T} = -\frac{1}{p-1}\left(\Lambda_0 - \sigma_\Pi\right) + \left[\left(\delta_{1,\pi} - \frac{1}{p-1}\widehat{\delta}_{1,r}\right)^\top K_Z^{-1}\left[I_{2\times 2} - \exp\left\{-K_Z\left(T-t\right)\right\}\right]\Sigma_Z\right]^\top + \sigma_\Pi. \tag{C.2.12}$$

C.3 Derivation of (4.3.13)

The closed-form specification of $X_t^{\text{opt}} M_t$ implicit in (4.3.8) reads as:

$$\begin{aligned} X_t^{\text{opt}} M_t &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}_\kappa^*\left(T,\omega\right)\widehat{\phi}_{T-t}\left(-\omega - i\kappa, g, j\right)\mathrm{d}\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}_\kappa^*\left(T,\omega\right)\widehat{Q}\left(t, Z_t, -\omega - i\kappa\right) M_t^{R^{\frac{p-i(-\omega-i\kappa)}{p-1}}} Y_t^{i(-\omega-i\kappa)}\mathrm{d}\omega, \end{aligned} \tag{C.3.13}$$

for all $t \in [0, T]$. An application of Itô's Lemma to $X_t^{\text{opt}} M_t$ in (C.3.13) suffices to conclude the proof. In particular, we find:

$$\begin{aligned} \mathrm{d}X_t^{\text{opt}} M_t = \frac{1}{2\pi}\int_{-\infty}^{\infty} & \widehat{f}_\kappa^*\left(T,\omega\right)\widehat{Q}\left(t, Z_t, -\omega - i\kappa\right) M_t^{R^{\frac{p-i(-\omega-i\kappa)}{p-1}}} Y_t^{i(-\omega-i\kappa)} \\ & \left[\left(\widehat{B}_{\widehat{Q}}\left(t, -\omega - i\kappa\right) + 2\widehat{C}_{\widehat{Q}}\left(t, -\omega - i\kappa\right) Z_t\right)^\top \Sigma_Z \right. \\ & \left. + \left(i\left[-\omega - i\kappa\right]\beta_t - \frac{p - i\left[-\omega - i\kappa\right]}{p-1}\Lambda_t^R\right)^\top\right]\mathrm{d}\omega\mathrm{d}W_t. \end{aligned} \tag{C.3.14}$$

Appendix D Proofs III

D.1 Derivation of (4.4.3)

We must determine $\bar{\alpha}$ and $\bar{\beta}$ such that the following holds:

$$\mathbb{E}\left[a_T\right] = \mathbb{E}\left[\widehat{a}_T\right] \quad \text{and} \quad \mathbb{E}\left[a_T Z_T\right] = \mathbb{E}\left[\widehat{a}_T Z_T\right]. \tag{D.1.1}$$

Therefore, we first determine the expectations for $\widehat{a}_T$. Its first moment reads:

$$\begin{aligned}\mathbb{E}\left[\widehat{a}_T\right] &= C\exp\left\{\bar{\alpha}\right\}\mathbb{E}\left[\exp\left\{\bar{\beta}^\top \int_0^T e^{-K_Z(T-s)}\Sigma_Z \mathrm{d}W_s\right\}\right] \\ &= C\exp\left\{\bar{\alpha} + \frac{1}{2}\bar{\beta}^\top \left(\int_0^T e^{-K_Z(T-s)}\left[e^{-K_Z(T-s)}\right]^\top \mathrm{d}s\right)\bar{\beta}\right\}.\end{aligned} \tag{D.1.2}$$

Now, define the change of measure from $\mathbb{P}$ to $\mathbb{W}$ (with $\mathbb{W} \sim \mathbb{P}$), induced by the following Radon-Nikodym derivative:

$$\left.\frac{\mathrm{d}\mathbb{W}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = e^{-\frac{1}{2}\bar{\beta}^\top \int_0^t e^{K_Z(T-s)}\left(e^{K_Z(T-s)}\right)^\top \mathrm{d}s + \bar{\beta}^\top \int_0^t e^{K_Z(T-s)}\Sigma_Z \mathrm{d}W_s}, \tag{D.1.3}$$

for all $t \in [0,T]$. Straightforwardly, under the $\mathbb{W}$ measure, the following process is a standard Brownian motion: $W_t^{\mathbb{W}} = W_t - \int_0^t \Sigma_Z^\top \left(e^{K_Z(T-s)}\right)^\top \bar{\beta}\mathrm{d}s$, for all $t \in [0,T]$. Using this change of measure, we find that $\mathbb{E}\left[\widehat{a}_T Z_T\right]$ reduces to:

$$\begin{aligned}\mathbb{E}\left[\widehat{a}_T Z_T\right] &= C\mathbb{E}\left[\widehat{a}_T\right]\underbrace{\mathbb{E}\left[\frac{\widehat{a}_T}{\mathbb{E}\left[\widehat{a}_T\right]}Z_T\right]}_{=\mathbb{E}^{\mathbb{W}}[Z_T]} \\ &= C\mathbb{E}\left[\widehat{a}_T\right]\left(\int_0^T e^{-K_Z(T-s)}\left[e^{K_Z(T-s)}\right]^\top \mathrm{d}s\right)\bar{\beta}.\end{aligned} \tag{D.1.4}$$

Then, we turn to the relevant expectations for a_T. Similar to the identity in (D.1.2), we are able to make use of the log-normality of $\frac{P_{T,T+i}}{\Pi_T}$ to derive its first moment. In concrete terms, its first moment is given by:

$$\mathbb{E}\left[a_T\right] = C\sum_{i=1}^{\tau_A}\mathbb{E}\left[\frac{P_{T,T+i}}{\Pi_T}\right], \tag{D.1.5}$$

where the expectation reads for all $i = 1, \ldots, \tau_A$:

$$\begin{aligned}\mathbb{E}\left[\frac{P_{T,T+i}}{\Pi_T}\right] = \exp\Bigg\{ & A^R(i) + \frac{1}{2}B^R(i)^\top \\ & \times\left(\int_0^T e^{-K_Z(T-s)}\left[e^{-K_Z(T-s)}\right]^\top \mathrm{d}s\right)B^R(i)\Bigg\}.\end{aligned} \tag{D.1.6}$$

Following the derivation in (D.1.4), introduce:

$$\left.\frac{\mathrm{d}\mathbb{W}_i}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t} = \left.\frac{\mathrm{d}\mathbb{W}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{F}_t,\bar{\beta}=B^R(i)}, \tag{D.1.7}$$

for all $t \in [0,T]$ and $i = 1,\ldots,\tau_A$. This process corresponds to a change of measure from $\mathbb{P}$ to $\mathbb{W}_i$ (with $\mathbb{W}_i \sim \mathbb{P}$), such that $W^{\mathbb{W}_i} = W_t - \int_0^t \Sigma_Z^\top \left(e^{K_Z(T-s)}\right)^\top B^R(i)\,\mathrm{d}s$ is for all $t \in [0,T]$ and $i = 1,\ldots,\tau_A$ a $\mathbb{W}_i$-standard Brownian motion. Using this change of measure, we are able to deduce that $\mathbb{E}[a_T Z_T]$ is specified as follows:

$$\begin{aligned}\mathbb{E}[a_T Z_T] &= C\sum_{i=1}^{\tau}\mathbb{E}\left[\frac{P_{T,T+i}}{\Pi_T}\right]\underbrace{\mathbb{E}\left[\frac{P_{T,T+i}/\Pi_T}{\mathbb{E}[P_{T,T+i}/\Pi_T]}Z_T\right]}_{=\mathbb{E}^{\mathbb{W}_i}[Z_T]}\\ &= C\sum_{i=1}^{\tau}\mathbb{E}\left[\frac{P_{T,T+i}}{\Pi_T}\right]\left(\int_0^T e^{-K_Z(T-s)}\left[e^{K_Z(T-s)}\right]^\top\mathrm{d}s\right)B^R(i)\,.\end{aligned} \tag{D.1.8}$$

Finally, (4.4.3) follows from solving the following system:

$$\begin{aligned}\bar{\alpha} + \frac{1}{2}\int_0^T \bar{\beta}^\top e^{-K_Z(T-s)}\left[e^{-K_Z(T-s)}\right]^\top\bar{\beta}\mathrm{d}s &= \log\frac{\mathbb{E}[a_T]}{C},\\ \left(\int_0^T e^{-K_Z(T-s)}\left[e^{K_Z(T-s)}\right]^\top\mathrm{d}s\right)\bar{\beta} &= \frac{\mathbb{E}[a_T Z_T]}{\mathbb{E}[a_T]}.\end{aligned} \tag{D.1.9}$$

D.2 Proof of Proposition 4.4.1

Define $\tilde{f}(T,H_T) = \frac{X_T^{\mathrm{opt}}}{\Pi_T}\frac{1}{Y_T} = (1-\mathcal{H}^{-1}(X_0)^{\frac{1}{p-1}}\frac{M_T^{R\,\frac{1}{p-1}}}{Y_T})\mathbb{1}_{\mathcal{A}_T}$, where $H_T = \log(M_T^{R^{-\frac{1}{p-1}}}Y_T)$. Then, by the definition of a CDF, we know that $F_{X/Y}(x) = \mathbb{E}[\mathbb{1}_{\{\tilde{f}(T,H_T)\le x\}} \mid H_0 = h]$ holds for all $x \in \mathbb{R}$. Resorting to an application of the Fourier transform, we derive for all $x \in [0,1)$:

$$\begin{aligned}F_{X/Y}(x) &= \mathbb{E}\left[\mathbb{1}_{\{\tilde{f}(T,H_T)\le x\}}\,\middle|\, H_0 = h\right]\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{f}_\kappa^*(T,x,\omega)\,\phi_T(-\omega - i\kappa, h)\,\mathrm{d}\omega.\end{aligned} \tag{D.2.10}$$

Here, $\tilde{f}^*_\kappa(T,\omega)$ is for some $\kappa < 0$, all $\omega \in \mathbb{R}$ and all $x \in [0,1)$ given by:

$$\begin{aligned}\tilde{f}^*_\kappa(T,x,\omega) &= \int_{-\infty}^{\infty} e^{(i\omega-\kappa)H}\mathbb{1}_{\left\{\tilde{f}(T,H)\le x\right\}}\mathrm{d}H \\ &= \int_{-\infty}^{\frac{\log\mathcal{H}^{-1}(X_0)}{p-1}} e^{(i\omega-\kappa)H}\mathrm{d}H + \int_{\frac{\log\mathcal{H}^{-1}(X_0)}{p-1}}^{\mathcal{H}_{x,p}} e^{(i\omega-\kappa)H}\mathrm{d}H \\ &= \frac{1}{i\omega-\kappa}e^{(i\omega-\kappa)\left[\frac{1}{p-1}\log\mathcal{H}^{-1}(X_0)-\log(1-x)\right]},\end{aligned} \tag{D.2.11}$$

where $\mathcal{H}_{x,p} = \frac{\log\mathcal{H}^{-1}(X_0)}{p-1} - \log(1-x)$. Clearly, $\tilde{f}^*_\kappa(T,x,\omega)$ is the Fourier transform of $e^{-\kappa H_T}\mathbb{1}_{\left\{\tilde{f}(T,H)\le x\right\}}$ with respect to H_T. Moreover, we have that ϕ_T is for all $\omega \in \mathbb{R}$ specified as follows: $\phi_T(\omega,h) = \int_{-\infty}^{\infty} e^{i\omega H}\phi(H,h)\,\mathrm{d}H = \mathbb{E}[M_T^{R^{-\frac{i\omega}{p-1}}} Y_T^{i\omega}]$. Note the similarity between ϕ_T and equation (B.1.31).

