Data driven approach to robust pricing, hedging and risk management and its dynamics in time

Johannes Christian Wolfgang Wiesel
St John's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2020

"When the facts change, I change my mind. What do you do, sir?" – *often attributed to John Maynard Keynes but probably due to Paul Samuelson.*

Acknowledgements

I would like to take the opportunity to thank my supervisor Jan Obłój for his continued support during the last 3.5 years. It has been a great pleasure working with him and I feel that I have profited enormously from his enthusiasm, guidance and experience. In particular I very much enjoyed the regular meetings with his working group formed of Anna Aksamit, Gaoyue Guo and Tongesok Lim, from whom I learned a lot.

Furthermore I would like to thank Rama Cont and Marcel Nutz for accepting to act as examiners for this thesis as well as for giving me the opportunity to present some of my preliminary results in the Oxford and Columbia mathematical finance seminars.

I would also like to express my gratitude to my coauthor Laurence Carassus for her guidance especially during the early stages of my DPhil, Daniel Bartl and Samuel Drapeau for many inspiring discussions, without which this DPhil thesis would be a lot less diverse, and Sam Cohen for discussing his creative mathematical (amongst others) ideas. A special thanks goes to Stephan Eckstein and Martin Tegner for sharing some of their vast knowledge about numerical implementations as well as to Matthias Löffler for providing insights into the state of the art of statistical modelling. Furthermore I am greatly indebted to Mathias Beiglböck for hosting me in the congenial environment of his research group in Vienna. In particular I would like to acknowledge the many constructive and stimulating discussions with Julio Backhoff and Gudmund Pammer.

I would also like to thank the members of the mathematical Finance and Stochastic Analysis Group in Oxford for making my time there so inspiring and enjoyable. Especially I would like to thank Matteo Burzoni for the many discussions on robust finance and Krzysztof Ciosmak for sharing his knowledge about optimal transport as well as Vlad Margarint and David Prömel for sharing their insights on everything else. I am also indebted to Christina Zou for reading an early version of this thesis.

I am deeply grateful to my family for their unconditional love and support. The orange rubber ball is still very much in use. Last but not least of course my heart belongs to Maddi.

Contents

List of Figures

Chapter 1

Introduction

The overaching goal of the results presented in this thesis is the development of a time-consistent data-driven approach to robust or model-independent mathematical finance. This relatively young research field is inspired by theoretical developments going back to Knight [1921] and aims to quantify uncertainty on the choice of pricing models, which had been largely ignored in the mathematical modelling of financial markets until the 2007/08 financial crisis and should be construed in stark contrast to other sources of uncertainty, which can be quantified within a specific pricing model (e.g. statistical uncertainty or miscalibration error). Within the last 20 years, a large body of research works in the area of robust mathematical finance has emerged, leading to the advancement of many mathematically challenging tools, such as the theory of quasi-sure and pathwise stochastic analysis. The majority of these works has been focused on describing market efficiency and pricing of options at time zero, i.e., today. In contrast, the motivation of the research presented in this thesis is to extend this framework to a setting, which is dynamic in time and takes new sources of information (e.g. new statistical estimates or changed market prices) into account as soon as they emerge. In this context, we establish new connections to the fields of optimal transport on the one hand and statistics on the other, with the ultimate goal to develop a universal toolbox for the implementation of robust and time-consistent trading strategies and risk assessment. To form a basis for this, we provide a unified framework for state of the art modelling in discrete time robust mathematical finance. This is presented in Chapter 2 and could serve as a point of reference for experts in the field and new researchers entering the area alike. Chapters 3 and 4 are concerned with robust no-arbitrage pricing bounds and with the robust superhedging price of a financial position. More concretely, Chapter 3 describes the evolution of robust superhedging prices through time via a dynamic programming principle. It also addresses the question how to identify a unique optimal superhedging strategy via a secondary utility maximisation problem. Chapter 4 then discusses several natural estimators for the $\mathbb{P}$-a.s. superhedging price and their robustness properties.

Chapter 5 is motivated by open research problems regarding the implementation of such a time-dynamic framework outlined above. In this context, we establish continuity properties of superhedging prices or more generally of martingale optimal transport cost functionals in Chapter 5, which provide the theoretical foundations for the computation of pricing bounds through a discretisation procedure. The new results developed on the way comprise theoretical approximation results as well as implementable algorithms for the construction of martingale measures with certain properties.

In the remainder of this section we will first give an informal introduction to discrete time financial markets. Subsequently we will summarize the main results presented in the different chapters of this thesis without going into the technical details. A more detailed introduction and literature review will be provided at the beginning of each individual Chapter 2-5. The chapters correspond to the articles Obłój and Wiesel [2018], Carassus et al. [2019], Obłój and Wiesel [2020] and Wiesel [2019].

1.1 Classical modelling approach in discrete time mathematical finance

Before we explain the concept of model uncertainty in mathematical finance in more detail, let us first quickly recap the nomenclature of the classical framework for modelling discrete time financial markets revelant for this thesis. This will mostly be an informal discussion. In particular we will try to make it as intuitive as possible and thus refer to Föllmer and Schied [2004] for references and a rigorous treatment of the topics introduced here.

Let us now detail how we model financial markets: we assume that we can trade at discrete time points, which we call $\{0, 1, \dots, T\}$ for simplicity. The usual convention is that today marks $t = 0$. Instead of trying to describe assets on this market via a system of deterministic relationships, we consider asset prices as stochastic processes. That is, we define a d-dimensional stochastic process $S = (S_t^i)_{t=0,\dots,T}^{i=1,\dots,d}$ on a fixed probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which models nominal prices of the relevant assets. The flow of information in this financial market is captured by a filtration $\mathbb{F} = (\mathcal{F}_t)_{t=1,\dots,T}$ and we assume that the prices S are observed, so S is $\mathbb{F}$-adapted. In particular we will refer to $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ as our market model in this section. We usually assume that there exists an additional asset $(S_t^0)_{t=0,\dots,T}$, which has a price equal to one for all times, i.e. $S_t^0 = 1 \; \forall \, t = 0, \dots, T$. The intuition for this is that this asset is a fixed-rate security with rate equal to zero, which is not too unrealistic nowadays. Alternatively we could define S as the price process discounted by S^0. We note however, that making this assumption in itself constitutes a crude abstraction and markets, where such a situation does not apply, are a well-researched topic in mathematical finance.

We can formally trade in the market described via another d-dimensional predictable process $H = (H_t^i)_{t=1,\dots,T}^{i=1,\dots d}$. The intuition is that at time zero we decide, how many units in the assets S we want to hold: we call this H_1. So at time one our gains and losses from trading are $H_1(S_1 - S_0)$. This of course only holds if we disregard market frictions, which will always be the case in this thesis. We take H to be self-financing. This means that we reinvest our wealth at time one, i.e., we choose H_1 such that $H_1 S_1 = H_2 S_1$. Continuing in this fashion we obtain accumulated gains and losses from trading after T periods of

$$\sum_{t=1}^{T} H_t(S_t - S_{t-1}).$$

Our common sense tells us that it should not be possible to create a sure profit out of zero initial capital by trading in the market. This concept of market efficiency is central to mathematical finance and is called No Arbitrage. Formally it means that the relationship

$$\mathbb{P}\left(\sum_{t=1}^{T} H_t(S_t - S_{t-1}) \geq 0\right) = 1 \quad \Rightarrow \quad \mathbb{P}\left(\sum_{t=1}^{T} H_t(S_t - S_{t-1}) = 0\right) = 1$$

holds. If this is the case for the market model $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$, then we say that the market is free of arbitrage. One of the most famous results in mathematical finance formulates a necessary and sufficient criterion for such markets. It is called the first fundamental theorem of asset pricing and states that a market is free of arbitrage if and only if there exists a probability measure $\mathbb{Q}$, which is equivalent to $\mathbb{P}$ and satisfies $\mathbb{E}_{\mathbb{Q}}[S_t|\mathcal{F}_{t-1}] = S_{t-1}$ for all $t = 1, \dots, T$. Such a measure $\mathbb{Q}$ is then called an equivalent martingale measure. Martingale measures can be interpreted as fair games, as on average, the gains and losses of a trading strategy H under $\mathbb{Q}$ will be zero, i.e. $\mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{T} H_t(S_t - S_{t-1})\right] = 0$.

So why is this theory useful? One reason is the following: assume that we want to sell an option, which is not liquidly traded in the market. That means that the price of the option is not explicitly known. Once we have established that the market modelled by $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ is efficient (i.e. free of arbitrage opportunities), we can try to find out the fair price for the option. Let us assume here that the option considered is of European type. Thus it can be modelled as a payoff $g(S)$, which we have to pay at time T. A natural way to find a price for g is to determine an initial capital x and a self-financing strategy H, such that

$$x + \sum_{t=1}^{T} H_t(S_t - S_{t-1}) = g(S) \qquad \mathbb{P}\text{-a.s.} \tag{1.1.1}$$

Such a strategy is called a hedging strategy, because, if we trade according to H we have eliminated any risk of not being able to pay g at terminal time $\mathbb{P}$-a.s. We also note here that taking expectations under any equivalent martingale measure $\mathbb{Q}$ yields $x = \mathbb{E}_{\mathbb{Q}}[g]$. Equivalent martingale measures $\mathbb{Q}$ are thus often thought of as pricing models for the market $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$.

As finding a self-financing strategy H satisfying (1.1.1) might not always be possible, we could instead demand the weaker relation

$$x + \sum_{t=1}^{T} H_t(S_t - S_{t-1}) \geq g(S) \qquad \mathbb{P}\text{-a.s.}$$

for x and H. In this case the strategy H is called a superhedging strategy with initial capital x. If we want to superhedge g, then we do not normally want to allocate more money for a superhedging strategy than strictly necessary, so we are interested in the quantity

$$\pi(g) = \inf\{x \in \mathbb{R} \mid \text{there exists } H \text{ s.t. } H \text{ is a superhedging strategy with initial capital } x\}.$$

Of course, if all market participants believe in the market model $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$, then nobody understanding how to trade in our market will pay $\pi(g)$ for it: buying the option from us will only really be attractive to them if they had to pay less than $\pi(g)$. In fact, it turns out that they can even make a $\mathbb{P}$-a.s. profit if they pay less than $-\pi(-g)$, the so-called subhedging price for the option g. Thus the range of sensible (or arbitrage free) prices for g is exactly given by $[-\pi(-g), \pi(g)]$. As we mentioned above there is a dual viewpoint for pricing the option g, namely via the equivalent martingale measures $\mathbb{Q}$, which can be thought of as pricing models, attributing a price of $\mathbb{E}_{\mathbb{Q}}[g]$ to the option g. The second famous result in mathematical finance we want to mention here is fittingly called the second fundamental theorem of asset pricing and states that, under No Arbitrage, the identity

$$\pi(g) = \sup_{\mathbb{Q} \text{ is an equivalent martingale measure}} \mathbb{E}_{\mathbb{Q}}[g]$$

holds, i.e. the superhedging price is equal to the highest model price. Let us point out here, that pricing via (super)-hedging does only depend on the underlying measure $\mathbb{P}$ via its null-sets, but there is in general no clear relationship between $\mathbb{P}$ and equivalent martingale measures $\mathbb{Q}$ otherwise. It is thus a natural question, how much one can say about $\pi(g)$ by observing the history of S. This is exactly the motivation for Chapter 4.

Instead of selling a specific option g and trying to superhedge it, we could also try to optimally invest our money in the market. The most famous way to formalise this is by looking at the expected utility of our trading gains and losses given by

$$\mathbb{E}_{\mathbb{P}}\left[U\left(x+\sum_{t=1}^{T}H_t(S_t-S_{t-1})\right)\right],$$

where U is a utility function, capturing the preferences of the investor: for example it is often assumed to be monotone and strictly concave, i.e. gaining more money does not make us equally more happy. For fixed initial capital x, we could then try and find a self-financing strategy H^*, which is a solution to

$$\max_{H \text{ self-financing}} \mathbb{E}_{\mathbb{P}}\left[U\left(x+\sum_{t=1}^{T}H_t(S_t-S_{t-1})\right)\right] \tag{1.1.2}$$

for given initial capital x. The problem (1.1.2) is called a utility maximisation problem and will feature in Chapter 3.

The notions of arbitrage and superhedging prices $\pi(g)$ will play an important role in Chapters 2-5 of this thesis. More specifically we will investigate these quantities when the underlying market model is uncertain. The concept of model uncertainty will thus be central to this thesis. It goes back to the economist Frank Knight[1], who states in [Knight, 1921, p. 20]:

> "It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all. We shall accordingly restrict the term "uncertainty" to cases of the non-quantitive type. It is this "true" uncertainty, and not risk, as has been argued, which forms the basis of a valid theory of profit and accounts for the divergence between actual and theoretical competition."

In essence, the key difference between model uncertainty (sometimes also called Knightian uncertainty or the unknown unknown) and model risk (sometimes also called the known unknown) is thus the fact, that model uncertainty cannot be quantified within a given model. How does the abstract concept of model uncertainty link to the market model $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ described above? To see this, we first have to acknowledge that $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ is always only an approximation to reality and is never explicitly known in real life. It can thus only be extrapolated (e.g. from past data or market expectations). In the rest of this thesis we will thus assume that the exact market model is not known. Instead we will introduce an advanced probabilistic setup, which includes different market models $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ simultaneously. This setup is called the robust or model-independent approach to mathematical

[1] (1885-1972)

finance. It thus solves the dilemma of not being able to quantify model uncertainty within a specific model by working on meta-level, which comprises many different models at the same time. Investigations in this area started approximately 20 years ago with research on drift uncertainty. This sparked many new developments both on the theoretical and practical side. Not least because of the last financial crisis, highlighting the need for new models, the field has become increasingly popular. We will provide a comprehensive literature review of research in this area in the introductions of Chapters 2 and 3.

Importantly the robust framework thus enables us to quantify how sensitive the original market model $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, S)$ is with respect to specific assumptions: this can be done by tracking the values of $\pi(g)$ or of the utility maximisation problem (1.1.2). In this thesis, Chapter 2 examines how different models capture model uncertainty, while Chapter 3 combines robust superhedging prices $\pi(g)$ and a robust utility maximisation problem similar to (1.1.2) in order to optimise superhedging strategies H. Chapter 5 contributes to the practical application of the above theory by rigorously justifying convergence of numerical algorithms for the computation of $\pi(g)$.

1.2 Chapter 2: A unified framework to modelling financial markets in discrete time

Over the last years, the literature in robust mathematical finance has put forward many notions of market efficiency and ways to produce no-arbitrage pricing bounds. The first main goal of this chapter is to provide an overaching understanding and unification of these different frameworks, highlighting their interrelations and dependencies. We focus here on a simple discrete-time setup without market frictions. Given this setting, most of the existing contributions work within the so-called quasi-sure approach or pathwise approach. Both approaches include as special cases the classical model-specific setting $(\Omega, \mathcal{F}, \mathbb{P})$ as well as the model-independent setting, where no model for the asset dynamics is specified. Furthermore, both allow to interpolate the modelling spectrum between these extreme cases by gradually changing the modelling framework. However, the quasi-sure framework uses families of probability measures $\mathcal{P}$ to articulate model uncertainty, while the pathwise framework employs a selection of scenarios Ω. The two approaches were captured in their full generality respectively in Bouchard and Nutz [2015] and in Burzoni et al. [2019] and both works have received a considerable attention in the research community. In particular, both approaches establish a Fundamental Theorem of Asset pricing of the form

$$\text{No Arbitrage} \Leftrightarrow \text{Existence of martingale measures } \mathbb{Q}$$

and a Superhedging Theorem of the form

$$\sup_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[g] = \inf\{x \mid x \text{ is the initial capital of a superhedging strategy of } g\}.$$

As a first contribution in this chapter we show that, under mild regularity assumptions, the two frameworks are equivalent in respect of these two theorems: results of one can be deduced from the other and vice-versa. More concretely the following metatheorem captures our approach:

Metatheorem. *Suppose we are in the quasi-sure setting with a given set of priors $\mathcal{P}$. Then, there exists a suitable selection of scenarios $\Omega^{\mathcal{P}}$ such that the pathwise result for $\Omega^{\mathcal{P}}$ implies the quasi-sure result for $\mathcal{P}$.*
Conversely, suppose we are given a selection of scenarios Ω. Then, there is a set of priors $\mathcal{P}^{\Omega}$ such that the quasi-sure result for $\mathcal{P}^{\Omega}$ implies the pathwise result for Ω.

In establishing the equivalence, we explain how to go from one to the other and back and link their respective no-arbitrage conditions. In essence this is achieved by reducing both frameworks to a class of representative models $\mathbb{P}$, for which the classical NA($\mathbb{P}$) condition holds if and only if robust no arbitrage in the corresponding framework is satisfied. A similar reduction argument yields the existence of extremal pricing models $\mathbb{P}$ which attain the robust superhedging price. This is done by utilising pointwise arguments for the geometry of quasi-sure supports.

Secondly, we provide a number of insights into existing results and extend these. In particular, we catalogue existing notions of (robust) no-arbitrage introduced in Riedel [2015], Acciaio et al. [2013], Davis and Hobson [2007], Bouchard and Nutz [2015], Burzoni et al. [2016, 2019], Blanchard and Carassus [2019], Bayraktar et al. [2014], Cox and Obłój [2011], Cox et al. [2016], Bartl [2019] and show how they relate to each other. We also investigate pathwise superhedging in detail. Given a prediction set of price paths Ω, we explore when a superhedging property with respect to the set of martingale measures supported on Ω may be extended to pathwise superhedging on Ω without changing the superhedging price. This is motivated by computational complexity, as it is in general expensive to determine the set of martingale measures supported on Ω explicitly. Alas, a general extension property turns out to be infeasible and we give several counterexamples to emphasize why this is the case. In fact, the requirement of (universal) measurability of trading strategies implies that

an extension property can only hold under strong continuity properties of superhedging prices combined with a certain structure of the underlying inefficient subset of Ω, on which pathwise arbitrage is possible.
Finally, we shed light on some of the technical assumptions in the previous works (e.g. analytic product structure assumptions, measurability of the hedging strategies).

1.3 Chapter 3: The robust superreplication problem: a dynamic approach

In this chapter we work in the discrete-time $\mathcal{P}$-quasi-sure framework of Bouchard and Nutz [2015] and consider an agent, who needs to hedge a liability g maturing at some future date $T > 0$. Our aim is to investigate the interplay between different modeling assumptions, which can be formulated in this setup: more concretely we model the beliefs used for assessing the risks, the beliefs used for an agent's investment decisions and the dynamics of agent's actions explicitly.
At the beginning of the trading period, the agent allocates capital equal to $\pi(g)$, the robust superhedging price. This guarantees that she will be able to cover the liability g $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$ by trading in the market. The formulation and properties of robust superhedging prices has been the subject of active research in the area of robust finance. Nevertheless, best to our knowledge, the focus has been mostly on the static problem, i.e. describing the initial capital $\pi(g)$ needed at time $t = 0$. On the contrary, our first goal of this chapter is to describe the evolution of the superhedging price through time. Denoting this evolution by $\pi_t(g)$, we establish the following dynamic programming principle:

Result (see Theorem 3.3.1)**.** *Under No $\mathcal{P}$-q.s. Arbitrage we have $\pi_t(g) = \mathcal{E}^t(g)$, where $\mathcal{E}^t(g)$ denotes the worst-case model price given as the supremum of expectations under measures in $\mathcal{P}$. The robust superhedging price $\pi_t(g)$ can also be construed as the quasi-sure concave envelope of $\pi_{t+1}(g)$ evaluated at the price S_t. Furthermore, there exists a superhedging strategy attaining $\pi_t(g)$, which is minimal in the sense of Föllmer and Kramkov [1997] .*

The above result is anticipated given the results presented in Chapter 2 and relies mainly on the technical assumptions on $\mathcal{P}$ together with observations made in Bouchard and Nutz [2015].

Our second main result departs from the observation, that, while the dynamics of the superhedging price $\pi_t(g)$ are uniquely determined, the corresponding robust superhedging strategies might well not be. In fact we demonstrate in a simple example, that, depending on the dynamics of S, the investor can choose if and when to consume trading gains. It is thus natural to determine the consumption stream (H^*, C^*), which is an optimiser of the problem

$$\sup_{(H,C)} \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{t=1}^{T} U(t, \Delta C_t)\right], \tag{1.3.1}$$

where the expected utility of inter-temporal consumption ΔC_t is optimised subject to a robust superhedging constraint for the initial capital $\pi(g)$. This utility maximisation problem is calculated under a different class of probability measures $\mathcal{P}^u$ representing the investors subjective market views, which are assumed to be consistent with $\mathcal{P}$. We establish the following result:

Result (see Theorems 3.4.3, 3.4.5). *(i) If the trader's utility functions $U(s,\cdot,\cdot)$ are bounded, concave and satisfy a Carathéodory condition, then there exists an optimising strategy (H^*, C^*) for* (1.3.1).

(ii) Assume that $U(t,\cdot)$ is as in (i) and strictly concave. If $\mathcal{P}^u$ is weakly compact and both $\mathcal{P}$ and $\mathcal{P}^u$ are continuous, then there exists a worst-case measure $\mathbb{P}^* \in \mathcal{P}^u$ such that

$$\sup_{(H,C)} \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{t=1}^{T} U(t, \Delta C_t)\right] = \sup_{(H,C)} \mathbb{E}_{\mathbb{P}^*} \left[\sum_{t=1}^{T} U(t, \Delta C_t)\right].$$

In this setting, the maximising strategy (H^*, C^*) is unique in the following sense: for any two maximising strategies $(H^1, C^1), (H^2, C^2)$ and for $1 \le t \le T$ we have $C_t^1 = C_t^2$ and $H_t^1 \Delta S_t = H_t^2 \Delta S_t$ $\mathbb{P}^*$-a.s.

We conclude the chapter with a number of examples highlighting necessity of our conditions in specific settings. We also give sufficient conditions, under which the $\mathcal{P}$-q.s. setting degenerates to the classical one: this is in particular the case, when both the set of priors $\mathcal{P}$ and the function g are essentially uniformly continuous.

1.4 Chapter 4: Statistical estimation of superhedging prices

In this chapter we consider the estimation of the risk of a financial position g. This is a topic of practical importance and one with a large body of literature, both theoretical and practical. However, to the best of our knowledge, the focus of existing works has been on a static risk assessment, i.e., financial positions are fixed and their risk is evaluated. In contrast, in this chapter we propose to consider a dynamic setting, in which an agent can trade and optimises her position to offset the risk of her liabilities. This perspective is not novel, indeed it underpins most of the mathematical finance literature, but has not been considered from a statistical perspective.

It may at first appear surprising that such a gap has been left open for a long time. This is largely due to the way two different communities evolved, thinking about objects in rather different terms. It could be seen as a contrast between risk assessment and trading activities of a bank or between the "physical measures" vs "risk-neutral measure" literature. In the mathematical finance community trading is at the heart of things, going back to the idea of Black and Scholes [1973b] of pricing via hedging. In practice, mareket models are calibrated to option prices ("forward looking data") and historical data is seldomly used directly. In contrast, a risk assessment group will typically only look at the historical prices to estimate extreme quantiles and use these to assess riskiness of positions. Our work shows how to use historical data, or both historical and option price data, in a consistent and coherent way to estimate the superhedging price, or more generally a risk measure-based price. The idea to use both kinds of data together was pioneered by Mykland [2003a]. In this work, historical returns are used to select a prediction set, i.e., the set of paths on which the superhedging property is required, and subsequently options are used as trading instruments in the computation of the cheapest superhedge. Our approach inherits from that perspective but takes a statistical viewpoint and evolves it into a dynamic and asymptotically consistent methodology. In this chapter we propose four new estimators of the superhedging price, establish their consistency and investigate their robustness with respect to random disturbances of the underlying data set. In this way we embed the research streams of robust statistics, optimal transport and robust or model-independent mathematical finance into a coherent framework.

More concretely, given stock returns $r_1, \dots, r_N$, this chapter aims to understand non-parametric estimation of the $\mathbb{P}$-a.s. superhedging price of an exotic option g. As a natural first step, we examine the simple plugin approach

$$\inf\{x \in \mathbb{R} \,|\exists H \in \mathbb{R}^d \text{ s.t. } x + H(r-1) \geq g(r) \; \forall r \in \{r_1, \dots, r_N\}\}.$$

We establish the following result:

Result (see Theorem 4.2.1)**.** *Let $r_1, \dots, r_N$ be realisations of a time-homogeneous ergodic Markov chain with invariant distribution $\mathbb{P}$. Then the plugin estimator is strongly consistent, but lacks robustness in the sense of Tukey-Hampel-Huber.*

In consequence we propose novel estimators, which overcome the shortcomings of the plugin approach. These new estimators are based on the dual formulation of the superhedging price

$$\inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H(r-1) \geq g(r) \ \mathbb{P}\text{-a.s}\} = \sup_{\mathbb{Q}\sim\mathbb{P}, \mathbb{E}_{\mathbb{Q}}[r]=1} \mathbb{E}_{\mathbb{Q}}[g].$$

In order to achieve statistical robustness and to obtain better control over the point estimates it is necessary to consider a larger class of martingale measures. A natural choice is

$$\pi_{\mathcal{Q}_N}(g) = \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}_{\mathbb{Q}}[g],$$

where $\mathcal{Q}_N$ is a subset of all martingale measures $\mathcal{M}$. The plugin estimator corresponds to taking $\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \hat{\mathbb{P}}_N\}$ and we suggest to replace it with

$$\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} : \exists \tilde{\mathbb{P}} \in B_N(\hat{\mathbb{P}}_N) \text{ s.t. } \mathbb{Q} \sim \tilde{\mathbb{P}}\},$$

where $B_N(\hat{\mathbb{P}}_N)$ is some "ball" in the space of probability measures around the empirical measure $\hat{\mathbb{P}}_N$. In this context the metric used on the space of probability measures plays a crucial role. Indeed, we have the following:

Result (see Theorems 4.3.3, 4.4.5)**.** *If $B_N(\hat{\mathbb{P}}_N)$ is a ball in Wasserstein infinity metric $\mathcal{W}^\infty$ and g is continuous, then under some regularity assumptions on the support of $\mathbb{P}$ the estimator $\pi_{\mathcal{Q}_N}(g)$ is strongly consistent and robust.*

In general however such $\mathcal{Q}_N$ is too large. Instead, the main insight is to consider a tradeoff between the size of the balls and the behaviour of martingale densities:

$$\hat{\mathcal{Q}}_N := \{\mathbb{Q} \in \mathcal{M} \mid \|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \leq k_N \text{ for some } \tilde{\mathbb{P}} \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)\},$$

where $B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)$ denotes the p-Wasserstein ball of radius ε_N around $\hat{\mathbb{P}}_N$ and $\varepsilon_N \to 0$ as well as $k_N \to \infty$. A suitable choice of ε_N and k_N can be found by taking confidence bounds in Wasserstein distance between the true and empirical measure of Fournier and Guillin [2015] into account. This gives:

Result (see Theorems 4.3.6, 4.4.2)**.** *The estimator $\pi_{\hat{\mathcal{Q}}_N}(g)$ is consistent for a bounded continuous or Lipschitz continuous function g. Furthermore $\pi_{\hat{\mathcal{Q}}_N}(g)$ is robust in a Hausdorff-Wasserstein metric.*

This also allows to study the cases when the estimator naturally extends to the setting of superhedging under model uncertainty about $\mathbb{P}$. We can furthermore extend the analysis to the case when risk is assessed not using the superhedging capital but rather via a generic risk measure ρ admitting a Kusuoka representation, see Kusuoka [2001]. We also address the related issue of estimating corresponding superhedging strategies and, in part, extend our results to a multiperiod discrete time setting.

1.5 Chapter 5: Continuity of the martingale optimal transport problem on the real line

This chapter considers the so-called martingale optimal transport problem, i.e. the minimisation problem

$$C(\mathbb{P}, \tilde{\mathbb{P}}) := \inf_{\pi \in \mathcal{M}(\mathbb{P}, \tilde{\mathbb{P}})} \int \ell(x_1, x_2)\, \pi(dx_1, dx_2).$$

This was introduced in Beiglböck et al. [2013] in discrete time and in Galichon et al. [2014] in continuous time. It constitutes a version of the optimal transport problem, which was first posed by Gaspard Monge [1781], but with additional linear constraints. In this chapter we focus on the case where $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are probability measures on $\mathbb{R}$. Here $\mathcal{M}(\mathbb{P}, \tilde{\mathbb{P}})$ is defined as the set of probability measures on $\mathbb{R} \times \mathbb{R}$, under which the canonical process (x_1, x_2) on $\mathbb{R} \times \mathbb{R}$ is a martingale with marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$ and $\ell : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a loss function. As mentioned in Section 1.2, this problem is often in duality to the problem of determining the cheapest price for a robust superhedge. From a practical perspective it is thus of fundamental importance to be able to efficiently compute the cost functional $C(\mathbb{P}, \tilde{\mathbb{P}})$ by numerical routines. Typically this is done by discretisation of the marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$. This motivates the question, if and under which conditions the map $(\mathbb{P}, \tilde{\mathbb{P}}) \mapsto C(\mathbb{P}, \tilde{\mathbb{P}})$ is continuous. Such a continuity property is well known for classical optimal transport (see e.g. [Villani, 2008, Theorem 5.20, p.77]). We affirmatively answer this question by establishing continuity properties of the projection of an arbitrary coupling of $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on to its (Wasserstein-)closest martingale counterpart. In particular we establish the following result:

Result (see Proposition 5.2.4). *Under a barycentre dispersion assumption on π we have*

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mathbb{P}, \tilde{\mathbb{P}})} \mathcal{W}^1(\pi, \tilde{\pi}) \leq \mathbb{E}_{\mathbb{P}} \left[\left| \int (x_2 - x_1)\, \pi_{x_1}(dx_2) \right| \right] \tag{1.5.1}$$

for a coupling $\pi \in \Pi(\mathbb{P}, \tilde{\mathbb{P}})$, where $\mathcal{W}^1$ denotes the 1*-Wasserstein distance*

$$\mathcal{W}^1(\pi, \tilde{\pi}) = \inf_{\gamma \in \Pi(\pi, \tilde{\pi})} \int |x_1 - y_1| + |x_2 - y_2|\, \gamma(dx, dy)$$

and $(\pi_{x_1})_{x_1 \in \mathbb{R}}$ is the disintegration of the coupling π with respect to its first coordinate.

Important examples of measures π, for which the above theorem holds, are the monotone (i.e. Hoeffding-Fréchet) and antitone couplings. The intuition of this theorem is that one can construct a minimising martingale measure by shifting probability mass of the disintegration π_{x_1} either only upwards or only downwards at each point $x_1 \in \mathbb{R}$. This explains why Jensen's inequality for $x \mapsto |x|$ is actually an equality in the above result.

However it is not the case that (1.5.1) holds for any $\pi \in \Pi(\mu, \nu)$. Nevertheless a uniform continuity property can be established:

Result (see Theorem 5.2.8). *For every $\delta > 0$ there exists a constant K such that the following holds: under a uniform integrability condition on $\tilde{\mathbb{P}}$, for every measure $\pi \in \Pi(\mathbb{P}, \tilde{\mathbb{P}})$, we have that*

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mathbb{P}, \tilde{\mathbb{P}})} \mathcal{W}^1(\pi, \tilde{\pi}) \leq \delta + K\, \mathbb{E}_{\mathbb{P}} \left| \int (x_2 - x_1)\, \pi_{x_1}(dx_2) \right|.$$

Through this characterisation we are then able to obtain the following stability of the martingale optimal transport problem:

Result (see Theorem 5.2.11). *Let $(\mathbb{P}^n)_{n \in \mathbb{N}}$, $(\tilde{\mathbb{P}}^n)_{n \in \mathbb{N}}$ be two sequences of measures with $\lim_{n\to\infty} \mathcal{W}^p(\mathbb{P}^n, \mathbb{P}) = 0$ and $\lim_{n\to\infty} \mathcal{W}^p(\tilde{\mathbb{P}}^n, \tilde{\mathbb{P}}) = 0$. Furthermore let $\ell : \mathbb{R}^2 \to \mathbb{R}$ be continuous and such that $|\ell(x_1, x_2)| \leq C(1 + |x_1|^p + |x_2|^p)$ for some $C \geq 0$. Then*

$$\lim_{n\to\infty} C(\tilde{\mathbb{P}}^n, \mathbb{P}^n) = C(\tilde{\mathbb{P}}, \mathbb{P}).$$

Additionally, we also give a new proof of the estimate

$$\inf_{\pi \in \mathcal{M}(\mathbb{P}, \tilde{\mathbb{P}})} \int |x_1 - x_2|\, \pi(dx_1, dx_2) \leq 2\mathcal{W}^1(\tilde{\mathbb{P}}, \mathbb{P})$$

first established in Jourdain and Margheriti [2018]. Lastly, the stability of martingale optimal transport combined with a discretisation result of martingale measures supported on a set Γ allows us to directly prove sufficiency of the monotonicity principle for martingale optimal transport of Beiglböck and Juillet [2016]. This is achieved by directly constructing finitely supported martingale measures, which attain a strictly lower cost as soon as $\pi \in \mathcal{M}(\mu, \nu)$ is not optimal for ℓ.

On a conceptual level, we offer new approximation results for measures in a metric closely related to the Wasserstein metric, namely the adapted or nested Wasserstein metric. We also provide two algorithms, which formalise the iterative construction of the martingale projections mentioned above.

Chapter 2

A Unified Framework for Robust Modelling of Financial Markets in discrete time

2.1 Introduction

Mathematical models of financial markets are of great significance in economics and finance and have played a key role in the theory of pricing and hedging of derivatives and of risk management. Classical models, going back to Samuelson [1965] and Black and Scholes [1973a] in continuous time, specify a fixed probability measure $\mathbb{P}$ to describe the asset price dynamics. They led to a powerful theory of complete, and later incomplete, financial markets. The original models have undergone a myriad of variations including, amongst others, local and stochastic volatility models and have been widely applied. However, they also faced important criticism for ignoring the issue of model uncertainty, particularly so in the wake of the 2007/08 financial crisis. Consequently, inspired by the theoretical developments going back to Knight [1921], new modelling approaches emerged which aim to address this fundamental issue. These can be broadly divided into two streams based on the so-called quasi-sure and pathwise approaches respectively.

The quasi-sure approach introduces a set of priors $\mathcal{P}$ representing possible market scenarios. These priors can be very different and $\mathcal{P}$ typically contains measures which are mutually singular. This presents significant mathematical challenges and led to the theory of quasi-sure stochastic analysis (see, e.g., Peng [2004], Denis and Martini [2006]). In discrete time, this framework was abstracted in Bouchard and Nutz [2015], which we call the quasi-sure formulation in the rest of this chapter. By varying the set of probability measures $\mathcal{P}$ between the "extreme" cases of one fixed probability measure, $\mathcal{P} = \{\mathbb{P}\}$, and that of considering all probability measures, $\mathcal{P} = \mathfrak{P}(X)$, this formulation allows for widely different specifications of market dynamics. The quasi-sure approach has been employed to consider model uncertainty along market frictions and other related problems, see e.g. Bayraktar and Zhou [2017], Bayraktar and Zhang [2016]. The pathwise approach addresses Knightian uncertainty in market modelling by describing the set of market scenarios in absence of a probability measure or any similar relative weighting of such scenarios. It is also referred to as the pointwise, or ω by ω, approach and it bears similarity to the way central banks carry out stress tests using scenario generators. In discrete time a suitable theory was obtained in Burzoni et al. [2019], based on earlier developments in Burzoni et al. [2017a, 2016]. The methodology builds on the notion of prediction sets introduced in Mykland [2003a] and used in continuous time in Hou and Obłój [2018]. The particular case of including all scenarios is often referred to as the model-independent framework and was pioneered in Davis and Hobson [2007] and Acciaio et al. [2013]. From here, a further model specification is carried out by including additional assumptions, which represent the different agents' beliefs. In this manner paths deemed impossible by all agents are eliminated. The remaining set of paths is then called the prediction set, or the model.

Both approaches, the quasi-sure and the pathwise, allow thus to interpolate between the two ends of the modelling spectrum, as identified by Merton [1973]: the model-independent and the model-specific settings (see Figure 2.1). In doing so, they allow to capture how their outputs change in function of adding or removing modelling assumptions, thus allowing to quantify the impact and risk that a given set of assumptions bear on the problem at hand, see Cont [2006]. Both approaches were successful in developing suitable notions of arbitrage

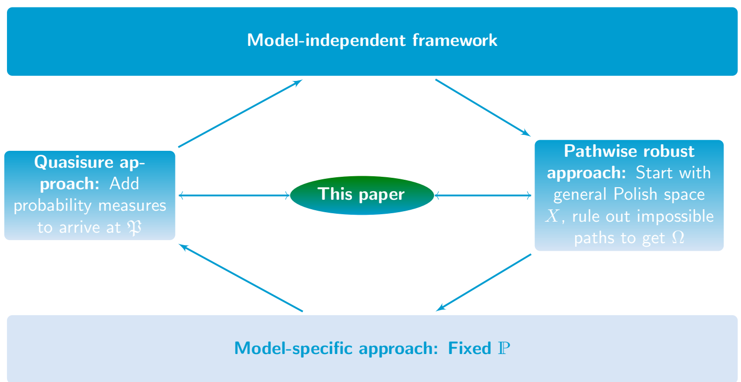

Figure 2.1: Different approaches to modelling financial markets

and extending the core results from the classical $\mathbb{P}$-a.s. setting to their more general context. In particular, in both approaches, it is possible to establish a Fundamental Theorem of Asset pricing of the form

$$\text{No Arbitrage} \Leftrightarrow \text{Existence of martingale measures } \mathbb{Q}$$

and a Superhedging Theorem of the form

$$\sup_{\mathbb{Q}} \mathbb{E}_{\mathbb{Q}}[g] = \inf\{x \mid x \text{ is the initial capital of a superhedging strategy of } g\}.$$

Our main contribution is to unify these two approaches to model uncertainty. We show that, under mild technical assumptions, the pathwise and quasi-sure Fundamental Theorems of Asset Pricing and Superhedging Dualities can be inferred from one another and are thus equivalent. Our statements follow a meta-structure outlined below:

Metatheorem. *Suppose we are in the quasi-sure setting with a given set of priors $\mathcal{P}$. Then, there exists a suitable selection of scenarios $\Omega^{\mathcal{P}}$ such that the pathwise result for $\Omega^{\mathcal{P}}$ implies the quasi-sure result for $\mathcal{P}$.*
Conversely, suppose we are given a selection of scenarios Ω. Then, there is a set of priors $\mathcal{P}^{\Omega}$ such that the quasi-sure result for $\mathcal{P}^{\Omega}$ implies the pathwise result for Ω.

Establishing such equivalence allows us to gain significant additional insights into the core objects in both approaches, as well as clarify links to the classical model-specific setting. In particular, when transposing the results from the pointwise to the quasi-sure setup, the key technical *analytic product structure* assumption in Bouchard and Nutz [2015], see Definition 2.2.1 below, is deduced naturally from the analyticity of the set of scenarios in Burzoni et al. [2019]. When establishing the Superhedging Theorem, we not only show that the pathwise superhedging price of g is equal to the quasi-sure one, but we also show that both are equal to the model-specific $\mathbb{P}$-superhedging price, where $\mathbb{P}$ depends on the setting, i.e., on $\mathcal{P}$ or equivalently on Ω, but also on the payoff g. Finally, the key implication in the proof of the robust Fundamental Theorem of Asset Pricing, i.e., *(5)* $\Rightarrow$ *(1)* in Theorem 2.2.7 below, is obtained by carefully constructing a suitable $\mathbb{P} \in \mathcal{P}$ which does not admit an arbitrage in the classical sense and hence admits an equivalent martingale measure.

Furthermore, we survey and relate the concepts of arbitrage used in both approaches. We provide an extensive list of arbitrage notions introduced and used across the literature on robust finance and establish clear relations between them. We also investigate in detail the notion of pathwise superhedging. As noted in Burzoni et al. [2017a], the pathwise superhedging duality does not hold for general claims g when superhedging on a general set Ω is required. Instead, one has to consider hedging on a smaller "efficient" set Ω^* (defined as the largest set supported by martingale measures and contained in Ω) to retain the pricing-hedging duality. We clarify when this is necessary and when one can extend the superhedging duality from Ω^* to Ω. Intuitively, since there are arbitrage opportunities on $\Omega \setminus \Omega^*$, one could try to superhedge the claim g on $\Omega \setminus \Omega^*$ without any additional cost by implementing an arbitrage strategy. We provide a number of counterexamples to show this idea is not feasible in general and link this to measurability constraints on arbitrage strategies, which were also encountered in Burzoni et al. [2016]. We then show that the above-mentioned intuition is only true for essentially uniformly continuous g under certain regularity conditions on Ω.

The rest of the chapter is organised as follows. Section 2.2 contains the main results. First, in Section 2.2.1, we introduce the general setup in which we work. We discuss different notions of (robust) arbitrage in Section 2.2.2. Then, in Section 2.2.3, we establish our version of the robust Fundamental Theorem of Asset Pricing which unifies the quasi-sure and pathwise perspectives. And in Section 2.2.4, we state a robust Superhedging Theorem. Section 2.3 presents complementary results on extending the superhedging duality from Ω^* to Ω without additional cost and on relations between two strong notions of pathwise arbitrage. Finally, Section 5.6 contains technical results and most of the proofs. In particular, we give the proofs of Theorems 2.2.6 and 2.2.7 in Section 2.4.1 and of Theorem 2.2.9 in Section 2.4.2.

2.2 Unified Framework for Robust Modelling of Financial Markets

2.2.1 Trading strategies and pricing measures

We use notation similar to Bouchard and Nutz [2015] and work in their setting, so we only recall the main objects of interest here and refer to Bouchard and Nutz [2015] and [Bertsekas and Shreve, 1978, Chapter 7] for technical details. Let $T \in \mathbb{N}$ and X_1 be a Polish space. We define for $t \in \{1, \dots T\}$ the Cartesian product $X_t := X_1^t$ and define $X := X_T$, with the convention that X_0 is a singleton. We denote by $\mathcal{B}(X)$ the Borel sets on X, by $\mathfrak{P}(X)$ the set of probability measures on $\mathcal{B}(X)$ and define the function $\text{proj}_t \colon X \to X_1$ which projects $\omega \in X$ to the t-th coordinate, i.e., $\text{proj}_t(\omega) = \omega_t$.
Next we specify the financial market. Let $d \in \mathbb{N}$, $\mathbb{F}$ an arbitrary filtration and let $S_t = (S_t^1, \dots, S_t^d) \colon X_t \to \mathbb{R}^d$ be Borel-measurable, $0 \le t \le T$, and adapted. All prices are given in units of a numeraire, S^0, which itself is thus normalised, $S_t^0 \equiv 1$, $0 \le t \le T$. Trading strategies $\mathcal{H}(\mathbb{F})$ are defined as the set of $\mathbb{F}$-predictable $\mathbb{R}^d$-valued processes. All trading is frictionless and self-financing. Given $H \in \mathcal{H}(\mathbb{F})$, we denote

$$H \circ S_t = \sum_{u=1}^{t} H_u \Delta S_u$$

with $H \circ S_t$ representing the cashflow at time t from trading using H. Above, and throughout, H is a row vector, S is a column vector and 1 denotes either a scalar or a column vector $(1, \dots, 1)^{\mathrm{T}}$. We let Φ denote the vector of payoffs of the statically traded assets

$\Phi = (\phi_\lambda : \lambda \in \Lambda)$, where Λ is some index set. For notational convenience, we often identify Φ with the set of its elements. We assume that each $\phi \in \Phi$ is Borel-measurable. When there are no statically traded assets we write $\Phi = 0$. These assets, which we think of as options, can only be bought or sold at time zero (without loss of generality at zero cost) and are held until maturity T. A trading position h can only hold finitely many of these assets, $h \in c_{00}(\Lambda)$ the space of sequences of reals indexed by Λ with only finitely many non-zero elements, and generates the payoff $h \cdot \Phi = \sum_{\lambda \in \Lambda} h_\lambda \phi_\lambda$ at time T. We call a pair $(h, H) \in c_{00}(\Lambda) \times \mathcal{H}(\mathbb{F})$ a semistatic trading strategy. The class of such strategies is denoted $\mathcal{H}_\Phi(\mathbb{F}) := c_{00}(\Lambda) \times \mathcal{H}(\mathbb{F})$. For technical reasons we also introduce the level sets of S, which are denoted by

$$\Sigma_t^\omega = \{\tilde{\omega} \in X \mid S_{0:t}(\omega) = S_{0:t}(\tilde{\omega})\}$$

for $t \in \{0, \dots, T\}$ and $\omega \in X_t$, where $S_{0:t} := (S_0, \dots S_t)$. Finally, we denote by $\mathbb{F}^0 = (\mathcal{F}_t^0)_{t=0,\dots,T}$ the natural filtration generated by S and let $\mathcal{F}_t^{\mathcal{U}}$ be the universal completion of $\mathcal{F}_t^0$, $t = 0, \dots, T$. Furthermore we write $(X, \mathcal{F}^{\mathcal{U}})$ for $(X_T, \mathcal{F}_T^{\mathcal{U}})$ and often consider $(X_t, \mathcal{F}_t^{\mathcal{U}})$ as a subspace of $(X, \mathcal{F}^{\mathcal{U}})$.

Within this setup, the literature on robust pricing and hedging adopts two approaches to model an agent's beliefs. One stream is scenario-based and proceeds by specifying a prediction set $\Omega \subseteq X$, which describes the possible price trajectories. The other stream proceeds by specifying a set of probability measures $\mathcal{P} \subseteq \mathfrak{P}(X)$, which determines the set of negligible outcomes. We refer to the latter as the quasi-sure approach, while the former is usually called the pathwise, or pointwise, approach. In both cases, the model specification may depend on the agent's market information as well as on her specific modelling assumptions. Changing the sets Ω or $\mathcal{P}$ can be seen as a natural way to interpolate between different beliefs. One of the principal aims of this chapter is to show that both model approaches are equivalent in terms of corresponding FTAPs and Superhedging prices.

In order to aggregate trading strategies on different level sets Σ_t^ω in a measurable way, we always assume in this chapter that Ω is analytic and $\mathcal{P}$ has the following structure:

Definition 2.2.1. *A set $\mathcal{P} \subseteq \mathfrak{P}(X)$ is said to satisfy the Analytic Product Structure condition (APS), if*

$$\mathcal{P} = \{\mathbb{P}_0 \otimes \dots \otimes \mathbb{P}_{T-1} \mid \mathbb{P}_t \textit{ is } \mathcal{F}_t^{\mathcal{U}}\textit{-measurable selector of } \mathcal{P}_t\},$$

where the sets $\mathcal{P}_t(\omega) \subseteq \mathfrak{P}(X_1)$ are nonempty, convex and

$$\textit{graph}(\mathcal{P}_t) = \{(\omega, \mathbb{P}) \mid \omega \in X_t,\ \mathbb{P} \in \mathcal{P}_t(\omega)\}$$

is analytic.

This structure facilitates a dynamic programming principle and allows to essentially paste together one-step results in order to establish their multistep counterparts.
In order to formulate a Fundamental Theorem of Asset pricing we need to define the dual objects to trading strategies: the pricing (martingale) measures. Given a set of measures $\mathcal{P}$, following Bouchard and Nutz [2015], we define

$$\begin{aligned}\mathcal{Q}_{\mathcal{P},\Phi} := \{\mathbb{Q} \in \mathfrak{P}(X) \mid & S \text{ is an } \mathbb{F}^{\mathcal{U}}\text{-martingale under } \mathbb{Q},\ \exists \mathbb{P} \in \mathcal{P} \text{ s.t. } \mathbb{Q} \ll \mathbb{P}, \\ & \mathbb{E}_{\mathbb{Q}}[\phi] = 0\ \forall \phi \in \Phi\},\end{aligned}$$

which, in the model-specific case $\mathcal{P} = \{\mathbb{P}\}$, is simply the familiar set of all martingale measures equivalent to $\mathbb{P}$. Within the pathwise approach, for a set $\Omega \subseteq X$ and a filtration $\mathbb{F}$, we define

$$\begin{aligned}\mathcal{M}^f_{\Omega,\Phi}(\mathbb{F}) := \{\mathbb{Q} \in \mathfrak{P}^f(X) \mid & S \text{ is an } \mathbb{F}\text{-martingale under } \mathbb{Q},\ \mathbb{Q}(\Omega) = 1, \\ & \mathbb{E}_{\mathbb{Q}}[\phi] = 0\ \forall \phi \in \Phi\},\end{aligned}$$

where $\mathfrak{P}^f(X)$ denotes the finitely supported Borel probability measures on $(X, \mathcal{B}(X))$. As a general convention, in this chapter we interpret the above sub- and super-scripts as restrictions on the sets of measures. When we drop some of them it is to indicate that these conditions are not imposed, e.g., $\mathcal{M}_\Omega(\mathbb{F})$ denotes all $\mathbb{F}$-martingale measures supported on Ω. Next let

$$\Omega^*_\Phi := \{\omega \in \Omega \mid \exists \mathbb{Q} \in \mathcal{M}^f_{\Omega,\Phi}(\mathbb{F}^0) \text{ s.t. } \mathbb{Q}(\omega) > 0\} = \bigcup_{\mathbb{Q} \in \mathcal{M}^f_{\Omega,\Phi}(\mathbb{F}^0)} \operatorname{supp}(\mathbb{Q})$$

with the same convention regarding sub- and super-scripts as above. We also define

$$\mathbb{F}^M := (\mathcal{F}^M_t)_{t \in \{0,\dots,T\}}, \quad \text{where } \mathcal{F}^M_t = \bigcap_{\mathbb{Q} \in \mathcal{M}_\Omega(\mathbb{F}^0)} \mathcal{F}^0_t \vee \mathcal{N}^{\mathbb{Q}}(\mathcal{F}^0_T),$$

$\mathcal{N}^{\mathbb{Q}}(\mathcal{F}^0_T) := \{N \subseteq A \in \mathcal{F}^0_T \mid \mathbb{Q}(A) = 0\}$ and $\mathcal{F}^M_t$ is the power set of Ω if $\mathcal{M}_\Omega(\mathbb{F}^0) = \emptyset$.

Remark 2.2.2. Note that $\mathbb{F}^0 \subseteq \mathbb{F}^{\mathcal{U}} \subseteq \mathbb{F}^M$ holds. All these filtrations generate the same martingale measures on Ω calibrated to Φ, which we denote by $\mathcal{M}_{\Omega,\Phi}$.

For $\mathbb{P} \in \mathfrak{P}(X)$, thus $\mathcal{N}^{\mathbb{P}} := \mathcal{N}^{\mathbb{P}}(\mathcal{F}^{\mathcal{U}})$ denotes the collection of its null sets. Likewise, given a family $\mathcal{P} \subset \mathfrak{P}(X)$, the collection of its polar sets if given by $\mathcal{N}^{\mathcal{P}} = \bigcap_{\mathbb{P} \in \mathcal{P}} \mathcal{N}^{\mathbb{P}}$. We say that a property holds $\mathcal{P}$-q.s. if it holds outside a $\mathcal{P}$-polar set.

2.2.2 Notions of Arbitrage

One of the most important underlying concepts in financial mathematics is the absence of arbitrage. In the literature on robust pricing and hedging many notions of arbitrage have been proposed to date. We present these here together in a unified manned and discuss their relative dependencies. To complement the picture, we establish some novel technical results. These are postponed to Section 2.3.2.

Definition 2.2.3. *Fix a filtration $\mathbb{F}$, a set $\mathcal{P}$, a set $\mathcal{S}$ of subsets of X and a set Ω. Recall that semistatic admissible trading strategies are given by $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$.*

- ***1pA***(Ω) *A One-Point Arbitrage (see Riedel [2015]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ such that $h \cdot \Phi + H \circ S_T \geq 0$ on Ω with strict inequality for some $\omega \in \Omega$.*

- ***OA***(Ω) *An Open Arbitrage (see Riedel [2015]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ such that $h \cdot \Phi + H \circ S_T \geq 0$ on Ω with strict inequality for some open subset of Ω.*

- ***SA***(Ω) *A Strong Arbitrage (see Acciaio et al. [2013]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ such that $h \cdot \Phi + H \circ S_T > 0$ on Ω.*

- ***USA***(Ω) *A Uniformly Strong Arbitrage (see Davis and Hobson [2007]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ such that $h \cdot \Phi + H \circ S_T \geq \varepsilon$ on Ω for some $\varepsilon > 0$.*

- ***A***$(\mathcal{P})$ *A $\mathcal{P}$-quasi-sure Arbitrage (see Bouchard and Nutz [2015]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ such that $h \cdot \Phi + H \circ S_T \geq 0$ holds $\mathcal{P}$-q.s. and $\mathbb{P}(h \cdot \Phi + H \circ S_T > 0) > 0$ for some $\mathbb{P} \in \mathcal{P}$. If $\mathcal{P} = \{\mathbb{P}\}$ a $\mathcal{P}$-quasi-sure Arbitrage is called a $\mathbb{P}$-arbitrage and is denoted **A**$(\mathbb{P})$.*

- ***CA***$(\mathcal{P})$ *A Classical Arbitrage in $\mathcal{P}$ (see Davis and Hobson [2007]) is a family of strategies $(h^\mathbb{P}, H^\mathbb{P})_{\mathbb{P} \in \mathcal{P}}$ such that, for all $\mathbb{P} \in \mathcal{P}$, $(h^\mathbb{P}, H^\mathbb{P})$ is a $\mathbb{P}$-arbitrage.*

- ***WA***$(\mathcal{P})$ *A Weak Arbitrage (see Blanchard and Carassus [2019]) is a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ which is a $\mathbb{P}$-arbitrage for some $\mathbb{P} \in \mathcal{P}$.*

- ***IntA***$(\mathcal{P})$ *An Interior Arbitrage (see Bayraktar et al. [2014]) is a sequence of strategies $(h^n, H^n) \in \mathcal{H}_\Phi(\mathbb{F})$ such that (h^n, H^n) is a $\mathcal{P}$-quasi-sure Arbitrage relative to option payoffs given by $\Phi + sign(h^n)/n$ for all n large enough.*

- ***WFLVR***(Ω) *A Weak Free Lunch With Vanishing Risk (see Cox and Obłój [2011], Cox et al. [2016]) is a sequence of strategies $(h^n, H^n) \in \mathcal{H}_\Phi(\mathbb{F})$ such that there* *exists a constant* $c \geq 0$ *and* $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ *with* $h^n \cdot \Phi + H^n \circ S_T \geq h \cdot \Phi + H \circ S_T - c$ *on* Ω *for all* $n \in \mathbb{N}$ *and*

$$\lim_{n \to \infty} (h^n \cdot \Phi + H^n \circ S_T) > 0 \quad \textit{on } \Omega.$$

***locA**($\mathcal{P}_t(\omega)$)* *A* (t, ω)*-local* $\mathcal{P}$*-quasi-sure Arbitrage (see Bartl [2019]) is a strategy* $H \in \mathbb{R}^d$ *such that* $H\Delta S_{t+1}(\omega) \geq 0$ $\mathcal{P}_t(\omega)$*-q.s. (where* $t \in \{0, \dots, T-1\}$ *and* $\omega \in X$*) and there exists* $\mathbb{P} \in \mathcal{P}_t(\omega)$ *such that* $\mathbb{P}(H\Delta S_{t+1} > 0) > 0$.

***A**($\mathcal{S}$)* *An Arbitrage de la Classe* $\mathcal{S}$ *(see Burzoni et al. [2016]) is a strategy* $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ *such that* $h \cdot \Phi + H \circ S_T \geq 0$ *on* Ω *and* $\{\omega \in \Omega \mid h \cdot \Phi + H \circ S_T > 0\} \supseteq \Gamma$ *for some* $\Gamma \in \mathcal{S}$.

When we want to stress the role of the filtration we include it as an argument, e.g., we write, e.g., ***SA***$(\Omega, \mathbb{F})$*. When the filtration is not specified it is implicitly taken to be* $\mathbb{F}^{\mathcal{U}}$*. We use a prefix* ***N*** *to indicate a negation of any of the above notions, e.g., we say that "**NA**($\mathcal{P}$) holds" when there does not exist a* $\mathcal{P}$*-quasi-sure arbitrage strategy, likewise* ***NUSA***(Ω) *denotes the absence of a uniformly strong arbitrage on* Ω*, etc.*

Lemma 2.2.4. *The following relations hold:*

1. $\mathbf{USA}(\Omega) \Rightarrow \mathbf{SA}(\Omega) \Rightarrow \mathbf{OA}(\Omega) \Rightarrow \mathbf{1pA}(\Omega)$.

2. $\mathbf{SA}(\Omega) \Rightarrow \mathbf{WFLVR}(\boldsymbol{\Omega})$.

3. $\mathbf{A}(\mathcal{P}) \Rightarrow \mathbf{A}(\mathbb{P})$ *for some* $\mathbb{P} \in \mathcal{P} \Leftrightarrow \mathbf{WA}(\mathcal{P})$.

4. $\mathbf{WA}(\mathcal{P}) \Leftarrow \mathbf{CA}(\mathcal{P}) \Leftrightarrow \mathbf{A}(\mathbb{P})$ *for all* $\mathbb{P} \in \mathcal{P}$.

5. $\mathbf{A}(\mathcal{P}) \Rightarrow \mathbf{IntA}(\mathcal{P})$.

6. *when* $\Phi = 0$ *then*
$\mathbf{A}(\mathcal{P}) \Leftrightarrow \mathbb{P}\left(\bigcup_{t=0}^{T-1} \{\omega \in X_t \mid \mathbf{locA}(\mathcal{P}_t(\omega)) \textit{ holds }\}\right) > 0$ *for some* $\mathbb{P} \in \mathcal{P}$.

Proof. Items *(1)-(4)* are immediate. Assertion *(6)* follows from [Bouchard and Nutz, 2015, Lemma 4.6, p.842]. For a strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F})$ satisfying

$$h \cdot \Phi + H \circ S_T \geq 0 \quad \mathcal{P}\text{-q.s.}$$

we have, for any $\varepsilon > 0$,

$$h \cdot (\Phi + \text{sign}(h)\varepsilon) + H \circ S_T = h \cdot \Phi + H \circ S_T + |h|_1\varepsilon \geq |h|_1\varepsilon \geq 0,$$

where $|h|_1 = \sum_{\lambda\in\Lambda} |h_\lambda|$. Absence of $\mathbf{IntA}(\mathcal{P})$ implies that there exists $\varepsilon > 0$ such that for any strategy as above we have

$$h \cdot \Phi + H \circ S_T = -|h|_1\varepsilon \quad \mathcal{P}\text{-q.s.},$$

so that $h = 0$ and absence of $\mathbf{A}(\mathcal{P})$ follows so *(5)* holds. □

$\mathbf{USA}(X)$ was first discussed in Davis and Hobson [2007], see also Cox and Obłój [2011] and Cox et al. [2016] for a definition of $\mathbf{USA}(\Omega)$ and $\mathbf{WFLVR}(\Omega)$, where $\Omega \subseteq X$. Note that if we take $(h, H) = (0, 0)$ in the definition of $\mathbf{WFLVR}(\Omega)$ and replace the pathwise inequalities by their $\mathbb{P}$-a.s. counterparts for some fixed $\mathbb{P} \in \mathfrak{P}(X)$, we recover a discrete version of the **NFLVR** condition of Delbaen and Schachermayer [1994].
$\mathbf{SA}((\mathbb{R}^d_+)^T)$ was used in Acciaio et al. [2013] in the canonical setup and $d = 1$. We refer to [Burzoni et al., 2019, Theorem 3] for a general FTAP connecting the notion of Strong and Uniformly Strong Arbitrage under the condition that there exists an option with a strictly convex super-linear payoff in the market. See also Bartl et al. [2017] for an equivalence result under marginal constraints. In Section 2.3.2 we discuss the connection between $\mathbf{SA}(\Omega)$ and $\mathbf{USA}(\Omega)$ without the above assumptions.
$\mathbf{A}(\mathcal{S})$ is a unifying concept since $\mathbf{1pA}(\Omega)$, $\mathbf{OA}(\Omega)$, $\mathbf{SA}(\Omega)$, $\mathbf{USA}(\Omega)$ and $\mathbf{A}(\mathcal{P})$ can all be seen as special cases of $\mathbf{A}(\mathcal{S})$, see [Burzoni et al., 2016, Section 4.6] for a detailed discussion. It was first defined in Burzoni et al. [2016] in a pathwise setting, see in particular the pathwise Fundamental Theorem of Asset pricing in [Burzoni et al., 2016, Theorem 2 & Section 4]. This extends the results obtained in Riedel [2015] who introduced $\mathbf{1pA}(\Omega)$ and $\mathbf{OA}(\Omega)$. $\mathbf{OA}(\Omega)$ is furthermore defined in the setup of Dolinsky and Soner [2014].
$\mathbf{A}(\mathcal{P})$ was introduced in the quasi-sure setting of Bouchard and Nutz [2015], where they prove a quasi-sure Fundamental Theorem of Asset pricing and Superhedging Theorem. From Lemma 2.2.4 above we see that the crucial distinction between $\mathbf{CA}(\mathcal{P})$ and $\mathbf{A}(\mathcal{P})$ is the aggregation of arbitrage strategies, which poses a fundamental technical difficulty overcome in Bouchard and Nutz [2015] by the specific (APS) structure of $\mathcal{P}$. We also note that $\mathbf{CA}(\mathcal{P})$ was actually referred to as *weak arbitrage* in Davis and Hobson [2007].
The notion of interior arbitrage $\mathbf{IntA}(\mathcal{P})$ was introduced, and called a *robust arbitrage*, by Bayraktar et al. [2014] in the context of transaction costs. Absence of $\mathbf{IntA}(\mathcal{P})$ is equivalent to absence of $\mathbf{A}(\mathcal{P})$ not only at the current prices of statically traded options Φ but also under all, sufficiently small, perturbations of their prices. This notion was also used in [Hou

and Obłój, 2018, Assumption 3.1]. It is equivalent to saying that the prices of the options Φ are strictly inside the region of their $\mathcal{P}$-q.s. no-arbitrage prices, thus avoiding the delicate issue of boundary classification. In general, $\mathbf{IntA}(\mathcal{P})$ does not imply $\mathbf{A}(\mathcal{P})$. To see this, take $\Phi = \{(S_T - K)^+\}$ for some $K > S_0$ and $\emptyset \neq \mathcal{P} \subseteq \{\mathbb{P} \in \mathfrak{P}(X) \mid \mathbb{P}(S_T \leq K) = 1\}$. Then there is no $\mathcal{P}$-q.s. arbitrage, while for every $\varepsilon > 0$ we have $(S_T - K)^+ + \varepsilon \geq \varepsilon > 0$ and thus $\mathbf{IntA}(\mathcal{P})$ holds.

Throughout the remainder of this chapter, unless otherwise stated, we take $\Lambda = \{1, \ldots, k\}$, i.e., we have a finite Φ with k statically traded options.

2.2.3 Robust Fundamental Theorem of Asset Pricing

The first Fundamental Theorem of Asset Pricing characterises absence of arbitrage in terms of existence of martingale (pricing) measures. In the classical discrete-time setting, this refers to the notion of $\mathbb{P}$-arbitrage. However, in a robust setting, there are many possible notions of arbitrage one can consider. If we adopt a strong notion of arbitrage, its absence should be equivalent to a weak statement, e.g., $\mathcal{M}_{\Omega,\Phi} \neq \emptyset$. This is often done in the pathwise literature, see Burzoni et al. [2019], and leads to a robust (multi-prior) version of the familiar Dalang-Morton-Willinger theorem.

Theorem 2.2.5 (Robust DMW Theorem)**.** *Let $\mathcal{P}$ be a set of probability measures satisfying (APS). Then there exists a universally measurable set of scenarios Ω with $\mathbb{P}(\Omega) = 1$ for all $\mathbb{P} \in \mathcal{P}$ and a filtration $\tilde{\mathbb{F}}$ with $\mathbb{F}^0 \subseteq \tilde{\mathbb{F}} \subseteq \mathbb{F}^{\mathcal{M}}$, such that the following are equivalent:*

(1) $\mathcal{Q}_{\mathcal{P},\Phi} \neq \emptyset$.

(2) $\mathbb{P}(\Omega^*_\Phi) > 0$ *for some* $\mathbb{P} \in \mathcal{P}$.

(3) $\mathcal{M}_{\Omega,\Phi} \neq \emptyset$.

(4) $\Omega^*_\Phi \neq \emptyset$.

(5) ***NSA***$(\Omega, \tilde{\mathbb{F}})$ *holds.*

Conversely, for an analytic set Ω there exists a set $\mathcal{P}$ satisfying (APS) such that for all $\omega \in \Omega$ there exists $\mathbb{P} \in \mathcal{P}$ with $\mathbb{P}(\{\omega\}) > 0$ and such that (1)-(5) are equivalent.

The above result follows from Theorem 2.2.7 below by setting $\mathcal{S} = \{\Omega\}$. To see its opposite twin we should adopt a weak notion of arbitrage, its absence thus being equivalent to a strong statement, e.g., for all $\mathbb{P} \in \mathcal{P}$ there exists $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$ such that $\mathbb{P} \ll \mathbb{Q}$. This route is most often taken in the quasi-sure literature, see Bouchard and Nutz [2015], and leads to the following version of the robust FTAP.

Theorem 2.2.6. *Let $\mathcal{P}$ be a set of probability measures satisfying (APS). Then there exists an analytic set of scenarios Ω with $\mathbb{P}(\Omega) = 1$ for all $\mathbb{P} \in \mathcal{P}$, such that the following are equivalent:*

(1) $\boldsymbol{N1pA}(\Omega^*_\Phi)$ *holds and* $\Omega = \Omega^*_\Phi$ *$\mathcal{P}$-q.s.*

(2) For all $\mathbb{P} \in \mathcal{P}$ there exists $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$ such that $\mathbb{P} \ll \mathbb{Q}$.

(3) $\boldsymbol{NA}(\mathcal{P}, \mathbb{F}^{\mathcal{U}})$ *holds.*

Conversely, if Ω is an analytic set, then there exists a set $\mathcal{P}$ of probability measures satisfying (APS) such that for all $\omega \in \Omega$ there exists $\mathbb{P} \in \mathcal{P}$ with $\mathbb{P}(\{\omega\}) > 0$ and such that the following are equivalent:

(1) $\boldsymbol{N1pA}(\Omega)$ *holds and* $\Omega = \Omega^*_\Phi$.

(2) For all $\mathbb{P} \in \mathcal{P}$ there exists $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$ such that $\mathbb{P} \ll \mathbb{Q}$.

(3) $\boldsymbol{NA}(\mathcal{P}, \mathbb{F}^{\mathcal{U}})$ *holds.*

Our proof of this theorem, given in Section 2.4.1, does not rely on the proof of (3) $\Rightarrow$ (2) given in Bouchard and Nutz [2015]. Instead we give pathwise arguments. In particular, given $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P}(\Omega \setminus \Omega^*_\Phi) > 0$ we explicitly construct a quasi-sure Arbitrage strategy using the Universal Arbitrage Aggregator of Burzoni et al. [2019]. This strengthens the results of Theorem 2.2.7 below. Indeed, using the fact that $\mathcal{P}$ satisfies (APS), it is possible to select $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$ for each $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P} \ll \mathbb{Q}$. Necessarily the support of each $\mathbb{P}$ is then concentrated on Ω^*_Φ.

Finally, we give our main abstract result, which establishes a pathwise and probabilistic characterisation of the absence of Arbitrage de la Classe $\mathcal{S}$. Its proof is presented in Section 2.4.1. As noted above, Arbitrage de la Classe $\mathcal{S}$ allows to consider many notions of arbitrage at once. Accordingly, the main result below implies Theorem 2.2.5 and can be strengthened to imply Theorem 2.2.6 as will be seen in Section 5.6.

Theorem 2.2.7. *Assume that $\mathcal{P}$ satisfies (APS) and $\mathcal{S} \subseteq \mathcal{B}(X)$ is such that*

$$\exists \{C_n\}_{n\in\mathbb{N}} \subseteq \mathcal{S} \text{ s.t. } \forall C \in \mathcal{S}\ \exists \{n_k\}_{k\in\mathbb{N}} \subseteq \mathbb{N} \text{ with } \mathbb{1}_{C_{n_k}} \uparrow \mathbb{1}_C\ (k \to \infty). \tag{2.2.1}$$

Then there exists a co-analytic set of scenarios Ω such that $\mathbb{P}(\Omega) = 1$ for all $\mathbb{P} \in \mathcal{P}$ and a filtration $\tilde{\mathbb{F}}$ with $\mathbb{F}^0 \subseteq \tilde{\mathbb{F}} \subseteq \mathbb{F}^{\mathcal{M}}$, such that the following are equivalent:

(1) For all $C \in \mathcal{S}$ with $C \subseteq \Omega$ there exists $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$ such that $\mathbb{Q}(C) > 0$.

(2) For all $C \in \mathcal{S}$ with $C \subseteq \Omega$ there exists $\mathbb{P} \in \mathcal{P}$ with $\mathbb{P}(\Omega^*_\Phi \cap C) > 0$.

(3) For all $C \in \mathcal{S}$ with $C \subseteq \Omega$ there exists $\mathbb{Q} \in \mathcal{M}_{\Omega,\Phi}$ such that $\mathbb{Q}(C) > 0$.

(4) $\{C \in \mathcal{B}(X) \mid C \subseteq \Omega \setminus \Omega^*_\Phi\} \cap \mathcal{S} = \emptyset$.

(5) There is no Arbitrage de la Classe $\mathcal{S}$ in $\mathcal{A}_\Phi(\tilde{\mathbb{F}})$ on Ω.

Conversely, for an analytic set Ω there exists a set $\mathcal{P}$ satisfying (APS), such that for all $\omega \in \Omega$ there exists $\mathbb{P} \in \mathcal{P}$ with $\mathbb{P}(\{\omega\}) > 0$ and such that (1)-(5) are equivalent.

Remark 2.2.8. Condition (2.2.1) was first stated in [Burzoni et al., 2016, Cor. 4.30 and the discussion thereafter]. It turns out that for the proof of Theorem 2.7 a weaker condition is sufficient: we only need the properties

$$\bigcup\{C \in \mathcal{S} \mid C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}\} \in \mathcal{B}(X) \tag{2.2.2}$$

and

$$\bigcup\{C \in \mathcal{S} \mid C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}\} \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}} \tag{2.2.3}$$

to hold, where we refer to Section 2.4.1 for a formal definition of $\Omega^{\mathcal{P}}$. Conditions (2.2.2) and (2.2.3) are compatibility conditions on Ω, $\mathcal{S}, \Phi$ and $\mathcal{P}$. Indeed, they assert that the (likely uncountable) union of "inefficient" subsets of $\Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^*_\Phi$ (modulo $\mathcal{P}$-polar sets), stays an "inefficient" subset (modulo $\mathcal{P}$-polar sets). If this condition is not satisfied for some arbitrary $\mathcal{P}$ and $\mathcal{S}$, then there is no reason why a set Ω for which *(2)* holds should exist. Take for example a collection $\mathcal{P}$ having densities and $\mathcal{S}$ a set of singletons in X. Then $\mathbb{P}(C) = 0$ for any $\mathbb{P} \in \mathcal{P}$ and any $C \in \mathcal{S}$ so the only Ω which could satisfy *(2)* is the empty set. We note that when $\mathcal{S} = \{C \subseteq X \mid C \text{ open}\}$ then (2.2.2) is always satisfied and (2.2.3) is satisfied as soon as X is separable. However, in general, conditions (2.2.2) and (2.2.3) may be hard to verify, which is why we provide (2.2.1) as an easier sufficient condition. Lastly, we remark that it is not straightforward to show that $\mathcal{S} = \{C \mid \mathbb{P}(C) > 0 \text{ for some } \mathbb{P} \in \mathcal{P}\}$ corresponding to $\mathbf{NA}(\mathcal{P})$ satisfies (2.2.3), which is why we give a direct proof of Theorem 2.2.6 in Section 5.6.

We note that the set Ω can in general not be assumed to be analytic. The implications *(1) ⇒ (2) ⇒ (3) ⇒ (4) ⇒ (5)* follow directly from the definitions. Apart from measurability considerations regarding Ω, the equivalence of *(3)*, *(4)* and *(5)* essentially follows from Burzoni et al. [2016]. Furthermore, given an analytic set Ω, we will simply define $\mathcal{P}$ as all the finitely supported probability measures on Ω. The analyticity of Ω then implies (APS) of $\mathcal{P}$. We then also have $\mathcal{Q}_{\mathcal{P},\Phi} = \mathcal{M}^f_{\Omega,\Phi}$ and equivalence of *(1)* and *(3)-(5)* follows from Burzoni et al. [2019]. In this context, the essential connection we make is the combination of pathwise and quasi-sure criteria as stated in *(2)*: for every $C \in \mathcal{S}$, the pathwise efficient subset $\Omega^*_\Phi \cap C$ is required to be "seen" by at least one measure $\mathbb{P}$ in the set $\mathcal{P}$.