As in (B.1.29), we then introduce the following system of ODE's:

$$\begin{aligned}\bar{A}'(t) &= -i\omega\bar{B}(t)^\top\Sigma_Z\bar{\Lambda}^R_{0,t} - \frac{1}{2}\bar{B}(t)^\top\bar{B}(t) + \bar{a}_t(\omega), \\ \bar{B}'(t) &= \left(-\frac{i\omega}{p-1}\Lambda_1^\top\Sigma_Z^\top + K_Z^\top - 2\bar{C}(t)^\top\right)\bar{B}(t) + \bar{b}_t(\omega), \\ \bar{C}'(t) &= 2\left(K_Z^\top - \frac{i\omega}{p-1}\Lambda_1^\top\Sigma_Z^\top\right)\bar{C}(t) - 2\bar{C}(t)^\top\bar{C}(t) - c_t(\omega).\end{aligned} \tag{D.2.12}$$

Here, $\bar{a}_t(\omega) = -a_t(\omega) - \mathrm{tr}\left(\bar{C}(t)\right)$ holds, with:

$$\begin{aligned}a_t(\omega) = i\omega\alpha_t &+ \frac{1}{2}i\omega\left[i\omega-1\right]\beta_t^\top\beta_t + \frac{i\omega}{p-1}\widehat{\delta}_{0,r} \\ &- \frac{1}{2}\frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1}-1\right]\Lambda_0^{R^\top}\Lambda_0^R + \frac{(i\omega)^2}{p-1}\beta_t^\top\Lambda_0^R.\end{aligned} \tag{D.2.13}$$

In addition to this, $\bar{b}_t(\omega) = -b_t(\omega) - 2i\omega\bar{C}(t)^\top\Sigma_Z\bar{\Lambda}^R_{0,t}$ and $\bar{\Lambda}^R_{0,t} = \frac{1}{p-1}\left[\Lambda_0 - \sigma_\Pi\right] + \beta_t$ hold for all $t \in [0,T]$ and $\omega \in \mathbb{R}$. In these expressions, we include the following specifications for $b_t(\omega)$ and $c_t(\omega)$:

$$\begin{aligned}b_t(\omega) &= \frac{-i\omega}{p-1}\left(-\widehat{\delta}_{1,r} + \left[\frac{-i\omega}{p-1}-1\right]\Lambda_1^\top\Lambda_0^R - i\omega\Lambda_1^\top\beta_t\right), \\ c_t(\omega) &= -\frac{1}{2}\frac{i\omega}{p-1}\left[-\frac{i\omega}{p-1}-1\right]\Lambda_1^\top\Lambda_1.\end{aligned} \tag{D.2.14}$$

The system of ODE's in (D.2.12) is given for the deterministic functions, $\bar{A} : [0,T] \times \mathbb{R} \to \mathbb{R}$, $\bar{B} : [0,T] \times \mathbb{R} \to \mathbb{R}^2$, and $\bar{C} : [0,T] \times \mathbb{R} \to \mathbb{R}^{2\times 2}$. Note that we suppress the dependencies of these functions on $\omega \in \mathbb{R}$ in (D.2.12). Furthermore, we emphasise that $\bar{A}(T) = 0$, $\bar{B}(T) = 0_2$ and $\bar{C}(T) = 0_{2\times 2}$ hold. Observe the similarity between this system and (B.3.49). Hence:

$$\begin{aligned} \phi_T(\omega, h) &= \int_{-\infty}^{\infty} e^{i\omega H} \phi(H, h)\, \mathrm{d}H = e^{\bar{A}(0,\omega)} Y_0^{i\omega}, \quad \text{where} \\ \bar{A}(t) &= -\int_t^T \left[-i\omega \bar{B}(t)^\top \Sigma_Z \bar{\Lambda}_{0,t}^R - \frac{1}{2} \bar{B}(t)^\top \bar{B}(t) + \bar{a}_t(\omega) \right] \mathrm{d}s. \end{aligned} \tag{D.2.15}$$

We derive $f_{X/Y}(x)$ for all $x \in (0,1)$ as follows:

$$\begin{aligned} f_{X/Y}(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial}{\partial x} \tilde{f}_\kappa^*(T, x, \omega)\, \phi_T(-\omega - i\kappa, h)\, \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{i\omega - \kappa}{1 - x} \tilde{f}_\kappa^*(T, x, \omega)\, \phi_T(-\omega - i\kappa, h)\, \mathrm{d}\omega. \end{aligned} \tag{D.2.16}$$

D.3 Proof of Corollary 4.4.2

Define $h \mapsto f_h(h)$ as the PDF of $H_T = \log M_T^{R^{-\frac{1}{p-1}}} Y_T$. Then, by the law of total probability for continuous functions, we have:

$$\begin{aligned} \mathbb{P}\left(\frac{X_T}{\Pi_T} \frac{1}{Y_T} \leq x \right) = &\int_{-\infty}^{\frac{\log \eta^{\text{opt}}}{p-1}} \mathbb{P}(0 \leq x)\, f_H(h)\, \mathrm{d}h \\ &+ \int_{\frac{\log \eta^{\text{opt}}}{p-1}}^{\infty} \mathbb{P}\left(1 - {\eta^{\text{opt}}}^{\frac{1}{p-1}} e^{-h} \leq x \right) f_H(h)\, \mathrm{d}h. \end{aligned} \tag{D.3.17}$$

Observe that this PDF is known for $\Lambda_1 = 0_{4\times 2}$. Moreover, for the sake of notational elegance, we make use of the fact that $\eta^{\text{opt}} = \mathcal{H}^{-1}(X_0)$. Suppose that $h \mapsto F_H(h)$ denotes the CDF corresponding to the latter random variable. Then, $F_{X/Y}(x)$ reads for all $x \in [0,1)$ as follows:

$$F_{X/Y}(x) = 1 - \mathbb{P}\left(e^{H_T} \leq \frac{1 - x}{{\eta^{\text{opt}}}^{\frac{1}{p-1}}} \right). \tag{D.3.18}$$

Using the fact that $M_T^{R^{-\frac{1}{p-1}}} Y_T$ is log-normally distributed (cf. Appendix B.2), we can immediately evaluate the latter probability as follows for all $x \in [0,1)$:

$$F_{X/Y}(x) = \Phi\left(\frac{\log \frac{\eta^{\mathrm{opt}^{\frac{1}{p-1}}}}{1-x} - \mathbb{E}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}}\right). \tag{D.3.19}$$

In this identity, we have that the following is true:

$$\begin{aligned}
\mathbb{E}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right] &= \log Y_0 + \int_0^T \left(\widehat{\nu}_s - \frac{1}{2}\left\|\beta_s + \frac{1}{p-1}\Lambda_0^R\right\|^2_{\mathbb{R}^4}\right) \mathrm{d}s, \\
\mathbb{V}\mathrm{ar}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right] &= \int_0^T \left\|\beta_s^\top + \frac{\Lambda_0^{R^\top}}{p-1} + \frac{\widehat{\delta}_{1,r}^\top}{p-1}\widehat{K}_{Z,t,T}\right\|^2_{\mathbb{R}^4} \mathrm{d}s.
\end{aligned} \tag{D.3.20}$$

Here, for notational purposes, we define $\widehat{K}_{Z,t,T} = K_Z^{-1}\left(I_{2\times 2} - e^{-K_Z[T-t]}\right)\Sigma_Z$ and $\widehat{\nu}_t = \alpha_t + \frac{\widehat{\delta}_{0,r}}{p-1} + \frac{1}{2}\frac{p}{(p-1)^2}\Lambda_0^{R^\top}\Lambda_0^R + \frac{1}{p-1}\beta_t^\top\Lambda_0^R$ for all $t \in [0,T]$.

Concerning $x \mapsto f_{X/Y}(x)$, we are able to derive the following:

$$\begin{aligned}
f_{X/Y}(x) &= \frac{\partial}{\partial x} F_{X/Y}(x) \\
&= \frac{\partial}{\partial x}\Phi\left(d_{0,T}(x)\right) = \phi\left(d_{0,T}(x)\right)\frac{\partial}{\partial x} d_{0,T}(x) \\
&= \frac{\phi\left(d_{0,T}(x)\right)}{(1-x)\sqrt{\mathbb{V}\mathrm{ar}\left[\log M_T^{R^{-\frac{1}{p-1}}} Y_T\right]}}, \quad \forall\, x \in (0,1).
\end{aligned} \tag{D.3.21}$$

5
Conclusion

In spite of the overarching theme, it is not easy to draw general conclusions from the combined chapters. All three core chapters touch upon duality techniques, yet in slightly different ways. We elaborately addressed these differences in Chapter 1. In fact, we argued that each chapter relates to a different part of Figure 1.1. As this figure visualises the technical mechanism underscoring duality, it is clear that the common grounds are far from trivially established. Even within the confines of portfolio theory, duality encloses a wide array of research domains. The variety of topics central to this dissertation exemplifies the latter. Therefore, apart from some unspecific general comments on the value of duality, we are not able to outline well-defined conclusive statements that apply to all chapters. For comments of the former kind, one can consult the introductory sections. Nevertheless, along the following three axes, we can examine the chapters in a considerably more general manner: (i) portfolio theory, (ii) *specific* duality domains, and (iii) the related outlook on future research. For this reason, we divide our conclusion into three parts. Each part concerns a particular duality-based and/or portfolio-linked avenue specific to one of the core chapters. Per chapter, this distinction enables us to investigate the implications of our findings for separate portfolio and/or duality areas. These areas are by construction defined in a broad sense. In addition to this, we are able to more precisely chalk out the corresponding outlook on future academic research. On the grounds of these partitioned conclusions, one can consequently obtain a more distinct perspective of this dissertation's contribution to the literature on duality and portfolio optimisation.

In accordance with the number of core chapters, we identify three areas of portfolio theory and interrelated duality techniques. For Chapter 2, the area of interest is covered by studies on dual-control approximate methods. These methods rely on duality machinery to develop and possibly improve approximations to optimal policy rules. In the context of finance and portfolio optimisation, these rules straightforwardly concern investment and/or savings decisions. For Chapter 3, the duality-linked domain abides by a more theoretical nature. Although the area of application pertains to habit formation, the topic focuses on the mathematical heart of duality: the dual formulation. In view of the formulation's intimate link to the former topic, we identify the area of interest as multiplicative habit formation. Despite this predominantly financial identification, we subsequently emphasise its relation to duality. For Chapter 4, duality plays a supplementary role. As addressed in the introduction, this chapter has a clear practical purpose. In particular, we resort to a sole application of martingale duality techniques to acquire optimal decision variables. These decision variables are relevant for investors with preferences that revolve around a reference level. By virtue of this applied relevance, the central theme can be defined as a treatment of reference-dependent preferences. Thereby, we aim to focus on the practical dimension of our results. Moreover, due to the peculiar character of reference-dependent preferences, we are able to specify interesting domains for duality-oriented research. The following sections respect the previously employed order. That is, we address the generalised themes chapter-by-chapter. Each section comprises of (i) a re-examination of the chapter-specific conclusions, (ii) a discussion of its contribution to the overarching theme, and (iii) a corresponding outlook on future research.

5.1 Dual-Control Methods

In Chapter 2, we developed an approximate dual-control method for constrained investment-consumption problems. The method relies on a three-fold procedure predicated on the duality relations unique to the artificial market. On the basis of this procedure, the routine manages to generate closed-form approximations to the optimal control variables. In order to measure the accuracy of these approximations, the method simply examines the magnitude of the duality gap. The approximating scheme has a wide range of application. It can be applied to multi-dimensional markets with general return dynamics. The corresponding trading constraints must attain values in a convex conic subset of $\mathbb{R}^N$. Moreover, the agent's preference qualification is allowed to be state-dependent and may include a benchmark process or reference level. Due to the additional incorporation of a labour income stream, the approximating framework can indeed be applied to an abundance of relevant problems. In the chapter at hand, we illustrated the method's potential accuracy in a three-dimensional market. More concretely, we relied on the environment proposed by Cocco et al. (2005). To model the preferences, we made use of power utility and the less well-known SAHARA function. Both functions included a strictly positive benchmark process. For a set of three different trading constraints, the method demonstrated to work notably well. The annual welfare losses were all smaller than 0.051% of the agent's initial endowment.

This dual-control method constitutes a generalised variant of the one developed by Bick et al. (2013). Their framework hinges on a similar approximating principle. However, whereas our method is able to cover a large number of constrained problems, theirs is more limited.[1] Due to the significant interest in models/situations of a more complicated type, it is clear that our approximate method forms an important addition to the literature on dual-control machinery. Dual-control methods are a relatively new phenomenon in the domain of finance and portfolio optimisation. So far, some studies have been concerned with this topic, but the field is far from saturated. The scarcity of studies applies to both theoretical contributions and applied papers. We attribute this scarcity to the abstract nature of duality per se and the fairly involved scheme underpinning dual-control methods. For these reasons, most studies prefer the conventional backward-induction techniques. Though understandable, in view of a.o. this dissertation, dual-control methods come with indisputable upsides. Compared to the conventional routines, Hambel et al. (2021) even show that dual-control methods can be more accurate. In addition to this, these duality-based schemes manage to render analytical insights and are consequently easy to implement. While backward-induction techniques are known to suffer from the curse of dimensionality, the aforementioned schemes are able to cope with large dimensions. In fact, for the examples considered in this dissertation, hardly any computational effort was required. With Chapter 2's results at hand, we therefore conclude that dual-control methods can be of utmost importance to the domain of portfolio optimisation.

On account of the latter, dual-control methods establish a promising branch for future research. In this regard, we are able to distinguish two interesting

[1] The configuration of their study assumes a uni-dimensional market model with standard CRRA preferences. The constraint set is held fixed at an example-specific version. Furthermore, the postulates for their return dynamics involve constant drift and diffusion terms. Despite these limitations, we note that their study pioneered the fundamental principle corresponding to the approximating scheme. That is, it introduced the idea of resorting to the artificial market to assemble analytical candidate solutions. Likewise, it introduced the duality gap as a financially relevant performance indicator.

areas: (i) applications of existing dual-control methods, and (ii) theoretical extensions of the corresponding approximating principle. In order to clarify the first item, let us restrict ourselves to the domain of Chapter 2. Our dual-control method enables one to near-optimally solve constrained investment-consumption problems in closed-form. Traditionally, studies on portfolio choice were required to fall back on purely numerical routines to obtain the optimal control variables. For this reason, there exists a great multitude of constrained problems for which there is very little known about the analytical structure of the related optimal solutions. With the help of our method, we are able to disclose this structure and thereby gain valuable economic/financial insights into the solutions' underlying dynamics. Problems that may in this respect be appealing to prospective studies are: Heston's stochastic volatility model, utility-maximisation for non-standard preferences in the presence of unhedgeable risk-drivers, and life-cycle research under pre-fixed borrowing and short-selling constraints. As for the second item, we note that the absence of closed-form solutions is not unique to "plain" constrained investment-consumption problems. Frameworks involving e.g. model and/or parameter uncertainty, ambiguity aversion, S-shaped utility functions, proportional transaction costs, or (multiplicative) habit formation may all suffer from the same analytical issues. For the majority of these setups, in the spirit of Figure 1.1, it is possible to outline a clearly defined duality mechanism. It could therefore be interesting to develop dual-control machinery suitable for these configurations. This may significantly improve our analytical understanding of these problems.

5.2 Multiplicative Habit Formation

In Chapter 3, we derived a dual formulation corresponding to the optimal consumption problem with multiplicative habit formation. The multiplicative habit component generates a value function that incorporates irremovable path-dependency. In addition to this, due to the same component, the objective is not fully concave. As a result of both attributes, applications of common Lagrangian techniques fail to identify a dual problem. To account for the path-dependent objective, we were required to resort to the less well-known notion of Fenchel duality. This duality theorem can be regarded as a generalised version of the ordinary Legendre transform. Unlike Legendre, Fenchel is able to cope with bounded linear maps as partial arguments of the optimisation target. By means of this generalised notion, we were able to deduce a dual formulation and prove that it satisfies strong duality. Naturally, the dual problem gave rise to analytical specifications of the duality relations. These relations were characterised by two identities that uncover the links between optimal consumption, the habit level, and the duality process. From these identities, we were able to concretise the true dynamics of the optimal control variables. To exemplify the practical dimension of this theoretical result, we developed a duality-based evaluation mechanism. It closely resembles the scheme unique to standard dual-control methods. After some minor modifications, the mechanism can indeed be employed as a dual-control routine. For the near-optimal analytical approximation proposed by van Bilsen et al. (2020a), we examined the evaluation mechanism. In agreement with the approximation's precision, the ensuing welfare losses were negligibly small.

The literature on habit formation is dominated by studies on additive models. In other words, for multiplicative setups there are not that many papers available. This disproportionate representation can be attributed to two main reasons. First, additive habit formation is mathematically easy to handle. Due to the isomorphism advanced by Schroder and Skiadas (2002), a large class of additive-linked problems can be solved in closed-form. Multiplicative habit models are considerably more difficult to solve.[2] Second, despite their nearly synchronous introduction to the literature, compared to additive setups, multiplicative models are a fairly recent phenomenon. In the '90s, the number of studies on additive habits rapidly increased after the contribution by Detemple and Zapatero (1991). For the multiplicative class, it took more than a decade after the high-impact paper by Abel (1990) for studies to pick up the fundamental idea. Note that the latter reason is closely linked to the

[2]Contrary to the additive configurations, multiplicative habit models suffer from several technical issues. These issues can all be related to the path-dependency and non-concave objective function addressed in the main text. Standard solution techniques are consequently not applicable to this type of habit formation. Even though certain numerical routines are able to cope with these atypical features, the complicating attributes make multiplicative models less appealing/accessible. More concretely, there does not exists a multiplicative-adapted user-friendly isomorphism. Note that we have demonstrated how this isomorphism ensues from the dual formulation in Chapter 1.

former one. In Chapter 1, we have addressed the salient economic/financial upsides of multiplicative habit formation. It is therefore unfortunate that the technical intricacies associated with multiplicative setups withhold researchers from studying these models. Nevertheless, to make the multiplicative domain more comprehensible and accessible, the dual formulation comes in handy. Its specification immediately opens doors to applications in the spirit of shadow prices or martingale duality machinery. On a more secondary/indirect level, the formulation can be employed to develop dual-control methods. In Chapter 2, we have lifted a corner of the veil corresponding to this dual-induced benefit. Hence, by means of the novel dual formulation, we have made a big leap forward on the subject of multiplicative-linked research potential.

As before, we are able to distinguish the preceding potential into two explicit branches for future research: (i) applications of our strong duality result, and (ii) theoretical extensions thereof to more involved frameworks. The former item pertains to the economic/financial upsides of multiplicative habits addressed in Chapter 1. In light of the many corresponding implications for applied research, we restrict ourselves to an overview of three possible ramifications. The first one ties in with Chapter 2, and concerns the development of dual-control methods. These methods can easily be predicated on the foundation of our evaluation mechanism.[3] The second one touches upon the design of a martingale duality framework adapted to multiplicative habit formation. For the additive setups, this framework is given by Schroder and Skiadas (2002)'s isomorphism. In this regard, the mere challenge for multiplicative configurations is the retrieval of a Lagrangian-like functional from the dual formulation. The third and last one is outlined by a more thorough economic inspection of the duality relations and the related FOCs. On the basis of the corresponding identities, it should be possible to derive in more concrete terms what the optimal decision variables analytically entail. In fact, it may even be possible to characterise optimal consumption in closed-form or as the solution to an FBSDE. With

[3]More concretely, at the end of our multiplicative treatment, we have shed light on the evaluation mechanism necessary for these methods. The remaining ingredient for a well-outlined dual-control routine is therefore a proper approximating principle. For this reason, we are required to examine the duality relations and accordingly construct approximate control variables. Some ideas for the creation of these approximations are: first-order Taylor expansions around a primal-linked point, a similar expansion applied to the optimal dual-control, or a log-normal postulate for optimal consumption.

regard to the second item, we note that our strong duality result rests on the assumption of market completeness. In a similar sense, we have identified the preference qualification as the "simple" CRRA utility function. To broaden its scope of application, one can enlarge this framework such that it accounts for e.g. unhedgeable risk, more exotic preferences, and/or proportional transaction costs. Extensions of this form open doors to novel dual-control methods. The duality relations for these extended setups may additionally endow us with relevant insights into the analytical structure of the optimal controls.

5.3 Reference-Dependent Preferences

In Chapter 4, we analysed an optimal investment problem over terminal wealth alone. We considered the matter against the background of a pension fund that attempts to meet its pre-defined liabilities. For this purpose, we introduced a utility-maximising agent with reference-dependent preferences of the LPM family. LPM operators are ordinarily employed as hedging criteria. However, in the confines of preference qualifications, they can be utilised as functions that accommodate a strong orientation towards some target. This target is typically referred to as the reference level and can be identified as the pension liabilities. Due to the person-specific and uncertain nature of the corresponding objective, the ensuing optimisation problem bears immediate relevance to DC pension schemes. On account of its frequent use in practice, we embedded the LPM problem in the financial environment proposed by Koijen et al. (2009). This model postulates an affine structure for the market prices of risk. As a consequence, it is not possible to derive the specific distributional features of the SPD process. To cope with these non-standard attributes, we resorted to an application of the Fourier transform. Based on this application, we were able to solve the LPM problem in closed-form. In our numerical study, we characterised the reference level as a life annuity. Our findings suggested that the LPM operator is able to notably improve the likelihood of achieving one's pension goals. Despite this outstanding outcome, the numerical results also demonstrated that the framework is highly dependent on the estimates for the market prices of risk. In addition to this, we showed that the numerically assessed portfolio strategies may be difficult to implement in reality.

The LPM operator marks a unique function in the domain of reference-dependent preferences. In the literature on reference levels, preference qualifications customarily incorporate convexity with respect to incurred losses. In more precise terms, studies in this field usually rely on S-shaped utility functions. While this convexity requirement can be aligned with the fundamental principles of prospect theory, a.o. Jarrow and Zhao (2006) argue that it may be relaxed. The most important feature according to them is established by the utility-related difference in evaluating equivalent gains and losses. For the LPM operator, this is true in a fairly elegant yet atypical manner. By assuming zero marginal utility for values of wealth above the reference level, the LPM criterion significantly overestimates losses. It particularly means that the agent is "satsified" when the target is achieved; the moment that wealth falls below the reference level, the same agent becomes extremely concerned. We stress that the LPM operator has only recently been identified as a reference-dependent preference qualification. Due to this relatively novel interpretation, our study forms an interesting addition to the applied literature on portfolio-linked prospect theory. By virtue of the former property, we also note that the operator is frequently used to model partial hedging problems. The literature on partial hedging is dominated by theoretical treatments. In consideration of our strong practical focus, the results of Chapter 4 constitute a valuable addition to this literature. Likewise, departing from this practical emphasis, our study generated useful insights for the pension industry. The positive impact on a fund's recovery potential is particularly relevant.