For a given $\mathcal{P}$, the set Ω in Theorem 2.2.7 can be explicitly constructed as the concatenation of the quasi-sure supports of $\mathcal{P}_t \circ \Delta(S_{t+1})^{-1}$. The main difficulty of the proof is then to show the implication *(5)* $\Rightarrow$ *(1)*, where one needs to establish existence of martingale measures $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}$, which are compatible with Ω and $\mathcal{S}$ in the sense of *(1)*. This, modulo measurable selection arguments, is achieved by finding an element $\mathbb{P} \in \mathcal{P}_t(\omega)$ such that zero is in the relative interior of the support of $\mathbb{P} \circ \Delta(S_{t+1})^{-1}$. Indeed, let us explain the main idea of the proof of *(5)* $\Rightarrow$ *(1)* based on the following example: assume $T = 1$, $d = 3$, $\Phi = 0$ and the set Ω^* is given by the grey polyhydron in Figure 2.2. Assume that the support of $\mathbb{P} \circ \Delta(S_1)^{-1}$ for a given measure $\mathbb{P} \in \mathcal{P}_0$ is given by the blue dot (see Figure 2.2). Then as $0 \in \mathrm{ri}(\Omega^*)$, we can find three additional points in Ω^*, such that zero is in the relative interior of the convex hull of the four points. By definition of Ω, the three additional points are in the support of some measures in $\mathcal{P}_0$, which we call $\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3$ in $\mathcal{P}_0$. As $\mathcal{P}_0$ is convex, it follows that

$$\tilde{\mathbb{P}} := \frac{\mathbb{P}_1 + \mathbb{P}_2 + \mathbb{P}_3 + \mathbb{P}}{4}$$

is an element of $\mathcal{P}_0$ as well, as visualised in Figure 2.3. Since zero is in the relative interior of the support of $\tilde{\mathbb{P}}$, one can now use results from Rokhlin [2008] to find a martingale measure $\mathbb{Q} \sim \tilde{\mathbb{P}}$, in particular $\mathbb{P} \ll \mathbb{Q}$.

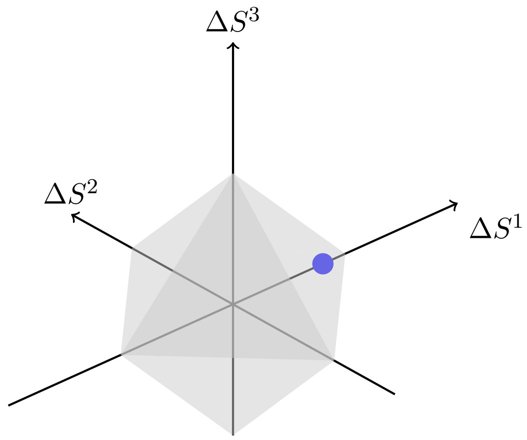

Figure 2.2: Construction of a martingale measure $\mathbb{Q} \gg \mathbb{P}$ for $d = 3$: the set Ω (grey) and $\mathrm{supp}(\mathbb{P} \circ (\Delta S_1)^{-1})$ (blue)

Note that this argument fundamentally relies on the convexity of $\mathcal{P}_t$. The analytic product structure assumption then grants suitable measurability for the concatenation procedure in the multiperiod case.

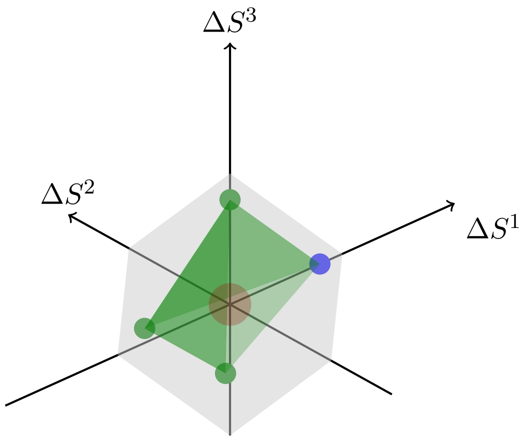

Figure 2.3: Construction of a martingale measure $\mathbb{Q} \gg \mathbb{P}$ for $d = 3$: finding a measure $\tilde{\mathbb{P}}$ such that $\mathrm{NA}(\tilde{\mathbb{P}})$ and $\tilde{\mathbb{P}} \gg \mathbb{P}$ holds.

2.2.4 Robust Superhedging Theorem

In this section we focus on the key result which characterises superhedging prices: the pricing-hedging duality, or the Superhedging Theorem. As before, we compare pathwise and quasi-sure superhedging approaches as extensions of the classical model-specific result, see [Föllmer and Schied, 2004, Chapter 5, Theorem 5.30].

For a set $\Omega \subseteq X$ we denote the pathwise superhedging price on Ω by

$$\pi_\Omega(g) := \inf\{x \in \mathbb{R} \mid \exists (h, H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}}) \text{ s.t. } x + h \cdot \Phi + (H \circ S_T) \geq g \text{ on } \Omega\}$$

and denote the $\mathcal{P}$-q.s. superhedging price by

$$\pi^{\mathcal{P}}(g) := \inf\{x \in \mathbb{R} \mid \exists (h, H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}}) \text{ s.t. } x + h \cdot \Phi + (H \circ S_T) \geq g \ \mathcal{P}\text{-q.s.}\}.$$

Take an analytic set Ω such that for all $\mathbb{P} \in \mathcal{P}$ we have $\mathbb{P}(\Omega^*_\Phi) = 1$. Using the Superhedging Theorems of Bouchard and Nutz [2015] and Burzoni et al. [2019] it is immediate that the following relationships hold for all upper semianalytic g:

$$\sup_{\mathbb{Q} \in \mathcal{M}_{\Omega,\Phi}} \mathbb{E}_\mathbb{Q}[g] = \pi_{\Omega^*_\Phi}(g) \geq \pi^{\mathcal{P}}(g) = \sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_\mathbb{Q}[g].$$

The above inequality is strict in general. An easy way to see this is to take $d = T = S_0 = 1$, $\Phi = 0$, $g(S_1) = \mathbb{1}_{\{S_1 = 0\}}$ and $\mathcal{P} = \{\frac{1}{2}\lambda|_{[0,2]}\}$, where $\lambda|_{[0,2]}$ denotes the Lebesgue measure on $[0, 2]$. Then $\Omega = \Omega^* = [0, 2]$ and the pathwise superhedging price is equal to $1/2$, while the quasi-sure superhedging price is equal to zero. In fact, to link the super-hedging and pathwise formulations, we have to choose a specific set $\Omega^{\mathcal{P}}_g$ which depends not only on $\mathcal{P}$ but also on g. We determine this set $\Omega^{\mathcal{P}}_g$ by reducing to superhedging under a fixed measure $\mathbb{P}^g$ as stated in the following theorem:

Theorem 2.2.9. *Let $\mathcal{P}$ be a set of probability measures satisfying (APS). Assume $\boldsymbol{NA}(\mathcal{P})$ holds and let $g : X \to \mathbb{R}$ be upper semianalytic. Then there exists a measure $\mathbb{P}^g = \mathbb{P}^g_0 \otimes \cdots \otimes \mathbb{P}^g_{T-1}$ and an $\mathcal{F}^{\mathcal{U}}$-measurable set $\Omega^{\mathcal{P}}_g$ with $\mathbb{P}(\Omega^{\mathcal{P}}_g) = 1$ for all $\mathbb{P} \in \mathcal{P}$, such that*

$$\sup_{\mathbb{Q} \in \mathcal{M}_{\Omega^{\mathcal{P}}_g,\Phi}} \mathbb{E}_{\mathbb{Q}}[g] = \pi_{(\Omega^{\mathcal{P}}_g)^*_\Phi}(g) = \pi^{\mathbb{P}^g}(g) = \pi^{\mathcal{P}}(g) = \sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g].$$

Conversely, let Ω be an analytic subset of X with $\Omega^*_\Phi \neq \emptyset$. For any set $\mathcal{P} \subseteq \mathfrak{P}(X)$, which satisfies (APS) and $\mathcal{N}^{\mathcal{P}} = \mathcal{N}^{\mathcal{M}^f_{\Omega,\Phi}}$, we have

$$\sup_{\mathbb{Q} \in \mathcal{M}_{\Omega,\Phi}} \mathbb{E}_{\mathbb{Q}}[g] = \pi_{\Omega^*_\Phi}(g) = \pi^{\mathcal{P}}(g) = \sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g].$$

In both cases, the value, if finite, is attained by a superhedging strategy $(h, H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}})$.

The proof of this result is postponed to Section 2.4.2. In particular, Theorem 2.2.9 lets us interpret robust superhedging prices $\pi^{\mathcal{P}}(g)$ as classical superreplication prices $\pi^{\mathbb{P}^g}(g)$ under an "extremal" measure $\mathbb{P}^g$. Determining such measures $\mathbb{P}^g$ is not straightforward in general. In a one-period case and for a continuous g, we can use the arguments in the proof of Lemma 2.4.10 to see that any measure $\mathbb{P}$ which attains the one-step quasisure support $\{\mathbb{P} \circ (\Delta S_T(\omega, \cdot))^{-1} \mid \mathbb{P} \in \mathcal{P}_{T-1}(\omega)\}$ can be chosen. To extend this result to the multiperiod-case, certain continuity properties of the maps $\omega \mapsto \mathcal{P}_t(\omega)$ have to be guaranteed: we refer to Proposition 3.3.7 for a sufficient condition.

2.3 Complementary results on superhedging and arbitrage

2.3.1 Extension of Pathwise Superhedging from Ω^* to Ω

The preceding results show that quasi-sure and pathwise superhedging are essentially equivalent. As $\mathcal{P}$-q.s. superhedging strategies might be difficult to compute and implement in practice, it might be preferable to work on a prediction set Ω using pathwise arguments. Given that determining Ω^* is computationally expensive as well, the quantity of interest is then the superhedging price on Ω and not on Ω^* seen in the duality results in Section 2.2.4. Thus, we would like to find sufficient conditions under which the superhedging strategy associated with $\pi_{\Omega^*_\Phi}(g)$ can be extended to Ω without any additional cost. The intuition is that as $\Omega \setminus \Omega^*$ describes non-efficient beliefs, we should be able to superhedge g on this set implementing an arbitrage strategy. It turns out that this intuition is not true in general. Indeed, we run into problems regarding measurability of these arbitrage strategies, which means that this procedure only works in special cases.

To simplify the analysis, throughout this section only, we assume that $\Phi = 0$ and $\omega \mapsto S_t(\omega)$ is continuous. The latter is satisfied, e.g., when $\omega \mapsto S_t(\omega)$ is the coordinate mapping, i.e., $S_t(\omega) = \omega_t$. In order to give some intuition and to identify necessary conditions for the sets Ω, Ω^* and the function g we first give two counterexamples:

Example 2.3.1. Let $d = 1$, $T = 1$ and $(\Omega, \mathcal{F}) = (\mathbb{R}_+ \setminus \{0\}, \mathcal{B}(\mathbb{R}_+ \setminus \{0\}))$. We set $S_0 = 2$ and $S_1(\omega) = 2 + \omega$. Then $\Omega^* = \emptyset$ and trivially

$$\begin{aligned} \pi_{\Omega^*}(1) =& \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ such that } x + H \circ S_T \geq 1 \text{ on } \Omega^*\} = -\infty, \\ \pi_{\Omega}(1) =& \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ such that } x + H \circ S_T \geq 1 \text{ on } \Omega\} = 1. \end{aligned}$$

Thus we have to assume that $\Omega^* \cap \Sigma_t^\omega \neq \emptyset$ in the remainder of this section. We also note that here $\mathbf{SA}(\Omega)$ holds whilst $\mathbf{USA}(\Omega)$ does not, see Section 2.3.2.

Example 2.3.2. Let $d = 2$, $T = 1$ and $\Omega = ((2, \infty) \times [0, \infty)) \cup (\{2\} \times [0, 7])$ and $\mathcal{F} = \mathcal{B}(\Omega)$. We set $S_0 = (2, 2)$ and $S_1(\omega) = \omega$. In particular, $\Delta S_1(\Omega)$ is not a closed set. Note that $\Omega^* = \{2\} \times [0, 7]$. Now, introduce the claim

$$g(S^1, S^2) = \Delta S_1^2 \mathbb{1}_{\{\Delta S_1^2 \leq 5\}} + 5(\Delta S_1^2 - 4)\mathbb{1}_{\{\Delta S_1^2 > 5\}}.$$

It is easy to see that $\pi_{\Omega^*}(g) = 0$ and any trading strategy $H_1 = (H_1^1, 1)$ with $H_1^1 \in \mathbb{R}$ is a superhedging strategy. We now claim that we cannot extend superhedging to $\Omega \setminus \Omega^*$ with initial capital zero. For this we show that even for initial capital one, there exist no superreplication strategies on Ω. Indeed for this we would need

$$1 + H_1^1 \Delta S_1^1 + H_1^2 \Delta S_1^2 \geq 5(\Delta S_1^2 - 4) \quad \text{on } \Delta S_1^1 > 0,\ \Delta S_1^2 > 5,$$

which is equivalent to

$$H_1^1 \geq \frac{(5 - H_1^2)\Delta S_1^2 - 21}{\Delta S_1^1}.$$

As $H_1^2 \in [4/5, 3/2]$, this means $H_1^1 \to \infty$ if ΔS_1^1 is arbitrarily close to 0 and ΔS_1^2 is sufficiently large. Even if we look at $\Omega = [2 + \varepsilon, \infty) \times [0, \infty) \cup \{2\} \times [0, 7]$ for some positive ε, then taking ΔS_1^2 arbitrarily large still leads to non-existence of superhedging strategies. In conclusion we will only consider compact sets $\Omega \cap \Sigma_t^\omega$ in the rest of this section, on which "arbitrage strategies are effective for superhedging" in a sense defined below. Furthermore, modifying the function g on $\Omega \setminus \Omega^*$ in the above example, we can easily construct situations, in which $\pi_\Omega(g) \neq \pi_{\Omega^*}(g)$ for discontinuous payoffs g. In conclusion we will also assume that g is continuous in this section.

We can also modify this example so that the $\Delta S_1(\Omega)$ is closed and there is no attainment of superreplication strategies for $\pi_\Omega(g)$. We stress that this is a fundamental difference to the case $\Omega = \Omega^*$, where attainment is always given (see Theorem 2.2.9). Namely, take $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x \in [2, \infty),\ 0 \leq y \leq 7 + \sqrt{x}\}$ with the other elements unchanged. Note that Ω^* did not change. Repeating the arguments above and looking at $\Delta S_1^1 = 1/n$ and $\Delta S_1^2 = 5 + 1/\sqrt{n}$ we find

$$H_1^1 \geq n\left(5 + \frac{5}{\sqrt{n}} - 5 - \frac{1}{\sqrt{n}}\right) = n\frac{4}{\sqrt{n}} \to \infty$$

for $n \to \infty$.

As we have seen in the examples above, in general it is necessary to assume that $\Omega^* \cap \Sigma_t^\omega \neq \emptyset$, $\omega \mapsto g(\omega)$ is continuous and that Ω is compact as well as "well suited for superhedging by arbitrage strategies". We first address the second point and show continuity of the one-step superhedging prices $\omega \mapsto \pi_{t,\Omega^*}(g)(\omega)$, which are defined via a dynamic programming approach:

Definition 2.3.3. *For a Borel-measurable $g : X \to \mathbb{R}$ we define the one-step superhedging prices*

$$\pi_{T,\Omega^*}(g)(\omega) := g(\omega),$$
$$\pi_{t,\Omega^*}(g)(\omega) := \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \ s.t. \ x + H\Delta S_{t+1}(v) \geq \pi_{t+1,\Omega^*}(g)(v) \ \forall v \in \Sigma_t^\omega \cap \Omega^*\},$$

where $0 \leq t \leq T-1$.

A sufficient condition for continuity of $\omega \mapsto \pi_{t,\Omega^*}(g)(\omega)$ is identified in Chapter 3 and relies on the following assumption:

Assumption 2.3.4. *The sets $\Sigma_t^\omega \cap \Omega^* \neq \emptyset$ and the sets $\Sigma_t^\omega \cap \Omega$ are compact for all $\omega \in \Omega$ and all $0 \leq t \leq T-1$. Furthermore, for all $0 \leq t \leq T-1$, the correspondence $\omega \twoheadrightarrow S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$ is uniformly continuous from (Ω, d_t^S) to the subsets of $\mathbb{R}^d$ endowed with the Hausdorff distance, and where $d_t^S(\omega, \tilde{\omega}) := \max_{s=0,\dots,t} |S_s(\omega) - S_s(\tilde{\omega})|$.*

We refer to Chapter 3 for a discussion and examples of sets Ω satisfying Assumption 2.3.4. The following lemma now follows from a direct application of Propsition 3.3.7.

Lemma 2.3.5. *Let $\omega \mapsto g(\omega)$ be continuous. Under Assumption 2.3.4 the one-step superhedging prices $\omega \mapsto \pi_{t,\Omega^*}(g)(\omega)$ are continuous for all $0 \leq t \leq T-1$.*

Secondly, Example 2.3.2 also shows, that it is important to identify the subset of $\Sigma_t^\omega \cap \Omega$, on which "arbitrage strategies are ineffective for superhedging" in the following sense:

Definition 2.3.6. *We denote by $proj_{\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)}(\Delta S_{t+1}(v))$ the orthogonal projection of $\Delta S_{t+1}(v)$ onto the linear subspace spanned by $\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$ and define the set A_t^ω as the collection of all $v \in \Sigma_t^\omega \cap \Omega$, for which $proj_{\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)}(\Delta S_{t+1}(v))$ is not an element of $\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$.*

For an illustration of the set A_t^ω see Figure 2.4. We now state an assumption ensuring compatibility of A_t^ω and $\Sigma_t^\omega \cap \Omega^*$:

Assumption 2.3.7. *For each level set the following is true: if a sequence of points* $(v_n) \subseteq A_t^\omega$ *converges to a point* $v \in span(\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*))$*, then necessarily* $v \in \Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$*.*

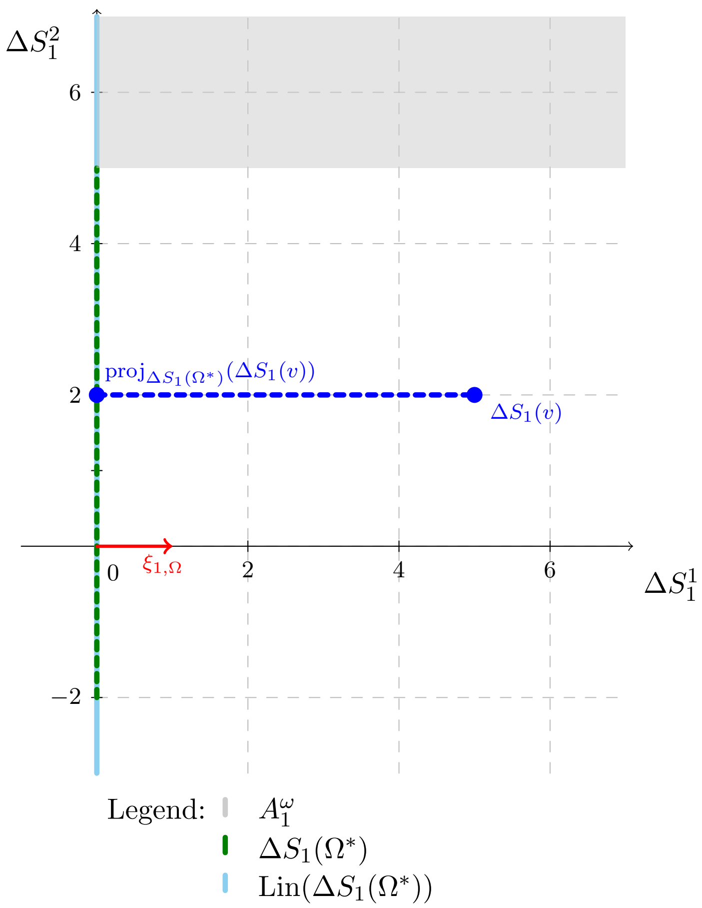

Figure 2.4: Example 2.3.2 with notation from Definition 2.3.6.

Alas, it turns out that while Assumptions 2.3.4 and 2.3.7 are sufficient to establish the equality $\pi_\Omega(g) = \pi_{\Omega^*}(g)$ for $d = 2$, it is not so for $d > 2$. It is linked with the notion of standard separators introduced in Burzoni et al. [2019], which are measurable selectors of pointwise arbitrage strategies. We refer the reader to [Burzoni et al., 2019, Proof of Lemma 1] and the discussion therein for a detailed definition. Here we formulate an example, in which the existence of two standard separators together with the measurability constraint on H_1 implies $\pi_\Omega(g) > \pi_{\Omega^*}(g)$:

Example 2.3.8. Let $d = 3$, $T = 1$ and $(\Omega, \mathcal{F}) = (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. We set $(S_0^1, S_0^2, S_0^3) = (2, 2, 2)$ and

$$S_1^1(\omega) = \begin{cases} 2 & \text{if } \omega \in \mathbb{R}_+ \setminus \mathbb{Q}, \\ 2.5 - \omega & \text{if } \omega \in \mathbb{Q} \cap [1/2, \infty), \\ 4 & \text{if } \omega \in \mathbb{Q} \cap [0, 1/2), \end{cases}$$

$$S_1^2(\omega) = \begin{cases} \omega & \text{if } \omega \in \mathbb{R}_+ \setminus \mathbb{Q}, \\ 0 & \text{if } \omega \in \mathbb{Q} \cap [1/2, \infty), \\ 2 & \text{if } \omega \in \mathbb{Q} \cap [0, 1/2), \end{cases} \quad S_1^3(\omega) = \begin{cases} 2 & \text{if } \omega \in \mathbb{R}_+ \setminus \mathbb{Q}, \\ 2 & \text{if } \omega \in \mathbb{Q} \cap [1/2, \infty), \\ 2 + \omega & \text{if } \omega \in \mathbb{Q} \cap [0, 1/2). \end{cases}$$

Then $\Omega^* = \mathbb{R}_+ \setminus \mathbb{Q}$ and using the notation of [Burzoni et al., 2019, Proof of Lemma 1] the standard separators are given by $\xi_{0,A_0} = (0, 0, 1)$ and $\xi_{0,A_1} = (-1, 0, 0)$. Next we define

$$g(S^1, S^2, S^3) = \Delta S_1^2 + |\Delta S_1^1| + |\Delta S_1^3|$$

We note that for $\Delta S_1^1 = \Delta S_1^3 = 0$ we have $g(S) = \Delta S_1^2$. So in particular to hedge g on Ω^* we need initial capital $\pi_{\Omega^*}(g) = 0$ and any hedging strategy satisfies $H_1 = (H_1^1, 1, H_1^3)$ for $H_1^1, H_1^3 \in \mathbb{R}$. For any such strategy to also superhedge on $\mathbb{Q} \cap [1/2, \infty)$ with initial capital $1/2$, H_1^1 has to satisfy in particular

$$1/2 + H_1^1 \Delta S_1^1 - 2H_1^2 \geq 0 \qquad \text{for } \Delta S_1^1 \leq -2,$$

so $H_1^1 \leq -3/4$ as $H_1^2 \in [1, 5/4]$. Lastly extending superhedging g on $\mathbb{Q} \cap [0, 1/2)$ gives the constraint

$$1/2 + 2H_1^1 + H_1^3 \Delta S_1^3 \geq 2.$$

Taking $\Delta S_1^3 = 0$ gives $H_1^1 \geq 3/4$, a contradiction. Thus we will assume that all one-point arbitrages can be reduced to a single standard separator.

Note that Example 2.3.8 can be easily altered to make $\Delta S_{t+1}(\Omega \cap \Sigma_t^\omega)$ compact by adding additional points. For clarity of exposition we have refrained from doing this but we conclude that Assumptions 2.3.4 and 2.3.7 are not sufficient for $d > 2$. We have to add a last assumption, which guarantees measurability of the corresponding Universal Arbitrage Aggregator. Intuitively it states, that locally, i.e., for every $0 \leq t \leq T-1$ and every $\omega \in \Omega$, there exists at most one arbitragable direction of the evolution of assets S, so that the first standard separator is already the Universal Arbitrage Aggregator:

Assumption 2.3.9. *For all $\omega \in \Omega$ and $0 \leq t \leq T-1$ we have $\xi_{t+1, \Omega \cap \Sigma_t^\omega} = H_t^*$, where H^* is the Universal Arbitrage Aggregator of Burzoni et al. [2019] for the set $\Sigma_t^\omega \cap \Omega$.*

Theorem 2.3.10. *Suppose that $\Phi = 0$ and $X \ni \omega \mapsto S_t(\omega)$ is continuous for all $1 \leq t \leq T$. For an analytic $\Omega \subseteq X$ satisfying Assumptions 2.3.4, 2.3.7 and 2.3.9 the Superhedging*

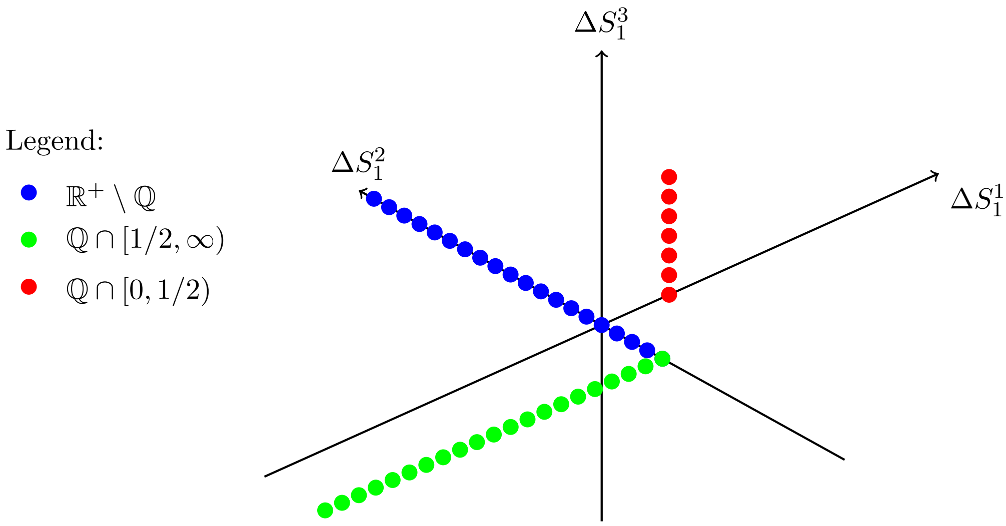

Figure 2.5: $\Omega \subseteq \mathbb{R}^3$ in Example 2.3.8

Duality of Burzoni et al. [2019] extends from Ω^* to Ω for all continuous $g : X \to \mathbb{R}$, i.e., we have

$$\begin{aligned}\sup_{\mathbb{Q}\in\mathcal{M}^f_\Omega} \mathbb{E}_\mathbb{Q}(g) &= \inf\{x \in \mathbb{R} \mid \exists H \in \mathcal{H}(\mathbb{F}^\mathcal{U}) \textit{ such that } x + H \circ S_T \geq g \textit{ on } \Omega^*\} \\ &= \inf\{x \in \mathbb{R} \mid \exists H \in \mathcal{H}(\mathbb{F}^\mathcal{U}) \textit{ such that } x + H \circ S_T \geq g \textit{ on } \Omega\}\end{aligned}$$

Proof. As before, we prove the claim by backward induction over $t = 0, \dots, T-1$. Let us now fix $\omega \in \Omega$. We assume $\Sigma_t^\omega \cap \Omega^* \neq \emptyset$ and $(\Sigma_t^\omega \cap \Omega) \setminus (\Sigma_t^\omega \cap \Omega^*) \neq \emptyset$, otherwise the claim is trivial. We first look at the case, where $\text{proj}_{\Delta S_{t+1}(\Sigma_t^\omega\cap\Omega^*)}(\Delta S_{t+1}(v))$ is an element of $\Delta S_{t+1}(\Sigma_{t+1}^\omega \cap \Omega^*)$, i.e., there exists $v' \in \Sigma_t^\omega \cap \Omega^*$ such that $\text{proj}_{\Delta S_{t+1}(\Sigma_t^\omega\cap\Omega^*)}(\Delta S_{t+1}(v)) = \Delta S_{t+1}(v')$. Note that by construction of Ω^* the standard separator $\xi_{t+1,\Omega\cap\Sigma_t^\omega}$ is orthogonal to $\text{span}(\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*))$. By definition of the superhedging price on Ω there exists an $\mathcal{F}_t^\mathcal{U}$-measurable strategy H_{t+1} such that

$$\pi_{t,\Omega^*}(g)(v') + H_{t+1}(\omega)\Delta S_{t+1}(v') \geq \hat{\pi}_{t+1}(g)(v') \quad \text{for all } v' \in \Sigma_t^\omega \cap \Omega^*, \tag{2.3.1}$$

where we can assume without loss of generality that $H_{t+1}(\omega) \in \text{span}(\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*))$. Now we fix $v \in \Sigma_t^\omega \cap \Omega$ and v' the corresponding orthogonal projection. Let $\varepsilon > 0$. As $\pi_{t+1,\Omega^*}(g)$ is uniformly continuous on $\Omega \cap \Sigma_t^\omega$, we can use [Vanderbei, 1997, Theorem 1] (in connection with Tietze's extension theorem to extend the domain to a convex set) in order to find $\delta > 0$ such that

$$\varepsilon + \pi_{t+1,\Omega^*}(g)(v') + \frac{S_{t+1}(v) - S_{t+1}(v')}{\delta/\varepsilon}\xi_{t+1,\Sigma_t^\omega\cap\Omega}(v) \geq \pi_{t+1,\Omega^*}(g)(v),$$

where δ is chosen such that for all $w, \tilde{w} \in \Sigma_t^\omega \cap \Omega$ we have $|\pi_{t+1,\Omega^*}(g)(w) - \pi_{t+1,\Omega^*}(g)(\tilde{w})| \leq \varepsilon$ whenever $|S_{t+1}(w) - S_{t+1}(\tilde{w}))| < \delta$. Note that $\Delta S_{t+1}(v) - \Delta S_{t+1}(v')$ is orthogonal to $H_{t+1}(\omega)$ and

$$\varepsilon + \pi_{t,\Omega^*}(g)(g)(v) + \left(H_{t+1}(\omega) + \frac{\xi_{t+1,\Sigma_t^\omega \cap \Omega}(v)}{\delta/\varepsilon}\right)(\Delta S_{t+1}(v') + \Delta S_{t+1}(v) - \Delta S_{t+1}(v'))$$
$$\geq \varepsilon + \pi_{t+1,\Omega^*}(g)(v') + \frac{S_{t+1}(v) - S_{t+1}(v')}{\delta/\varepsilon}\xi_{t+1,\Sigma_t^\omega \cap \Omega}(v) \geq \pi_{t+1,\Omega^*}(g)(v).$$

Next we use the assumption that A_t^ω is bounded and has no points of convergence in $\operatorname{span}(\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*))$ outside the set $\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$. Also, the continuous functions $\pi_{t+1,\Omega^*}(g)$ and $H_{t+1}\Delta S_{t+1}$ are uniformly continuous and bounded on A_t^ω, as $\Omega \cap \Sigma_t^\omega$ is comact. There exists $\delta > 0$ such that for all $v \in \Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)$, $\tilde{v} \in A_t^\omega$ with $|v - \tilde{v}| \leq \delta$ (5.3.1) implies

$$\varepsilon + \pi_{t,\Omega^*}(g)(\omega) + H_{t+1}(\omega)\Delta S_{t+1}(\tilde{v}) \geq \pi_{t+1,\Omega^*}(g)(\tilde{v}).$$

By assumption there exists $\tilde{\delta} > 0$ such that $\operatorname{dist}(\tilde{v}, \operatorname{span}(\Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*))) > \tilde{\delta}$ for all $\tilde{v} \in A_t^\omega$ with $\operatorname{dist}(\tilde{v}, \Delta S_{t+1}(\Sigma_t^\omega \cap \Omega^*)) > \delta$. Define $\pi_{\max} = \sup_{v \in A_t^\omega} \pi_{t+1,\Omega^*}(g)(v) < \infty$ and $C = \inf_{v \in A_t^\omega} H_{t+1}(\omega)\Delta S_{t+1}(v) + \pi_{t,\Omega^*}(g)(\omega) > -\infty$. Now we note that

$$\varepsilon + \pi_{t+1,\Omega^*}(g)(\tilde{v}) + H_{t+1}(\tilde{v})\Delta S_{t+1}(\tilde{v})$$
$$+ \frac{|\pi_{\max}| + |C|}{\tilde{\delta}}\xi_{t+1,\Sigma_t^\omega \cap \Omega}(v)\Delta S_{t+1}(\tilde{v}) \geq \pi_{t+1,\Omega^*}(g)(\tilde{v})$$

for all $\tilde{v} \in A_t^\omega$. This concludes the proof. □

2.3.2 Comparison of Strong and Uniformly Strong Arbitrage

We take now a closer look at the notions $\mathbf{SA}(\Omega)$ and $\mathbf{USA}(\Omega)$ and establish their equivalence in specific market setups. Clearly every Uniformly Strong Arbitrage is a Strong Arbitrage. In general the opposite assertion is not true: take for example $d = 1$, $S_0 = 1$, $S_1(\omega) = \omega$, $\Omega = (1, 2]$, then every $H_1 > 0$ is a Strong Arbitrage, but there do not exist any Uniformly Strong Arbitrages. This simple example can be generalised: a one-period market in the canonical setting with $S_0 = 1$ and an open convex set Ω such that $\{1\} \cap \Omega = \emptyset$ and $1 \in \bar{\Omega}$ admits a Strong Arbitrage but exhibits no Uniformly Strong Arbitrages. On the level of superhedging prices a Uniformly Strong Arbitrage on Ω corresponds to $\pi_\Omega(0) = -\infty$. For a financial market which exhibits a Strong Arbitrage but no uniformly Strong Arbitrages, the Pricing-Hedging duality cannot hold (as there are no martingale measures supported on Ω) but $\pi_\Omega(0) = 0$. In conclusion, the difference between Uniformly Strong Arbitrage and

Strong Arbitrage can be seen as a property describing the boundary of the prediction set Ω and thus manifests itself in the boundary behaviour of the superhedging functional

$$S_0 \mapsto \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H(S_1 - S_0) \geq 0 \text{ on } \Omega\}.$$

As it is an upper semicontinuous function, it takes the value zero on the boundary of Ω, while its lower semicontinuous version takes the value $-\infty$. Nevertheless the two notions agree in specific cases, which we now explore.

We assume the canonical setting $X_1 = \mathbb{R}^d_+$, $S_0(\omega) = s_0$ and set $\mathcal{F}_t = \mathcal{F}^0_t$ for all $0 \leq t \leq T$. In this section we allow for countably many statically traded options, $\Lambda = \mathbb{N}$, but only of European type, $\Phi = \{\phi_n = \phi_n(S_T) \mid n \in \mathbb{N}\}$, with real-valued continuous payoffs and a common maturity T. We write $c_{00} = c_{00}(\mathbb{N})$ for simplicity. We fix a closed subset $\Omega \subseteq (\mathbb{R}^d_+)^T$ and recall that martingale measures on Ω calibrated to Φ are denoted by $\mathcal{M}_{\Omega,\Phi}(\mathbb{F})$. We define $|S(\omega)|_1 := \sum_{t=1}^T \sum_{k=1}^d |S^k_t(\omega)|$ and denote by $C^b_{|S|_1}(\Omega)$ the space of real-valued continuous functions $f : \Omega \mapsto \mathbb{R}$ such that

$$\sup_{\omega\in\Omega} \frac{|f(\omega)|}{|S(\omega)|_1 \vee 1} < \infty.$$

Finally, we define the calibrated supermartingale measures as

$$\mathcal{SM}_{\Omega,\Phi}(\mathbb{F}) := \{\mathbb{Q} \in \mathfrak{P}(\Omega) \mid \mathbb{E}_{\mathbb{Q}}[\phi_n] \leq 0 \ \forall n \in \mathbb{N},\ \mathbb{E}_{\mathbb{Q}}[S_t|\mathcal{F}_{t-1}] \leq S_{t-1} \text{ a.s. } \forall t \leq T\}.$$

The following theorem can be seen as a unification of [Acciaio et al., 2013, Theorem 1.3], [Cox and Obłój, 2011, Prop. 2.2, p.6] and [Bartl et al., 2017, Cor. 4.6]. We also refer to [Burzoni et al., 2019, Thm. 3], who extend [Acciaio et al., 2013, Thm. 1.3] under the assumption $\Omega = \Omega^*$ and to [Burzoni et al., 2017b, Thm. C.5] for a general discussion in the case $\Phi = 0$. In contrast to the work of Acciaio et al. [2013], we do not need to assume the existence of a convex superlinear payoff g, which might be artificial in some settings, but explicitly enforce tightness of martingale measures through the **WFLVR**(Ω) condition.

Theorem 2.3.11. *The following hold:*

1. ***SA***$(\Omega) \Leftrightarrow$ ***USA***(Ω).
2. *Assume $\phi_n \in C^b_{|S|_1}(\Omega)$, no short-selling in any of the assets and*

$$\lim_{|S_T|_1\to\infty} \frac{(\phi_n(S_T))^-}{|S_T|_1} = 0$$

for all $n \in \mathbb{N}$. Then

$$\boldsymbol{SA}(\Omega) \Leftrightarrow \boldsymbol{USA}(\Omega) \Leftrightarrow \boldsymbol{WFLVR}(\Omega) \Leftrightarrow \mathcal{SM}_{\Omega,\Phi}(\mathbb{F}) = \emptyset.$$

3. *As in (2) assume that* $\phi_n \in C^b_{|S|_1}(\Omega)$. *Furthermore assume that for every sequence* $(\omega_n)_{n\in\mathbb{N}}$ *with* $\lim_{n\to\infty}|S(\omega_n)|_1 = \infty$, *there exists a sequence* $(h^k, H^k)_{k\in\mathbb{N}}$ *of trading strategies, a constant* $C > 0$ *and a sequence* $(p^k)_{k\in\mathbb{N}}$ *such that*

   - $\lim_{k\to\infty}\lim_{n\to\infty}\frac{(h^k\cdot\Phi+H^k\circ S_T)(\omega_n)}{|S(\omega_n)|_1\vee 1} > 0$ *and*
   - $|h^k\cdot\Phi + H^k\cdot S_T| \le C(|S|_1 \vee 1)$ *on* Ω *for all* $k \in \mathbb{N}$,
   - $\lim_{k\to\infty}(h^k\cdot\Phi + H^k\circ S_T)(\omega) = -\lim_{k\to\infty} p^k$ *for all* $\omega \in \Omega$.

   Then

$$\boldsymbol{SA}(\Omega) \Leftrightarrow \boldsymbol{USA}(\Omega) \Rightarrow \boldsymbol{WFLVR}(\Omega) \Leftrightarrow \mathcal{M}_{\Omega,\Phi}(\mathbb{F}) = \emptyset,$$

   but in general $\boldsymbol{WFLVR}(\Omega)$ *does not imply* $\boldsymbol{SA}(\Omega)$.

Remark 2.3.12. 1. The case $|\Phi| < \infty$ is covered in Bouchard and Nutz [2015], Burzoni et al. [2019]), while the case $|\Phi| = \infty$ is not. The basic idea in both works is to inductively construct a martingale measure calibrated to a finite number of options.

2. Contrary to the case $|\Phi| < \infty$ (see [Burzoni et al., 2019, proof of Theorem 1, p.1050]), the set $\mathcal{M}_{\Omega,\Phi}(\mathbb{F})$ might not necessarily contain any finitely supported martingale measures.

3. An example showing that **WFLVR**(Ω) does not imply **SA**(Ω) is given in [Cox and Obłój, 2011, Prop. 2.2].

4. A special but important case of *(3)* is $T = 1, d = 1$ and $\phi_n(S_1) = (S_1 - n)^+ - p^n$, where $p^n \ge 0$. In this case we can set $H^k = 0$, $h^k = e_k$ for all $k \in \mathbb{N}$, where e_k is the kth unit vector and note that

$$\lim_{k\to\infty}\lim_{n\to\infty}\frac{(h^k\cdot\Phi + H^k(S_1 - S_0))(\omega_n)}{|S_1(\omega_n)|_1 \vee 1} = \lim_{k\to\infty}\lim_{n\to\infty}\frac{(S_1(\omega_n) - k)^+ - p^k}{|S_1(\omega_n)|_1} = 1 > 0$$

   and

$$\lim_{k\to\infty}(S_1(\omega) - k)^+ - p^k = -\lim_{k\to\infty} p^k,$$

   in particular all three conditions in *(3)* are satisfied.

Proof. For simplicity of exposition we only give the proof for $T = 1$. This conveys the important ideas, while the multiperiod case extends these via a dynamic programming approach and can found in Section 2.4.3.
Regarding *(1)*, clearly **USA**$(\Omega) \Rightarrow$ **SA**(Ω), so we show **SA**$(\Omega) \Rightarrow$ **USA**(Ω). Let $(h, H) \in \mathbb{R}^k \times \mathbb{R}^d$ be a Strong Arbitrage. We show that it is actually a Uniformly Strong Arbitrage. For $x \in \mathbb{R}^d_+$ we denote by $|x|_1 := \sum_{i=1}^d x^i$ the ℓ_1-norm of x and define the compact set

$K = [0, s_0^1 + 2|s_0|_1] \times [0, s_0^2 + 2|s_0|_1] \times \cdots \times [0, s_0^d + 2|s_0|_1]$. Then, as $S_1 \mapsto h \cdot \Phi(S_1) + H(S_1 - S_0)$ is continuous and positive on the compact set $\Omega \cap K$, there exists $\varepsilon > 0$ such that

$$h \cdot \Phi(S_1) + H \cdot (S_1 - S_0) \geq \varepsilon \quad \text{on } K \cap \Omega.$$

Scaling (h, H) suitably we can without loss of generality assume take $\varepsilon = 2|s_0|_1$. Let $\mathbf{e} = (1, \ldots, 1)$ be the row unit vector in $\mathbb{R}^d$. Then

$$h \cdot \Phi(S_1) + (H + \mathbf{e}) \cdot (S_1 - S_0) \geq 2|s_0|_1 - |s_0|_1 = |s_0|_1 \quad \text{on } K \cap \Omega. \tag{2.3.2}$$

Furthermore on $\Omega \setminus K$ we have

$$h \cdot \Phi(S_1) + (H + \mathbf{e}) \cdot (S_1 - S_0) \geq \mathbf{e} \cdot (S_1 - S_0) \geq 2|s_0|_1 - |s_0|_1 = |s_0|_1.$$

Now we show *(2)*. Clearly the relation $\mathbf{USA}(\Omega) \Rightarrow \mathbf{WFLVR}(\Omega)$ holds and by *(1)* also $\mathbf{SA}(\Omega) \Leftrightarrow \mathbf{USA}(\Omega)$. Further, $\mathbf{WFLVR}(\Omega)$ readily implies $\mathcal{SM}_{\Omega,\Phi}(\mathbb{F}) = \emptyset$ since otherwise if $\mathbb{Q} \in \mathcal{SM}_{\Omega,\Phi}(\mathbb{F})$ then, by Fatou's lemma,

$$0 \geq \liminf_{n \to \infty} \mathbb{E}_{\mathbb{Q}}[h^n \cdot \Phi + H^n(S_1 - S_0)] \geq \mathbb{E}_{\mathbb{Q}}[\liminf_{n \to \infty} h^n \cdot \Phi + H^n(S_1 - S_0)] > 0,$$

a contradiction. Next we show $\mathbf{NUSA}(\Omega) \Rightarrow (\mathcal{SM}_{\Omega,\Phi}(\mathbb{F}) \neq \emptyset)$ following closely the argument in [Acciaio et al., 2013, proof of Prop. 2.3 and Theorem 1.3, pp. 240-242]. We denote by c_{00}^+ the subset of all non-negative sequences in c_{00}. We define the set

$$K := \left\{ h \cdot \Phi(S_1) + H(S_1 - S_0) \;\middle|\; (h, H) \in c_{00}^+ \times \mathbb{R}_+^d \right\} \subseteq C^b_{|S|_1}(\Omega).$$

Note that K is convex and non-empty. Furthermore denote the positive cone of $C^b_{|S|_1}(\Omega)$ by

$$C_{++}(\Omega) = \left\{ f \in C^b_{|S|_1}(\Omega) \;\middle|\; \inf_{\omega \in \Omega} \frac{f(\omega)}{|S(\omega)|_1 \vee 1} > 0 \right\}.$$

By $\mathbf{NUSA}(\Omega)$ we have $K \cap C_{++}(\Omega) = \emptyset$. An application of Hahn-Banach theorem yields existence of a positive measure $\mu = \mu^r + \mu^s$ such that

$$\int_\Omega \frac{f}{|S|_1 \vee 1} \, d\mu > 0 \quad \text{for all } f \in C_{++}(\Omega),$$
$$\int_\Omega \frac{f}{|S|_1 \vee 1} \, d\mu \leq 0 \quad \text{for all } f \in K.$$

We now aim to show that the normalised measure $\mathbb{Q}$ given by

$$d\mathbb{Q} := \frac{1}{|S|_1 \vee 1} \left(\int \frac{1}{|S|_1 \vee 1} \, d\mu^r \right)^{-1} d\mu^r$$

is an element of $\mathcal{SM}_{\Omega,\Phi}$. For this let us first assume that $\mu^r = 0$. Then

$$\int_\Omega \frac{\mathbf{e}(S_1 - S_0)}{|S|_1 \vee 1}\, d\mu = \int_\Omega \frac{|S|_1 - |S_0|_1}{|S|_1 \vee 1}\, d\mu^s = \int_\Omega 1\, d\mu^s > 0$$

as μ is positive, which is a contradiction. As $\int_\Omega \frac{(\phi_n(S_1))^-}{|S|_1 \vee 1}\, d\mu^s = 0$, we conclude

$$\int_\Omega \frac{\phi_n(S_1)}{|S|_1 \vee 1}\, d\mu^r \le \int_\Omega \frac{\phi_n(S_1)}{|S|_1 \vee 1}\, d\mu \le 0 \quad \text{for all } n \in \mathbb{N}.$$

Furthermore

$$\int_\Omega \frac{S_1 - S_0}{|S|_1 \vee 1}\, d\mu^r = 0$$

and thus $\mathbf{NUSA}(\Omega) \Rightarrow \mathcal{SM}_{\Omega,\Phi} \neq \emptyset$ follows.

Lastly we show *(3)*. For this we follow the same construction as in *(2)*. In particular redefining

$$K := \left\{ h \cdot \Phi(S_1) + H(S_1 - S_0) \;\middle|\; (h, H) \in c_{00} \times \mathbb{R}^d \right\} \subseteq C^b_{|S|_1}(\Omega).$$

we note that again by $\mathbf{NUSA}(\Omega)$ we have $K \cap C_{++}(\Omega) = \emptyset$. Thus all that is left to show is $\mu^s = 0$. Let us assume towards a contradiction $\mu^s \neq 0$ and take $(h^k, H^k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \int_\Omega \frac{h^k \cdot \Phi(S_1) + H^k(S_1 - S_0)}{|S|_1 \vee 1} d\mu^s > 0. \tag{2.3.3}$$

Then by symmetry of K and the same reasoning as in *(2)* we have

$$\int_\Omega \frac{h^k \cdot \Phi(S_1) + H^k(S_1 - S_0)}{|S|_1 \vee 1}\, d\mu = 0 \quad \text{for all } k \in \mathbb{N}. \tag{2.3.4}$$

Using (2.3.3) and (2.3.4)

$$\lim_{k\to\infty} \int_\Omega \frac{h^k \cdot \Phi(S_1) + H^k(S_1 - S_0)}{|S|_1 \vee 1} d\mu^r = - \lim_{k\to\infty} \int_\Omega \frac{h^k \cdot \Phi(S_1) + H^k(S_1 - S_0)}{|S|_1 \vee 1} d\mu^s < 0. \tag{2.3.5}$$

Note that for a sequence $(p^k)_{n\in\mathbb{N}}$ with

$$\lim_{k\to\infty} h^k \cdot \Phi(S_1) + H^k(S_1 - S_0) = - \lim_{k\to\infty} p^k$$

for all $\omega \in \Omega$ we need to have by no $\mathbf{WFLVR}(\Omega)$ that $\lim_{k\to\infty} p^k = 0$, so the LHS of (2.3.5) is equal to zero, a contradiction. □

2.4 Technical results and proofs

2.4.1 Proof of Theorems 2.2.6 and 2.2.7

We start with the following technical observation:

Proposition 2.4.1. *Let Ω be analytic. Then the FTAP of Bouchard and Nutz [2015] implies:*

$$\boldsymbol{N1pA}(\Omega, \mathbb{F}^{\mathcal{U}}) \Leftrightarrow \Omega = \Omega^*_\Phi$$

Proof. Set $\hat{\mathcal{P}} := \mathfrak{P}^f(\Omega)$. To apply the FTAP of Bouchard and Nutz [2015] we only need to show that $\hat{\mathcal{P}}_t(\omega) := \mathfrak{P}^f(\mathrm{proj}_{t+1}(\Omega \cap \Sigma_t^\omega))$ has analytic graph: we therefore fix $n \in \mathbb{N}$ and consider the Borel measurable function

$$\Sigma\colon X^n \to X^n \times \left(\mathbb{R}^d_+\right)^{(t+1)n} \qquad (\omega^1, \dots \omega^n) \mapsto (\omega^1, \dots, \omega^n, S_{0:t}(\omega^1), \dots, S_{0:t}(\omega^n))$$

and note that the image $\Sigma(\Omega^n)$ is analytic, since Ω is analytic and the image of an analytic set under a Borel measurable map as well as the Cartesian product of analytic sets is analytic (see [Bertsekas and Shreve, 1978, Prop. 7.38 & 7.40, p. 165]). Next we consider the continuous function

$$\begin{aligned} F\colon X^n \times \left(\mathbb{R}^d_+\right)^{(t+1)} &\to X^n \times \left(\mathbb{R}^d_+\right)^{(t+1)n} \\ (\omega^1, \dots \omega^n, x) &\mapsto (\omega^1, \dots, \omega^n, x, \dots, x). \end{aligned}$$

Note that

$$F\left(X^n \times \left(\mathbb{R}^d_+\right)^{(t+1)}\right) \cap \Sigma(\Omega^n)$$

is analytic and as projections of analytic sets are analytic

$$A_n := \{(\omega, \tilde{\omega}_1, \dots, \tilde{\omega}_n) \mid \omega \in \Omega,\ \tilde{\omega}_i \in \mathrm{proj}_{t+1}(\Omega \cap \Sigma_t^\omega),\ i = 1, \dots, n\}$$

is analytic as well. Let $\Delta_n \subseteq \mathbb{R}^n$ denote the simplex. Since the functions

$$\begin{aligned} G : A_n \times \Delta_n &\to \Omega \times \mathfrak{P}(X_n) \times \Delta_n \\ (\omega, \tilde{\omega}_1, \dots \tilde{\omega}_n, \lambda_1, \dots, \lambda_n) &\mapsto (\omega, \delta_{\tilde{\omega}_1}, \dots, \delta_{\tilde{\omega}_n}, \lambda_1, \dots, \lambda_n) \end{aligned}$$

and

$$\begin{aligned} H : \Omega \times \mathfrak{P}(X_n) \times \Delta_n &\to \Omega \times \mathfrak{P}(X_1) \\ (\omega, \delta_{\tilde{\omega}_1}, \dots, \delta_{\tilde{\omega}_n}, \lambda_1, \dots, \lambda_n) &\mapsto \left(\omega, \sum_{i=1}^n \delta_{\tilde{\omega}_i} \lambda_i\right) \end{aligned}$$

are continuous, it follows that $\operatorname{graph}\left(\hat{\mathcal{P}}_t\right) = \bigcup_{n\in\mathbb{N}} H(G(A_n \times \Delta_n))$ is analytic.
Take now $\omega \in \Omega$ and $\mathbb{P} \in \mathfrak{P}^f(\Omega)$ such that $\mathbb{P}(\{\omega\}) > 0$. By the FTAP of Bouchard and Nutz [2015] there exists $\mathbb{Q} \in \mathcal{M}$ such that $\mathbb{Q} \ll \tilde{\mathbb{P}}$ for some $\tilde{\mathbb{P}} \in \mathfrak{P}^f(\Omega)$, $\mathbb{E}_{\mathbb{Q}}[\phi_j] = 0$ for all $j = 1, \dots, k$ and $\mathbb{P} \ll \mathbb{Q}$. In particular $\mathbb{Q} \in \mathcal{M}^f_\Omega$ and $\mathbb{Q}(\{\omega\}) > 0$.
Lastly assume that $\Omega = \Omega^*_\Phi$ and fix $\mathbb{P} \in \hat{\mathcal{P}}$ such that $\operatorname{supp}(\mathbb{P}) = \{\omega_1, \dots, \omega_n\}$ for some $n \in \mathbb{N}$. We can find $\mathbb{Q}_1, \dots, \mathbb{Q}_n \in \mathcal{M}^f_\Omega$ such that $\mathbb{Q}_i(\{\omega_i\}) > 0$ for $i = 1, \dots, n$. Then $\mathbb{Q} := 1/n \sum_{i=1}^n \mathbb{Q}_i \in \mathcal{M}^f_{\Omega,\Phi}$ and $\mathbb{Q}(\{\omega_i\}) > 0$ for $i = 1, \dots, n$, i.e., $\mathbb{P} \ll \mathbb{Q}$. □

We now give a complete proof of the quasi-sure FTAP in Bouchard and Nutz [2015] using results from Burzoni et al. [2019]. We first look at the case $\Phi = 0$ and start with an auxiliary lemma:

Lemma 2.4.2. *Let $t \in \{1, \dots, T\}$ and $\Omega \subseteq X_t$ be analytic. Then the conditional standard separator of Burzoni et al. [2019] denoted by $\xi_{t,\Omega}$ is $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable.*

Proof. We shortly recall arguments from Burzoni et al. [2019][proof of Lemma 1]: let us define the multifunction

$$\psi_{t,\Omega} : \omega \in X \twoheadrightarrow \{\Delta S_t(\tilde{\omega}) \mid \tilde{\omega} \in \Sigma^\omega_{t-1} \cap \Omega\} \subseteq \mathbb{R}^d.$$

Then $\psi_{t,\Omega}$ is an $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable multifunction. Indeed, for $O \subseteq \mathbb{R}^d$ open we have

$$\{\omega \in X \mid \psi_{t,\Omega}(\omega) \cap O \neq \emptyset\} = S^{-1}_{0:t-1}(S_{0:t-1}((\Delta S_t)^{-1}(O) \cap \Omega)).$$

As ΔS_t is Borel measurable $(\Delta S_t)^{-1}(O) \in \mathcal{F}^0_t$. Also as intersections, projections and preimages of analytic sets are analytic (see [Bertsekas and Shreve, 1978, Prop. 7.35 & Prop. 7.40]), we find that $\{\omega \in X \mid \psi_{t,\Omega}(\omega) \cap O \neq \emptyset\}$ is analytic and in particular $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable. Let $\mathbb{S}^d$ be the unit sphere in $\mathbb{R}^d$, then by preservation of measurability also the multifunction

$$\psi^*_{t,\Omega}(\omega) := \{H \in \mathbb{S}^d \mid H \cdot y \geq 0 \text{ for all } y \in \psi_{t,\Omega}(\omega)\}$$

is $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable and closed-valued. Let $\{\xi^n_{t,\Omega}\}_{n\in\mathbb{N}}$ be its $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable Castaing representation. The conditional standard separator is then defined as

$$\xi_{t,\Omega} = \sum_{n=1}^{\infty} \frac{1}{2^n} \xi^n_{t,\Omega}.$$

□

Remark 2.4.3. We recall that this separator has the property that it aggregates all one-dimensional One-point Arbitrages on $\Sigma_{t-1}^{\omega} \cap \Omega$ in the sense that

$$\{\omega \in X \mid \xi(\omega) \cdot \Delta S_t(\omega) > 0\} \subseteq \{\omega \in X \mid \xi_{t,\Omega}(\omega) \cdot \Delta S_t(\omega) > 0\}$$

for every measurable selector ξ of $\psi_{t,\Omega}^*$.

Proof of Theorem 2.2.6 for $\Phi = 0$. We start by proving the first part of Theorem 2.2.6, i.e., we are given a set of measures $\mathcal{P}$ satisfying (APS) and we need to construct $\Omega = \Omega^{\mathcal{P}}$ such that *(1)-(3)* are equivalent. We define for $\omega \in X_{t-1}$

$$\tilde{\chi}_{\mathcal{F}_{t-1}^0}(\omega) = \bigcap \{A \subseteq \mathbb{R}^d \text{ closed } \mid \mathbb{P}(\Delta S_t(\omega, \cdot) \in A) = 1 \ \forall \ \mathbb{P} \in \mathcal{P}_{t-1}(\omega)\}.$$

Then $\tilde{\chi}_{\mathcal{F}_{t-1}^0}$ is closed valued and $\mathbb{P}(\Delta S_t(\omega, \cdot) \in \tilde{\chi}_{\mathcal{F}_{t-1}^0}(\omega)) = 1$ for all $\mathbb{P} \in \mathcal{P}_{t-1}(\omega)$ and all $\omega \in X_{t-1}$. Evidently

$$\begin{aligned}
\tilde{\chi}_{\mathcal{F}_{t-1}^0}(\omega) &= \{x \in \mathbb{R}^d \mid \forall \varepsilon > 0 \ \mathbb{P}(\Delta S_t(\omega, \cdot) \in B(x, \varepsilon)) > 0 \text{ for some } \mathbb{P} \in \mathcal{P}_{t-1}(\omega)\} \\
&= \overline{\bigcup_{\mathbb{P} \in \mathcal{P}_{t-1}(\omega)} \operatorname{supp}(\mathbb{P} \circ \Delta S_t(\omega, \cdot)^{-1})}.
\end{aligned}$$

Also it follows from [Bouchard and Nutz, 2015, Lemma 4.3, page 840], that $\tilde{\chi}_{\mathcal{F}_{t-1}^0}$ is analytically measurable. We quickly repeat their argument: let us define

$$l : X_{t-1} \times \mathfrak{P}(X_1) \to \mathfrak{P}(\mathbb{R}^d) \quad l(\omega, \mathbb{P}) = \mathbb{P} \circ \Delta S_t(\omega, \cdot)^{-1}.$$

Then l is Borel measurable. Next we consider

$$\mathcal{R} : X_{t-1} \twoheadrightarrow \mathfrak{P}(\mathbb{R}^d) \quad \mathcal{R}(\omega) := l(\omega, \mathcal{P}_{t-1}(\omega)) = \{\mathbb{P} \circ \Delta S_t(\omega, \cdot)^{-1} \mid \mathbb{P} \in \mathcal{P}_{t-1}(\omega)\}.$$

Since its graph is analytic, it follows that for $O \subseteq \mathbb{R}^d$ open

$$\begin{aligned}
\{\omega \in X_{t-1} \mid \tilde{\chi}_{\mathcal{F}_{t-1}^0}(\omega) \cap O \neq \emptyset\} &= \{\omega \in X_{t-1} \mid R(O) > 0 \text{ for some } R \in \mathcal{R}(\omega)\} \\
&= \operatorname{proj}_{X_{t-1}} \{(\omega, R) \in \operatorname{graph}(\mathcal{R}) \mid R(O) > 0\}
\end{aligned}$$

is analytic as $R \mapsto R(O)$ is Borel.

We also note that for $\varepsilon > 0$ the function $x \mapsto R(B_\varepsilon(x))$ is continuous, so $(x, R) \mapsto R(B_\varepsilon(x))$ is Borel and

$$\begin{aligned}
\operatorname{graph}(\tilde{\chi}_{\mathcal{F}_{t-1}^0}) &= \{(\omega, x) \in (X_{t-1} \times \mathbb{R}^d) \mid x \in \tilde{\chi}_{\mathcal{F}_{t-1}^0}(\omega)\} \\
&= \bigcap_{\varepsilon \in \mathbb{Q}_+} \operatorname{proj}_{X_{t-1} \times \mathbb{R}^d} \Big(\{(\omega, R, x) \in (\operatorname{graph}(\mathcal{R}) \times \mathbb{R}^d) \mid R(B_\varepsilon(x)) > 0\} \Big)
\end{aligned}$$

is analytic. Now we define

$$U = \{\omega \in X_t \mid \Delta S_t(\omega) \in \tilde{\chi}_{\mathcal{F}^0_{t-1}}(\omega)\}.$$

Then

$$U = \mathrm{proj}_{X_t}(\mathrm{graph}(\Delta S_t) \cap \mathrm{graph}(\tilde{\chi}_{\mathcal{F}^0_{t-1}}))$$

is analytic and by Fubini's theorem $\mathbb{P}(U) = 1$ holds for all $\mathbb{P} \in \mathcal{P}$. We now set

$$\Omega^{\mathcal{P}} = \bigcap_{t=1}^{T} \left\{\omega \in X_t \mid \Delta S_t(\omega) \in \tilde{\chi}_{\mathcal{F}^0_{t-1}}(\omega)\right\},$$

which is again analytic and $\mathbb{P}(\Omega^{\mathcal{P}}) = 1$ for all $\mathbb{P} \in \mathcal{P}$.

Having defined $\Omega^{\mathcal{P}}$ we can now begin to prove equivalence of *(1)-(3)*. If *(2)* holds then *(3)* follows immediately by a contradiction argument, so we now show the more involved implications *(3)* $\Rightarrow$ *(1)* and *(1)* $\Rightarrow$ *(2)*. Let us start with the proof of *(3)* $\Rightarrow$ *(1)*: we assume that there exists $\hat{\mathbb{P}} \in \mathcal{P}$ such that $\hat{\mathbb{P}}\left(\Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^*\right) > 0$. We want to find $H \in \mathcal{H}(\mathbb{F}^{\mathcal{U}})$ and $\tilde{\mathbb{P}} \in \mathcal{P}$ such that $H \circ S_T \geq 0$ $\mathcal{P}$-q.s and $\tilde{\mathbb{P}}(H \circ S_T > 0) > 0$. For this we take $t = T-1$ and assume that

$$\hat{\mathbb{P}}\left(\{\omega \in \mathrm{proj}_{0:T-1}(\Omega^{\mathcal{P}}) \mid \text{there is a One-point Arbitrage on } \Sigma^{\omega}_{T-1} \cap \Omega^{\mathcal{P}}\}\right) > 0.$$

Let us now fix $\omega \in \{\mathrm{proj}_{0:T-1}(\Omega^{\mathcal{P}}) \mid \text{there is a One-point Arbitrage on } \Sigma^{\omega}_{T-1} \cap \Omega^{\mathcal{P}}\}$. Denote by $\xi_{T,\Omega^{\mathcal{P}}}$ the $\mathcal{F}^{\mathcal{U}}_{T-1}$-measurable standard separator of Lemma 2.4.2. Now we define for each $\mathbb{P} \in \mathcal{P}_{T-1}(\omega)$ the push-forward of $\mathbb{P}$ as

$$\mathbb{P}_{\Delta S_T(\omega,\cdot)}(A) = \mathbb{P}(\Delta S_T(\omega,\cdot) \in A),$$

where $A \in \mathcal{B}(\mathbb{R}^d)$. We note that by definition

$$\mathbb{P}_{\Delta S_T(\omega,\cdot)}\left(\tilde{\chi}_{\mathcal{F}^0_{T-1}}(\omega)\right) = 1$$

holds for all $\mathbb{P} \in \mathcal{P}_{T-1}(\omega)$. With a slight abuse of notation we recall the set

$$B^1(\omega) := \{\omega' \in \mathrm{proj}_T(\Sigma^{\omega}_{T-1} \cap \Omega^{\mathcal{P}}) \mid \xi_{T,\Omega^{\mathcal{P}}}(\omega) \cdot \Delta S_T(\omega,\omega') > 0\}$$

from [Burzoni et al., 2019, proof of Lemma 1, Step 1] and note that for all $\mathbb{P} \in \mathcal{P}_{T-1}(\omega)$

$$\begin{aligned}\mathbb{P}\left(\{\omega' \in \mathrm{proj}_T(\Sigma^{\omega}_{T-1} \cap \Omega^{\mathcal{P}}) \mid \xi_{T,\Omega^{\mathcal{P}}}(\omega) \cdot \Delta S_T(\omega,\omega') > 0\}\right)\\ = \mathbb{P}_{\Delta S_T(\omega,\cdot)}(\{x \in \mathbb{R}^d \mid \xi_{T,\Omega^{\mathcal{P}}}(\omega) \cdot x > 0\})\end{aligned}$$

follows. Clearly the set $\{x \in \mathbb{R}^d \mid \xi_{T,\Omega^{\mathcal{P}}}(\omega) \cdot x > 0\}$ is open in $\mathbb{R}^d$, thus by definition of $\tilde{\chi}_{\mathcal{F}^0_{T-1}}(\omega)$ there is a $\tilde{\mathbb{P}} \in \mathcal{P}_{T-1}(\omega)$ such that

$$\tilde{\mathbb{P}}_{\Delta S_T(\omega,\cdot)}(\{x \in \mathbb{R}^d \mid \xi_{T,\Omega^{\mathcal{P}}}(\omega) \cdot x > 0\}) > 0$$

or there are no One-point Arbitrages on $\Sigma^\omega_{T-1} \cap \Omega^{\mathcal{P}}$. To finish the proof of *(3) ⇒ (1)* we need to select $\tilde{\mathbb{P}}$ in a measurable way and this follows by standard arguments: Define the correspondence $\Psi : \mathbb{R}^d \times X_{T-1} \twoheadrightarrow \mathfrak{P}(X_1)$ by

$$\Psi(H,\omega) = \{\mathbb{P} \in \mathcal{P}_{T-1}(\omega) \mid \mathbb{E}_{\mathbb{P}}[H \cdot \Delta S_T(\omega,\cdot)]^+ > 0\}.$$

This function has analytic graph by arguments in [Nutz, 2016, proof of Lemma 3.4, p.11], so we can employ the Jankov-von-Neumann theorem (cf. [Bertsekas and Shreve, 1978, Proposition 7.49, page 182]) to find a universally measurable kernel

$$\mathbb{P}'_{T-1} : \mathbb{R}^d \times X_{T-1} \to \mathfrak{P}(X_1)$$

such that $\mathbb{P}'_{T-1}(H,\omega) \in \mathcal{P}_{T-1}(\omega)$ for all $(H,\omega) \in \mathbb{R}^d \times X_{T-1}$ and $\mathbb{P}_{T-1}(H,\omega) \in \Psi(H,\omega)$ on $\{\Psi(H,\omega) \neq \emptyset\}$. Then also the kernel

$$\omega \mapsto \tilde{\mathbb{P}}_{T-1}(\omega) := \mathbb{P}'_{T-1}(\xi_{T,\Omega^{\mathcal{P}}}(\omega),\omega)$$

is universally measurable. Defining $\tilde{\mathbb{P}} := \hat{\mathbb{P}}|_{X_{T-1}} \otimes \tilde{\mathbb{P}}_{T-1}$, which is the product measure formed from the restriction of $\hat{\mathbb{P}}$ to X_{T-1} and $\tilde{P}_{T-1}$ gives $\tilde{\mathbb{P}}(\xi_{T,\Omega^{\mathcal{P}}} \cdot \Delta S_T > 0) > 0$. This proves *(3) ⇒ (1)* by backward induction.

Lastly we show *(1) ⇒ (2)*: let us assume $\mathbb{P}((\Omega^{\mathcal{P}})^*) = 1$ for all $\mathbb{P} \in \mathcal{P}$. Note that by the arguments given in the proof of *(3) ⇒ (1)* this means that

$$\bigcup_{t=1}^{T}\{\mathrm{proj}_{0:t-1}(\Omega^{\mathcal{P}}) \mid \text{there is a One-point Arbitrage on } \Sigma^\omega_{t-1} \cap \Omega^{\mathcal{P}}\}$$

is a $\mathcal{P}$-polar set, so in particular $0 \in \mathrm{ri}(\tilde{\chi}_{\mathcal{F}^0_{t-1}}(\omega))$ for all $t = 1,\dots,T$ and $\mathcal{P}$-q.e. $\omega \in X$. Here $\mathrm{ri}(\tilde{\chi}_{\mathcal{F}^0_{t-1}}(\omega))$ denotes the relative interior of the convex hull of $\tilde{\chi}_{\mathcal{F}^0_{t-1}}(\omega)$. Let $\hat{\mathbb{P}} \in \mathcal{P}$ be fixed. We define for an arbitrary $\mathbb{P} \in \mathcal{P}$ and $\omega \in X_{t-1}$ the support of $\mathbb{P}_{t-1}(\omega) \circ \Delta S_t^{-1}(\omega,\cdot)$ conditioned on $\mathcal{F}^0_{t-1}$ as

$$\chi^{\mathbb{P}}_{\mathcal{F}^0_{t-1}}(\omega) = \{x \in \mathbb{R}^d \mid \mathbb{P}_{t-1}(\omega)(\Delta S_t(\omega,\cdot) \in B_\varepsilon(x)) > 0 \text{ for all } \varepsilon > 0\}.$$

Using selection arguments which are explained below, we can now find measurable selectors $\mathbb{P}_{(0,1)},\dots,\mathbb{P}_{(0,d)},\mathbb{P}_{(1,1)},\dots,\mathbb{P}_{(T-1,d)}$ such that

$$\mathbb{P}_{(t,1)}(\omega),\dots,\mathbb{P}_{(t,d)}(\omega) \in \mathcal{P}_t(\omega)$$

and $\mathbb{P}_{(0,1)}, \dots, \mathbb{P}_{(T-1,d)}$ fulfil the following property: define

$$\tilde{\mathbb{P}}_t(\omega) = \frac{1}{d+1}\left(\hat{\mathbb{P}}_t(\omega) + \sum_{i=1}^{d} \mathbb{P}_{(t,i)}(\omega)\right)$$

for $t = 0, \dots, T-1$ and every $\omega \in X_t$. Then for $\tilde{\mathbb{P}} = \tilde{\mathbb{P}}_0 \otimes \cdots \otimes \tilde{\mathbb{P}}_{T-1}$ we have

$$0 \in \operatorname{ri}\left(\chi^{\tilde{\mathbb{P}}}_{\mathcal{F}^0_{t-1}}\right) \quad \tilde{\mathbb{P}}\text{-a.s. for all } 1 \le t \le T,$$

where $\operatorname{ri}\left(\chi^{\tilde{\mathbb{P}}}_{\mathcal{F}^0_{t-1}}\right)$ denotes the relative interior of the convex hull of $\chi^{\tilde{\mathbb{P}}}_{\mathcal{F}^0_{t-1}}$.
We note that since $\mathcal{P}_t(\omega)$ is convex, we have $\tilde{\mathbb{P}}_t(\omega) \in \mathcal{P}_t(\omega)$ for $\omega \in X_t$ and by definition $\hat{\mathbb{P}} \ll \tilde{\mathbb{P}}$ holds. Now it follows from [Rokhlin, 2008, Theorem 1, page 1], that there exists a martingale measure $\mathbb{Q}$ equivalent to $\tilde{\mathbb{P}}$. The fact that $\tilde{\mathbb{P}} \in \mathcal{P}$ implies $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P}}$, which shows the claim.
We now present the measurable selection argument: we fix $t \in \{1, \dots, T\}$. Note that for all $\omega \in \operatorname{proj}_{0:t-1}((\Omega^{\mathcal{P}})^*)$ we conclude $0 \in \operatorname{ri}(\Delta S_t(\omega, \Sigma^{\omega}_{t-1} \cap (\Omega^{\mathcal{P}})^*))$ by definition of $(\Omega^{\mathcal{P}})^*$, which implies by [Bonnice and Reay, 1969, Theorem D, p.1] that there exist $\mathbb{P}^1, \dots, \mathbb{P}^d \in \mathcal{P}_{t-1}(\omega)$, which might not be pairwise distinct, s.t.

$$0 \in \operatorname{ri}\left(\operatorname{supp}\left(\frac{\hat{\mathbb{P}}_{t-1}(\omega) + \mathbb{P}^1 + \cdots + \mathbb{P}^d}{d+1} \circ \Delta S_t(\omega, \cdot)^{-1}\right)\right).$$

Note that $\omega \mapsto \hat{\mathbb{P}}_{t-1}(\omega)$ is universally measurable. We define the correspondence $\rho : \mathfrak{P}(X_1)^{d+1} \twoheadrightarrow \mathbb{R}^d$ by

$$\rho : (\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) = \operatorname{supp}\left(\frac{\mathbb{P}^0 + \mathbb{P}^1 + \cdots + \mathbb{P}^d}{d+1} \circ \Delta S_t(\omega, \cdot)^{-1}\right).$$

Note that for $O \subseteq \mathbb{R}^d$ open we have

$$\begin{aligned}\{(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) \mid \rho(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) \cap O \neq \emptyset\} \\ = \bigcup_{i=0}^{d} \{(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) \mid \mathbb{P}^i \circ \Delta S_t(\omega, \cdot)^{-1}(O) > 0\}.\end{aligned}$$

Since $\mathbb{P} \mapsto \mathbb{P}(O)$ is Borel measurable, we conclude that ρ is weakly measurable. Let us denote by $\mathbb{S}^d$ the unit sphere in $\mathbb{R}^d$. By preservation of measurability (cf. [Rockafellar and Wets, 2009, Exercise 14.12, page 653]) it follows that the correspondence $\Psi : \mathfrak{P}(X_1)^{d+1} \twoheadrightarrow \mathbb{R}^d$

$$\Psi(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) = \{H \in \mathbb{S}^d \mid H \cdot y \ge 0 \text{ for all } y \in \rho(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d)\}$$

is weakly measurable. Then also the correspondence $\tilde{\Psi} : \mathfrak{P}(X_1)^{d+1} \twoheadrightarrow \mathbb{R}^d$

$$\begin{aligned}\tilde{\Psi}(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d) =& \{H \in \mathbb{S}^d \mid H \cdot y \le 0 \text{ for all } y \in \rho(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d)\} \\ &\cap \Psi(\mathbb{P}^0, \mathbb{P}^1, \dots, \mathbb{P}^d)\end{aligned}$$

is weakly measurable and closed-valued. Let V be a countable base of $\mathbb{R}^d$. The set

$$\begin{aligned}
&\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid \tilde{\Psi}(\omega,\mathbb{P}^1,\dots,\mathbb{P}^d)=\Psi(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d)\} \\
=&\bigcap_{O:O\in V}(\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid \Psi\cap O\neq\emptyset\}\cap\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid \tilde{\Psi}\cap O\neq\emptyset\} \\
&\cup\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid \Psi\cap O=\emptyset\}\cap\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid \tilde{\Psi}\cap O=\emptyset\})
\end{aligned}$$

is weakly measurable. Note that for an arbitrary convex set $C\subseteq\mathbb{R}^d$ the relationship

$$0\in\mathrm{ri}(C)\Leftrightarrow(\forall\ H\in\mathbb{S}^d\text{ s.t. }H\cdot x\geq 0\ \forall x\in C\ \Rightarrow\ H\cdot x=0\ \forall x\in C)$$

holds. Let

$$A:=\{(\mathbb{P}^0,\mathbb{P}^1,\dots,\mathbb{P}^d) \mid 0\in\mathrm{ri}\left(\rho(\mathbb{P}^0,\mathbb{P}^1,\dots\mathbb{P}^d)\right)\ \text{for } i=1,\dots,d\}.$$

Then from the above arguments it follows that A is Borel and in particular the set-valued mapping

$$A(\omega,\mathbb{P}^0):=\{(\mathbb{P}^1,\dots,\mathbb{P}^d) \mid 0\in\mathrm{ri}\left(\rho(\mathbb{P}^0,\mathbb{P}^1,\dots\mathbb{P}^d)\right),\ \mathbb{P}^i\in\mathcal{P}_{t-1}(\omega)\text{ for } i=1,\dots,d\}$$

has analytic graph. We can now employ the Jankov-von-Neumann theorem (cf. Bertsekas and Shreve [1978], Proposition 7.49, page 182) to find universally measurable kernels $\mathbb{P}^i_{t-1}:X_{t-1}\to\mathfrak{P}(X_1)$ such that for every $\omega\in X_{t-1}$ we have $\mathbb{P}^i_{t-1}(\omega)\in\mathcal{P}_{t-1}(\omega)$ and

$$0\in\mathrm{ri}\left(\rho(\hat{\mathbb{P}}_{t-1}(\omega),\mathbb{P}^1_{t-1},\dots\mathbb{P}^d_{t-1})\right).$$

This concludes the proof of *(1)* $\Rightarrow$ *(2)*.
The second part of Theorem 2.2.6 follows immediately from Proposition 2.4.1. □

Before continuing the proof of Theorem 2.2.6 let us first give a short remark on the measurability of the arbitrage strategies involved in the proof of above:

Remark 2.4.4. By the FTAP of Burzoni et al. [2019] there exists a filtration $\tilde{\mathbb{F}}$ with $\mathbb{F}^0\subseteq\tilde{\mathbb{F}}\subseteq\mathbb{F}^M$ such that there is no Strong Arbitrage in $\mathcal{H}(\tilde{\mathbb{F}})$ on $\Omega^{\mathcal{P}}$. More concretely there exists an $\mathcal{H}(\tilde{\mathbb{F}})$- and thus $\mathcal{H}(\mathbb{F}^M)$-measurable arbitrage aggregator H^*. So in particular if $\mathbb{P}(\Omega^{\mathcal{P}}\setminus(\Omega^{\mathcal{P}})^*)>0$ for some $\mathbb{P}\in\mathcal{P}$, then H^* is an $\mathcal{H}(\tilde{\mathbb{F}})$-measurable $\mathcal{P}$-q.s arbitrage. In general the inclusion $\tilde{\mathbb{F}}\subseteq\mathbb{F}^{\mathcal{U}}$ does not hold. This is why we need to construct a new $\mathbb{F}^{\mathcal{U}}$-measurable arbitrage strategy, which captures the arbitrages essential for $\mathcal{P}$. More generally, in this chapter we manage to avoid using projectively measurable sets, which were essential for the arguments in Burzoni et al. [2019]. In fact, all our trading strategies are universally measurable without invoking the axiom of projective determinacy.
Furthermore, we hope that by constructing an explicit arbitrage strategy in the proof of

(3) $\Rightarrow$ *(1)* we can clarify the proof of Burzoni et al. [2016], Theorem 4.23, pp. 42-46 (in particular Burzoni et al. [2016][A.3]) by offering a similar to the above (but much simpler) reasoning for the case $\mathcal{P} = \{\mathbb{P}\}$. Introducing a measurable separator ξ it is apparent that j_z in [Burzoni et al., 2016, p.44] can always be chosen equal to one in our setting. Also the resulting strategy $H^{\mathbb{P}}$ therein can be chosen universally measurable.