In Chapter 4, we were predominantly concerned with the practical implications of the LPM-optimal solutions. As a consequence, duality only served a complementary purpose. For generalised frameworks, duality can play a more prominent role on the subject of prospect theory. However, the convexity features of the corresponding preference qualifications complicate immediate applications of duality methods. For this reason, a promising branch for future theoretical research concerns concavification and interlinked duality techniques. The latter would boil down to the development of a duality framework suitable for reference-dependent preferences of a more general type. Closely related to this class of problems is the idea of probability weighting. Prospect theory empirically demonstrates that individuals tend to overestimate small probability events. To incorporate this phenomenon into a utility-maximising setup, one simply includes a weighting function in the target of optimisation. The resulting objective is typically non-concave and complicated to handle. It is therefore interesting to develop a duality configuration appropriate to probability weighting setups. Consistent with the previous two themes, the construction of these duality frameworks naturally gives rise to novel dual-control methods. On a more applied note, we are able to identify the following three branches for future research. Within the borders of Chapter 4, it might be worth it to study different target-oriented preference qualifications. One could, for example, measure the differences in performance for dual-CRRA, SAHARA, and/or HARA-based kinked utility functions. Similarly, it may render useful insights to employ other model specifications. The environment proposed by Koijen et al. (2009) does not account for e.g. stochastic volatility or generally defined reference levels. Last, as uttered at the end of Chapter 4, the LPM-based portfolio strategies are hard to implement and highly dependent on the estimates for the market prices of risk. For these two reasons, it makes perfect sense to allow for trading constraints and/or robustness procedures.

Bibliography

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. *The American Economic Review*, *80*(2), 38–42.

Alvarez-Cuadrado, F., Monteiro, G., & Turnovsky, S. J. (2004). Habit formation, catching up with the Joneses, and economic growth. *Journal of economic growth*, *9*(1), 47–80.

Balter, A. (2016). Model uncertainty: The effect on robustness, estimation and stochastic optimisation. *Dissertation*.

Balter, A., Beijering, L., Janssen, P., de Jong, F., Joseph, A., Kamma, T., & Pelsser, A. (2020). Investing for retirement with an explicit benchmark. *Netspar Design Paper*.

Bardhan, I. (1994). Consumption and investment under constraints. *Journal of Economic Dynamics and Control*, *18*(5), 909–929.

Basak, S. (1995). A general equilibrium model of portfolio insurance. *The Review of Financial Studies*, *8*(4), 1059–1090.

Basak, S., & Chabakauri, G. (2012). Dynamic hedging in incomplete markets: A simple solution. *Review of financial studies*, *25*(6), 1845–1896.

Basak, S., Pavlova, A., & Shapiro, A. (2007). Optimal asset allocation and risk shifting in money management. *The Review of Financial Studies*, *20*(5), 1583–1621.

Basak, S., & Shapiro, A. (2001). Value-at-risk-based risk management: Optimal policies and asset prices. *The review of financial studies*, *14*(2), 371–405.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., & Sugden, R. (1997). A test of the theory of reference-dependent preferences. *The Quarterly Journal of Economics*, *112*(2), 479–505.

Battauz, A., De Donno, M., & Ortu, F. (2011). Intertemporal asset pricing and the marginal utility of wealth. *Journal of Mathematical Economics*, *47*(2), 227–244.

Battauz, A., De Donno, M., & Ortu, F. (2015a). Envelope theorems in Banach lattices and asset pricing. *Mathematics and Financial Economics*, *9*(4), 303–323.

Battauz, A., De Donno, M., & Sbuelz, A. (2015b). Kim and Omberg revisited: The duality approach. *Journal of Probability and Statistics*, *2015*.

Berkelaar, A. B., Kouwenberg, R., & Post, T. (2004). Optimal portfolio choice under loss aversion. *Review of Economics and Statistics*, *86*(4), 973–987.

Bernard, C., & Ghossoub, M. (2010). Static portfolio choice under cumulative prospect theory. *Mathematics and Financial Economics*, *2*(4), 277–306.

Bichuch, M., & Guasoni, P. (2018). Investing with liquid and illiquid assets. *Mathematical Finance*, *28*(1), 119–152.

Bick, B., Kraft, H., & Munk, C. (2013). Solving constrained consumption–investment problems by simulation of artificial market strategies. *Management Science*, *59*(2), 485–503.

Bodie, Z., Detemple, J. B., Otruba, S., & Walter, S. (2004). Optimal consumption–portfolio choices and retirement planning. *Journal of Economic Dynamics and Control*, *28*(6), 1115–1148.

Borwein, J. M., & Zhu, Q. J. (2004). Techniques of variational analysis.

Bouchard, B., Touzi, N., & Zeghal, A. (2004). Dual formulation of the utility maximization problem: The case of nonsmooth utility. *The Annals of Applied Probability*, *14*(2), 678–717.

Brandt, M. W., Goyal, A., Santa-Clara, P., & Stroud, J. R. (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *The Review of Financial Studies*, *18*(3), 831–873.

Brennan, M. J., & Xia, Y. (2002). Dynamic asset allocation under inflation. *The Journal of Finance*, *57*(3), 1201–1238.

Brown, D. B., & Smith, J. E. (2011). Dynamic portfolio optimization with transaction costs: Heuristics and dual bounds. *Management Science*, *57*(10), 1752–1770.

Brown, D. B., Smith, J. E., & Sun, P. (2010). Information relaxations and duality in stochastic dynamic programs. *Operations Research*, *58*(4-part-1), 785–801.

Browne, S. (1999). Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. *Finance and Stochastics*, *3*(3), 275–294.

Browne, S. (2013). Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark. In *Handbook of the fundamentals of financial decision making: Part ii* (pp. 711–730). World Scientific.

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, *107*(2), 205–251.

Campi, L., & Schachermayer, W. (2006). A super-replication theorem in kabanov's model of transaction costs. *Finance and Stochastics*, *10*(4), 579–596.

Carr, P., & Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of computational finance*, *2*(4), 61–73.

Carroll, C. D. (2000). Solving consumption models with multiplicative habits. *Economics Letters*, *68*(1), 67–77.

Carroll, C. D., Overland, J., & Weil, D. N. (1997). Comparison utility in a growth model. *Journal of Economic Growth*, *2*(4), 339–367.

Carroll, C. D., Overland, J., & Weil, D. N. (2000). Saving and growth with habit formation. *American Economic Review*, *90*(3), 341–355.

Chan, Y. L., & Kogan, L. (2002). Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices. *Journal of Political Economy*, *110*(6), 1255–1285.

Chen, A., Hieber, P., & Nguyen, T. (2019). Constrained non-concave utility maximization: An application to life insurance contracts with guarantees. *European Journal of Operational Research*, *273*(3), 1119–1135.

Chen, A., Pelsser, A., & Vellekoop, M. (2011). Modeling non-monotone risk aversion using SAHARA utility functions. *Journal of Economic Theory*, *146*(5), 2075–2092.

Chen, A., & Vellekoop, M. (2017). Optimal investment and consumption when allowing terminal debt. *European Journal of Operational Research*, *258*(1), 385–397.

Chen, D. H., Beetsma, R. M., & van Wijnbergen, S. (2020). Unhedgeable inflation risk within pension schemes. *Insurance: Mathematics and Economics*, *90*, 7–24.

Choi, J. H., Sirbu, M., & Zitkovic, G. (2013). Shadow prices and well-posedness in the problem of optimal investment and consumption with transaction costs. *SIAM Journal on Control and Optimization*, *51*(6), 4414–4449.

Choi, J., & Jonsson, M. (2009). Partial hedging in financial markets with a large agent. *Applied Mathematical Finance*, *16*(4), 331–346.

Cocco, J. F., Gomes, F. J., & Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, *18*(2), 491–533.

Constantinides, G. M. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy*, *98*(3), 519–543.

Corrado, L., & Holly, S. (2011). Multiplicative habit formation and consumption: A note. *Economics Letters*, *113*(2), 116–119.

Cox, J. C., & Huang, C.-F. (1989). Optimal consumption and portfolio policies when asset prices follow a diffusion process. *Journal of economic theory*, *49*(1), 33–83.

Cox, J. C., & Huang, C.-F. (1991). A variational problem arising in financial economics. *Journal of Mathematical Economics*, *20*(5), 465–487.

Cuoco, D. (1997). Optimal consumption and equilibrium prices with portfolio constraints and stochastic income. *Journal of Economic Theory*, *72*(1), 33–73.

Cuoco, D., & Cvitanić, J. (1998). Optimal consumption choices for a 'large' investor. *Journal of Economic Dynamics and Control*, *22*(3), 401–436.

Cvitanic, J. (2000). Minimizing expected loss of hedging in incomplete and constrained markets. *SIAM Journal on Control and Optimization*, *38*(4), 1050–1066.

Cvitanić, J. (1997). Optimal trading under constraints. In *Financial mathematics* (pp. 123–190). Springer.

Cvitanić, J., Goukasian, L., & Zapatero, F. (2003). Monte Carlo computation of optimal portfolios in complete markets. *Journal of Economic Dynamics and Control*, *27*(6), 971–986.

Cvitanić, J., & Karatzas, I. (1992). Convex duality in constrained portfolio optimization. *The Annals of Applied Probability*, 767–818.

Cvitanić, J., & Karatzas, I. (1993). Hedging contingent claims with constrained portfolios. *The Annals of Applied Probability*, 652–681.

Cvitanić, J., & Karatzas, I. (1996). Hedging and portfolio optimization under transaction costs: A martingale approach 1 2. *Mathematical finance*, *6*(2), 133–165.

Cvitanić, J., Pham, H., & Touzi, N. (1999). A closed-form solution to the problem of super-replication under transaction costs. *Finance and Stochastics*, *3*(1), 35–54.

Cvitanić, J., Schachermayer, W., & Wang, H. (2001). Utility maximization in incomplete markets with random endowment. *Finance and Stochastics*, *5*(2), 259–272.

Dai, Q., & Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, *63*(3), 415–441.

Delbaen, F., & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, *300*(1), 463–520.

de Palma, A., & Prigent, J.-L. (2008). Utilitarianism and fairness in portfolio positioning. *Journal of Banking & Finance*, *32*(8), 1648–1660.

de Palma, A., & Prigent, J.-L. (2009). Standardized versus customized portfolio: A compensating variation approach. *Annals of Operations Research*, *165*(1), 161.

Detemple, J. B. (2014). Portfolio selection: A review. *Journal of Optimization Theory and Applications*, *161*(1), 1–21.

Detemple, J. B., Garcia, R., & Rindisbacher, M. (2003). A Monte Carlo method for optimal portfolios. *The Journal of Finance*, *58*(1), 401–446.

Detemple, J. B., & Rindisbacher, M. (2010). Dynamic asset allocation: Portfolio decomposition formula and applications. *The Review of Financial Studies*, *23*(1), 25–100.

Detemple, J. B., & Zapatero, F. (1991). Asset prices in an exchange economy with habit formation. *Econometrica: Journal of the Econometric Society*, 1633–1657.

Detemple, J. B., & Karatzas, I. (2003). Non-addictive habits: Optimal consumption-portfolio policies. *Journal of Economic Theory*, *113*(2), 265–285.

Donnelly, C., Khemka, G., & Lim, W. (2022). Investing for retirement: Terminal wealth constraints or a desired wealth target? *European Financial Management*.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical finance*, *6*(4), 379–406.

Dufresne, D. (2008). Sums of lognormals. *Actuarial Research Conference*, 1–6.

Dybvig, P. H., & Liu, H. (2010). Lifetime consumption and investment: Retirement and constrained borrowing. *Journal of Economic Theory*, *145*(3), 885–907.

Edwards, W. (1955). The prediction of decisions among bets. *Journal of Experimental Psychology*, *50*(3), 201.