To prove the first part of Theorem 2.2.6 for the case $\Phi \neq 0$ we recall the following notion from Burzoni et al. [2019]:

Definition 2.4.5 (Burzoni et al. [2019], Def. 4)**.** *A pathspace partition scheme $\mathcal{R}(\alpha^*, H^*)$ of Ω is a collection of trading strategies $H_1, ..., H_\beta \in \mathcal{H}(\mathbb{F}^{\mathcal{U}})$, $\alpha_1, \dots, \alpha_\beta \in \mathbb{R}^k$ and arbitrage aggregators $\tilde{H}_0, \dots, \tilde{H}_\beta$ for some $1 \leq \beta \leq k$ such that*

1. *the vectors α_i, $1 \leq i \leq \beta$ are linearly independent,*
2. *for any $i \leq \beta$*

$$\alpha_i \cdot \Phi + H_i \circ S_T \geq 0 \quad \text{on } A^*_{i-1},$$

   where $A_0 = \Omega$, $A_i := \{\alpha_i \cdot \Phi + H_i \circ S_T = 0\} \cap A^*_{i-1}$,

3. *for any $i = 0, \dots, \beta$, $\tilde{H}_i$ is an Arbitrage Aggregator for A_i,*
4. *if $\beta < k$, then either $A_\beta = \emptyset$ or for any $\alpha \in \mathbb{R}^k$ linearly independent from $\alpha_1, \dots, \alpha_\beta$ there does not exist H such that*

$$\alpha \cdot \Phi + (H \circ S_T) \geq 0 \quad \text{on } A^*_\beta.$$

Definition 2.4.6 (Burzoni et al. [2019], Def. 5)**.** *A pathspace partition scheme $\mathcal{R}(\alpha^*, H^*)$ is successful if $A^*_\beta \neq \emptyset$.*

We quote the following results:

Lemma 2.4.7 (Burzoni et al. [2019], Lemma 5)**.** *For any $\mathcal{R}(\alpha^*, H^*)$, $A^*_i = \Omega^*_{\{\alpha^j \cdot \Phi \mid j \leq i\}}$. Moreover, if $\mathcal{R}(\alpha^*, H^*)$ is successful, then $A^*_\beta = \Omega^*_\Phi$.*

Lemma 2.4.8 (Burzoni et al. [2019], Proof of Theorem 1 for $\Phi \neq 0$)**.** *A pathspace partition scheme $\mathcal{R}(\alpha^*, H^*)$ is successful if and only if $\Omega^*_\Phi \neq \emptyset$.*

We now complete the first part of the proof of Theorem 2.2.6 for the case $\Phi \neq 0$:

Proof of Theorem 2.2.6 for $\Phi \neq 0$. The existence of $\Omega^{\mathcal{P}}$ and *(1)* $\Rightarrow$ No Strong Arbitrage in $\mathcal{A}_\Phi(\tilde{\mathbb{F}})$ on $\Omega^{\mathcal{P}}$, *(2)* $\Rightarrow$ *(3)* follow exactly as before. We now argue that *(1)* $\Rightarrow$ *(2)* holds in the spirit of [Bouchard and Nutz, 2015, Theorem 5.1, p. 850], by induction over the number e of options available for static trading. In particular we can assume without loss of generality that there exists a random variable $\varphi \geq 1$ such that $|\phi_j| \leq \varphi$ for all $j = 1, \dots, k$ and consider the set $\mathcal{Q}_\varphi = \{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P}} \mid \mathbb{E}_\mathbb{Q}[\varphi] < \infty\}$ in order to avoid integrability issues. So let us assume there are $e \geq 0$ traded options $\phi_1, \dots, \phi_e$, for which *(1)* $\Rightarrow$ *(2)* holds. We introduce an additional option $g = \phi_{e+1}$ and assume $\mathbb{P}\left((\Omega^{\mathcal{P}})^*_{\{\phi_1,\dots,\phi_{e+1}\}}\right) = 1$ for all $\mathbb{P} \in \mathcal{P}$. Then clearly $\mathbb{P}\left((\Omega^{\mathcal{P}})^*_{\{\phi_1,\dots,\phi_e\}}\right) = 1$ for all $\mathbb{P} \in \mathcal{P}$ and by the induction hypothesis there is no arbitrage in the market with options $\{\phi_1, \dots, \phi_e\}$ available for static trading. Let $\mathbb{P} \in \mathcal{P}$. Then by exactly the same arguments as in [Bouchard and Nutz, 2015, proof of Theorem 5.1(a)] we can use convexity of $\mathcal{Q}_\varphi$ and Theorem 2.2.9 to find a measure $\mathbb{Q} \in \mathcal{Q}_\varphi$, such that $\mathbb{P} \ll \mathbb{Q}$ and $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\{\phi_1,\dots,\phi_{e+1}\}}$, so *(2)* holds.
Lastly it remains to show *(3)* $\Rightarrow$ *(1)*. Let us thus assume there exists $\hat{\mathbb{P}} \in \mathcal{P}$ such that $\hat{\mathbb{P}}(\Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^*_\Phi) > 0$. We want to find $(h, H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}})$ and $\tilde{\mathbb{P}} \in \mathcal{P}$ such that $h \cdot \Phi + H \circ S_T \geq 0$ $\mathcal{P}$-q.s and $\tilde{\mathbb{P}}(h \cdot \Phi + H \circ S_T > 0) > 0$. We use the properties of a pathspace partition scheme $\mathcal{R}(\alpha^*, H^*)$ recalled above. We define

$$m = \min(k \in \{0, \dots, \beta\} \mid \tilde{\mathbb{P}}(A_k \setminus A^*_k) > 0 \text{ for some } \tilde{\mathbb{P}} \in \mathcal{P})$$
$$\tilde{m} = \min(k \in \{1, \dots, \beta\} \mid \tilde{\mathbb{P}}(A^*_{k-1} \setminus A_k) > 0 \text{ for some } \tilde{\mathbb{P}} \in \mathcal{P}),$$

where $A_0 = \Omega^{\mathcal{P}}$. If $\tilde{m} \leq m$ then we select the strategy $(\alpha_{\tilde{m}}, H_{\tilde{m}}) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}})$ which satisfies $H_{\tilde{m}} \circ S_T + \alpha_{\tilde{m}} \cdot \Phi \geq 0$ on $A^*_{\tilde{m}-1}$. We note that $\mathbb{P}(A^*_{\tilde{m}-1}) = 1$ for all $\mathbb{P} \in \mathcal{P}$ by definition of m, $\tilde{m}$ and $\{H_{\tilde{m}} \circ S_T + \alpha_{\tilde{m}} \cdot \Phi > 0\} = A^*_{\tilde{m}-1} \setminus A_{\tilde{m}}$, so that $\tilde{\mathbb{P}}(H_{\tilde{m}} \circ S_T + \alpha_{\tilde{m}} \cdot \Phi > 0) > 0$ for some $\tilde{\mathbb{P}} \in \mathcal{P}$. If $\tilde{m} > m$, then $\mathbb{P}(A_m) = 1$ for all $\mathbb{P} \in \mathcal{P}$, $\tilde{\mathbb{P}}(A_m \setminus A^*_m) > 0$ for some $\tilde{\mathbb{P}} \in \mathcal{P}$, so we can argue as in the proof of Proposition 2.2.6 for $\Phi = 0$ *(3)* $\Rightarrow$ *(1)* using a standard separator and measurable selection of a measure in $\mathcal{P}$.
As before, the second part of Theorem 2.2.6 follows immediately from Proposition 2.4.1. This concludes the proof. □

Proof of Theorem 2.2.7. We recall the analytic set $\Omega^{\mathcal{P}}$ from the proof of Theorem 2.2.6 for $\Phi = 0$ and the sets $\{C_n\}_{n\in\mathbb{N}}$ from (2.2.1). Now we define

$$B := \bigcup_{n\in\mathbb{N}} \{C_n \mid \mathbb{P}(C_n \cap (\Omega^{\mathcal{P}})^*_\Phi) = 0 \text{ for all } \mathbb{P} \in \mathcal{P}\} \in \mathcal{B}(X).$$

We claim that (2.2.1) implies that

$$\begin{aligned} B \cap (\Omega^{\mathcal{P}})^*_\Phi &= \bigcup \{C \in \mathcal{S} \mid \mathbb{P}(C \cap (\Omega^{\mathcal{P}})^*_\Phi) = 0 \text{ for all } \mathbb{P} \in \mathcal{P}\} \cap (\Omega^{\mathcal{P}})^*_\Phi \\ &= \bigcup \{C \in \mathcal{S} \mid C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}\} \cap (\Omega^{\mathcal{P}})^*_\Phi. \end{aligned}$$

Indeed, clearly $B \subseteq \bigcup\{C \in \mathcal{S} \mid C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}\}$. Now assume towards a contradiction that there exists

$$\omega \in \left(\bigcup\{C \in \mathcal{S} \mid C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}\} \cap (\Omega^{\mathcal{P}})^*_\Phi\right) \setminus (B \cap (\Omega^{\mathcal{P}})^*_\Phi).$$

In particular $\omega \in C$ for some $C \in \mathcal{S}$ such that $C \cap (\Omega^{\mathcal{P}})^*_\Phi \in \mathcal{N}^{\mathcal{P}}$. By (2.2.1) there exists $n_0 \in \mathbb{N}$ such that $C_{n_0} \subseteq C$ and $\omega \in C_{n_0}$. This implies $\omega \in B$ and thus shows the claim.
Let us now first assume that $\Phi = 0$ and set

$$\Omega := \Omega^{\mathcal{P}} \setminus ((\Omega^{\mathcal{P}})^* \cap B) \in \mathcal{F}^{\mathcal{U}}. \tag{2.4.1}$$

By assumption we have $\mathbb{P}(\Omega) = 1$ for all $\mathbb{P} \in \mathcal{P}$. By definition of the $()^*$ operation

$$\Omega^* = ((\Omega^{\mathcal{P}})^* \setminus B)^*_\Phi = (\Omega^{\mathcal{P}} \setminus B)^*$$

follows. To see the above equality, take a martingale measure $\mathbb{Q} \in \mathcal{M}_\Omega$ and assume that $\mathbb{Q}(\Omega \setminus (\Omega^{\mathcal{P}} \setminus B)) > 0$. As $\Omega \setminus (\Omega^{\mathcal{P}} \setminus B) = \Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^* \cap B$ we conclude that $\mathbb{Q}(\Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^*) > 0$. Since any calibrated martingale measure supported on a subset of Ω is in $\mathcal{M}_{\Omega^{\mathcal{P}}}$ this leads to a contradiction to the definition of $(\Omega^{\mathcal{P}})^*$. Also, $\Omega^{\mathcal{P}} \setminus B = \Omega^{\mathcal{P}} \cap B^c$ is the intersection of two analytic sets, so we conclude that Ω^* is analytic. Lastly, by definition of $\Omega^{\mathcal{P}}$ we conclude $\Omega^* = (\Omega^{\mathcal{P}})^*$ $\mathcal{P}$-q.s..

The implications *(1)* $\Rightarrow$ *(2)* $\Rightarrow$ *(3)* $\Rightarrow$ *(4)* $\Rightarrow$ *(5)* follow directly from the definition. Thus we only need to show *(5)* $\Rightarrow$ *(1)*. Let us fix $C \in \mathcal{S}$ such that $C \subseteq \Omega$. No Arbitrage de la Classe $\mathcal{S}$ on Ω implies that $\Omega^* \cap C \neq \emptyset$. From (2.4.1) we thus conclude that $\mathbb{P}((\Omega^{\mathcal{P}})^* \cap C) > 0$ for some $\mathbb{P} \in \mathcal{P}$. As $\Omega^* = (\Omega^{\mathcal{P}})^*$ $\mathcal{P}$-q.s. this implies $\mathbb{P}(\Omega^* \cap C) > 0$. Using a construction similar to the proof of Proposition 2.2.6 for the case $\Phi = 0$, we can find a measure $\tilde{\mathbb{P}} \in \mathcal{P}$ such that $\tilde{\mathbb{P}}(C) > 0$ and 0 is in the interior of the conditional support of $\tilde{\mathbb{P}}(\cdot|\Omega^*)$. By [Rokhlin, 2008, Theorem 1], we conclude that there exists a martingale measure $\mathbb{Q} \in \mathcal{Q}_{\mathcal{P}}$ equivalent to $\tilde{\mathbb{P}}(\cdot|\Omega^*)$, in particular $\mathbb{Q}(C) > 0$. The case $\Phi \neq 0$ can be treated similarly. □

2.4.2 Proof of Theorem 2.2.9

We first show that the quasi-sure superhedging theorem of Bouchard and Nutz [2015] implies the second part of Theorem 2.2.9.

Proposition 2.4.9. *Let Ω be an analytic subset of X and $\Omega^*_\Phi \neq \emptyset$. Let the set $\mathcal{P}$ satisfy (APS) and $\mathcal{N}^{\mathcal{P}} = \mathcal{N}^{\mathcal{M}^f_{\Omega,\Phi}}$ Then $\mathcal{N}^{\mathcal{Q}_{\mathcal{P},\Phi}} = \mathcal{N}^{\mathcal{M}^f_{\Omega,\Phi}}$ and for an upper semianalytic function $g : X \to \mathbb{R}$*

$$\sup_{\mathbb{Q} \in \mathcal{M}^f_{\Omega,\Phi}} \mathbb{E}_{\mathbb{Q}}[g] = \pi_{\Omega^*_\Phi}(g) = \pi^{\mathcal{P}}(g) = \sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g]. \tag{2.4.2}$$

Proof. That $\mathcal{Q}_{\mathcal{P},\Phi}$ and $\mathcal{M}^f_{\Omega,\Phi}$ have the same polar sets follows by the definition of Ω^*_Φ and [Burzoni et al., 2019, Lemma 2]. We now show (2.4.2): consider

$$\mathcal{P}^\Omega := \mathfrak{P}^f(\Omega^*_\Phi).$$

Note that there is no $\mathcal{M}^f_{\Omega,\Phi}$-q.s. arbitrage iff there is no $\mathcal{P}^\Omega$-q.s arbitrage.
We now show that Ω^*_Φ is analytic if Ω is analytic. Recall the set $\mathfrak{P}_{Z,\Phi}$ from Lemma 5.4 of Burzoni et al. [2017a], page 13 defined by

$$\mathfrak{P}_{Z,\Phi} := \left\{ \mathbb{P} \in \mathfrak{P}^f(X) \mid \exists \mathbb{Q} \in \mathcal{M}^f_{X,\Phi} \text{ such that } \frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{c(\mathbb{P})}{1+Z} \right\},$$

where $Z = \max_{i=1,\dots,d} \max_{t=0,\dots,T} S^i_t$ and $c(\mathbb{P}) = (\mathbb{E}_\mathbb{P}[1+Z]^{-1})^{-1}$. Burzoni et al. [2017a] show that the set

$$\{(\omega, \mathbb{P}) \mid \omega \in X^*,\ \mathbb{P} \in \mathfrak{P}^\omega\}$$

is analytic, where $\mathfrak{P}^\omega = \{\mathbb{P} \in \mathfrak{P}_{Z,\Phi} \mid \mathbb{P}(\{\omega\}) > 0\}$. Their results furthermore imply (see [Burzoni et al., 2017a, Remark 5.6])

$$\{(\omega, \mathbb{P}) \mid \omega \in X^*,\ \mathbb{P} \in \mathfrak{P}^\omega\} \cap \Big(\Omega \times \mathfrak{P}^f(\Omega)\Big)$$

is analytic and the projection of the above set to the first coordinate is exactly Ω^*_Φ, so Ω^*_Φ is analytic. We note $\omega \mapsto \mathcal{P}^\Omega_t(\omega) = \mathfrak{P}^f(\mathrm{proj}_{t+1}(\Sigma^\omega_t \cap \Omega^*_\Phi))$ has analytic graph by exactly the same argument as in the proof of Proposition 2.4.1 replacing Ω by Ω^*_Φ. The result now follows from the Superhedging Theorem of Bouchard and Nutz [2015] and the definition of $\mathcal{M}^f_{\Omega,\Phi}$. □

We now show that the classical $\mathbb{P}$-a.s. one-step superhedging duality can be deduced by means of pathwise reasoning:

Lemma 2.4.10. *Let $t \in \{0, \dots, T-1\}$ and $g : X_{t+1} \to \mathbb{R}$ be $\mathcal{F}^\mathcal{U}_{t+1}$-measurable. Let $\mathbb{P} \in \mathfrak{P}(X_1)$ and fix $\omega \in X_t$ such that $\boldsymbol{NA}(\mathbb{P})$ holds for the one-period model $(S_t(\omega), S_{t+1}(\omega, \cdot))$. Then*

$$\sup_{\mathbb{Q} \sim \mathbb{P},\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_\mathbb{Q}[g(\omega, \cdot)] = \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega, \cdot) \geq g(\omega, \cdot)\ \mathbb{P}\text{-a.s.}\}.$$

Proof. As g is $\mathcal{F}^\mathcal{U}_{t+1}$-measurable, by [Bertsekas and Shreve, 1978, Lemma 7.27, p.173] there exists a Borel-measurable function $\tilde{g} : (\mathbb{R}^d)^{t+1} \to \mathbb{R}$ such that $g(\omega) = \tilde{g}(S_{0:t+1}(\omega))$ for $\mathbb{P}$-a.e. $\omega \in X_{t+1}$. Assume first that $S_{t+1} \mapsto \tilde{g}(S_{0:t}(\omega), S_{t+1})$ is continuous. Define $\chi^\mathbb{P} := \mathrm{supp}(\mathbb{P} \circ$

$\Delta S_{t+1}(\omega,\cdot)^{-1})$. Then as $\mathbf{NA}(\mathbb{P})$ holds $\chi^{\mathbb{P}} = (\chi^{\mathbb{P}})^*$ and thus by [Burzoni et al., 2019, Theorem 2] and continuity of $S_{t+1} \mapsto \tilde{g}(S_{0:t}(\omega), S_{t+1})$ as well as $S_{t+1} \mapsto H(S_{t+1} - S_t(\omega))$

$$
\begin{aligned}
\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[g(\omega,\cdot)] &\le \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge g(\omega,\cdot)\ \mathbb{P}\text{-a.s.}\} \\
&= \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge \tilde{g}(S_{0:t}(\omega),\cdot) \text{ on } \chi^{\mathbb{P}}\} \\
&= \sup_{\mathbb{Q}\in\mathcal{M}^f_{\chi^{\mathbb{P}}}} \mathbb{E}_{\mathbb{Q}}[\tilde{g}(S_{0:t}(\omega),\cdot)] \\
&= \sup_{\mathbb{Q}\sim\mathbb{P}\circ\Delta S_{t+1}(\omega,\cdot)^{-1},\ \mathbb{Q}\in\mathcal{M}_{\mathbb{R}^d}} \mathbb{E}_{\mathbb{Q}}[\tilde{g}(S_{0:t}(\omega),\cdot)] \\
&= \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[g(\omega,\cdot)].
\end{aligned}
$$

If $S_{t+1} \mapsto \tilde{g}(S_{0:t}(\omega), S_{t+1})$ is Borel-measurable, then by Lusin's theorem (see [Cohn, 1980, Theorem 7.4.3, p.227]) there exists an increasing sequence of compact sets $(K_n)_{n\in\mathbb{N}}$ such that $K_n \subseteq \chi^{\mathbb{P}}$, $\mathbb{P}\circ\Delta S_{t+1}(\omega,\cdot)^{-1}(K_n^c) \le 1/n$ and $\tilde{g}(S_{0:t}(\omega),\cdot)|_{K_n}$ is continuous. In particular there exists $n_0\in\mathbb{N}$ such that for all $n \ge n_0$ we have $K_n = (K_n)^*$. By the above argument

$$
\begin{aligned}
&\inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge \tilde{g}(S_{0:t}(\omega),\cdot) \text{ on } K_n\} \\
&\quad = \sup_{\mathbb{Q}\sim\mathbb{P}\circ\Delta S_{t+1}(\omega,\cdot)^{-1}\ \ \mathbb{Q}\in\mathcal{M}_{K_n}} \mathbb{E}_{\mathbb{Q}}[g(\omega,\cdot)]
\end{aligned} \tag{2.4.3}
$$

holds for $n \ge n_0$. The claim now follows by taking suprema in $n\in\mathbb{N}$ on both sides of (2.4.3):

$$
\begin{aligned}
&\inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge g(\omega,\cdot)\ \mathbb{P}\text{-a.s.}\} \\
= &\sup_{n\in\mathbb{N}} \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge \tilde{g}(S_{0:t}(\omega),\cdot) \text{ on } K_n\} \\
= &\sup_{n\in\mathbb{N}} \sup_{\mathbb{Q}\sim\mathbb{P}\circ\Delta S_{t+1}(\omega,\cdot)^{-1}\ \ \mathbb{Q}\in\mathcal{M}_{K_n}} \mathbb{E}_{\mathbb{Q}}[\tilde{g}(S_{0:t}(\omega),\cdot)] \\
\le &\sup_{\mathbb{Q}\sim\mathbb{P}\circ\Delta S_{t+1}(\omega,\cdot)^{-1}\ \ \mathbb{Q}\in\mathcal{M}_{\mathbb{R}^d}} \mathbb{E}_{\mathbb{Q}}[\tilde{g}(S_{0:t}(\omega),\cdot)] \\
\le &\inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge g(\omega,\cdot)\ \mathbb{P}\text{-a.s.}\}.
\end{aligned}
$$

□

Using this one-step duality result under fixed $\mathbb{P}$ and (APS) of $\mathcal{P}$ we now prove the first part of Theorem 2.2.9, which is restated in the following proposition:

Proposition 2.4.11. *Let $\mathbf{NA}(\mathcal{P})$ hold and let $g: X \to \mathbb{R}$ be upper semianalytic. Then there exists a measure $\mathbb{P}^g = \mathbb{P}^g_0 \otimes \cdots \otimes \mathbb{P}^g_{T-1}$ and an $\mathcal{F}^{\mathcal{U}}$-measurable set $\Omega^{\mathcal{P}}_g$ with $\mathbb{P}(\Omega^{\mathcal{P}}_g) = 1$ for all $\mathbb{P}\in\mathcal{P}$, such that*

$$
\pi^{\mathcal{P}}(g) = \pi^{\hat{\mathbb{P}}}(g) = \pi_{(\Omega^{\mathcal{P}}_g)^*_{\Phi}}(g) = \sup_{\mathbb{Q}\in\mathcal{M}_{\Omega^{\mathcal{P}}_g,\Phi}} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\mathbb{Q}\in\mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g].
$$

Proof. We note that by $\mathbf{NA}(\mathcal{P})$ and Theorem 2.2.6 the difference $\Omega^{\mathcal{P}} \setminus (\Omega^{\mathcal{P}})^*_{\Phi}$ is $\mathcal{P}$-polar. We first take $\Phi = 0$. Recall the definition of the one-step functionals given in [Bouchard and Nutz, 2015, Lemma 4.10, p. 846]

$$\begin{aligned}
\mathcal{E}_T(g)(\omega) &= g(\omega) \\
\mathcal{E}_t(g)(\omega) &= \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)], \qquad t = 0, \dots, T-1.
\end{aligned}$$

By (APS) and upper semianalyticity of g, every $\mathcal{E}_t(g)$ is upper semianalytic. We show recursively that for every $t = 0, \dots, T-1$ and for $\mathcal{P}$-q.e. $\omega \in X_t$ there exists a measure $\mathbb{P} \in \mathfrak{P}(X_1)$ such that $\mathbf{NA}(\mathbb{P})$ holds and

$$\sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] = \sup_{\mathbb{Q} \sim \mathbb{P},\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)].$$

Note that by measurable selection arguments detailed in Bouchard and Nutz [2015] and construction of $\Omega^{\mathcal{P}}$ we conclude that for $\mathcal{P}$-q.e. $\omega \in X_t$ the properties $\mathbf{NA}(\mathcal{P}_t(\omega))$ and $\mathbb{P}(\mathrm{proj}_{t+1}(\Omega^{\mathcal{P}} \cap \Sigma_t^{\omega})) = 1$ hold for all $\mathbb{P} \in \mathcal{P}_t(\omega)$. We now fix $t \in \{0, \dots, T-1\}$ and $\omega \in X_t$ such that $\mathbf{NA}(\mathcal{P}_t(\omega))$ and $\mathbb{P}(\mathrm{proj}_{t+1}(\Omega^{\mathcal{P}} \cap \Sigma_t^{\omega})) = 1$ for all $\mathbb{P} \in \mathcal{P}_t(\omega)$ holds. Note that there exists a sequence $(\mathbb{P}_n)_{n \in \mathbb{N}}$ such that $\mathbb{P}_n \in \mathcal{P}_t(\omega)$ for all $n \in \mathbb{N}$ and

$$\sup_{\mathbb{Q} \ll \mathbb{P}_n,\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] \uparrow \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] \quad (n \to \infty).$$

We see from the proof of Theorem 2.2.6 for $\Phi = 0$ in Section 2.4.1 that under $\mathbf{NA}(\mathcal{P}_t(\omega))$ and for a fixed $\mathbb{P} \in \mathcal{P}_t(\omega)$, we can always find $\tilde{\mathbb{P}} \in \mathcal{P}_t(\omega)$ such that $\mathbb{P} \ll \tilde{\mathbb{P}}$ and $\mathbf{NA}(\tilde{\mathbb{P}})$ holds. Thus we can assume without loss of generality that $\mathbf{NA}(\mathbb{P}_n)$ holds for all $n \in \mathbb{N}$. Define $\hat{\mathbb{P}}_n := \sum_{k=1}^n 2^{-k}/(1-2^{-n})\mathbb{P}_k \in \mathcal{P}_t(\omega)$ as well as $\mathbb{P}_t^g(\omega) := \sum_{k=1}^{\infty} 2^{-k}\mathbb{P}_k$ and note that $\mathbf{NA}(\hat{\mathbb{P}}_n)$ as well as $\mathbf{NA}(\mathbb{P}_t^g(\omega))$ hold for all $n \in \mathbb{N}$. Furthermore

$$\begin{aligned}
\mathcal{E}_t(g)(\omega) &= \sup_{n \in \mathbb{N}} \sup_{\mathbb{Q} \ll \mathbb{P}_n,\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] \le \sup_{n \in \mathbb{N}} \sup_{\mathbb{Q} \sim \hat{\mathbb{P}}_n,\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] \\
&\le \sup_{\mathbb{Q} \sim \mathbb{P}_t^g(\omega),\ \mathbb{Q} \in \mathcal{M}_{X_1}} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega, \cdot)] \\
&= \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega, \cdot) \ge \mathcal{E}_{t+1}(g)(\omega, \cdot))\ \mathbb{P}_t^g(\omega)\text{-a.s.}\},
\end{aligned}$$

where the last equality follows from Lemma 2.4.10. Define

$$\begin{aligned}
\pi_t^{\mathbb{P}_t^g(\omega)}(g) &= \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega, \cdot) \ge \mathcal{E}_{t+1}(g)(\omega, \cdot)\ \mathbb{P}_t^g(\omega)\text{-a.s.}\} \\
\pi^{\mathcal{P}_t(\omega)}(g) &= \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H\Delta S_{t+1}(\omega, \cdot) \ge \mathcal{E}_{t+1}(g)(\omega, \cdot)\ \mathcal{P}_t(\omega)\text{-q.s.}\}.
\end{aligned}$$

Clearly $\pi^{\mathbb{P}_t^g(\omega)}(g) \le \pi^{\mathcal{P}_t(\omega)}(g)$. Now assume towards a contradiction that the inequality is strict and set $\varepsilon := \pi^{\mathcal{P}_t(\omega)}(g) - \pi^{\mathbb{P}_t^g(\omega)}(g) > 0$. Furthermore note that for a sequence of compact sets $(K_n)_{n\in\mathbb{N}}$ such that $K_n \uparrow \mathbb{R}^d$ we have

$$\begin{aligned}\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g) &:= \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x+H\Delta S_{t+1}\geq \mathcal{E}_{t+1}(g)\\ &\qquad\qquad \mathbb{P}^g_t(\omega)(\cdot|\Delta S_{t+1}(\omega,\cdot)\in K_n)\text{-a.s.}\}\\ &\uparrow \pi^{\mathbb{P}^g_t(\omega)}(g) \quad (n\to\infty),\\ \pi^{\mathcal{P}_t(\omega)|_{K_n}}(g) &:= \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ s.t. } x+H\Delta S_{t+1}\geq \mathcal{E}_{t+1}(g)\ \mathcal{P}_t(\omega)|_{K_n}\text{-q.s.}\}\\ &\uparrow \pi^{\mathcal{P}_t(\omega)}(g) \quad (n\to\infty),\end{aligned}$$

where

$$\mathcal{P}_t(\omega)|_{K_n} := \{\mathbb{P}(\cdot|\Delta S_{t+1}(\omega,\cdot)\in K_n) \mid \mathbb{P}\in\mathcal{P}_t(\omega),\ \mathbb{P}(\Delta S_{t+1}(\omega,\cdot)\in K_n)>0\}.$$

Choose $n\in\mathbb{N}$ large enough, such that $\pi^{\mathcal{P}_t(\omega)|_{K_n}}(g)-\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)>3\varepsilon/4$. Denote by $\mathcal{H}_{K_n}$ the closed set of $H\in\mathbb{R}^d$ such that

$$\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)+\varepsilon/2+H\Delta S_{t+1}(\omega,\cdot)\geq \mathcal{E}_{t+1}(g)(\omega,\cdot)) \quad \mathbb{P}^g_t(\omega)(\cdot|\Delta S_{t+1}(\omega,\cdot)\in K_n)\text{-a.s.}$$

Then for every $H\in\mathcal{H}_{K_n}$ there exists $\mathbb{P}^H_n\in\mathcal{P}_t(\omega)$ such that

$$\mathbb{P}^H_n(\{\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)+\varepsilon/2+H\Delta S_{t+1}(\omega,\cdot)<\mathcal{E}_{t+1}(g)(\omega,\cdot))\}\cap\{\Delta S_{t+1}(\omega,\cdot)\in K_n\})>0.$$

Note that there exists a countable sequence $(H^k_n)_{k\in\mathbb{N}}$, which is dense in $\mathcal{H}_{K_n}$. In particular for every $H\in\mathbb{R}^d$ such that

$$\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)+\varepsilon/4+H\Delta S_{t+1}(\omega,\cdot)\geq \mathcal{E}_{t+1}(g)(\omega,\cdot))\ \mathbb{P}^g_t(\omega)(\cdot|\Delta S_{t+1}(\omega,\cdot)\in K_n)\text{-a.s.}$$

there exists $k\in\mathbb{N}$ such that

$$\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)+\varepsilon/2+H^k_n\Delta S_{t+1}(\omega,\cdot)\geq \mathcal{E}_{t+1}(g)(\omega,\cdot))\ \mathbb{P}^g_t(\omega)(\cdot|\Delta S_{t+1}(\omega,\cdot)\in K_n)\text{-a.s.}$$

Set now $\mathbb{P}^n=\sum_{k=1}^\infty 2^{-k}\mathbb{P}^{H^k_n}_n\in\mathfrak{P}(X_1)$ and note that for all $n\in\mathbb{N}$ large enough

$$\pi^{\frac{1}{2}(\mathbb{P}^g_t(\omega)+\mathbb{P}^n)|_{K_n}}(g)-\pi^{\mathbb{P}^g_t(\omega)|_{K_n}}(g)\geq\varepsilon/4.$$

Taking $K_n\uparrow\mathbb{R}^d$ we have in particular

$$\mathcal{E}_t(g)(\omega)\leq \sup_{\mathbb{Q}\sim\mathbb{P}^g_t(\omega),\ \mathbb{Q}\in\mathcal{M}_{X_1}}\mathbb{E}_\mathbb{Q}[g] < \lim_{n\to\infty}\sup_{\mathbb{Q}\sim\frac{1}{2}(\mathbb{P}^g_t(\omega)+\mathbb{P}^n)|_{K_n},\ \mathbb{Q}\in\mathcal{M}_{X_1}}\mathbb{E}_\mathbb{Q}[g]\leq\mathcal{E}_t(g)(\omega),$$

a contradiction. Thus

$$\mathcal{E}_t(g)(\omega)=\pi^{\mathcal{P}_t(\omega)}(g)=\pi^{\mathbb{P}^g_t(\omega)}(g).$$

As $0\in\operatorname{ri}(\operatorname{supp}((\mathbb{P}^g_t(\omega)))$ a natural universally measurable candidate for a superhedging strategy $\omega\mapsto H_{t+1}(\omega)$ is the right derivative $\lim_{\varepsilon\in\mathbb{Q},\ \varepsilon\downarrow 0}(\mathcal{E}^{\varepsilon e_i}_t(g)(\omega)-\mathcal{E}_t(g)(\omega))/\varepsilon$ where

$\mathcal{E}_t^{\varepsilon e_i}(g)(\omega)$ is the superhedging price for the Borel-measurable stock $(S_t+\varepsilon e_i)$, $i=1,\dots,d' \leq d$ instead of S_t. This is a pointwise limit of differences of upper seminanalytic functions and thus universally measurable. For $\omega \in X_t$ such that this quantity does not exist, we set $H_{t+1}(\omega)=0$. Furthermore in order to show that the map $\omega \mapsto \mathbb{P}_t^g(\omega)$ can be chosen to be universally measurable we first note that in [Bouchard and Nutz, 2015, Lemma 4.8, p.843] the set

$$\{(\mathbb{Q},\mathbb{P}) \in \mathfrak{P}(\mathrm{proj}_{t+1}(\Sigma_t^\omega)) \times \mathfrak{P}(\mathrm{proj}_{t+1}(\Sigma_t^\omega)) \mid \\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0,\ \mathbb{P} \in \mathcal{P}_t(\omega),\ \mathbb{Q} \ll \mathbb{P}\}$$

is analytic. Thus we can apply the Jankov-von-Neumann selection theorem (see [Bertsekas and Shreve, 1978, Proposition 7.50, p.184]) to find $1/n$-optimisers $(\mathbb{Q}_t^n(\omega), \mathbb{P}_t^n(\omega))$ for $\mathcal{E}_t(g)(\omega)$ and the claim follows. The case $\Phi \neq 0$ can be handled by induction as in the proof of Theorem 2.2.6 for $\Phi \neq 0$.
In conclusion we have found a strategy $(h,H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}})$ such that

$$\sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g] + h\cdot\Phi + (H \circ S_T) \geq g \quad \mathcal{P}\text{-q.s.}$$

We now define

$$\Omega_g^{\mathcal{P}} = \Omega^{\mathcal{P}} \cap \left\{ \omega \in X \;\middle|\; \sup_{\mathbb{Q} \in \mathcal{Q}_{\mathcal{P},\Phi}} \mathbb{E}_{\mathbb{Q}}[g] + h\cdot\Phi(\omega) + (H \circ S_T)(\omega) \geq g(\omega) \right\} \in \mathcal{F}^{\mathcal{U}}.$$

This concludes the proof. □

Remark 2.4.12. By $\mathbf{NA}(\mathcal{P})$ Proposition 2.4.11 implies for $g=0$

$$\begin{aligned} 0 &= \inf\{x \in \mathbb{R} \mid \exists (h,H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}}) \text{ such that } x + h\cdot\Phi + (H\circ S_T) \geq 0\ \mathcal{P}\text{-q.s.}\} \\ &= \inf_{\tilde{g} \in \mathfrak{E}_0} \inf\{x \in \mathbb{R} \mid \exists (h,H) \in \mathcal{H}_\Phi(\mathbb{F}^{\mathcal{U}}) \text{ such that } x + h\cdot\Phi + (H\circ S_T) \geq \tilde{g} \ \text{ on } (\Omega_{\tilde{g}}^{\mathcal{P}})_\Phi^*\} \\ &= \inf_{\tilde{g} \in \mathfrak{E}_0} \sup_{\mathbb{Q} \in \mathcal{M}_{\Omega_{\tilde{g}}^{\mathcal{P}},\Phi}} \mathbb{E}_{\mathbb{Q}}[\tilde{g}], \end{aligned}$$

where we define

$$\mathfrak{E}_0 = \{\tilde{g} : X \to (-\infty, 0]\ \mathcal{F}^{\mathcal{U}}\text{-measurable} \mid \tilde{g} = 0\ \mathcal{P}\text{-q.s.}\}.$$

A similar result was obtained by Burzoni et al. [2017b] in a more general setup. Aggregating the martingale measures corresponding to all $\tilde{g}$ (and thus to all $\mathcal{P}$-polar sets) to achieve a result comparable to Bouchard and Nutz [2015] in a setup without using (APS) of $\mathcal{P}$ remains an open problem.

2.4.3 Proof of Theorem 2.3.11 for $T > 1$

Let us recall the notational conventions introduced in Section 2.3.2. In this section we additionally assume that all trading strategies $H \in \mathcal{H}(\mathbb{F})$ are bounded and continuous. This can be done without loss of generality as was observed in Acciaio et al. [2013].

Proof of Theorem 2.3.11 for $T > 1$. We only need to prove *(1)* for $T > 1$. This follows by backwards induction: indeed, let $H \in \mathcal{H}(\mathbb{F})$ be a strong arbitrage. Let us define the compact sets

$$K_t = \{(\omega_0, \omega_1, \dots, \omega_T) \in \Omega \mid \omega_0 = s_0,\ \omega_{t+1}^i \in [0, (\omega_t^i + 2(|\omega_t|_1 \vee |s_0|_1))] \text{ for all } 1 \le i \le d\} \subseteq (\mathbb{R}^d)^t$$

for $0 \le t \le T-1$. Consider the correspondence

$$\Xi_t : \omega \twoheadrightarrow \Sigma_t^\omega \cap K_t.$$

Then for all closed $C \subseteq X$

$$\{\omega \in \Omega \mid \Xi_t(\omega) \cap C \neq \emptyset\} = S_{0:t}^{-1}(S_{0:t}(C \cap K_t))$$

is Borel measurable as the projection of compact sets is compact. Consequently Ξ_t is weakly measurable. As Ξ_t is closed-valued, there exists a Castaing-representation $\{\xi_t^n \mid n \in \mathbb{N}\}$ of Ξ_t (see [Rockafellar and Wets, 1998, Theorem 14.5(a), p.646]). Now we take $t = T-1$. For ω we define

$$\Delta \xi_{T-1}^n(\omega) := \xi_{T-1}^n(\omega) - S_{T-1}(\omega)$$

and set

$$\psi_{T-1}(\omega) := \inf_{n \in \mathbb{N}} H_T(\omega) \Delta(\xi_{T-1}^n(\omega)) = \inf_{\tilde{\omega} \in \Xi_{T-1}(\omega)} H_T(\omega) \Delta S_T(\tilde{\omega}),$$

which is Borel-measurable. By Blackwell's theorem (cf. [Cohn, 1980, Theorem 8.6.7, p.291]), ψ_{T-1} is then also $\mathcal{F}_{T-1} = \mathcal{F}_{T-1}^0$-measurable. We are now in a position to construct a modified uniformly strong arbitrage strategy $\tilde{H}$: for this we define the set $A_{T-1} = \{\omega \in \Omega \mid H_1(\omega)\Delta S_1(\omega) \le 0, \dots, H_{T-1}(\omega)\Delta S_{T-1}(\omega) \le 0\}$ and set for any $\mathcal{F}_{T-1}$-measurable function f_{T-1}

$$\begin{aligned}
\tilde{H}_T(\omega, f_{T-1}) = &\sum_{n=1}^\infty 2((|S_{T-1}(\omega)|_1 \vee |s_0|_1) + f_{T-1}^-) n H_T(\omega) \mathbb{1}_{\{1/(n-1) \ge \psi_{T-1}(\omega) > 1/n\}} \\
&+ \mathbf{e}(1 + f_{T-1}^- / |S_{T-1}(\omega)|_1) \quad \text{on } \Omega \cap \{\psi_{T-1}(\omega) > 0\}, \\
\tilde{H}_T(\omega, f_{T-1}) = 0 \quad &\text{otherwise.}
\end{aligned}$$

As H is a strong arbitrage strategy and K_{T-1} is compact, we note that $A_{T-1} \subseteq \{\psi_{T-1} > 0\}$ and as in the one-period case

$$\tilde{H}_T(\omega, f_{T-1})\Delta S_T(\omega, \cdot) \geq |s_0|_1 + f_{T-1}^-(\omega) \quad \text{on } \Omega$$

if $\psi_{T-1}(\omega) > 0$. Next we define

$$\perp_{T-2}(\omega) = \{\tilde{\omega} \in \Sigma_{T-2}^{\omega} \cap \Omega \mid H_{T-1}(\omega)\Delta S_{T-1}(\tilde{\omega}) = 0\}.$$

We set

$$A_{T-2} := \{\omega \in \Omega \mid H_1(\omega)\Delta S_1(\omega) \leq 0, \dots, H_{T-2}(\omega)\Delta S_{T-2}(\omega) \leq 0\}.$$

Let us now fix $\omega \in A_{T-2}$, such that $\{\tilde{\omega} \in \Sigma_{T-2}^{\omega} \cap \Omega \mid H_{T-1}(\tilde{\omega})\Delta S_{T-1}(\tilde{\omega}) > 0\} \neq \emptyset$. As $A_{T-2} \in \mathcal{F}_{T-2}$, we will write $\omega = S_{0:T-2}(\omega)$ from now on to shorten notation, as long as the meaning is unambiguous. In order to prepare the induction step, we first have to treat $\omega \in \Omega$ for which $H_{T-1}(\omega)\Delta S_{T-1}(\omega)$ is close to zero. More concretely we aim is to show that $\psi_{T-1}(\omega, \tilde{\omega}_{T-1})$ is bounded from below by a strictly positive number for all $(\omega, \tilde{\omega}_{T-1})$ in a neighbourhood of $\perp_{T-2}(\omega) \cap \Xi_{T-2}(\omega)$.
For this we note that compactness of $S_{0:T-1}(\Xi_{T-2}(\omega))$ implies that

$$\tilde{\omega}_{T-1} \mapsto \Xi_{T-1}(\omega, \tilde{\omega}_{T-1})$$

is uniformly continuous in Hausdorff distance on $S_{T-1}(\Xi_{T-2}(\omega))$: indeed take a sequence $(\omega, \omega_{T-1}^n)_{n\in\mathbb{N}}$ such that $(\omega, \omega_{T-1}^n) \in S_{0:T-1}(\Xi_{T-2}(\omega))$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}(\omega, \omega_{T-1}^n) = (\omega, \omega_{T-1}) \in S_{0:T-1}(\Xi_{T-2}(\omega))$. Assume towards a contradiction that there exists $\varepsilon > 0$ such that for every $n \in \mathbb{N}$ there exists $(\omega, \omega_{T-1}^n, \omega_T^n) \in \Xi_{T-1}((\omega, \omega_{T-1}^n))$ such that

$$\inf_{(\omega, \omega_{T-1}, \tilde{\omega}_T) \in \Xi_{T-1}(\omega, \omega_{T-1})} \left|(\omega, \omega_{T-1}, \tilde{\omega}_T) - (\omega, \omega_{T-1}^n, \omega_T^n)\right| \geq \varepsilon.$$

Then after a subsequence $\lim_{n\to\infty}(\omega, \omega_{T-1}^n, \omega_T^n) = (\omega, \omega_{T-1}, \omega_T) \in \Xi_{T-1}(\omega, \omega_{T-1})$, a contradiction. This shows the claim.
Now we fix ω_{T-1} such that $(\omega, \omega_{T-1}) \in \perp_{T-2}(\omega) \cap \Xi_{T-2}(\omega)$. By the one-period case there exists $\varepsilon > 0$ such that

$$H_T(\omega, \omega_{T-1})\Delta S_T((\omega, \omega_{T-1}, \cdot)) \geq \varepsilon \quad \text{on } \Xi_{T-1}(\omega, \omega_{T-1}),$$

(recall that $\omega \in A_{T-2}$). By uniform continuity of $\tilde{\omega}_{T-1} \mapsto \Xi_{T-1}(\omega, \tilde{\omega}_{T-1})$ and $\tilde{\omega}_{T-1} \mapsto H_T(\omega, \tilde{\omega}_{T-1})$ on $S_{T-1}(\Xi_{T-2}(\omega))$ there exists $\delta > 0$ such that

$$H_T(\omega, \tilde{\omega}_{T-1})\Delta S_T(\omega, \tilde{\omega}_{T-1}, \cdot) \geq \varepsilon/2 \quad \text{on } \Xi_{T-1}(\omega, \tilde{\omega}_{T-1})$$

for all $\tilde{\omega}_{T-1}$ with $|\tilde{\omega}_{T-1}-\omega_{T-1}|\le\delta$. By compactness of $S_{0:T-1}(\Xi_{T-2}(\omega))$ there exists $\tilde{\delta}>0$, $\tilde{\varepsilon}>0$ such that for all $(\omega,\tilde{\omega}_{T-1})\in S_{0:T-1}(\Xi_{T-2}(\omega)\cap A_{T-2})$ with

$$\inf_{(\omega,\omega_{T-1})\in\Xi_{T-2}(\omega)\cap\perp_{T-2}(\omega)}|(\omega,\omega_{T-1})-(\omega,\tilde{\omega}_{T-1})|\le\tilde{\delta}$$

we have $H_T(\omega,\tilde{\omega}_{T-1})\Delta S_T(\omega,\tilde{\omega}_{T-1},\cdot)\ge\tilde{\varepsilon}$ on $\Xi_{T-1}(\omega,\tilde{\omega}_{T-1})$. This shows that $\psi_{T-1}(\omega,\tilde{\omega}_{T-1})>0$ for $(\omega,\tilde{\omega}_{T-1})$ in a neighbourhood of $\perp_{T-2}(\omega)\cap\Xi_{T-2}(\omega)$.
Now we define

$$\psi_{T-2}(\omega)=\inf_{\{(\omega,\tilde{\omega}_{T-1})\in\Xi_{T-2}(\omega)\ |\ \psi_{T-1}(\omega,\tilde{\omega}_{T-1})\le 0\}}H_{T-1}(\omega)\Delta S_{T-1}(\omega,\tilde{\omega}_{T-1}),$$

which is $\mathcal{F}_{T-2}$-measurable. By the arguments above we find $\psi_{T-2}(\omega)\in\{-\infty\}\cup(0,\infty)$ on $\Xi_{T-2}(\omega)\cap A_{T-2}$. Next we define for an $\mathcal{F}_{T-2}$ measurable function f_{T-2}

$$\begin{aligned}\tilde{H}_{T-1}(\omega,f_{T-2})&=\sum_{n=1}^{\infty}2((|S_{T-2}(\omega)|_1\vee|s_0|_1)+f_{T-2}^-)nH_{T-1}(\omega)\mathbb{1}_{\{1/(n-1)\ge\psi_{T-2}(\omega)>1/n\}}\\&\quad+\mathbf{e}(1+f_{T-2}^-/|S_{T-2}(\omega)|_1)\quad\text{on } A_{T-2}\\\tilde{H}_{T-1}(\omega,f_{T-2})&=0\quad\text{otherwise.}\end{aligned}$$

and note that

$$\tilde{H}_{T-1}(\omega,f_{T-2})\Delta S_{T-1}(\omega,\cdot)\ge|s_0|_1+f_{T-2}^-\qquad\text{for }(\omega,\cdot)\in A_{T-2}\cap\{\psi_{T-2}>0\}.$$

Thus in particular

$$\tilde{H}_{T-1}(\omega,f_{T-2})\Delta S_{T-1}(\omega)+\tilde{H}_T(\omega,f_{T-2}+\tilde{H}_{T-1}(\omega,f_{T-2})\Delta S_{T-1}(\omega))\Delta S_T(\omega)\ge|s_0|_1$$

for $\omega\in\Omega\cap A_{T-2}$. This concludes the induction step. The claim now follows setting

$$\begin{aligned}f_1(\cdot)&:=\tilde{H}_1(\cdot)\Delta S_1(\cdot),\\f_2(\cdot)&:=\sum_{t=1}^{2}\tilde{H}_t(\cdot,f_{t-1}(\cdot))\Delta S_t(\cdot),\\&\ \ \vdots\\f_{T-1}(\cdot)&:=\sum_{t=1}^{T}\tilde{H}_t(\cdot,f_{t-1}(\cdot))\Delta S_t(\cdot)\end{aligned}$$

and $A_0=\Omega$.
The rest of the proof follows as in Section 2.3.2 replacing S_1 by S_T.

□

Chapter 3

The robust superreplication problem: a dynamic approach

3.1 Introduction

We consider a discrete time financial market and an agent who needs to hedge a liability g maturing at a future date T in a robust and risk-conservative way. Our focus is on the interplay between the beliefs used for assessing the risks, the beliefs used for agent's investment decisions and the dynamics of agent's actions. For simplicity we assume away other factors and consider an agent who can trade in a dynamic way with no constraints or frictions in d assets available in the market at prices which are exogenous. More precisely, following the approach of Samuelson [1969] and Black and Scholes [1973a], risky assets are modelled as stochastic processes and their behaviour specified by a probability measure. However, unlike the classical uni-prior approach which fixes one such measure $\mathbb{P}$, we consider a multi-prior framework and work simultaneously under a whole family of measures $\mathbb{P} \in \mathcal{P}$. This offers a robust approach which accounts for model ambiguity, also referred to as *Knightian uncertainty* after Knight [1921].

The price to pay for a robust modelling view comes through specificity of outputs: while the uni-prior setting might generate a unique fair price for a derivate contract, a multi-prior setting will typically generate a relatively wide interval of no-arbitrage prices, a tradeoff first identified in the seminal paper of Merton [1973]. We consider a trader who, due to regulatory requirements or internal risk management reasons, is required to hedge g in a risk-conservative way relative to $\mathcal{P}$. This means that initially she has to allocate capital equal to $\pi(g)$, the superhedging price of g, i.e., the price of cheapest trading strategies which are guaranteed to cover the liability g under all $\mathbb{P} \in \mathcal{P}$. There might be many such cheapest superhedging strategies and the trader can pick any one of them to follow until time T. This is a conservative and non-linear risk assessment: the capital the trader would be allowed to borrow against a long position in g is $-\pi(-g)$ and is typically significantly lower than $\pi(g)$.

The superhedging price $\pi(g)$ can be characterised theoretically and has been considered in a number of papers, see Bouchard and Nutz [2015] and the discussion below. To the best of our knowledge, the focus of most of these works has been on the static problem: the problem today for the horizon T. In contrast, in this chapter we want to focus on the dynamics of the robust pricing and hedging problem *through time*. We ask how $\pi(g)$ changes *over time* and how the trader should act optimally through time. Clearly, tomorrow she will see new prices in the market and will be able to recompute the superhedging price. If the new price is lower, she will be able to unwind her old position, buy a new position and be left with a surplus. She could then consume this (e.g., pay into her credit line if the initial capital was borrowed) or invest further if she believes the market offers suitable opportunities.

Our first main contribution is to describe the evolution of $\pi_t(g)$ - the superhedging price at time t of the liability g at maturity T. We work in the setting of Bouchard and Nutz [2015] and consider an abstract set of priors $\mathcal{P}$, possibly large and in particular not dominated by a single probability measure. The measures $\mathbb{P} \in \mathcal{P}$ are represented as compositions of one-step kernels and to establish the dual characterisation of $\pi_0(g)$ Bouchard and Nutz [2015] have essentially proven a dynamic programming principle for the dual objects. We prove that $(\pi_t(g))_{0\leq t\leq T}$ satisfy a dynamic programming principle, and that $\pi_t(g)$ can be seen as a concave envelope of $\pi_{t+1}(g)$ evaluated at today's prices. To the best of our knowledge, this was first suggested in the robust setting by Dupire [2010]. We also characterise $\pi_t(g)$ as the wealth of a minimal superhedging strategy in the sense of Föllmer and Kramkov [1997]. These results provide natural robust extensions of classical uni-prior results, see Föllmer and Schied [2004], including a robust version of the algorithm in Carassus et al. [2006]. Further, considering $\mathcal{P}$ which corresponds to the pointwise robust setting of Burzoni et al. [2019], we show that $\pi_t(g)$ corresponds to the uni-prior superhedging price for an extreme $\mathbb{P} \in \mathcal{P}$. Proving our results in the robust setting requires rather lengthy and technical arguments. This is mainly due to delicate measurability questions.

Our second main contribution is to consider an optimal investment problem for a trader who is rolling over her robust superhedge. This is phrased as a problem of robust maximisation of expected utility of inter-temporal consumption subject to a robust superhedging constraint. Here the robust constraint means that superhedging has to be satisfied $\mathbb{P}$-a.s. for all $\mathbb{P} \in \mathcal{P}$. The robust utility maximisation means that we consider a max-min problem, where minimisation is over $\mathbb{P} \in \mathcal{P}^u$. We argue that the latter problem should be considered with respect to a different set of priors $\mathcal{P}^u \subseteq \mathcal{P}$ than the former problem. Measures

$\mathbb{P} \in \mathcal{P}^u$ no longer have to capture all regulatory or institutional risk views but rather represent trader's subjective views on market dynamics. In particular $\mathcal{P}^u$ does not have to be arbitrage free. Under suitable assumptions on the trader's utility functions, we show that optimal investment and consumption strategies exist and further specify when, and in what sense, these may be unique. We provide examples to illustrate that neither existence nor uniqueness have to hold when our assumptions are not satisfied.

Throughout, we work in the setup of Bouchard and Nutz [2015] who extended the classical uni-prior theory of pricing and hedging in discrete time to the robust mutli-prior case, introducing a suitable notion of no-arbitrage, proving a robust version of the fundamental theorem of pricing and hedging and establishing a robust pricing-hedging duality. Numerous authors have since adopted their setup and worked on robust extensions of the classical problems in quantitative finance such as pricing and hedging of American options, utility maximisation or transaction cost theory to name just a few examples, see Nutz [2016], Blanchard and Carassus [2017], Aksamit et al. [2018], Bayraktar and Zhou [2017], Bouchard et al. [2019] and the references therein. We note that alternative ways to address model uncertainty are possible, including the pathwise, or pointwise, approach developed in Davis and Hobson [2007], Acciaio et al. [2013], Burzoni et al. [2016, 2017a, 2019] among others. Whilst the resulting robust framework for pricing and hedging is equipped with different notions of arbitrage and different fundamental theorems, it was shown in Chapter 2 to be equivalent to the multi-prior approach. Thus, on an abstract level, there is no loss of generality in our choice to adopt the multi-prior approach of Bouchard and Nutz [2015]. It is important however that we work in discrete time. While in the classical setup no-arbitrage theory, including dynamic understanding of the superhedging price, is well developed in continuous time, see Föllmer and Kramkov [1997], Delbaen and Schachermayer [2006], in the robust setting an extension of abstract no-arbitrage theory, as developed in Bouchard and Nutz [2015] or Burzoni et al. [2019], to the continuous time is still open. This is despite a body of works which have achieved either particular or generic steps towards such a goal, large enough so that we can not do it justice in this introduction but refer to Avellaneda et al. [1996], Lyons [1995], Denis and Martini [2006], Cox and Obłój [2011], Denis and Kervarec [2007], Epstein and Ji [2014], Biagini et al. [2017], Hou and Obłój [2018], Beiglböck et al. [2017], Bartl et al. [2020] and the references therein.

We note that d may be large and our assets may include both primary and derivate assets. Indeed, one way of making robust outputs more specific is by including more traded assets in the analysis. This was the original motivation behind the works on the robust pricing and hedging in continuous time, going back to Hobson [1998.], where one typically assumes that the market prices of European options on the underlying assets co-maturing with our liability g are known. Here, we consider an abstract general setup and allow any d-tuple of traded assets, for a finite d. We may expect that the level of uncertainty regarding different assets may differ and this would be reflected in $\mathcal{P}$. It is important that all the assets are traded dynamically, as this is necessary to obtain a dynamic programming principle for the superhedging prices. This is without loss of generality, as discussed in Aksamit et al. [2018], in the sense that any setup of the kind which is specified in Bouchard and Nutz [2015] where some assets are only available for trading at time zero can be lifted to a setup with dynamic trading in all assets in a way which does not introduce arbitrage and does not affect time-zero superhedging prices, see Remark 3.3.3 below.

The remainder of the chapter is organised as follows. The next section introduces and discusses our modelling framework. Section 3.3 presents the results characterising the dynamics of the superhedging price. We then specialise, in Section 3.3.2, to the pathwise setting when $\mathcal{P}$ contains all measures with specified supports. This allows for a more intuitive interpretation of the results, easier proofs and explicit examples. Section 3.4 then considers the secondary utility maximisation problem for a trader who dynamically re-balances her superhedging strategy and states the existence and uniqueness results for the optimal investment and consumption strategies. Finally, proofs are presented in the last three sections.

3.2 Models of Financial markets

In this section we set up the multi-prior modelling framework and give introductory definitions. Future dynamics of financial assets are modelled using probability measures but, unlike the classical case where one such measure is fixed, we typically work simultaneously under all $\mathbb{P}$ from a large family of measures $\mathcal{P}$. Our market has d traded assets, these could be stocks or options, but importantly all are traded dynamically. We do not consider statically traded assets, i.e., only available for buy-and-hold trading, as then the superhedging prices typically can not admit a dynamic programming principle across all times, see Aksamit et al. [2018].

3.2.1 Uncertainty modelling

We work in the setting of Bouchard and Nutz [2015] to which we refer for details and motivation. We only recall the main objects of interest here and refer to Bertsekas and Shreve [1978][Chapter 7] for technical details. Let Ω be a Polish space and denote by Ω^t its t-fold Cartesian product. We define the price process S of discounted prices of d traded stocks as a Borel measurable map $S_t(\omega) = (S_t^1(\omega), \dots, S_t^d(\omega)) : \Omega^T \to \mathbb{R}_+^d$ for every $\omega = (\omega_0, \dots, \omega_T)$ with the convention $S_0(\omega) = s_0 \in \mathbb{R}_+^d$ and $T \in \mathbb{N}$ is the time horizon.

Prices are specified in discounted units and we have a riskless asset with price equal to 1 for all $0 \leq t \leq T$. Furthermore let $\mathfrak{P}(\Omega^t)$ be the set of all probability measures on $\mathcal{B}(\Omega^t)$, the Borel-σ-algebra on Ω^t. We denote by $\mathcal{F}_t^{\mathcal{U}}$ the universal completion of $\mathcal{B}(\Omega^t)$. We often consider $(\Omega^t, \mathcal{F}_t^{\mathcal{U}})$ as a subspace of $(\Omega^T, \mathcal{F}_T^{\mathcal{U}})$ and write $\mathbb{F}^{\mathcal{U}} = (\mathcal{F}_t^{\mathcal{U}})_{t=0,\dots,T}$. In the rest of the chapter, we will use the same notation for $\mathbb{P} \in \mathfrak{P}(\Omega^T)$ and for its (unique) extension to $\mathcal{F}_T^{\mathcal{U}}$, and we write supp($\mathbb{P}$) for the support of $\mathbb{P}$. For a given $\mathcal{P} \subseteq \mathfrak{P}(\Omega^T)$, $N \subset \Omega^T$ is called a $\mathcal{P}$-polar set if for all $\mathbb{P} \in \mathcal{P}$ there exists some $A_{\mathbb{P}} \in \mathcal{B}(\Omega^T)$ such that $\mathbb{P}(A_{\mathbb{P}}) = 0$ and $N \subset A_{\mathbb{P}}$. We say that a property holds $\mathcal{P}$-quasi-surely (q.s.), if it holds outside a $\mathcal{P}$-polar set. Finally we say that a set is of $\mathcal{P}$-full measure if its complement is a $\mathcal{P}$-polar set.
To give a probabilistic description of the market we consider a family of random sets $\mathcal{P}_t : \Omega^t \twoheadrightarrow \mathfrak{P}(\Omega)$, for all $0 \leq t \leq T-1$. The set $\mathcal{P}_t(\omega)$ can be seen as the set of all possible models for the $t+1$-th period given the path $\omega \in \Omega^t$ at time t. In order to aggregate trading strategies on different paths in a measurable way, we assume here that the sets $\mathcal{P}_t$ have the following property:

Assumption 3.2.1. *The set $\mathcal{P}$ has Analytic Product Structure (APS), i.e.,*

$$\mathcal{P} = \{\mathbb{P}_0 \otimes \cdots \otimes \mathbb{P}_{T-1} \mid \mathbb{P}_t \textit{ is an } \mathcal{F}_t^{\mathcal{U}}\textit{-measurable selector of } \mathcal{P}_t\},$$

where the sets $\mathcal{P}_t(\omega) \subseteq \mathcal{P}(\Omega)$ are nonempty, convex and

$$\textit{graph}(\mathcal{P}_t) = \{(\omega, \mathbb{P}) \mid \omega \in \Omega^t,\ \mathbb{P} \in \mathcal{P}_t(\omega)\}$$

is analytic.

The fact that graph($\mathcal{P}_t$) is analytic (see Bertsekas and Shreve [1978][Chapter 7] for a definition) allows for an application of the Jankov-von-Neumann theorem ([Bertsekas and Shreve, 1978, Prop. 7.49, p.182]), which guarantees the existence of universally measurable selectors $\mathbb{P}_t : \Omega^t \to \mathfrak{P}(\Omega)$. Here $\mathbb{P}_0 \otimes \cdots \otimes \mathbb{P}_{T-1}$ denotes the T-fold application of Fubini's theorem, which defines a measure on $\mathfrak{P}(\Omega^T)$. Indeed, analyticity of the graph of $\mathcal{P}_t$ is of paramount importance for the preservation of measurability properties. For example the proof of a quasi-sure superreplication theorem (see [Bouchard and Nutz, 2015, Lemma 4.10]) uses the fact that if $X_{t+1} : \Omega^{t+1} \to \mathbb{R}$ is upper semianalytic, then $\sup_{\mathbb{P} \in \mathcal{P}_t(\omega)} \mathbb{E}_{\mathbb{P}}[X_{t+1}(\omega, \cdot)]$ remains upper semianalytic. Apart from Assumption 3.2.1, we make no specific assumptions on the set of priors $\mathcal{P}$. It is neither assumed to be dominated by a given reference probability measure nor to be weakly compact. Some concrete examples, including when $\mathcal{P}_t(\omega)$ are non-compact random sets, are discussed in Section 3.3.2.

3.2.2 Trading

Trading strategies are represented by $\mathbb{F}^{\mathcal{U}}$-predictable d-dimensional processes $H := \{H_t\}_{1\le t\le T}$ where for all $1 \le t \le T$, H_t represents the investor's holdings in each of the d assets at time t. The set of trading strategies is denoted by $\mathcal{H}(\mathbb{F}^{\mathcal{U}})$. Investors are allowed to consume and their cumulative consumption is represented by an $\mathbb{R}$-valued $\mathbb{F}^{\mathcal{U}}$-adapted process $C = \{C_t\}_{1\le t\le T}$, $C_0 = 0$ and which is assumed to be non-decreasing: $C_t \le C_{t+1}$ $\mathcal{P}$-q.s. The set of cumulative consumption processes is denoted by $\mathcal{C}$. We will use the notation $\Delta S_t = S_t - S_{t-1}$ and $\Delta C_t = C_t - C_{t-1}$ for $1 \le t \le T$. Given an initial wealth $x \in \mathbb{R}$, a trading portfolio H and a cumulative consumption process C, the wealth process $V^{x,H,C}$ is governed by

$$\begin{aligned} V_0^{x,H,C} &= x \\ V_t^{x,H,C} &= V_{t-1}^{x,H,C} + H_t \Delta S_t - \Delta C_t \quad \text{for } 1 \le t \le T. \end{aligned} \tag{3.2.1}$$

The condition $C = 0$ means that the portfolio H is self-financing and in this case we write $V^{x,H}$ instead of $V^{x,H,0}$.

We are interested in superhedging of a (European) contingent claim and therefore adapt the presentation of Föllmer and Kramkov [1997] to the robust framework. A (European) contingent claim is represented by an $\mathcal{F}_T^{\mathcal{U}}$-measurable random variable g and the set of superhedging strategies for g is denoted by

$$\mathcal{A}(g) := \left\{ (x, H, C) \in \mathbb{R} \times \mathcal{H}(\mathbb{F}^{\mathcal{U}}) \times \mathcal{C} \;\middle|\; V_T^{x,H,C} \ge g \ \mathcal{P}\text{-q.s.} \right\}. \tag{3.2.2}$$

Definition 3.2.2. *The superreplication price $\pi(g)$ of an $\mathcal{F}_T^{\mathcal{U}}$-measurable random variable g is the minimal initial capital needed for superhedging g, i.e.,*

$$\pi(g) := \inf \left\{ x \in \mathbb{R} \mid \exists (H, C) \in \mathcal{H}(\mathbb{F}^{\mathcal{U}}) \times \mathcal{C} \ \textit{such that}\ (x, H, C) \in \mathcal{A}(g) \right\}, \tag{3.2.3}$$

with $\pi(h) = +\infty$ if $\mathcal{A}(g) = \emptyset$. A superhedging strategy $(x^*, H^*, C^*) \in \mathcal{A}(g)$ is called minimal *if for all $(x, H, C) \in \mathcal{A}(g)$ $V_t^{x,H,C} \ge V_t^{x^*,H^*,C^*}$ $\mathcal{P}$-q.s. for all $0 \le t \le T$.*

It is easy to see that $x^* = \pi(g)$ for any minimal superhedging strategy $(x^*, H^*, C^*) \in \mathcal{A}(g)$.

3.2.3 No-arbitrage condition and Pricing measures

We recall the no-arbitrage condition introduced in Bouchard and Nutz [2015].

Definition 3.2.3. *We say there is no $\mathcal{P}$-quasi-sure arbitrage (NA($\mathcal{P}$)) in the market if for all $H \in \mathcal{H}(\mathbb{F}^{\mathcal{U}})$ with $V_T^{0,H} \ge 0$ $\mathcal{P}$-q.s. we have $V_T^{0,H} = 0$ $\mathcal{P}$-q.s.*

The above definition gives an intuitive extension of the classical no-arbitrage condition, specified under a fixed probability measure $\mathbb{P}$, to the multi-prior case of family of probability measures $\mathcal{P}$. The intuition is justified by the FTAP generalisation proved by [Bouchard and Nutz, 2015, Theorem 4.5]: under Assumption 3.2.1 (recall that S is Borel-adapted) NA($\mathcal{P}$) is equivalent to the fact that for all $\mathbb{P} \in \mathcal{P}$, there exists some $\mathbb{Q} \in \mathcal{Q}$ such that $\mathbb{P} \ll \mathbb{Q}$ where

$$\mathcal{Q} := \{\mathbb{Q} \in \mathfrak{P}(\Omega^T) \mid \exists\, \mathbb{P} \in \mathcal{P},\ \mathbb{Q} \ll \mathbb{P} \text{ and } S \text{ is a martingale under } \mathbb{Q}\}. \tag{3.2.4}$$

We note that convexity of $\mathcal{P}_t(\omega)$ in Assumption 3.2.1 is necessary for the above mentioned FTAP generalisation to hold: take for example $S_0 = 0, T = 1$ and $S_1(\omega) = \omega$, where $\Omega = \mathbb{R}$ and define $\mathcal{P} = \{\delta_{-1}, \delta_1\}$. Then there is no $\mathcal{P}$-q.s. arbitrage in the market but there exists no martingale measure $\mathbb{Q}$, which is absolutely continuous wrt. δ_{-1} or δ_1. We refer also to [Bouchard and Nutz, 2015, Lemmata 3.3 and 3.5].

Remark 3.2.4. By the same token, further results, e.g., on the Superhedging Theorem or the worst-case expected utility maximisation (see Nutz [2016], Blanchard and Carassus [2017], Bartl [2019] and Neufeld and Sikic [2018]) provide more evidence supporting the view that NA($\mathcal{P}$) is a well-chosen extension of the classical no-arbitrage assumption. However, the price to pay when using NA($\mathcal{P}$) is related to technical measurability issues arising when one considers a one step version of the NA($\mathcal{P}$) (see (3.2.5) below). In Bartl [2019] a stronger version of Definition 3.2.3 is introduced which states that (3.2.5) below is satisfied for all $\omega \in \Omega^t$. In Blanchard and Carassus [2017], a stronger version of no-arbitrage is proposed (sNA($\mathcal{P}$)) which states that there is no-arbitrage in the classical sense for all measures $\mathbb{P} \in \mathcal{P}$. In both cases some of the measurability issues are simplified. Finally, different approaches to model uncertainty may lead to fundamentally different notions of arbitrage. In the pathwise approach, one typically asks that some subset of paths supports a feasible model – this is in contrast to the multi-prior setup in this chapter where essentially *all* $\mathbb{P} \in \mathcal{P}$ are assumed to be feasible models. In consequence, the no-arbitrage conditions in the pathwise approach, e.g., model independent arbitrage as in Davis and Hobson [2007], Cox and Obłój [2011], Acciaio et al. [2013] or Arbitrage de la classe $\mathcal{S}$ (see Burzoni et al. [2016]), are much weaker than NA($\mathcal{P}$), i.e., their notions of arbitrage are much stronger than the $\mathcal{P}$-q.s. arbitrage. To wit, negation of sNA($\mathcal{P}$) above gives that there is a classical arbitrage for at least one $\mathbb{P} \in \mathcal{P}$ while Davis and Hobson [2007] say that there is a *weak arbitrage opportunity* if there is a classical arbitrage under *all* $\mathbb{P} \in \mathcal{P}$.