El Karoui, N., & Jeanblanc-Picqué, M. (1998). Optimization of consumption with labor income. *Finance and Stochastics*, *2*(4), 409–440.

El Karoui, N., Peng, S., & Quenez, M. C. (1997). Backward stochastic differential equations in finance. *Mathematical finance*, *7*(1), 1–71.

El Karoui, N., & Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM journal on Control and Optimization*, *33*(1), 29–66.

Fenton, L. (1960). The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, *8*(1), 57–67.

Föllmer, H., & Kabanov, Y. M. (1997). Optional decomposition and Lagrange multipliers. *Finance and Stochastics*, *2*(1), 69–81.

Föllmer, H., & Leukert, P. (2000). Efficient hedging: Cost versus shortfall risk. *Finance and Stochastics*, *4*(2), 117–146.

Fuhrer, J. C. (2000). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review*, *90*(3), 367–390.

Gali, J. (1994). Keeping up with the Joneses: Consumption externalities, portfolio choice, and asset prices. *Journal of Money, Credit and Banking*, *26*(1), 1–8.

Gao, J., Zhou, K., Li, D., & Cao, X. (2017). Dynamic mean-LPM and mean-CVaR portfolio optimization in continuous-time. *SIAM Journal on Control and Optimization*, *55*(3), 1377–1397.

Gomes, F., & Michaelides, A. (2003). Portfolio choice with internal habit formation: A life-cycle model with uninsurable labor income risk. *Review of Economic Dynamics*, *6*(4), 729–766.

Gómez, J.-P. (2007). The impact of keeping up with the Joneses behavior on asset prices and portfolio choice. *Finance Research Letters*, *4*(2), 95–103.

Gómez, J.-P., Priestley, R., & Zapatero, F. (2009). Implications of keeping-up-with-the-Joneses behavior for the equilibrium cross section of stock returns: International evidence. *The Journal of Finance*, *64*(6), 2703–2737.

Gómez, J.-P., & Zapatero, F. (2003). Asset pricing implications of benchmarking: A two-factor CAPM. *The European Journal of Finance*, *9*(4), 343–357.

Guasoni, P., & Wang, G. (2020). Consumption in incomplete markets. *Finance and Stochastics*, 1–40.

Haliassos, M., & Michaelides, A. (2003). Portfolio choice and liquidity constraints. *International Economic Review*, *44*(1), 143–177.

Hambel, C., Kraft, H., & Munk, C. (2021). Solving life-cycle problems with biometric risk by artificial insurance markets. *Scandinavian Actuarial Journal*, 1–21.

Han, N.-W., & Hung, M.-W. (2012). Optimal asset allocation for DC pension plans under inflation. *Insurance: Mathematics and Economics*, *51*(1), 172–181.

Harlow, W. V., & Rao, R. K. (1989). Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence. *Journal of Financial and Quantitative Analysis*, *24*(3), 285–311.

Haugh, M. B., & Kogan, L. (2007). Duality theory and approximate dynamic programming for pricing American options and portfolio optimization. *Handbooks in operations research and management science*, *15*, 925–948.

Haugh, M. B., Kogan, L., & Wang, J. (2006). Evaluating portfolio policies: a duality approach. *Operations Research*, *54*(3), 405–418.

He, H., & Pages, H. F. (1993). Labor income, borrowing constraints, and equilibrium asset prices. *Economic Theory*, *3*(4), 663–696.

He, X. D., & Zhou, X. Y. (2011). Portfolio choice under cumulative prospect theory: An analytical treatment. *Management Science*, *57*(2), 315–331.

Hershey, J. C., & Schoemaker, P. J. (1985). Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science*, *31*(10), 1213–1231.

Hou, S., Krabichler, T., & Wunsch, M. (2022). Deep partial hedging. *Journal of Risk and Financial Management*, *15*(5), 223.

Hugonnier, J., Kramkov, D., et al. (2004). Optimal investment with random endowments in incomplete markets. *The Annals of Applied Probability*, *14*(2), 845–864.

Jarrow, R., & Zhao, F. (2006). Downside loss aversion and portfolio management. *Management Science*, *52*(4), 558–566.

Jonsson, M., & Sircar, K. R. (2002). Partial hedging in a stochastic volatility environment. *Mathematical Finance*, *12*(4), 375–409.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.

Kallsen, J., & Muhle-Karbe, J. (2010). On using shadow prices in portfolio optimization with transaction costs. *The Annals of Applied Probability*, *20*(4), 1341–1358.

Kamma, T., & Pelsser, A. (2022a). Dual formulation of the optimal consumption problem with multiplicative habit formation. *Working paper*.

Kamma, T., & Pelsser, A. (2022b). Investing towards an exogenous reference level using a lower partial moments criterion. *Working Paper*.

Kamma, T., & Pelsser, A. (2022c). Near-optimal asset allocation in financial markets with trading constraints. *European Journal of Operational Research*, *297*(2), 766–781.

Karatzas, I., & Kou, S. G. (1996). On the pricing of contingent claims under constraints. *The annals of applied probability*, 321–369.

Karatzas, I., Lehoczky, J. P., & Shreve, S. E. (1987). Optimal portfolio and consumption decisions for a small investor on a finite horizon. *SIAM Journal on Control and Optimization*, *25*(6), 1557–1586.

Karatzas, I., Lehoczky, J. P., Shreve, S. E., & Xu, G.-L. (1991a). Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal on Control and Optimization*, *29*(3), 702–730.

Karatzas, I., Ocone, D. L., & Li, J. (1991b). An extension of Clark's formula. *Stochastics: An International Journal of Probability and Stochastic Processes*, *37*(3), 127–131.

Karatzas, I., & Shreve, S. E. (1998). *Methods of mathematical finance* (Vol. 39). Springer.

Karatzas, I., & Shreve, S. E. (2012). *Brownian motion and stochastic calculus* (Vol. 113). Springer Science & Business Media.

Karatzas, I., Žitković, G., et al. (2003). Optimal consumption from investment and random endowment in incomplete semimartingale markets. *The Annals of Probability*, *31*(4), 1821–1858.

Keppo, J., Meng, X., & Sullivan, M. G. (2007). A computational scheme for the optimal strategy in an incomplete market. *Journal of Economic Dynamics and Control*, *31*(11), 3591–3613.

Kim, J.-H. (2012). Shortfall risk minimization: The dual approach. *The Pure and Applied Mathematics*, *19*(2), 179–192.

Kim, T. S., & Omberg, E. (1996). Dynamic nonmyopic portfolio behavior. *The Review of Financial Studies*, *9*(1), 141–161.

Klein, I., & Rogers, L. C. (2007). Duality in optimal investment and consumption problems with market frictions. *Mathematical Finance*, *17*(2), 225–247.

Koijen, R. S., Nijman, T. E., & Werker, B. J. (2009). When can life cycle investors benefit from time-varying bond risk premia? *The Review of Financial Studies*, *23*(2), 741–780.

Koo, H. K. (1998). Consumption and portfolio selection with labor income: A continuous time approach. *Mathematical Finance*, *8*(1), 49–65.

Kozicki, S., & Tinsley, P. A. (2002). Dynamic specifications in optimizing trend-deviation macro models. *Journal of Economic Dynamics and Control*, *26*(9-10), 1585–1611.

Krabichler, T., & Wunsch, M. (2021). Hedging goals. *arXiv preprint arXiv:2105.07915*.

Kraft, H., & Munk, C. (2011). Optimal housing, consumption, and investment decisions over the life cycle. *Management Science*, *57*(6), 1025–1041.

Kramkov, D. (1996). Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets. *Probability Theory and Related Fields*, *105*(4), 459–479.

Kramkov, D., & Schachermayer, W. (1999). The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Annals of Applied Probability*, 904–950.

Lakner, P., & Nygren, M. (2006). Portfolio optimization with downside constraints. *Mathematical Finance*, *16*(2), 283–299.

Leitner, J. (2008). Optimal portfolios with lower partial moment constraints and lpm-risk-optimal martingale measures. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, *18*(2), 317–331.

Li, C., Scaillet, O., & Shen, Y. (2020). Decomposition of optimal dynamic portfolio choice with wealth-dependent utilities in incomplete markets. *Swiss Finance Institute Research Paper*, *0*(20-22).

Li, T., Wang, S., & Yang, J. (2021). Robust consumption and portfolio choices with habit formation. *Economic Modelling*, *98*, 227–246.

Lim, A. E., & Wong, B. (2010). A benchmarking approach to optimal asset allocation for insurers and pension funds. *Insurance: Mathematics and Economics*, *46*(2), 317–327.

Lim, B. H., & Choi, U. J. (2009). Optimal consumption and portfolio selection with portfolio constraints. *International Sciences*, *4*, 293–309.

Liu, J. (2006). Portfolio selection in stochastic environments. *The Review of Financial Studies*, *20*(1), 1–39.

Longstaff, F. A. (2001). Optimal portfolio choice and the valuation of illiquid securities. *The Review of Financial Studies*, *14*(2), 407–431.

Luenberger, D. G. (1997). *Optimization by vector space methods*. John Wiley & Sons.

Ma, J., Li, W., & Zheng, H. (2017). Dual control Monte-Carlo method for tight bounds of value function in regime switching utility maximization. *European Journal of Operational Research*, *262*(3), 851–862.

Ma, J., Li, W., & Zheng, H. (2020). Dual control Monte-Carlo method for tight bounds of value function under Heston stochastic volatility model. *European Journal of Operational Research*, *280*(2), 428–440.

Markowitz, H. (1952). The utility of wealth. *Journal of Political Economy*, *60*(2), 151–158.

Marzilli Ericson, K. M., & Fuster, A. (2011). Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *The Quarterly Journal of Economics*, *126*(4), 1879–1907.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, *3*(4), 373–413.

Munk, C. (2008). Portfolio and consumption choice with stochastic investment opportunities and habit formation in preferences. *Journal of Economic Dynamics and Control*, *32*(11), 3560–3589.

Munro, A., & Sugden, R. (2003). On the theory of reference-dependent preferences. *Journal of Economic Behavior & Organization*, *50*(4), 407–428.

Muraviev, R. (2011). Additive habit formation: Consumption in incomplete markets with random endowments. *Mathematics and Financial Economics*, *5*(2), 67.

Nawrocki, D., & Viole, F. (2014). Behavioral finance in financial market theory, utility theory, portfolio theory and the necessary statistics: A review. *Journal of Behavioral and Experimental Finance*, *2*, 10–17.

Nualart, D. (2006). *The malliavin calculus and related topics* (Vol. 1995). Springer.

Nygren, L. M., & Lakner, P. (2012). Partial hedging using malliavin calculus. *Journal of Mathematical Finance*, *2*, 203–213.

Ocone, D. L., & Karatzas, I. (1991). A generalized Clark representation formula, with application to optimal portfolios. *Stochastics: An International Journal of Probability and Stochastic Processes*, *34*(3-4), 187–220.

O'Donoghue, T., & Sprenger, C. (2018). Reference-dependent preferences. In *Handbook of behavioral economics: Applications and foundations 1* (pp. 1–77, Vol. 1). Elsevier.

Owen, M. P., & Žitković, G. (2009). Optimal investment with an unbounded random endowment and utility-based pricing. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, *19*(1), 129–159.

Pelsser, A. (2019). Kalman filter estimation of the KNW model. *Available at SSRN 3885556*.

Pham, H. (2000). Dynamic L^p-hedging in discrete time under cone constraints. *SIAM Journal on Control and Optimization*, *38*(3), 665–682.

Pham, H. (2002). Minimizing shortfall risk and applications to finance and insurance problems. *The Annals of Applied Probability*, *12*(1), 143–172.

Pliska, S. R. (1986). A stochastic calculus model of continuous trading: Optimal portfolios. *Mathematics of Operations Research*, *11*(2), 371–382.

Polkovnichenko, V. (2007). Life-cycle portfolio choice with additive habit formation preferences and uninsurable labor income risk. *The Review of Financial Studies*, *20*(1), 83–124.

Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.

Rogers, L. C. (2003). Duality in constrained optimal investment and consumption problems: A synthesis. In *Paris-princeton lectures on mathematical finance 2002*. Springer.

Rogers, L. C. (2013). *Optimal Investment*. Springer.

Rogers, L. C., & Williams, D. (2000). *Diffusions, Markov processes and martingales: Volume 2, Itô calculus* (Vol. 2). Cambridge university press.

Rogers, L. C., & Zaczkowski, P. (2015). Optimal investment: Bounds and heuristics. *Journal of Computational Finance*, *19*(2), 1–28.

Samuelson, P. A. (1969). Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, *51*(3), 239–246.

Sangvinatsos, A., & Wachter, J. A. (2005). Does the failure of the expectations hypothesis matter for long-term investors? *The Journal of Finance*, *60*(1), 179–230.

Schachermayer, W. (2001). Optimal investment in incomplete markets when wealth may become negative. *Annals of Applied Probability*, 694–734.

Schroder, M., & Skiadas, C. (2002). An isomorphism between asset pricing models with and without linear habit formation. *The Review of Financial Studies*, *15*(4), 1189–1221.

Schröder, P., Schober, P., & Wittum, G. (2013). Dimension-wise decompositions and their efficient parallelization. *Electronic version of an article published in Recent Developments in Computational Finance, Interdisciplinary Mathematical Sciences*, *14*, 445–472.

Sekine, J. (2004). Dynamic minimization of worst conditional expectation of shortfall. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, *14*(4), 605–618.

Tepla, L. (2000). Optimal portfolio policies with borrowing and shortsale constraints. *Journal of Economic Dynamics and Control*, *24*(11-12), 1623–1639.

Tepla, L. (2001). Optimal investment with minimum performance constraints. *Journal of Economic Dynamics and Control*, *25*(10), 1629–1645.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061.

Van Binsbergen, J. H., Brandt, M. W., & Koijen, R. S. (2008). Optimal decentralized investment management. *The Journal of Finance*, *63*(4), 1849–1895.

van Bilsen, S., Bovenberg, A. L., & Laeven, R. J. (2020a). Consumption and portfolio choice under internal multiplicative habit formation. *Journal of Financial and Quantitative Analysis*, *55*(7), 2334–2371.

van Bilsen, S., & Laeven, R. J. (2020). Dynamic consumption and portfolio choice under prospect theory. *Insurance: Mathematics and Economics*, *91*, 224–237.

van Bilsen, S., Laeven, R. J., & Nijman, T. E. (2020b). Consumption and portfolio choice under loss aversion and endogenous updating of the reference level. *Management Science*, *66*(9), 3927–3955.

Vanderbei, R. J. (2014). *Linear programming*. Springer.

Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *Journal of Financial and Quantitative Analysis*, *37*(1), 63–91.

Wagner, N. (2002). On a model of portfolio selection with benchmark. *Journal of Asset Management*, *3*(1), 55–65.

Weiss, F. (2020). A numerical approach to solve consumption-portfolio problems with predictability in income, stock prices, and house prices. *Mathematical Methods of Operations Research*, 1–49.

Xia, J. (2005). Mean–variance portfolio choice: Quadratic partial hedging. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, *15*(3), 533–538.

Xu, G.-L., & Shreve, S. E. (1992). A duality method for optimal consumption and investment under short-selling prohibition. I. general market coefficients. *The Annals of Applied Probability*, 87–112.

Yang, L., Pelsser, A., Vellekoop, M., et al. (2021). The recovery potential for underfunded pension plans. *Netspar Industry Series*.

Yener, H. (2015). Maximizing survival, growth and goal reaching under borrowing constraints. *Quantitative Finance*, *15*(12), 2053–2065.

Yogo, M. (2008). Asset prices under habit formation and reference-dependent preferences. *Journal of Business & Economic Statistics*, *26*(2), 131–143.

Yu, X. (2015). Utility maximization with addictive consumption habit formation in incomplete semimartingale markets. *The Annals of Applied Probability*, *25*(3), 1383–1419.

Zank, H. (2010). On probabilities and loss aversion. *Theory and Decision*, *68*(3), 243–261.

Summary

In this addendum, we provide a summary of this dissertation. We first address the main overarching theme. Thereafter, we elaborate in a recapitulating manner on the separate chapters central to the research output of this dissertation. Consistent with its title, the theme of this doctoral thesis concerns *Duality Methods for Stochastic Optimal Control Problems in Finance*. In particular, all three core chapters touch upon convex duality against the economic/financial background of portfolio optimisation. The literature on duality for investment-consumption problems can be classified in accordance with the following three categories: (i) applied studies, (ii) theoretical studies, and (iii) mixtures of the previous two. This dissertation covered all three categories. Concretely, in Chapter 2, we dealt with item (iii), and developed a dual-control mechanism suitable for acquiring analytical near-optimal solutions to constrained investment-consumption problems. In Chapter 3, covering item (ii), we derived a dual formulation corresponding to an optimal consumption problem involving multiplicative habit formation. In Chapter 4, which addressed item (i), we made use of duality techniques to derive optimal policy rules for a pension fund that offers a DC scheme. These three studies jointly constitute the research-based nucleus of this thesis, i.e. its “core”. As an introduction to this core, Chapter 1 expanded on the duality-linked theme in a rather general sense. The introductory chapter thereby aimed to highlight what duality theoretically entails and why it is practically useful. In addition to this, it supplied brief synopses of the academic content addressed by the preceding three chapters. Ultimately, in Chapter 5, we concluded this dissertation. Therein, we specifically focused on the contribution of the academically relevant output to the literature on economic/financial duality. As Chapters 1 and 5 serve complementary roles with regard to the remainder of this thesis, we subsequently summarise Chapters 2, 3 and 4 at greater length. For smaller variants of the following summaries, one can consult the abstracts at the beginning of each core chapter. Correspondingly, for more extensive overviews, the distinct paragraphs provided in the third section of Chapter 1 may be useful.

Chapter 2

In Chapter 2, we developed a dual-control method applicable to a broad class of constrained utility-maximisation problems. Its mechanism rests on a generalisation of the approximating routine proposed by Bick et al. (2013). In order to adequately outline the details underpinning this scheme, we are obliged to look more closely at duality in the context of portfolio optimisation. Due to the inclusion of trading restrictions, constrained investment-consumption problems are difficult to solve. The mathematical complexity associated with a derivation of optimal solutions is almost entirely attributable to the non-uniqueness of equivalent martingale measures. Namely, since the market is constrained and therefore incomplete, there exist infinitely many martingale measures. The most optimal or "least-favourable" martingale measure can be determined by an appropriate minimisation procedure of the dual formulation. However, in most cases, the ensuing first-order conditions cannot be solved in closed-form. In fact, for particular specifications of the conic constraint set, it is not even possible to derive such conditions. As the optimal dual controls give rise to optimal primal rules, the latter phenomena directly encumber a derivation of the optimal decision variables. That is, unless one is able to analytically spell out the dual-optimal martingale measure, closed-form expressions for the primal-optimal investment-consumption strategies are not available. For this reason, most constrained utility-maximisation problems lack analytical tractability and are solved by means of computationally demanding numerical machinery. The general absence of analytical solutions and the interrelated need for numerically intense approaches are two major issues in the applied domain of constrained utility-maximisation.

Our dual-control method deals with the aforementioned issues and manages to generate near-optimal closed-form solutions in a highly efficient way. To this end, it makes use of three-fold approximating scheme. First, it recognises that all analytical nuisance stems from the dual. Therefore, it analytically approximates the optimal dual controls by a modified minimisation procedure of the dual formulation. In particular, it restricts the set of dual controls to a tractable analogue and optimises the dual problem accordingly. The resulting dual approximations bring forth closed-form candidates for the optimal primal controls. To make these candidates *admissible* in the primal environment, they have to be slightly adjusted. In other words, the "raw" candidate solutions generally fail to satisfy the trading/liquidity constraints. Consequently, in the second step, our method projects the candidate solutions into the admissibility set to arrive at near-optimal controls that are primal-feasible. In the third and final step, the approximating routine measures the accuracy of these approximate solutions by a financial evaluation of the corresponding duality gap. As a result, the dual-control method concretely renders analytical approximate policy rules that are accompanied by a "hard" guarantee concerning their accuracy. In the numerical illustrations, the method proved to work well. For the examples under scrutiny, the approximating method resulted in annual welfare losses smaller than 0.051% of the agent's initial endowment. Our conclusive statement on the possible accuracy of the dual-control method is supported by the variety of examined trading constraints and the technical complexity of both the financial environment and the preference qualifications.

Chapter 3

In Chapter 3, we studied the optimal consumption problem with multiplicative habit formation. The habit-linked literature is dominated by studies on additive models. In these models, an agent is assumed to derive utility from the *difference* between consumption and the habit level. While additive configurations are easy to handle in a mathematical sense, they lack economic relevance. As most utility functions only admit strictly positive arguments, consumption is namely required to exceed the habit level at all times. For this reason, in additive frameworks, the habit component is typically interpreted as a subsistence level. Such identifications are plausible from macro-related perspectives, wherein one examines the optimal consumption patterns of nations or large-scaled populations. However, in the confines of a micro-linked setup suitable for e.g. individuals and/or households, similar interpretations cannot be upheld. More specifically, the endogeneity of the habit level complicates such economic/financial identifications of the habit level. Therefore, when studying individuals and/or households, the habit component is generally characterised as a person-specific standard of living. The artificial lower bound imposed upon consumption correspondingly implies that the utility-maximising agent is obliged to consume at least as much as his/her standard of living. Even though this corresponds to a fairly ideal situation, it is not realistic. Adverse shifts in the financial circumstances can always urge a person to scale down consumption below the level to which he/she has become accustomed. Hence, despite the mathematical elegance involved with solving additive problems, they are not economically/financially relevant for all environments.

To arrive at a setup that manages to relax the unnatural lower bound imposed upon consumption, one can make use of multiplicative habit models. In these models, an agent is assumed to derive utility from the *ratio* of consumption to the habit level. This ratio is strictly positive for any budge-feasible consumption strategy. Due to the latter property, the ratio can be incorporated into most conventional preference qualifications. As a consequence, in multiplicative frameworks, consumption is not required to exceed a peculiar lower bound. On account of the relaxation of this bound, the aforementioned frameworks gain a significant amount of economic relevance. The habit-linked configuration is now amenable to micro-related situations consistent with small households and/or individuals. In spite of the ensuing economic advantages, the multiplicative habit models come at a high technical cost. By virtue of the mathematically complicated objective function, consumption problems involving multiplicative habit formation cannot be solved in closed-form. Most studies on multiplicative models therefore resort to numerical applications or the design of approximate solutions. Ordinarily, parts of the mathematical complexity related to the optimal consumption problem are addressed or facilitated by the dual formulation. However, for this problem, there is no dual problem known. In Chapter 3, we filled this gap in the literature, and made an entire branch of dual-related applications accessible, by deriving a corresponding dual formulation. We did so by means of a "concavification" procedure and the less well-known notion of

Fenchel duality. This strong duality result gave rise to a myriad of interesting implications. In our study, we exclusively focused on the duality relations and an evaluation mechanism commonly associated with dual-control methods.