The one step version of the NA($\mathcal{P}$) is the following: for $\omega \in \Omega^t$ fixed we say that NA($\mathcal{P}_t(\omega)$) condition holds if for all $H \in \mathbb{R}^d$

$$H\Delta S_{t+1}(\omega, \cdot) \geq 0\ \mathcal{P}_t(\omega)\text{-q.s.} \quad \Rightarrow \quad H\Delta S_{t+1}(\omega, \cdot) = 0\ \mathcal{P}_t(\omega)\text{-q.s.} \tag{3.2.5}$$

It is proved in [Bouchard and Nutz, 2015, Theorem 4.5] that under the assumption that S is Borel measurable and $\mathcal{P}$ has (APS), the condition NA($\mathcal{P}$) is equivalent to the fact that for all $0 \leq t \leq T-1$, there exists some $\mathcal{P}$-full measure set $\Omega^t_{NA} \in \mathcal{F}^{\mathcal{U}}_t$, such that for all $\omega \in \Omega^t_{NA}$, NA($\mathcal{P}_t(\omega)$) holds. We also introduce the one-step versions of the set $\mathcal{Q}$:

$$\mathcal{Q}_t(\omega) = \{\mathbb{Q} \in \mathfrak{P}(\Omega) \mid \exists\, \mathbb{P} \in \mathcal{P}_t(\omega) \text{ such that } \mathbb{Q} \ll \mathbb{P} \text{ and } \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega, \cdot)] = 0\}.$$

As is shown in [Bouchard and Nutz, 2015, Lemma 4.8], $\mathcal{Q}_t$ has an analytic graph. An application of the Jankov-von Neumann Theorem and Fubini's Theorem shows that we have

$$\mathcal{Q} = \{\mathbb{Q}_0 \otimes \cdots \otimes \mathbb{Q}_{T-1} \mid \mathbb{Q}_t \text{ is } \mathcal{F}^{\mathcal{U}}_t\text{-measurable selector of } \mathcal{Q}_t \text{ for all } 0 \leq t \leq T-1\}. \tag{3.2.6}$$

3.3 Existence and characterisation of minimal superhedging strategies

The Superhedging Theorem, also known as the pricing-hedging duality, is one of the fundamental results in the classical setting of $\mathcal{P} = \{\mathbb{P}\}$, see Föllmer and Schied [2004], Föllmer and Kramkov [1997] and the references therein. One of the main results in Bouchard and Nutz [2015] was its extension to the multi-prior case:

$$\pi(g) = \sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[g]. \tag{3.3.1}$$

While this duality is important and theoretically pleasing, its use for computations may be hampered by lack of a tractable characterisation of the set $\mathcal{Q}$. One of our aims is to give a more algorithmic approach to the above duality. To this end, we establish a suitable dynamic programming principle (DPP) for the superhedging price and also show existence of minimal superhedging strategies in the spirit of Föllmer and Kramkov [1997]. This leads to a robust generalisation of the algorithm in Carassus et al. [2006] and gives a way to handle computation of superhedging prices and, importantly, strategies.

3.3.1 Main Result

To state our main result we need to introduce some further notation. For an upper semi-analytic function $g : \Omega^T \to \mathbb{R}$ let $\{\pi_t(g)\}_{0 \leq t \leq T}$ denote the one step superhedging prices $\pi_t(g) : \Omega^t \to \overline{\mathbb{R}}$ given by

$$\begin{aligned} \pi_T(g)(\omega) &= g(\omega), \qquad \text{and for } 0 \leq t \leq T-1 \\ \pi_t(g)(\omega) &= \inf\{x \mid \exists H \in \mathbb{R}^d \text{ such that } x + H\Delta S_{t+1}(\omega, \cdot) \geq \pi_{t+1}(g)(\omega, \cdot)\ \mathcal{P}_t(\omega)\text{-q.s.}\}. \end{aligned} \tag{3.3.2}$$

Note that the above superhedging prices can be construed as concave envelopes. Indeed, with a slight abuse of notation we denote the one-step quasi-sure concave envelope $\widehat{f}$: $\Omega^t \times \mathbb{R}^d_+ \to \mathbb{R}$ by

$$\widehat{f}(\omega, s) = \inf\{u(s) \mid u : \mathbb{R}^d_+ \to \mathbb{R} \text{ concave}, u(S_{t+1}(\omega, \cdot)) \geq f(\omega, \cdot) \ \mathcal{P}_t(\omega)\text{-q.s.}\}$$

for $t \in \{1, \ldots, T\}$ and an upper semianalytic function $f : \Omega^t \times \Omega \to \mathbb{R}$, where we recall that a concave function is closed, if its superlevel sets are closed. We note that the pointwise infimum of closed concave functions is closed and concave (see e.g. [Bertsekas and Shreve, 1978, Prop. 7.32, p.148] and [Rockafellar, 1970, Theorem 5.5, p.35]). Furthermore, as a consequence of the Fenchel-Moreau theorem every closed concave function can be written as the pointwise infimum of linear functions. Thus the equality

$$\pi_t(g)(\omega) = \widehat{\pi_{t+1}(g)}(\omega, S_t(\omega)), \quad \omega \in \Omega^t, \quad 0 \leq t \leq T-1 \tag{3.3.3}$$

holds and the one-step superhedging prices can be obtained by iteratively taking concave envelopes in the coordinates of Ω.

Let us now define the corresponding dual expressions for the one step case. For $\omega \in \Omega^t$ and $f : \Omega^t \times \Omega \to \overline{\mathbb{R}}$, we define $\mathcal{E}_t(f) : \Omega^t \to \overline{\mathbb{R}}$ by

$$\mathcal{E}_t(f)(\omega) = \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[f(\omega, \cdot)]. \tag{3.3.4}$$

Furthermore, for measurable $g : \Omega^T \to \mathbb{R}$, we define the sequences of operators

$$\mathcal{E}^T(g) = g \quad \text{and} \quad \mathcal{E}^t(g) = \mathcal{E}_t \circ \mathcal{E}^{t+1}(g), \quad 0 \leq t \leq T-1. \tag{3.3.5}$$

With notation at hand, we can state our first main result which gives existence of minimal superhedging strategies and establishes a Dynamic Programming Principle for $\pi_t(g)$ and $\mathcal{E}^t(g)$.

Theorem 3.3.1. *Let Assumption 3.2.1 and NA($\mathcal{P}$) hold. Let $g : \Omega^T \to \mathbb{R}$ be an upper semianalytic function. Then:*

(i) there exists a minimal superhedging strategy in $\mathcal{A}(g)$;

(ii) for any minimal superhedging strategy $(x^*, H^*, C^*) \in \mathcal{A}(g)$, its value satisfies

$$V_t^{x^*,H^*,C^*} = \pi_t(g) = \mathcal{E}^t(g) \ \ \mathcal{P}\text{-q.s.}, \quad 0 \leq t \leq T. \tag{3.3.6}$$

In particular,

$$x^* = \pi(g) = \pi_0(g) = \mathcal{E}^0(g).$$

Remark 3.3.2. Perhaps surprisingly the proof of the above result is technically involved and is thus relegated to Section 3.6. However in the special case of the canonical setting $\Omega = \mathbb{R}^d_+$, $S_t(\omega) = \omega_t$ and $\mathcal{P} = \{\mathbb{P} \in \mathfrak{P}(X) \mid \text{supp}(\mathbb{P}) \text{ is finite}\}$ for an analytic set $X \subseteq \Omega^T$, the underlying arguments are quite intuitive and simple. The crucial equality $\pi_t(g) = \mathcal{E}_t(\pi_{t+1}(g))$ is proved directly using the concave envelope characterisation (3.3.3), see also Beiglböck and Nutz [2014] and the references therein. Indeed, it follows from Proposition 2.4.1 that $\mathcal{P}$ satisfies Assumption 3.2.1 in this case and

$$\mathcal{Q} = \{\mathbb{Q} \in \mathfrak{P}(X) \mid \text{supp}(\mathbb{Q}) \text{ is finite and } S \text{ is a martingale under } \mathbb{Q}\},$$

see also [Bouchard and Nutz, 2015, Example 1.2, p.827] for $X = (\mathbb{R}^d)^T$ and [Lange, 1973, Cor. 4.6, p.151] for a locally compact X. Let $\omega = (\omega_1, \dots, \omega_t) \in \Omega^t$. Using Jensen's inequality

$$\begin{aligned} \mathcal{E}_t(f)(\omega) = \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[f(\omega, \cdot)] &\leq \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\widehat{f}(\omega, \cdot)] \\ &\leq \sup_{\mathbb{Q} \in \mathcal{Q}_t(\omega)} \widehat{f}(\omega, \mathbb{E}_{\mathbb{Q}}[\cdot]) = \widehat{f}(\omega, \omega_t), \end{aligned} \tag{3.3.7}$$

where $\mathbb{E}_{\mathbb{Q}}[\cdot] = \int_{\mathbb{R}^d_+} y \mathbb{Q}(dy)$. To establish the " $\geq$ "-inequality, it suffices to observe that

$$s \mapsto \sup_{\mathbb{Q} \ll \mathbb{P} \text{ for some } \mathbb{P} \in \mathcal{P}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\cdot] = s} \mathbb{E}_{\mathbb{Q}}[f(\omega, \cdot)]$$

is concave and dominates $f(\omega, \cdot)$ on $S_{t+1}(\Sigma_t^\omega)$, where $\Sigma_t^\omega := \{\tilde{\omega} \in X \mid (\tilde{\omega}_1, \dots, \tilde{\omega}_t) = \omega\}$. While concavity is clear in general (see [Beiglböck and Nutz, 2014, Lemma 2.2]), the domination property crucially relies on the fact that the set $\{\mathbb{Q} \ll \mathbb{P} \text{ for some } \mathbb{P} \in \mathcal{P}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\cdot] = s\}$ contains the Dirac measures at points $s \in S_{t+1}(\Sigma_t^\omega)$. For a general set $\mathcal{P}$ this is not true: for example in the case $\mathcal{P} = \{\mathbb{P}\}$ for some $\mathbb{P} \in \mathfrak{P}(\Omega)$ in general the set $\{\mathbb{Q} \ll \mathbb{P},\ \mathbb{E}_{\mathbb{Q}}[\cdot] = s\}$ is non-empty only for s in the relative interior of the convex hull of the support of $\mathbb{P}$ (see [Föllmer and Schied, 2004, Theorem 1.48, p.29]).

Remark 3.3.3. We stress that the dynamic programming principle in Theorem 3.3.1 depends on the fact that the universe of traded assets is the same for all $t \in [0, T]$. As recalled in the introduction, in the literature on robust pricing and hedging focusing on computing $\pi_0(g)$ only, it is often assumed that some additional options are available for buy-and-hold (static) trading only. Pricing measures in $\mathcal{Q}$ are then required to be calibrated to these time zero prices and the duality (3.3.1) still holds. Such a setup can be embedded in ours by assuming the options are traded dynamically but taking a set of priors $\mathcal{P}$ which does not impose any assumptions about their dynamics other than these resulting from no arbitrage in the initial setup. An admissible pricing measure in the original setup can be used to define dynamic options' prices via conditional expectations and can thus be lifted to a martingale measure in the extended setup. It follows from the superhedging theorem that the time zero superhedging price in both setups agree, we refer to Aksamit et al. [2018], where this embedding is considered in detail. This equivalence is lost for $t > 0$, as the agent is locked-in at time zero into a buy and hold position over $[0, T]$. In practice thus, the process $(S_t)_{0 \le t \le T}$ in this article represents the vector of liquidly traded assets.

3.3.2 Canonical space: concave envelopes and computation of the superhedging price

In this subsection we work on the canonical space, i.e. we set $\Omega = \mathbb{R}^d_+$ and $S_t(\omega) = (\omega_t^1, \dots, \omega_t^d)$. In particular $g(S_1(\omega), \dots, S_T(\omega)) = g(\omega)$ holds.
We start by developing in more detail the special case when $\mathcal{P}$ is obtained by specifying the support for feasible moves of the stock prices. This captures the pathwise approach but is also natural in the quasi-sure framework as $\mathrm{NA}(\mathcal{P})$ and $\pi(g)$ only depend on the polar sets of $\mathcal{P}$. More precisely we give the following definition:

Definition 3.3.4. *Given correspondences $f_t : \Omega^t \twoheadrightarrow \mathbb{R}^d$, $0 \le t \le T-1$, we define $\mathcal{P}_t(\omega) : \Omega^t \twoheadrightarrow \mathfrak{P}(\Omega)$ by*

$$\mathcal{P}_t(\omega) = \{\mathbb{P} \in \mathfrak{P}(\Omega) \mid \mathrm{supp}(\mathbb{P}) \subseteq f_t(\omega)\}.$$

We say that the sequence of random sets $(\mathcal{P}_t)_{0 \le t \le T-1}$ is generated by $\{f_t\}_{0 \le t \le T-1}$.

Recall that a correspondence $f : \Omega^t \twoheadrightarrow \mathbb{R}^d$ is called measurable if $\{\omega \in \Omega^t \mid f(\omega) \cap O \neq \emptyset\} \in \mathcal{B}(\Omega^t)$ for all open sets $O \subseteq \mathbb{R}^d$. We refer to [Rockafellar and Wets, 1998, 14.A, p.643ff.] for the theory of measurable correspondences.

Lemma 3.3.5. *Let $(\mathcal{P}_t)_{0 \le t \le T-1}$ be generated by measurable, closed valued correspondences $\{f_t\}_{0 \le t \le T-1}$. Then $\mathcal{P}_t$ has Borel measurable graph for all $0 \le t \le T-1$.*

Under the assumptions of Lemma 3.3.5 we can then define $\mathcal{P} \subseteq \mathfrak{P}(\Omega^T)$ satisfying (APS) as in Assumption 3.2.1 since $\mathfrak{P}_t(\omega)$ is clearly non-empty and convex for all $0 \le t \le T-1$ and $\omega \in \Omega^t$.

Proof. By assumption the graph of f_t is $\mathcal{B}(\Omega^t) \otimes \mathcal{B}(\mathbb{R}^d) = \mathcal{B}((\mathbb{R}^d)^{t+1})$-measurable for all $t \in \{0, \dots T-1\}$ (see [Rockafellar and Wets, 1998, Theorem 14.8, p.648]). Thus by [Bertsekas and Shreve, 1978, Cor. 7.25.1, p.134] $\mathfrak{P}(\mathrm{graph}(f_t))$ is Borel as well. Define the map

$$D : \Omega^t \times \mathfrak{P}(\mathbb{R}^d_+) \to \mathfrak{P}(\Omega^{t+1}),\ (\omega, \mathbb{P}) \mapsto \delta_\omega \otimes \mathbb{P}$$

and note that D is a homeomorphism from $\Omega^t \times \mathfrak{P}(\mathbb{R}^d_+)$ to $\{\delta_\omega \otimes \mathbb{P} \mid \omega \in \Omega^t,\ \mathbb{P} \in \mathfrak{P}(\mathbb{R}^d_+)\}$. Indeed, take a sequence $(\omega_n, \mathbb{P}_n) \in \Omega^t \times \mathfrak{P}(\mathbb{R}^d_+)$ such that $(\omega_n, \mathbb{P}_n)$ converges to $(\omega, \mathbb{P})$ in the product topology. Denote by $\mathcal{L}^1_b(\Omega^{t+1})$ the bounded 1-Lipschitz functions on Ω^{t+1}. Then

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{f\in\mathcal{L}^1_b(\Omega^{t+1})} \left| \int_{\Omega^{t+1}} f d(\delta_{\omega_n} \otimes \mathbb{P}_n) - \int_{\Omega^{t+1}} f d(\delta_\omega \otimes \mathbb{P}) \right| \\
&\leq \lim_{n\to\infty} \left(|\omega_n - \omega| + \sup_{f\in\mathcal{L}^1_b(\Omega^{t+1})} \left| \int_{\Omega^{t+1}} f(\omega, \cdot) d\mathbb{P}_n - \int_{\Omega^{t+1}} f(\omega, \cdot) d\mathbb{P} \right| \right) = 0,
\end{aligned}$$

so $\delta_{\omega_n} \otimes \mathbb{P}_n$ converges weakly to $\delta_\omega \otimes \mathbb{P}$, see [Huber and Ronchetti, 2009, Cor. 2.18, p.35]. Continuity of the inverse map follows directly from the definition of weak convergence of measures. Note also that a homeomorphism maps Borel sets to Borel sets. As

$$\mathfrak{P}(\text{graph}(f_t)) \cap \{\delta_\omega \otimes \mathbb{P} \mid \omega \in \Omega^t,\ \mathbb{P} \in \mathfrak{P}(\mathbb{R}^d)\}$$

is Borel-measurable, applying the inverse map D^{-1} we conclude that

$$\text{graph}(\mathcal{P}_t) = D^{-1}(\mathfrak{P}(\text{graph}(f_t)) \cap \{\delta_\omega \otimes \mathbb{P} \mid \omega \in \Omega^t,\ \mathbb{P} \in \mathfrak{P}(\mathbb{R}^d)\})$$

is Borel. □

In fact, for such a set $\mathcal{P}$ the condition $\text{NA}(\mathcal{P}_t(\omega))$ is equivalent to $0 \in \text{ri}(f_t(\omega) - S_t(\omega))$, where $\text{ri}(A)$ denotes the relative interior of the convex hull of A. For a proof of this result in a more general setup, see Theorem 2.2.6 This deterministic condition is called No Pointwise Arbitrage in Burzoni et al. [2019] and can be checked without resorting to the use of probability measures.

The following definition further characterises closed-valued correspondences $\{f_t\}_{0\leq t\leq T-1}$ and is needed to identify an important subclass of sets $\{\mathcal{P}_t\}_{0\leq t\leq T-1}$ generated by $\{f_t\}_{0\leq t\leq T-1}$:

Definition 3.3.6. *A closed-valued correspondence $f_t : \Omega^t \to \mathbb{R}^d$ is called uniformly continuous if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\omega, \omega' \in \Omega^T$ such that $|\omega' - \omega| \leq \delta$ we have $d_H(f_t(\omega), f_t(\omega')) \leq \varepsilon$, where*

$$d_H(A, B) := \max \left(\sup_{v\in A} \inf_{\tilde{v}\in B} |v - \tilde{v}|, \sup_{\tilde{v}\in B} \inf_{v\in A} |v - \tilde{v}| \right)$$

denotes the Hausdorff metric on closed subsets A, B of Ω.

Uniformly continuous correspondences are in particular continuous (see [Rockafellar and Wets, 1998, Def. 5.4, p.152]) and thus measurable ([Rockafellar and Wets, 1998, Theorem 5.7, p.154]). It turns out, that when the correspondences fulfil this continuity condition and are compact-valued, the $\mathcal{P}$-q.s. superhedging price of a continuous payoff g coincides with the $\mathbb{P}$-a.s. superhedging price of g for every $\mathbb{P}$ with support equal to the paths generated by the correspondences $\{f_t\}_{0\leq t\leq T-1}$:

Proposition 3.3.7. *Suppose $(\mathcal{P}_t)_{0\le t\le T-1}$ is generated by compact-valued, uniformly continuous correspondences $\{f_t\}_{0\le t\le T-1}$ and that NA($\mathcal{P}$) holds. Furthermore assume that the function $g:\Omega^T\to\mathbb{R}$ is continuous. Take any measure $\mathbb{P}=\mathbb{P}_0\otimes\cdots\otimes\mathbb{P}_{T-1}$ such that*

$$supp(\mathbb{P}_t(\omega)) = f_t(\omega), \quad 0\le t\le T-1,\ \omega\in\Omega^t.$$

Then, for all $0\le t\le T-1$ and $\omega\in\Omega^t$,

$$\pi_t(g)(\omega) = \inf\{x\in\mathbb{R} \mid \exists H\in\mathbb{R}^d \text{ such that } x + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(g)(\omega,\cdot)\ \mathbb{P}\text{-a.s}\}. \tag{3.3.8}$$

and $\omega\mapsto\pi_t(g)(\omega)$ is continuous.

The proof of the above result is relegated to Section 3.5.
We now apply this result to a one-dimensional case of particular interest, as in Carassus and Vargiolu [2018], where it is easy to explicitly compute the minimal superhedging prices:

Proposition 3.3.8. *Assume that for all $0\le t\le T-1$, $d_{t+1}<1<u_{t+1}$ and that the (random) sets $\mathcal{P}_t$ are given by*

$$\mathcal{P}_t(\omega) = \{\mathbb{P}\in\mathfrak{P}(\mathbb{R}) \mid supp(\mathbb{P})\subset[\omega_t d_{t+1}, \omega_t u_{t+1}]\},$$

where $\omega=(\omega_1,\dots,\omega_t)\in\Omega^t$. Then NA($\mathcal{P}$) holds. Let $g:\mathbb{R}^T\to\mathbb{R}$ be convex. Then

$$\begin{aligned}\pi_T(g) &= g\\ \pi_t(g)(\omega) &= \alpha_{t+1}\pi_{t+1}(g)(\omega,\omega_t u_{t+1}) + (1-\alpha_{t+1})\pi_{t+1}(g)(\omega,\omega_t d_{t+1}),\end{aligned} \tag{3.3.9}$$

where $\alpha_t := \frac{1-d_t}{u_t-d_t}$, $1\le t\le T$.

Proof. Noting that $f_t(\omega)=[\omega_t d_{t+1},\omega_t u_{t+1}]$ is a uniformly continuous compact-valued correspondence, the graph of $\mathcal{P}_t$ is clearly non-empty, convex and Borel measurable for $0\le t\le T-1$ by Lemma 3.3.5. As $0\in\text{ri}(f_t(\omega)-S_t(\omega)) = \text{ri}([-\omega_t(1-d_{t+1}),\omega_t(u_{t+1}-1))$, NA($\mathcal{P}$) holds. We prove by induction that $\pi_t(g)$ satisfies (3.3.9) and is convex. This is clear for $t=T$. Now we assume that for some $0\le t\le T-1$, $\pi_{t+1}(g)$ is convex. As $\mathcal{P}_t(\omega)$ contains the Dirac measures on $[\omega_t d_{t+1},\omega_t u_{t+1}]$ we conclude that

$$\pi_t(g)(\omega) = \inf\{x\in\mathbb{R} \mid \exists H \text{ s.t. } x + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(g)(\omega,\cdot) \text{ on } [\omega_t d_{t+1},\omega_t u_{t+1}]\}.$$

Note that for a convex function f on $[\omega_t d_{t+1},\omega_t u_{t+1}]$, its concave envelope $\widehat{f}$ is simply the affine function intersecting the points $f(\omega_t d_{t+1})$ and $f(\omega_t u_{t+1})$. As $\pi_t(g)(\omega)$ is the pointwise concave envelope of the convex function $\pi_{t+1}(g)(\omega,\cdot)$, it can thus be written as the unique convex combination of the extreme points of $\pi_{t+1}(g)(\omega,\cdot)$ on the interval $[\omega_t d_{t+1},\omega_t u_{t+1}]$ which preserves the barycentre ω_t. Thus, we obtain (3.3.9) for t. Clearly $\pi_t(g):\mathbb{R}^t\to\mathbb{R}$ is then a convex combination of convex functions and thus also a convex function. □

It is insightful to observe that the above superreplication price corresponds to the actual replication price in a Cox-Ross-Rubinstein model of Cox et al. [1979] where the stock price evolves on a binomial tree with $S_{t+1} \in \{d_{t+1}S_t, u_{t+1}S_t\}$.

3.4 Maximising expected utility of consumption in $\mathcal{A}(g)$

3.4.1 Main results

In Theorem 3.3.1 above, we characterised the superhedging prices $\pi_t(g)$ and introduced ways for computing minimal superhedging strategies. However, these are typically non-unique. Indeed, as we see from (3.3.3), if the concave envelope $\widehat{f(\omega, \cdot)}$ of a function $f : \Omega^{t+1} \to \mathbb{R}$ is not differentiable at ω_t, every point $H \in \mathbb{R}^d$ in its superdifferential constitutes a minimal superhedging strategy, see also Example 3.4.7 below. To select the "best" among minimal superhedging strategies we propose a secondary optimisation problem of robust maximisation of expected utility with intermediate consumption, given by

$$\sup_{(H,C)\in\mathcal{A}_x} \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{s=1}^{T} U(s, \Delta C_s) \right], \tag{3.4.1}$$

where $\mathcal{A}_x$ is the set of investment-consumption strategies which superhedge $g : \Omega^T \to \mathbb{R}$, i.e.

$$\mathcal{A}_x := \{(H, C) \in \mathcal{H}(\mathbb{F}^{\mathcal{U}}) \times \mathcal{C} \mid V_T^{x,H,C} \geq g \ \mathcal{P}\text{-q.s.}\}$$

and the set $\mathcal{P}^u \subseteq \mathfrak{P}(\Omega^T)$ fulfils the following condition:

Assumption 3.4.1. *$\mathcal{P}^u$ satisfies (APS) and $\mathcal{P}^u \subseteq \mathcal{P}$.*

The set $\mathcal{P}^u$ represents the subjective views of an investor. While superhedging with respect to $\mathcal{P}$ reflects the necessity to satisfy certain regulatory and risk requirements, $\mathcal{P}^u$ is used to express individual preferences for the optimisation problem (3.4.1) and does not need to satisfy any further requirements than those of Assumption 3.4.1, e.g. NA($\mathcal{P}^u$) can fail. On the other hand the assumption $\mathcal{P}^u \subseteq \mathcal{P}$ guarantees that (3.4.1) is well defined (cf. Assumption 3.4.2). In Theorems 3.4.3 and 3.4.5 below, we show that (3.4.1) is well posed and admits an optimiser which, under suitable assumptions, is unique.

The assumptions imposed on the utility functions $U(t, \cdot, \cdot)$ are in line with those in Nutz [2016]:

Assumption 3.4.2. *For $t = 1, \dots, T$ the utility function $U(t, \cdot, \cdot) : \Omega^t \times [0, \infty) \to \mathbb{R}$ is lower semianalytic and bounded from above. Furthermore*

1. $\omega \mapsto U(t, \omega, x)$ is bounded from below for each $x > 0$.

2. $x \mapsto U(t, \omega, x)$ *is non-decreasing, concave and continuous for each* $\omega \in \Omega^t$.

We believe that boundedness assumptions on utility functions which we make here could be weakened, similarly to Blanchard and Carassus [2017]. However, due to the overall length and already technical character of proofs, we decided to leave this extension for further research.
We remark that by 2. in Assumption 3.4.2 it is sufficient to consider investment-consumption strategies which hedge g, i.e. for which $V_T^{x,H,C} = g$, since the superhedging surplus can be consumed at terminal time.
Note that by Assumption 3.4.2 and standard results on Carathéodory functions (see [Aliprantis and Border, 2006, Lemma 4.51, p. 153]) we conclude that $U(t, \cdot, \cdot)$ is $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}_+)$-measurable. We set $U(t, x, \omega) = -\infty$ for $x < 0$ and often write $U(t, x)$ instead of $U(t, x, \omega)$.

Theorem 3.4.3. *Let* $U(t, \cdot, \cdot)$ *be given for* $1 \leq t \leq T$ *and let* $NA(\mathcal{P})$, *Assumptions 3.2.1, 3.4.1 and 3.4.2 hold. Then for any Borel* $g : \Omega^T \to \mathbb{R}$, *there exists* $(H^*, C^*) \in \mathcal{A}_\pi$ *such that*

$$\inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{s=1}^{T} U(s, \Delta C_s^*) \right] = \sup_{(H,C) \in \mathcal{A}_\pi} \inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{s=1}^{T} U(s, \Delta C_s) \right],$$

where $\pi = \pi(g)$ *is the* $\mathcal{P}$*-q.s. superhedging price of* g.

In order to obtain uniqueness of the above maximiser $(\hat{H}, \hat{C})$, we again switch to the canonical setup $\Omega^T = (\mathbb{R}_+^d)^T$, $S_t(\omega) = \omega_t$. In line with Denis and Kervarec [2007] we strengthen assumptions on the utility functions $U(t, \cdot, \cdot)$ and also assume weak compactness of the set $\mathcal{P}^u$. This enables us to show existence of a "worst-case" measure $\hat{\mathbb{P}} \in \mathcal{P}^u$, in analogy to the argumentation in Schied and Wu [2005]. In fact, Example 3.4.8 below shows, that one cannot expect uniqueness of maximizers in general, if $\mathcal{P}^u$ is not weakly closed.

Assumption 3.4.4. *For* $1 \leq t \leq T$ *the non-random utility functions* $U(t, \cdot)$ *satisfy Assumption 3.4.2 and are bounded. The mapping* $x \mapsto U(t, x)$ *is strictly concave, non-decreasing and continuous. Furthermore, for* $0 \leq t \leq T-1$ *and* $\mathcal{P}^u$*-q.e* $\omega \in \Omega^t$ *the set* $\mathcal{P}_t^u(\omega)$ *is weakly compact and the sets* $\mathcal{P}$ *and* $\mathcal{P}^u$ *fulfil the following continuity criteria:*

(i) If $\omega, \tilde{\omega} \in \Omega^t$ *and* $\varepsilon > 0$, *then there exists* $\delta > 0$ *such that for* $|\omega - \tilde{\omega}| \leq \delta$ *and for every* $\mathbb{P} \in \mathcal{P}_t^u(\omega)$ *there exists* $\tilde{\mathbb{P}} \in \mathcal{P}_t^u(\tilde{\omega})$ *such that* $d_L(\mathbb{P}, \tilde{\mathbb{P}}) \leq \varepsilon$, *where*

$$d_L(\mathbb{P}, \tilde{\mathbb{P}}) = \inf\{\varepsilon \geq 0 \mid \mathbb{P}(A) \leq \tilde{\mathbb{P}}(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(\Omega)\}$$

denotes the Levy metric on $\mathfrak{P}(\Omega)$ *and* $A^\varepsilon = \{\omega \in \Omega \mid \exists \tilde{\omega} \in A \text{ such that } |\omega - \tilde{\omega}| < \varepsilon\}$.

(ii) The map $f_t(\omega) := supp(\mathcal{P}_t(\omega))$ is uniformly continuous in the sense of Definition 3.3.6, where

$$supp(\mathcal{P}_t(\omega)) = \bigcap \{A \subseteq \Omega \text{ closed} \mid \mathbb{P}(A) = 1 \text{ for all } \mathbb{P} \in \mathcal{P}_t(\omega)\}$$

is the quasi-sure support of $\mathcal{P}_t(\omega)$ for $\omega \in \Omega^t$.

Theorem 3.4.5. *In the setup of Theorem 3.4.3 assume further that Assumption 3.4.4 holds and that the functions $\pi_t(g) : \Omega^t \to \mathbb{R}$ are continuous for all $1 \le t \le T$. Then there exists a probability measure $\mathbb{P}^* \in \mathcal{P}^u$ such that*

$$\sup_{(H,C)\in\mathcal{A}_\pi} \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}\left[\sum_{s=1}^T U(s,\Delta C_s)\right] = \sup_{(H,C)\in\mathcal{A}_\pi} \mathbb{E}_{\mathbb{P}^*}\left[\sum_{s=1}^T U(s,\Delta C_s)\right].$$

Furthermore, the maximising strategy $(H^*, C^*) \in \mathcal{A}_\pi$ is unique in the following sense: for any two maximising strategies $(H^1, C^1), (H^2, C^2) \in \mathcal{A}_\pi$ and for $1 \le t \le T$ we have $C^1_t = C^2_t$ and $H^1_t \Delta S_t = H^2_t \Delta S_t$ $\mathbb{P}^*$-a.s.

Remark 3.4.6. In the setting of Theorem 3.4.5 one can interchange the order of the supremum and the infimum. Indeed

$$\begin{aligned}\inf_{\mathbb{P}\in\mathcal{P}^u} \sup_{(H,C)\in\mathcal{A}_\pi} \mathbb{E}_{\mathbb{P}}\left[\sum_{s=1}^T U(s,\Delta C_s)\right] &\le \sup_{(H,C)\in\mathcal{A}_\pi} \mathbb{E}_{\mathbb{P}^*}\left[\sum_{s=1}^T U(s,\Delta C_s)\right] \\ &= \sup_{(H,C)\in\mathcal{A}_\pi} \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}\left[\sum_{s=1}^T U(s,\Delta C_s)\right] \\ &\le \inf_{\mathbb{P}\in\mathcal{P}^u} \sup_{(H,C)\in\mathcal{A}_\pi} \mathbb{E}_{\mathbb{P}}\left[\sum_{s=1}^T U(s,\Delta C_s)\right].\end{aligned}$$

The proofs of Theorems 3.4.3 and 3.4.5 are given in Section 3.7. We first establish Theorem 3.4.3 in the one-period case ($T = 1$) and then extend it to the general multi-step setting and consider the uniqueness.

3.4.2 Examples and comments

To illustrate the above results, we discuss several examples. We start with a simple example for non-uniqueness of minimal superhedging strategies.

Example 3.4.7 (Non-Uniqueness of minimal superhedging strategies, maximizers and $\mathbb{P}^*$ in Theorem 3.4.5). We take $\Omega = \mathbb{R}_+$, where $d = 1$ and $T = 2$ as well as $s_0 = 2$. Furthermore $S_t(\omega) = \omega_t$ for $t = 1, 2$ and

$$\mathcal{P}_t(\omega) = \{\mathbb{P} \in \mathfrak{P}(\mathbb{R}_+)\}, \quad t = 0, 1.$$

We want to superhedge the running minimum at time 2, i.e. $g(\omega) = \underline{S}_2(\omega)$. Clearly $\mathcal{Q}_t(\omega) = \{\mathbb{Q} \in \mathfrak{P}(\mathbb{R}_+) \mid \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega, \cdot)] = 0\}$ for all $\omega \in \Omega^t$ and $t = 0, 1$. Besides it is easy to see that

$$\sup_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[g] = s_0 = 2,$$

so we have some degree of freedom to choose our superhedging strategy $H \in \mathcal{H}(\mathbb{F}^{\mathcal{U}})$. As it turns out we can choose any $H_1 \in [0, 1]$, which gives a wealth of $2 + H_1(S_1 - 2)$ at time 1. Indeed, we need to have $2 + H_1(S_1 - 2) \geq 2$ for $S_1 \geq 2$, which is fulfilled iff $H_1 \geq 0$ and $2 + H_1(S_1 - 2) \geq S_1$ for $S_1 < 2$, which is fulfilled as long as $H_1 \leq 1$. For time 2 we have

$$H_2(\omega) \in \begin{cases} [0, H_1] & \text{if } S_1(\omega) \geq 2, \\ \left[0, \frac{2}{S_1(\omega)} + \frac{H_1}{S_1(\omega)}(S_1(\omega) - 2)\right] & \text{if } S_1(\omega) < 2. \end{cases}$$

To see this we argue for different cases. We have to find conditions such that

$$2 + H_1(S_1 - 2) + H_2(S_2 - S_1) \geq \min(2, S_1, S_2)$$

holds. If $S_2 \geq S_1$ this gives the necessary condition $H_2 \geq 0$. If $S_2 < S_1$ we distinguish the cases (a) $2 \leq S_2 < S_1$ and (b) $S_2 < 2 \leq S_1$ and (c) $S_2 < S_1 < 2$. In case (a) we have

$$2 + H_1(S_1 - 2) + H_2(S_2 - S_1) \geq 2,$$

which yields

$$H_2 \leq \inf_{S_2 \in [2, S_1]} H_1(2 - S_1)/(S_2 - S_1) = H_1.$$

In case (b) and (c) we have

$$2 + H_1(S_1 - 2) + H_2(S_2 - S_1) \geq S_2,$$

which yields

$$H_2 \leq \frac{S_2 - 2 + H_1(2 - S_1)}{S_2 - S_1}.$$

Note that

$$\left(\frac{x - a}{x - b}\right)' = \frac{a - b}{(x - b)^2}.$$

Here $a - b := 2 + H_1(S_1 - 2) - S_1 = (2 - S_1)(1 - H_1)$. Thus in case (b) we have

$$a - b < 0$$

and

$$\inf_{S_2 \in [0, 2]} \frac{S_2 - 2 + H_1(2 - S_1)}{S_2 - S_1} = H_1.$$

In case (c) we have $a-b \geq 0$ thus

$$\inf_{S_2 \in [0,S_1]} \frac{S_2 - 2 + H_1(2-S_1)}{S_2 - S_1} = \frac{2}{S_1} + \frac{H_1}{S_1}(S_1 - 2).$$

Note also that the superhedging cost at time 1 is given by

$$\pi_1(g)(\omega) = \sup_{\mathbb{Q} \in \mathcal{Q}_1(\omega)} \mathbb{E}_{\mathbb{Q}}[g(\omega, \cdot)] = \begin{cases} 2 & \text{if } S_1(\omega) \geq 2, \\ S_1(\omega) & \text{if } S_1(\omega) < 2. \end{cases}$$

So according to (3.2.1) and (3.3.2) we can consume

$$C_1(\omega) \in \begin{cases} [0, H_1(S_1(\omega) - 2)] & \text{if } S_1(\omega) \geq 2, \\ [0, (H_1(\omega) - 1)(S_1(\omega) - 2)] & \text{if } S_1(\omega) < 2 \end{cases}$$

at time 1.