Chapter 4

In Chapter 4, we analysed an optimal terminal wealth problem from an applied point of view. More concretely, we examined a utility-maximising pension fund that operates in conformity with a DC scheme. Terminal wealth setups can easily be identified with the individual-specific nature of conventional DC plans. On the grounds of the personally oriented specification of most preference functions, a participant's attitude towards risk can be included in a very precise manner. This possibility is of significant importance to DC providers, as the participating agents are generally required to carry all retirement-linked risk. The target of optimisation consequently outlines a person-specific function adapted to the preference/risk profile of a unique individual. Another important attribute inherent in a great majority of DC setups is the notion of underfunded starting positions. Individuals typically enter DC schemes with relatively high expectations regarding their retirement wealth. These practically unrealistic outlooks on pension goals result in an initial "mismatch" between the participant's contributions and his/her expectations. Translated into financial jargon, this mismatch can be characterised as an underfunding situation. That is, the pension fund is not in possession of sufficient funds to risk-neutrally cover the pension liabilities/goals. Even if the participants' prospects are adapted to reality, pension funds are still confronted with challenges concerning these retirement goals. These challenges may arise due to a.o. detrimental changes in the economic circumstances. In the context of utility-maximisation, the underfunding positions can easily be accommodated. Given the correspondingly realistic model setup, in Chapter 4, we aimed to answer the following practically relevant question: *Is it possible to increase the likelihood of achieving one's pension goals using target-oriented preferences?*

To be able to answer this question, we considered the LPM operator as a goal-based preference function. This operator essentially specifies a mathematical criterion suitable for problems in the domain of partial hedging. At the same time, its technical definition includes a parameter that accounts for one's personal risk tolerance. By reason of its specification as a hedging criterion, the LPM operator is strongly target-oriented. In conjunction with the person-specific nature implied by the preference parameter, the LPM function is exceptionally appropriate for modelling DC frameworks. Consistent with its specification as a hedging criterion, the LPM operator incorporates a so-called reference level. This reference level is unique to individuals and can be modelled as an explicit retirement goal. For this reason, in our study, we identified the reference level as a person-specific life annuity. We investigated the corresponding utility-maximisation problem in the financial environment proposed by Koijen et al. (2009). Their market model is employed by a.o. the Dutch central bank (DNB) and therefore constitutes a financially meaningful framework. In this environment, the market prices of risk are assumed to be affine in a mean-reverting stochastic process. As an immediate consequence, it is not possible to derive the exact distributional features of the stochastic deflator process. Due to its significant impact on the general optima in the area of continuous-time portfolio optimisation, the latter complicates an analytical retrieval of closed-form solutions. Nevertheless, using inverse Fourier techniques, we were able to derive analytical expressions for the optimal policy rules. Furthermore, we managed to disentangle the distributional properties of retirement wealth in closed-form. Our numerical results demonstrated that the LPM operator is able to significantly improve the likelihood of achieving one's pension goals. Despite this potentially great performance, we also showed that the optimal policy rules are highly sensitive to the estimates for the market prices of risk and may be difficult to implement in reality.

Valorisation

In this attachment, we address the possible impact of this dissertation on both society and the non-academic industry. Despite the thoroughly technical nature of the overarching theme, each core chapter can be related to a topic of practical importance. In fact, all topics concern optimisation problems with regard to portfolio choice and/or savings decisions. The latter phenomena constitute crucial parts in the lives of individuals and institutional investors alike. We are therefore able to evaluate the societal and/or industry-linked impact of this dissertation along a wide array of practical dimensions. In an attempt to classify the corresponding domains of impact, we narrow the subjects of this valorisation down to three interlinked and substantial fields: (i) the pension industry, (ii) asset-liability management, and (iii) the (re-)insurance industry. We believe that this categorisation is meaningful, as most individuals are either directly or indirectly affected by *at least* one of the aforementioned domains. In addition to this, for a great majority of people, the mere connection to portfolio choice problems is established through one of the preceding items. Against the former background, it is noteworthy that this research has actively contributed to the Dutch debate on pension reforms. As a companion paper to Chapters 2, 3 and 4, we have co-authored an industry-oriented article that was conducive to the new Dutch pension agreement.[1] On the grounds of the pension

[1] This research was financially supported by the Network for Studies on Pensions, Aging and Retirement (NETSPAR). NETSPAR forms a platform on which researchers and practitioners contribute to both academic and pension-related discourse. NETSPAR thereby aims to bridge the gap between academia and the industry. Under the umbrella of the NETSPAR-linked theme "*Design of Pension Contracts in Incomplete Markets and under Uncertainty*", we have written this dissertation. Parts of this research have accordingly been presented at numerous NETSPAR seminars and conferences. The article mentioned in the main text concerns Balter et al. (2020) and appeared in NETSPAR's *Design Series*. Their *Design Series* consists of articles that bear relevance to the Dutch pension debate. Subsequently, we address our article's content in more detail.

industry's close ties to a.o. asset-liability management and (re-)insurance, the foregoing contribution exemplifies this dissertation's impact on the remaining two domains. In the sequel, we visit the separate fields and elaborate on the impact associated with the core chapters.

Pension Industry

All chapters included in this dissertation address problems that are relevant to the pension industry. We first focus on the distinct chapters and then comment on the previously mentioned article. Pension funds are in general concerned with acquiring the best possible replacement ratios. In doing so, they have to deal with multiple sources of unhedgeable risk. The ensuing incompleteness can be attributed to e.g. mortality risk or extremely long-dated/illiquid cash-flows. In addition to this, most pension funds in the EU-area are legally obliged to keep up with clearly defined solvency requirements. These requirements pose direct restrictions upon the funds' feasible set of policy rules. The corresponding situation can therefore be modelled by an ordinary constrained terminal wealth or utility-maximisation framework. In that regard, it is clear that the topic central to Chapter 2 becomes highly relevant. We recall that this chapter introduces an approximate dual-control method suitable for constrained optimal control problems. Its mechanism manages to generate near-optimal approximations to the optimal decision variables in closed-form. The advantages associated with this approximating routine are highly beneficial to the pension industry. It concretely enables pension funds to reduce the real-time computational effort required to implement their investment policies. For similar reasons, the dual-control method furnishes analytically tractable and useful insights regarding the impact of market incompleteness on their executed policy rules. Moreover, the mechanism endows a fund with an explicit framework appropriate for effectively managing non-traded risk-drivers. Due to the cumbersome nature of market incompleteness, specifically in the pension-linked context, such a framework constitutes a valuable asset.

In Chapters 3 and 4, we studied utility-maximisation problems involving different types of reference levels. Chapter 3 concerned a theoretical treatment of multiplicative habit formation. The reference level was accordingly defined as an endogenous or internal process. In accordance with the concept of habit formation, we specified utility with respect to consumption over the entirety of the trading interval. For the corresponding optimal consumption problem, we derived a dual formulation. This primarily theoretical result gave rise to numerous applications/insights of practical importance. From a pension fund's perspective, the most meaningful attributes are given by the semi-analytical specification of optimal consumption and the dual-induced evaluation mechanism. To this end, it is important to note that the habit formation framework can be employed to study the optimal consumption/savings behaviour required to ensure person-specific satisfaction with regard to one's standard of living.[2] In a pension-linked configuration, this setup consequently allows for the exact calculation of individually optimal deposits into e.g. a DC scheme. Due to the semi-analytical specification of the optimal consumption process, the fund is able to infer more precisely how these optimal deposits are affected by changes in the financial circumstances. This closed-form element therefore enables a fund to raise realistic expectations on the subject of the participant's defined contributions. The evaluation mechanism can be utilised to gauge the approximate closed-form contributions' accuracy. Since it is difficult to handle problems of this kind analytically, the preceding evaluation routine saves time and opens doors to the construction of more tractable deposit-related policies. The latter touches upon the creation and implementation of dual-control methods adapted to setups involving multiplicative habit formation.

In Chapter 4, we made use of an exogenous or externally defined reference level. Moreover, instead of consumption over the trading interval, the agent was assumed to derive utility from terminal wealth alone. The exogenous nature of the reference level fits well in the confines of a terminal wealth problem.

[2]We stress that this point of view is unique to the multiplicative branch within the literature on habit formation. In additive setups, an agent is required to keep consumption above the habit level at all times. As a consequence, it is hard to interpret the habit component as a standard of living. Adverse changes in the financial circumstances are very likely to negatively affect an individual's savings behaviour. One may realistically be required to scale down consumption below the level to which he/she has become accustomed. Due to the evident relation to pension contributions, it is clear that the additive framework is too restrictive for individual-specific pension schemes. The multiplicative setup is in that respect useful, as it allows for consumption below the habit component. This attribute makes the setup amenable to interpretations provided in the main text.

Specifically in the context of a DC scheme, the latter configuration admits a variety of pertinent interpretations. In the chapter of interest, we predominantly concentrated on the situation for a pension fund offering a DC scheme. For this purpose, we identified the reference level as an individual-specific life annuity. Moreover, to model the agent's preferences, we relied on the LPM operator. This operator outlines a partial hedging criterion adjusted for the agent's risk-tolerance. The target of optimisation correspondingly accommodates a strong orientation towards the predefined reference level. By means of this unique operator, we studied whether it is possible to improve the likelihood of achieving one's pension goals. In consideration of the new Dutch pension agreement, this study is highly relevant for two concrete reasons. First, in conformity with the new agreement, retirement wealth is adjusted based on the fund's performance over the life-cycle. Furthermore, the participant's contributions are more or less held fixed throughout the entire accumulation phase. The nature of the ensuing scheme closely resembles the DC configuration at the heart of our chapter. Second, in line with the corresponding shift in risk from the employer to the employee, the new agreement puts more emphasis on the participant-specific preferences. Qualifications adapted to an individual's personal risk profile should generate retirement-linked outcomes adequate for his/her specific situation. The LPM framework clearly allows for a person-specific identification of preferences, whilst retaining the possibility of favourable pension outcomes. Apart from some technical downsides[3], Chapter 4 indeed demonstrates that the LPM criterion can significantly improve the likelihood of achieving one's desired pension target. Chapter 4's setup could therefore be interesting to pension funds operating in a.o. the Dutch second pillar.

The industry-oriented paper accompanying Chapters 2, 3 and 4 is closely linked to the previous topic. This article is written by Balter et al. (2020) and carries the title "*Investing for Retirement with an Explicit Benchmark*". It concretely studies the impact of a goal-based utility function on the recovery potential of

[3]These downsides concern the optimal policy rules. The numerical results of Chapter 4 namely revealed that the optimal solutions are highly sensitive to the estimates for the market prices of risk. Moreover, in light of particular solvency requirements, the numerically assessed investment strategies are difficult to implement in practice. Both downsides can easily be handled by slight modifications of the optimisation framework. We have addressed potential modifications in Chapter 5.

a pension fund. For this purpose, the paper relies on a DC setup similar to Chapter 4. The primary differences consist in the preference qualifications and the postulates underpinning the financial environment. Whereas the former chapter resorts to the market proposed by Koijen et al. (2009), the industry-oriented article utilises an ordinary uni-dimensional Black-Scholes model. The corresponding stylised market suffices to appropriately convey the benefits associated with the target-oriented preference function. Unlike the aforementioned chapter, the article makes use of a novel utility qualification: the dual-CRRA function. As the name suggests, this preference qualification constitutes a minor modification of the ordinary CRRA operator. It incorporates two CRRA functions with different coefficients of relative risk-aversion. Depending on whether retirement wealth exceeds a prefixed benchmark, utility is derived from either of the two CRRA operators. This newly defined preference paradigm correspondingly allows one to explicitly characterise an individual's preferences around a reference level or target. For this reason, a pension fund is merely required to estimate/calibrate two preference-linked parameters, i.e. the separate coefficients of risk-aversion. In the context of a DC scheme, we identified the reference level as a person-specific life annuity. Moreover, we studied a participant who becomes notably less risk-averse when wealth falls below the reference level. This risk-related behavior approaches the gamble-for-resurrection phenomenon unique to prospect theory. The numerical results suggested that the dual-CRRA function is capable of substantially improving a pension fund's recovery potential. This improvement was based on relative performance with respect to the ordinary CRRA operator. Even though our study can be embedded in a larger body of similar NETSPAR-linked contributions, it was part of an explicit discussion that led to the new pension agreement. The clause on more individual-specific and preference-linked investment strategies can particularly be related to the article at hand.

Asset-Liability Management

Even though asset-liability management, henceforth ALM, constitutes a crucial practice within the pension industry, we are able to highlight some unique corresponding aspects of impact related to this dissertation. For this reason, it is noteworthy that the practice of ALM is not unique to pension funds. In the domains of, for instance, (re-)insurance, banking or trading, ALM plays a prominent role as well. We are best able to stress the impact of this dissertation on the domain of ALM along the lines of Chapters 2 and 4. In Chapter 2, we developed a dual-control method suitable for utility-maximisation frameworks involving convex trading constraints. Constraints of this kind can be modelled to account for solvency requirements and/or (partially) unhedgeable risk-drivers. For example, by enforcing restrictions upon investments in well-defined volatility derivatives, one arrives at a model with non-traded volatility risk. Similar reasoning applies to e.g. mortality risk or inflation risk. By the same token, the former type of constraints can be employed to keep an investor from taking (extremely) large and/or short positions in any of the traded assets. As practitioners in the field of ALM are generally confronted with such restrictions/requirements, our duality framework bears significant relevance to this domain. The latter also held true for the pension industry. However, ALM as such entails a larger body of (un)hedgeable risk factors. In addition to this, the ALM-specific interpretation of the utility-maximisation framework considerably differs from the pension-related one. Consistent with the configuration of Chapter 4, due to the possibly person-specific nature of a utility function, the most obvious pension-linked interpretation pertains to a DC setup. Therein, terminal wealth should be identified as retirement wealth, and the reference level as a person-specific pension goal. To make the setup amenable to ALM in a broader sense, this interpretation has to be generalised. Terminal wealth ought to be identified as the asset process, and the reference level(s) as the liability process. The preference function can correspondingly be specified in a "risk-neutral" manner by means of e.g. hedging operators or mean-variance criteria. Note here that our dual-control method applies to generally defined state-dependent utility qualifications incorporating exogenous benchmark variables governed by broadly specified semi-martingales.

The relevance of Chapter 2 for the extensive field of ALM is straightforward and resembles the pension-linked importance. More specifically, the dual-control method endows asset-liability managers with a tractable and time-efficient tool for calculating/implementing the optimal investment strategies. Both the efficiency and the tractability are attributable to the closed-form nature of the ensuing near-optimal policy rules. These features save time and allow the managers to clearly communicate the details underpinning their strategy-linked choices. Furthermore, due to the solutions' analytical tractability, the "black box" surrounding the conventional numerical applications disappears. In other words, the exact impact of particular model specifications on the approximate decision variables is clear and explainable. Similarly, comparative analyses amongst a set of dual-induced strategies is fairly uncomplicated and consequently facilitates the construction of unique policy rules. On a slightly less fundamental level, Chapter 4's impact on ALM can be situated in the use of goal-based hedging criteria. We recall that this chapter studied the possibility of improving a pension fund's recovery potential using a strongly target-oriented LPM operator. Akin to the preceding dual-control method, this setup can easily be adapted to a more general ALM problem. Under a modified identification of the terminal wealth process as well as the reference level, the setup can be aligned with general ALM frameworks. As a result, the aforementioned recovery potential coincides with a solvency ratio quantifying the degree up to which an institution is able to meet its liabilities. Given the numerically verified positive impact of the LPM operator on the recovery potential, it is clear that this target-oriented function is also able to positively affect the preceding solvency ratio. By means of different hedging benchmarks, our study consequently suggests that asset-liability managers are capable of improving their results/performance. On account of the endogeneity of the reference level in Chapter 3, the link of this chapter to ALM is not clearly visible. Nevertheless, for situations wherein asset-liability managers are required to withdraw capital from their asset process(es), the multiplicative dual formulation can come in handy. Particularly in the spirit of possible dual-control mechanisms, this framework can be employed to facilitate numerical computations in an analytical-friendly manner. As this closely resembles the conceptual impact of the approximating routine developed in Chapter 2, we do not elaborate on the technical details.

Insurance Industry

Great parts of the (re-)insurance industry are concerned with the design and related pricing of products. Well-known products crucial to the life and non-life sectors are health insurances and car insurances, respectively.[4] More abstract and technical products are handled by re-insurers, which may entail e.g. the evaluation of options on risk carried by an insurer. In pricing these products, the insurance industry as well as the re-insurance industry are obliged to deal with a great body of unhedgeable risk-drivers. For life-linked setups, products straightforwardly hinge upon a.o. mortality/longevity risk. Likewise, in configurations relevant for the non-life branch, claim sizes and corresponding frequencies are typically subject to unhedgeable sources of uncertainty. Due to the involvement of (partially) non-traded risk-drivers, the aforementioned pricing process is highly nontrivial. In agreement with our analysis on optimal investment in the presence of trading constraints, this process generally requires computationally demanding applications that lack analytical tractability. Put differently, the risk-neutral evaluation of insurance-linked products can be time-consuming and may pose mathematical challenges. In an attempt to tackle both issues in a relatively understandable manner, the dual-control method central to Chapter 2 proves useful. Moreover, to improve the performance of replicating strategies associated with particular products, the results in Chapter 4 are helpful. We note that the technical finding outlined in Chapter 3 is not compatible with conventional pricing schemes. By virtue of the endogeneity of internal habit components, the framework cannot be reconciled with most evaluation methods. Therefore, in the sequel, we solely elaborate on the precise impact of Chapters 2 and 4 on the (re-)insurance industry.

On account of the universal mechanism underscoring pricing routines, we subsequently do not distinguish between insurers and re-insurers. For the same reason, the impact of this dissertation on general pricing techniques reaches beyond the insurance industry. Investment banks or private investors can possibly benefit from our research as well. However, in view of the fact that (re-)insurance companies occupy a substantial part of the market for financial/actuarial products, we confine ourselves to an impact-related

[4]In this regard, we deem it noteworthy that almost all EU-citizens are in possession of at least a health insurance. In, for example, the Netherlands and Germany, having a health insurance is required by law. Similarly, although not necessarily enforced by legislation, most individuals in possession of a car have an automobile insurance. In addition to these widespread products, a great amount of people wish to purchase e.g. mortgage-linked insurances, life insurances, or personal liability insurances. This stresses the omnipresence of insurance products and their importance for society.

assessment of this sector alone. The latter choice is corroborated by the large amount of individuals who are actively involved in the purchasing process of insurance-linked products, e.g. health insurances or car insurances.

We have already addressed the possible computational and mathematical issues involved with the pricing procedure of insurance-linked products. The burdensome nature of this process is principally attributable to the partial or full non-tradeability of certain risk-drivers. In the confines of utility-maximisation, the dual-control method developed in Chapter 2 is capable of coping with this unhedgeability in a tractable and efficient way. For the upsides of this dual-control method, one can visit the preceding sections or Chapter 1. Hence, to underline the relevance of this duality mechanism for pricing schemes, we must disclose the link between utility-maximisation and risk-neutral evaluation techniques. As it is debatable whether investors at insurance companies can be classified as risk averse or risk-seeking individuals/agents, preference qualifications do not appear to be the greatest targets of optimisation. Nevertheless, the generality of the utility operators included in our dual-control framework allows for more “risk-neutral” objective functions. Examples of such operators include, but are not limited to, the LPM criterion from Chapter 4 or concavified variants of the celebrated mean-variance function. We recall that the state-dependent preference qualifications in Chapter 2 may incorporate exogenous benchmark processes or reference levels. Therefore, under the additional identification of these reference levels as insurance products, the utility-maximisation problem reduces to a setup suitable for finding the best replicating strategies. These replicating strategies would correspond to a fixed initial endowment. To find the “best” price, the (re-)insurer can determine this endowment in such a manner that the replacement ratio exceeds 100% with a probability of, say, 99%. The latter implies that, in 99% of the cases, the near-optimal analytical replicating strategy generates a proper (partial) hedge against the uncertainty induced by the product. As a result, the corresponding price seems appropriate. The outcomes reported in Chapter 4 demonstrate that a pricing approach of this kind is able to render viable outcomes.

In a similar fashion, one can adapt our dual-control mechanism to utility indifference pricing techniques. These setups do not necessarily depend on "risk-neutral" characterisations of the preference functions. In fact, this pricing technique makes explicit use of Inada-type utility functions in an attempt to compute fair evaluations of partially non-traded insurance products. Utility indifference pricing works along the lines of two interrelated problems as follows. In the first problem, one simply solves for the optimal trading strategy. Utility is here derived from terminal wealth alone. Taking into account this is far from straightforward in the presence of non-traded risk, our dual-control method comes in useful. In the second problem, one also solves for the optimal trading strategy. However, the agent is endowed with his/her initial capital *minus* a constant amount of monetary units. In addition to this, utility is derived from terminal wealth *plus* the insurance product. For reasons addressed around the first problem, our dual-control method may be utilised to arrive at tractable near-optimal solutions for the relevant policy rules. According to the principle of indifference pricing, the fair price for the preceding product is equal to the aforementioned amount of monetary units. This amount must namely be determined such that the objective functions of both problems are equal to each other. Indifference pricing typically relies on utility functions from the exponential family to derive closed-form expressions. Other utility functions in general pose problems with regard to an analytical retrieval of optimal replicating strategies and/or indifference prices. Our approximating routine enlarges this narrow class of applications. The (re-)insurance industry can consequently employ our dual-control routine to find tractable indifference prices in an efficient way for a considerably larger class of (more realistic) utility choices. This shows how our research facilitates and improves the financial/actuarial fair pricing of insurance products. The fair nature benefits both the (re-)insurers and the large number of insured agents.

Curriculum Vitae

Thijs Kamma was born on 25 February, 1997 in Amsterdam, the Netherlands. Between 2008 and 2014, he attended high school at Stella Maris College in Meerssen. After receiving his Gymnasium diploma, he studied Econometrics and Operations Research at Maastricht University. In August 2017, he obtained his B.Sc. degree. During his undergraduate studies, he spent one semester at Universität Mannheim, Germany, as an exchange student. In August 2018, he obtained his M.Sc. degree with a specialisation in Actuarial Science with distinction (*cum laude*). His master thesis entitled "Investing for General Utility: Results for the Black-Scholes-Hull-White Model". After graduation, Thijs joined the Department of Quantitative Economics as a Ph.D. Candidate in September 2018, under the supervision of Prof. dr. Antoon Pelsser and dr. Thomas Post. The results of his research are presented in this dissertation. Thijs presented his work at various seminars and international conferences, such as the International Congress on Insurance, Mathematics and Economics (IME), the Vienna Congress on Mathematical Finance (VCMF), the Actuarial and Financial Mathematics Conference (AFMath), the Bernoulli IMS One World Symposium, the Conference on Methods for Actuarial Sciences and Finance (MAF), the Winter Seminar on Mathematical Finance, the World Congress of the Bachelier Finance Society, the European Actuarial Journal Conference (EAJ), and Netspar meetings at different places in the Netherlands. In October 2022, he joined the Chair of Mathematical Finance at the Technische Universität München as a postdoctoral researcher.