We now show that if Assumption 3.4.4 is not satisfied (namely $\mathcal{P}^u$ does not fulfil Assumption 3.4.4.1.), then Theorem 3.4.5 is not true in general. For this we specify the set $\mathcal{P}^u$ and iteratively solve the optimization problem (3.4.1): We set $U(2,\omega,x) = U(1,\omega,x) = U(x)$ for some bounded concave, non-decreasing and continuous function $U : \mathbb{R}_+ \to \mathbb{R}_+$ as well as $\mathcal{P}_1^u(S_1) = \{\delta_{S_1}\}$ for $S_1 > 2$ and $\mathcal{P}_1^u(S_1) = \{\delta_{S_1+1}\}$ for $S_1 \leq 2$. Note that $\mathcal{P}_1^u$ obviously violates Assumption 3.4.4.1. We obtain the following optimal one-step prices, where we use notation from Section 3.7.2: for $S_1 > 2$ and $x \geq 2$ we find

$$\begin{aligned} U_1(S_1, x) &= \sup_{(H,c) \in \mathcal{A}_{1,x}(S_1)} \left(\mathbb{E}_{\delta_{S_1}}[U(x + H(S_2 - S_1) - \underline{S_2} - c)] + U(c) \right) \\ &= \sup_{(H,c) \in \mathcal{A}_{1,x}(S_1)} (U(x - 2 - c) + U(c)) = 2U\left(\frac{x-2}{2}\right) \end{aligned}$$

with $c = (x-2)/2$ and some $0 \leq H \leq \min(\frac{x/2+1}{S_1}, \frac{x/2-1}{S_1-2})$. Indeed, note that by concavity

$$U(x - 2/2) = U(\frac{x - 2 - c + c}{2}) \geq 1/2U(x - 2 - c) + 1/2U(c).$$

For H we need to have

$$\frac{x}{2} + 1 + H(S_2 - S_1) \geq \min(2, S_2),$$

which is always satisfied for $S_2 \geq S_1$ and $H \geq 0$. For $S_2 < S_1$ we have in case (a): $2 \leq S_2 < S_1$. Then we need to have

$$\frac{x}{2} + 1 + H(S_2 - S_1) \geq 2$$

or

$$H \leq (1 - x/2)/(S_2 - S_1),$$

which is minimised for $S_2 = 2$ giving $(x/2 - 1)/(S_1 - 2)$. Case (b): $S_2 < 2 < S_1$, which yields

$$H \leq (S_2 - x/2 - 1)/(S_2 - S_1).$$

The RHS is minimised either for $S_2 = 2$ or $S_2 = 0$ depending on $S_1 > x/2+1$ or $S_1 \leq x/2+1$. For $S_1 \leq 2$ and $x \geq S_1$ we thus have

$$\begin{aligned} U_1(S_1, x) &= \sup_{(H,c)\in\mathcal{A}_{1,x}(S_1)} \Big(\mathbb{E}_{\delta_{S_1+1}}[U(x + H(S_2 - S_1) - \underline{S_2} - c)] + U(c)\Big) \\ &= \sup_{(H,c)\in\mathcal{A}_{1,x}(S_1)} (U(x + H - S_1 - c) + U(c)) \geq U(0) + U(1) \end{aligned}$$

with $H = x/S_1$ and $c = 0$: for H we need to have

$$x - c + H(S_2 - S_1) \geq \min(S_1, S_2),$$

which is always satisfied for $S_2 \geq S_1$ if $H \geq 0$ (and $x - c \geq S_1$). For $S_2 < S_1$ we have $S_2 < S_1 \leq 2$. This gives

$$x - c + H(S_2 - S_1) \geq S_2,$$

which yields

$$H \leq (S_2 - x + c)/(S_2 - S_1).$$

The RHS is minimised for $S_2 = 0$ if $x - c \geq S_1$ giving $H \leq (x - c)/S_1$.
Setting $\mathcal{P}_0^u = \{\delta_x \mid x \in \mathbb{R}_+\}$ we obtain

$$\begin{aligned} U_0(2) &= \sup_{H\in\mathcal{A}_{0,2}} \inf_{\mathbb{P}\in\mathcal{P}_0^u} \mathbb{E}_{\mathbb{P}}[U_1(S_1, 2 + H(S_1 - 2))] \\ &= \sup_{H\in\mathcal{A}_{0,2}} \inf_{\mathbb{P}\in\mathcal{P}_0^u} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{S_1>2\}} 2U\left(\frac{2 + H(S_1 - 2) - 2}{2}\right)\right. \\ &\quad \left. + \mathbb{1}_{\{S_1\leq 2\}} U_1(2 + H(S_1 - 2))\right] \\ &= 2U(0). \end{aligned}$$

Indeed, the first term is trivially minimised in S_1 by $S_1 \downarrow 2$, while the second term is always greater equal $U(0) + U(1)$. Note that by the proof of Theorem 3.4.5 under Assumption 3.4.4 there would exist $\hat{\mathbb{P}} \in \mathcal{P}_0^u$ such that

$$U_0(2) = \sup_{H\in\mathcal{A}_{0,2}} \mathbb{E}_{\hat{\mathbb{P}}}[U_1(S_1, x + H\Delta S_1)].$$

On the contrary, in our case there exists no $\hat{\mathbb{P}} \in \mathcal{P}_0^u$ such that

$$U_0(2) = 2U(0) = \mathbb{E}_{\hat{\mathbb{P}}}\left[\mathbb{1}_{\{S_1>2\}} 2U\left(\frac{S_1 - 2}{2}\right) + \mathbb{1}_{\{S_1<2\}} U_1(S_1, 2)\right]$$

as the RHS is strictly greater than $2U(0)$ for all $\hat{\mathbb{P}} \in \mathcal{P}_0^u$. Indeed, the first term is maximised for $H = 1$ while the second term is decreasing in H, so maximised for $H = 0$. Thus Theorem 3.4.5 does not hold.

The next example shows that we cannot expect to have uniqueness of maximizers without assuming some closedness property of $\mathcal{P}^u$.

Example 3.4.8 (Non-uniqueness of maximisers for non-closed $\mathcal{P}^u$). Let $T = 1$, $d = 2$, $\Omega = \mathbb{R}^2$, $\mathcal{P} = \mathfrak{P}(\mathbb{R}^2_+)$, $S_t(\omega) = \omega_t$ and $S_0 = (1, 1)$. Consider $g = \min(S^1_1, S^2_1)$. Then $\pi(g) = 1$ and H_1 is of the form

$$H_1 = \begin{pmatrix} \lambda \\ 1-\lambda \end{pmatrix},$$

where $\lambda \in [0, 1]$. Take

$$\mathcal{P}^u = \{\mathbb{P}_n\}_{n=1}^{\infty} \quad \text{where } \mathbb{P}_n = \frac{\delta_{\{S^1_1 = n - \frac{1}{n},\ S^2_1 = n + \frac{1}{n}\}}}{2} + \frac{\delta_{\{S^1_1 = 0,\ S^2_1 = 0\}}}{2}.$$

Then clearly $\mathcal{P}^u$ is not closed. We note that for $H \in \mathcal{A}_1$

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_n}\left[U\left(1 + H\Delta S_1 - g\right)\right] &= \frac{1}{2}U\left(\lambda\left(n - \frac{1}{n}\right) + (1-\lambda)\left(n + \frac{1}{n}\right) - \left(n - \frac{1}{n}\right)\right) \\ &\quad + \frac{1}{2}U(0) = \frac{1}{2}U\left((1-\lambda)\frac{2}{n}\right) + \frac{1}{2}U(0) \downarrow U(0),\ n \to \infty.\end{aligned}$$

Thus we conclude

$$\sup_{H \in \mathcal{A}_1} \inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1 + H\Delta S_1 - g)] = U(0),$$

in particular

$$H \mapsto \inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1 + H\Delta S_1 - g)] = U(0)$$

is constant and thus the maximizer is not unique.

Finally, we illustrate that even with a compact $\mathcal{P}^u$ we can not strengthen the sense in which the optimisers are unique in Theorem 3.4.5.

Example 3.4.9 (On uniqueness property of maximisers). We consider a one-step version of Example 3.4.7: $T = 1$, $d = 1$, $\Omega = \mathbb{R}_+$, $S_t(\omega) = \omega_t$, $s_0 = 2$, $g(S) = \underline{S_1}$, $\mathcal{P} = \mathfrak{P}(\mathbb{R}_+)$. We have $\pi(g) = 2$. We also set $\mathcal{P}^u = \{\delta_2\}$, where δ_2 is defined by

$$\delta_2(S_t = 2 \text{ for all } t = 0, 1) = 1.$$

Furthermore let $U(\cdot) = U(1, \cdot, \cdot)$ such that the conditions of Theorem 3.4.5 are satisfied. The optimisers are then non-unique in the sense that (3.4.1) is equal to $U(0)$ and is attained for every $H \in [0, 1]$ but are unique in the sense of Theorem 3.4.5 since $H\Delta S_1 = 0$ δ_2-a.s. for all $H \in \mathbb{R}$.

We now provide the proofs of Proposition 3.3.7, Theorem 3.3.1 and of Theorems 3.4.3, 3.4.5. These proofs require a number of technical lemmata which are established alongside the main proofs.

3.5 Proof of Proposition 3.3.7

Proof. Fix $\omega \in \Omega^{T-1}$ and $\varepsilon > 0$. Recall that g is continuous and $\{f_t\}_{0\le t\le T-1}$ are compact-valued. Note that the set

$$B := \{(\tilde{\omega}, \tilde{v}) \in \Omega^{T-1} \times \mathbb{R}^d \mid \operatorname{dist}((\omega, f_{T-1}(\omega)), (\tilde{\omega}, \tilde{v})) \le 1\}$$

is compact, thus g is uniformly continuous on B, i.e. there exists $\delta \in (0,1)$ such that $|g(\omega, v) - g(\tilde{\omega}, \tilde{v})| \le \varepsilon/3$ for $|(\omega, v) - (\tilde{\omega}, \tilde{v})| \le \delta$ for $v \in f_{T-1}(\omega)$, $(\tilde{\omega}, \tilde{v}) \in B$. This implies $\sup_{\{\tilde{\omega}|\ |\omega-\tilde{\omega}|\le 1\}} \pi_{T-1}(g)(\tilde{\omega}) < \infty$ and that for all $\tilde{\omega} \in \Omega^{T-1}$ with $|\omega - \tilde{\omega}| \le 1$ there exists $H_T(\tilde{\omega}) \in \mathbb{R}^d$ such that

$$\varepsilon/3 + \pi_{T-1}(g)(\tilde{\omega}) + H_T(\tilde{\omega})\Delta S_T(\tilde{\omega}, \cdot) \ge g(\tilde{\omega}, \cdot) \quad \text{on } f_{T-1}(\tilde{\omega}) \tag{3.5.1}$$

or equivalently the inequality (3.5.1) holds $\mathcal{P}_{T-1}(\tilde{\omega})$-q.s.
Note that by the uniform continuity of the correspondence f_{T-1} for any $\tilde{\omega}$ close to ω and for any $v \in f_{T-1}(\omega)$ there exists $\tilde{v} \in f_{T-1}(\tilde{\omega})$ which is close to v, thus $|(\omega, v) - (\tilde{\omega}, \tilde{v})|$ is small. Furthermore we show below that $H_T(\tilde{\omega})$ can be chosen bounded uniformly in $\tilde{\omega}$ for all $\tilde{\omega}$ close to ω. Thus, for some δ_1 determined below, $|\omega - \tilde{\omega}| \le \delta_1$ implies

$$\begin{aligned}\varepsilon + \pi_{T-1}(g)(\tilde{\omega}) + H_T(\tilde{\omega})\Delta S_T(\omega, v) &\ge \varepsilon + \pi_{T-1}(g)(\tilde{\omega}) + H_T(\tilde{\omega})\Delta S_T(\tilde{\omega}, \tilde{v}) - \varepsilon/3 \\ &\ge \varepsilon/3 + g(\tilde{\omega}, \tilde{v}) \ge g(\omega, v),\end{aligned} \tag{3.5.2}$$

and thus $\pi_{T-1}(g)(\omega) \le \pi_{T-1}(g)(\tilde{\omega}) + \varepsilon$. Exchanging the roles of ω and $\tilde{\omega}$ concludes the proof of continuity of $\omega \mapsto \pi_{T-1}(\omega)$.
We now argue that there exists $\delta_0 > 0$ and $C > 0$ such that $|H_T(\tilde{\omega})| < C$ for all $\tilde{\omega} \in \Omega^{T-1}$ with $|\omega - \tilde{\omega}| \le \delta_0$ and $H_T(\tilde{\omega}) \in \operatorname{span}(f_{T-1}(\tilde{\omega}) - S_{T-1}(\tilde{\omega}))$. Assume towards a contradiction this is not the case, i.e. there exists a sequence $(\tilde{\omega}^N)_{N\in\mathbb{N}}$ with $|\omega - \tilde{\omega}^N| \le 1/N$, $H_T(\tilde{\omega}^N) \in \operatorname{span}(f_{T-1}(\tilde{\omega}^N) - S_{T-1}(\tilde{\omega}^N))$ for all $N \in \mathbb{N}$ and $\lim_{N\to\infty} |H_T(\tilde{\omega}^N)| = \infty$. After passing to a subsequence (without relabelling) $\tilde{H}^N := H_T(\tilde{\omega}^N)/|H_T(\tilde{\omega}^N)| \to \tilde{H}$ with $|\tilde{H}| = 1$. Note that as $f_{T-1}(\tilde{\omega}^N)$ converges in Hausdorff distance to $f_{T-1}(\omega)$ and as $f_{T-1}(\omega)$ is compact, it follows by the same arguments as above that $\sup_{f_{T-1}(\tilde{\omega}^N)} g(\tilde{\omega}^N, \cdot)$ and $\pi_{T-1}(g)(\tilde{\omega}^N)$ are bounded uniformly in $N \in \mathbb{N}$. Thus dividing (3.5.1) by $|H_T(\tilde{\omega}^N)|$ and taking limits we get

$$\tilde{H}\Delta S_T(\omega, \cdot) \ge 0 \quad \text{on } f_{T-1}(\omega).$$

By $\text{NA}(\mathcal{P}_{T-1}(\omega))$ this yields $\tilde{H}\Delta S_T(\omega, \cdot) = 0$ on $f_{T-1}(\omega)$. As $\tilde{H} \in \operatorname{span}(f_{T-1}(\omega) - S_{T-1}(\omega))$, $\tilde{H} = 0$ follows, a contradiction.
Now we choose $\delta_1 \le \delta_0$ such that for $|\omega - \tilde{\omega}| \le \delta_1$ we have

$$d_H((\omega, f_{T-1}(\omega)), (\tilde{\omega}, f_{T-1}(\tilde{\omega})) \le \min(\delta, \varepsilon/(3C))$$

and see that (3.5.2) holds. The proof of continuity of $\omega \mapsto \pi_t(g)(\omega)$ for $1 \le t \le T-2$ follows by backward induction using dynamic programming principle and the same arguments as above. Lastly, as for any $\mathbb{P} \in \mathfrak{P}(\mathbb{R}^d)$ such that $\mathrm{supp}(\mathbb{P}) = f_{t-1}(\omega)$

$$\pi_{t-1}(g)(\omega) + H_t(\omega)\Delta S_t(\omega, \cdot) \ge \pi_t(g)(\omega, \cdot) \quad \mathbb{P}\text{-a.s.}$$

implies

$$\pi_{t-1}(g)(\omega) + H_t(\omega)\Delta S_t(\omega, \cdot) \ge \pi_t(g)(\omega, \cdot) \quad \text{ on } f_{t-1}(\omega),$$

the claim follows. □

Remark 3.5.1. Note that the proof of boundedness of $H_T(\tilde{\omega})$ above does not require that $f_{T-1}(\tilde{\omega})$ is compact-valued as the argument above remains valid on any compact subset K of $f_{T-1}(\omega)$ such that 0 is contained in the relative interior of the convex hull of $K - S_{T-1}(\omega)$. Such a compact set exists e.g. by Tchakaloff's theorem (see [Beiglböck and Nutz, 2014, Theorem 5.1]).

3.6 Proof of Theorem 3.3.1

Lemma 3.6.1. *Let NA($\mathcal{P}$) hold. Assume that g is upper semianalytic and $g > -\infty$ $\mathcal{P}$-q.s. Then $\Omega_g^t = \{\omega \in \Omega^t \mid \mathcal{E}^t(g)(\omega) > -\infty\}$ is an analytic set and $\mathcal{E}^t(g) > -\infty$ $\mathcal{P}$-q.s. for all $0 \le t \le T$.*

Proof. Using [Bouchard and Nutz, 2015, Lemma 4.10] recursively, $\mathcal{E}^t(g)$ is upper semianalytic for all $0 \le t \le T$. As $\Omega_g^t = \bigcup_{n\ge1}\{\mathcal{E}^t(g) \ge -n\}$, Ω_g^t is an analytic set. We prove by backwards induction that $\mathcal{E}^t(g) > -\infty$ $\mathcal{P}$-q.s. This is clearly true for $t = T$ and assume first that $t \le T-1$. From [Bertsekas and Shreve, 1978, Proposition 7.50 p184] (recall (3.3.4) and that $\mathcal{Q}_t$ has an analytic graph), there exists an $\mathcal{F}_t^{\mathcal{U}}$-measurable function $\mathbb{Q}_\varepsilon : \Omega^t \to \mathfrak{P}(\Omega)$, such that $\mathbb{Q}_\varepsilon(\omega) \in \mathcal{Q}_t(\omega)$ for all $\omega \in \Omega^t$ and

$$\mathcal{E}^t(g)(\omega) \ge \mathbb{E}_{\mathbb{Q}_\varepsilon}[\mathcal{E}^{t+1}(g)(\omega, \cdot)] \ge \begin{cases} \mathcal{E}^t(g)(\omega) - \varepsilon & \text{if } \mathcal{E}^t(g)(\omega) < \infty, \\ \frac{1}{\varepsilon} & \text{otherwise} \end{cases}$$

and $(\Omega_g^t)^c \subset \{\mathbb{Q}_\varepsilon(\mathcal{E}^{t+1}(g)(\omega, \cdot) = -\infty) > 0\}$. Assume now towards a contradiction that there exists $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P}((\Omega_g^t)^c) > 0$. Changing $\mathbb{Q}_\varepsilon$ on a $\mathbb{P}$-nullset, we can assume that $\mathbb{Q}_\varepsilon$ is Borel-measurable. Then as in the proof of [Bouchard and Nutz, 2015, Lemma 4.8], one can show that there exists a $\mathcal{F}_t^{\mathcal{U}}$-measurable correspondence $\tilde{\mathbb{P}}_t : \Omega^t \to \mathfrak{P}(\Omega)$ such that $\tilde{\mathbb{P}}_t(\omega) \in \mathcal{P}_t(\omega)$ and $\mathbb{Q}_\varepsilon(\omega) \ll \tilde{\mathbb{P}}_t(\omega)$ $\mathcal{P}$-q.s. In particular setting $\tilde{\mathbb{P}} = \mathbb{P}|_{\mathcal{F}_t} \otimes \tilde{\mathbb{P}}_t$ we conclude that $\tilde{\mathbb{P}} \in \mathcal{P}$ and $\tilde{\mathbb{P}}(\mathcal{E}^{t+1}(g) = -\infty) > 0$. This concludes the proof. □

Remark 3.6.2. Recall the set $\Omega^t_{\mathrm{NA}} = \{\omega \in \Omega^t \mid \mathrm{NA}(\mathcal{P}_t(\omega)) \text{ holds}\}$, which is universally measurable and of $\mathcal{P}$-full measure (see [Bouchard and Nutz, 2015, Lemma 4.6, p.842]). Let $\omega \in \Omega^t_{\mathrm{NA}}$. From [Bouchard and Nutz, 2015, Lemma 4.1], we know that $\mathcal{E}^t(g)(\omega) = -\infty$ implies that $\{\mathcal{E}^{t+1}(g)(\omega,\cdot) = -\infty\}$ is not $\mathcal{P}_t(\omega)$-polar.

Lemma 3.6.3. *If $g : \Omega^T \to \mathbb{R}$ is upper semianalytic, then $\pi_t(g)$ is upper semianalytic for all $0 \le t \le T-1$.*

Proof. We proceed by induction. As $\pi_T(g) = g$ the claim is true for $t = T$. Assume now the $\pi_{t+1}(g)$ is upper semianalytic for some $t \in \{0, \dots, T-1\}$. We show that the claim is true for t. Indeed for all $a \in \mathbb{R}$

$$\begin{aligned}
&\{\omega \in \Omega^t \mid \pi_t(g) < a\} \\
&= \{\omega \in \Omega^t \mid \exists H \in \mathbb{R}^d,\ \varepsilon > 0 \\
&\qquad \text{such that } \forall \mathbb{P} \in \mathcal{P}_t(\omega)\ \mathbb{P}(a - \varepsilon + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(g)(\omega,\cdot)) = 1\} \\
&= \{\omega \in \Omega^t \mid \sup_{\varepsilon \in \mathbb{Q}_+} \sup_{H \in \mathbb{Q}^d} \inf_{\mathbb{P} \in \mathcal{P}_t(\omega)} \mathbb{P}(a - \varepsilon + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(g)(\omega,\cdot)) \ge 1\}
\end{aligned}$$

As the function $(\omega, \mathbb{P}, H, \varepsilon) \mapsto \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{a-\varepsilon+H\Delta S_{t+1}(\omega,\cdot)\ge \pi_{t+1}(g)(\omega,\cdot)\}}\right]$ is lower semianalytic, the same holds true for $\omega \mapsto \sup_{\varepsilon\in\mathbb{Q}_+} \sup_{H\in\mathbb{Q}^d} \inf_{\mathbb{P}\in\mathcal{P}_t(\omega)} \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{a-\varepsilon+H\Delta S_{t+1}(\omega,\cdot)\ge\pi_{t+1}(g)(\omega,\cdot)\}}\right]$ (see [Bertsekas and Shreve, 1978, Lemma 7.30, p.177, Prop. 7.47, p.180]), thus the set above is coanalytic. To complete the proof, we argue why

$$\begin{aligned}
&\{\omega \in \Omega^t \mid \exists H \in \mathbb{R}^d,\ \varepsilon > 0 \text{ such that } a - \varepsilon + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(\omega,\cdot)\ \mathcal{P}_t(\omega)\text{-q.s.}\} \\
&\subseteq \{\omega \in \Omega^t \mid \exists H \in \mathbb{Q}^d,\ \varepsilon \in \mathbb{Q}_+ \text{ such that } a - \varepsilon + H\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(\omega,\cdot)\ \mathcal{P}_t(\omega)\text{-q.s.}\} :
\end{aligned}$$

Fix $\omega \in \Omega^t$, $\tilde{H} \in \mathbb{R}^d$, $\varepsilon > 0$ such that $a - \varepsilon + \tilde{H}\Delta S_{t+1}(\omega,\cdot) \ge \pi_{t+1}(\omega,\cdot)$ $\mathcal{P}_t(\omega)$-q.s. Take $\tilde{\varepsilon} \in \mathbb{Q}_+$ such that $0 < \tilde{\varepsilon} < \varepsilon/2$ and $H \in [0,\infty)^d$ such that

$$H^1 + \cdots + H^d \le \frac{\varepsilon/2}{\max_{1\le i\le d} S^i_t(\omega)}.$$

It follows that for $\mathcal{P}_t(\omega)$-q.e. $\omega' \in \Omega$

$$\begin{aligned}
a - \tilde{\varepsilon} + (H + \tilde{H})\Delta S_{t+1}(\omega,\omega') &\ge a - \varepsilon/2 + \tilde{H}\Delta S_{t+1}(\omega,\omega') + H\Delta S_{t+1}(\omega,\omega') \\
&\ge \pi_{t+1}(g)(\omega,\omega') + \varepsilon/2 - HS_t(\omega) \\
&\ge \pi_{t+1}(g)(\omega,\omega').
\end{aligned}$$

In particular the above inequality is valid for some H such that $\tilde{H} + H \in \mathbb{Q}^d$. ☐

Proof of Theorem 3.3.1. Recall the set $\Omega^t_{\mathrm{NA}} = \{\omega \in \Omega^t \mid \mathrm{NA}(\mathcal{P}_t(\omega)) \text{ holds}\}$, which is universally measurable and of $\mathcal{P}$-full measure (see [Bouchard and Nutz, 2015, Lemma 4.6, p.842]). Let

$$\Omega_{\mathrm{NA},g} := \{\omega \in \Omega^T \mid \omega \in \Omega^t_{\mathrm{NA}} \cap \Omega^t_g \text{ for all } 0 \leq t \leq T-1\},$$

where the definition of Ω^t_g is given in Lemma 3.6.1. Then by Lemma 3.6.1 and [Bouchard and Nutz, 2015, Lemma 4.6, p. 842] $\Omega_{\mathrm{NA},g}$ is universally measurable and of $\mathcal{P}$-full measure. Let $\omega \in \Omega_{\mathrm{NA},g}$. By [Bouchard and Nutz, 2015, Lemma 4.10] (replacing the last condition by $\mathcal{E}^t(g)(\omega) > -\infty$), there exists a universally measurable function $\hat{H}_{t+1}$ such that

$$\mathcal{E}^t(g)(\omega) + \hat{H}_{t+1}(\omega)\Delta S_{t+1}(\omega,\cdot) \geq \mathcal{E}^{t+1}(g)(\omega,\cdot) \quad \mathcal{P}_t(\omega)\text{-q.s.} \tag{3.6.1}$$

To see that

$$\pi_t(g) = \mathcal{E}^t(g) \quad \mathcal{P}\text{-q.s.} \tag{3.6.2}$$

for $0 \leq t \leq T$ we argue by backwards induction. Indeed the claim is true by definition for $t = T$. Now we assume that the claim is true for $t+1 \in \{1, \dots, T\}$. By [Bouchard and Nutz, 2015, eq. (4.8) in Lemma 4.8, p.843] the correspondence

$$\mathcal{H}_t(\omega) = \{(\mathbb{Q}, \mathbb{P}) \in \mathfrak{P}(\Omega) \times \mathfrak{P}(\Omega) \mid \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)] = 0,\ \mathbb{P} \in \mathcal{P}_t(\omega),\ \mathbb{Q} \ll \mathbb{P}\}$$

has analytic graph. By [Bertsekas and Shreve, 1978, Prop. 7.47, p. 179, Prop. 7.48, p. 180, Prop. 7.50, p.184] $(\omega, \mathbb{Q}, \mathbb{P}) \mapsto \mathbb{E}_{\mathbb{Q}}[\mathcal{E}_{t+1}(g)(\omega,\cdot)]$ and $(\omega, \mathbb{Q}, \mathbb{P}) \mapsto \mathbb{E}_{\mathbb{Q}}[\pi_{t+1}(g)(\omega,\cdot)]$ are upper seminanalytic functions and there exists sequences $(\hat{\mathbb{P}}_n, \hat{\mathbb{Q}}_n)_{n\in\mathbb{N}}$ and $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n\in\mathbb{N}}$ of $\mathcal{F}^{\mathcal{U}}_t$-measurable selectors of $\mathcal{H}_t$ such that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}_{\hat{\mathbb{Q}}_n(\omega)}[\mathcal{E}^{t+1}(g)(\omega,\cdot)] &= \sup_{(\mathbb{Q},\mathbb{P})\in\mathcal{H}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}^{t+1}(g)(\omega,\cdot)] = \mathcal{E}^t(g)(\omega),\\
\lim_{n\to\infty} \mathbb{E}_{\bar{\mathbb{Q}}_n(\omega)}[\pi_{t+1}(g)(\omega,\cdot)] &= \sup_{(\mathbb{Q},\mathbb{P})\in\mathcal{H}_t(\omega)} \mathbb{E}_{\mathbb{Q}}[\pi_{t+1}(g)(\omega,\cdot)] = \mathcal{E}_t(\pi_{t+1}(g))(\omega).
\end{aligned}$$

Define $\mathbb{P}_n(\omega) = (\hat{\mathbb{P}}_n(\omega) + \bar{\mathbb{P}}_n(\omega))/2 \in \mathcal{P}_t(\omega)$ and $\tilde{\mathbb{P}}_t(\omega) = \sum_{n=1}^\infty 2^{-n}\mathbb{P}_n(\omega)$. Then $\tilde{\mathbb{P}}_t(\omega) \in \mathfrak{P}(\Omega)$ for all $\omega \in \Omega^t$, $\omega \mapsto \tilde{\mathbb{P}}_t(\omega)$ is $\mathcal{F}^{\mathcal{U}}_t$-measurable and $\hat{\mathbb{P}}_n(\omega), \bar{\mathbb{P}}_n(\omega), \mathbb{P}_n(\omega)$ are absolutely continuous with respect to $\tilde{\mathbb{P}}_t(\omega)$. Furthermore for $\omega \in \Omega^t_{\mathrm{NA}}$

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}_n(\omega)}[\mathcal{E}^{t+1}(g)(\omega,\cdot)] &\leq \sup_{\mathbb{Q}\ll\tilde{\mathbb{P}}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}^{t+1}(g)(\omega,\cdot)]\\
&\leq \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d\\
&\qquad \text{such that } x + H\Delta S_{t+1}(\omega,\cdot) \geq \mathcal{E}^{t+1}(g)(\omega,\cdot)\ \tilde{\mathbb{P}}_t(\omega)\text{-a.s.}\}\\
&\leq \pi_t(\mathcal{E}^{t+1}(g))(\omega) = \mathcal{E}_t(\mathcal{E}^{t+1}(g))(\omega) = \mathcal{E}^t(g)(\omega),
\end{aligned}$$

where the third inequality follows from the fact that $\mathbb{P}_n(\omega) \in \mathcal{P}_t(\omega)$ for $n \in \mathbb{N}$ and the first equality follows from [Bouchard and Nutz, 2015, Theorem 3.4] as $\omega \in \Omega^t_{\text{NA}}$. The same reasoning applies with $\pi_{t+1}(g)(\omega, \cdot)$ instead of $\mathcal{E}^{t+1}(g)(\omega, \cdot)$ and letting $n \to \infty$ we conclude

$$\sup_{\mathbb{Q} \ll \tilde{\mathbb{P}}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}^{t+1}(g)(\omega, \cdot)] = \mathcal{E}^t(g)(\omega),$$
$$\sup_{\mathbb{Q} \ll \tilde{\mathbb{P}}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0} \mathbb{E}_{\mathbb{Q}}[\pi_{t+1}(g)(\omega, \cdot)] = \pi_t(g)(\omega) = \mathcal{E}_t(\pi_{t+1})(g)(\omega).$$

Fix now $\mathbb{P} \in \mathcal{P}$ and define $\tilde{\mathbb{P}} = \mathbb{P}|_{\mathcal{F}^{\mathcal{U}}_t} \otimes \tilde{\mathbb{P}}_t$. Then as $\mathbb{P}_n(\omega) \in \mathcal{P}_t(\omega)$ the induction assumption implies that $\mathcal{E}^{t+1}(g) = \pi_{t+1}(g)$ holds $\tilde{\mathbb{P}}$-a.s. and thus for $\tilde{\mathbb{P}}$-a.e. $\omega \in \Omega^t$ we have

$$\begin{aligned}
\mathcal{E}^t(g)(\omega) &= \sup_{\mathbb{Q} \ll \tilde{\mathbb{P}}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0} \mathbb{E}_{\mathbb{Q}}[\mathcal{E}^{t+1}(g)(\omega, \cdot)] \\
&= \sup_{\mathbb{Q} \ll \tilde{\mathbb{P}}_t(\omega),\ \mathbb{E}_{\mathbb{Q}}[\Delta S_{t+1}(\omega,\cdot)]=0} \mathbb{E}_{\mathbb{Q}}[\pi_{t+1}(g)(\omega, \cdot)] \\
&= \mathcal{E}_t(\pi_{t+1}(g))(\omega) = \pi_t(g)(\omega).
\end{aligned}$$

This concludes the proof of (3.6.2).
Let $(x, H, C) \in \mathcal{A}(g)$. Now we show that

$$V^{x,H,C}_t \geq \pi_t(g) \quad \mathcal{P}\text{-q.s.} \tag{3.6.3}$$

This is clearly true at $t = T$. Fix some $1 \leq t \leq T$ and assume that (3.6.3) holds true for t. Then

$$V^{x,H,C}_{t-1} + H_t\Delta S_t \geq V^{x,H,C}_t \geq \pi_t(g) \quad \mathcal{P}\text{-q.s.}$$

Noting that $V^{x,H,C}_{t-1}$ is $\mathcal{F}^{\mathcal{U}}_{t-1}$-measurable and $\pi_t(g)$ is upper seminanalytic and using the same reasoning as in [Bouchard and Nutz, 2015, proof of Lemma 4.10, pp.846-848] we conclude that for $\omega \in \Omega^{t-1}$ in a $\mathcal{P}$ full-measure set

$$V^{x,H,C}_{t-1}(\omega) + H_t(\omega)\Delta S_t(\omega, \cdot) \geq \pi_t(g)(\omega, \cdot) \quad \mathcal{P}_{t-1}(\omega)\text{-q.s.} \tag{3.6.4}$$

Thus $V^{x,H,C}_{t-1}(\omega) \geq \pi_{t-1}(g)(\omega)$ by (3.3.2) and (3.6.3) is proved for $t-1$. Next we define the consumption process $\hat{C}$. Let $\mathbb{P} = \mathbb{P}_0 \otimes \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_{T-1} \in \mathcal{P}$, where $\mathbb{P}_t \in \mathcal{P}_t(\omega)$ for all $0 \leq t \leq T-1$. Then using (3.6.1) and Fubini's Theorem (recall [Bertsekas and Shreve, 1978, Proposition 7.45 p175]), we get that

$$\mathcal{E}^{t-1}(g) + \hat{H}_t\Delta S_t \geq \mathcal{E}^t(g) \quad \mathcal{P}\text{-q.s.} \tag{3.6.5}$$

for a universally measurable function $\hat{H}_t : \Omega^t \to \mathbb{R}^d$. Using (3.6.5) recursively,

$$\mathcal{E}^0(g) + \sum_{u=1}^{t} \hat{H}_u\Delta S_u \geq \mathcal{E}^t(g) \quad \mathcal{P}\text{-q.s.} \tag{3.6.6}$$

follows. Now we set $\hat{C}_t = \mathcal{E}^0(g) + \sum_{u=1}^t \hat{H}_u \Delta S_u - \mathcal{E}^t(g)$. Then $\hat{C}_t(\omega,\cdot) - \hat{C}_{t-1}(\omega) = \mathcal{E}^{t-1}(g)(\omega) - \mathcal{E}^t(g)(\omega,\cdot) + \hat{H}_t(\omega)\Delta S_t(\omega,\cdot) \geq 0$ $\mathcal{P}_{t-1}(\omega)$-q.s. and using again Fubini's Theorem $\hat{C}_t - \hat{C}_{t-1} \geq 0$ $\mathcal{P}$-q.s. Thus $\hat{C} = (\hat{C}_t)_{0\leq t\leq T}$ is a cumulative consumption process.
Now we prove that $\pi(g) = \pi_0(g)$. Let (x, H) such that $V_T^{x,H} \geq g$ $\mathcal{P}$-q.s. Then as $V_{T-1}^{x,H} + H_T \Delta S_T \geq g$ $\mathcal{P}_{T-1}$-q.s. it follows as in (3.6.4)

$$V_{T-1}^{x,H}(\omega) + H_T(\omega)\Delta S_T(\omega,\cdot) \geq g(\omega,\cdot)\ \mathcal{P}_{T-1}(\omega)\text{-q.s.}$$

for all for $\omega \in \Omega^{T-1}$ in an $\mathcal{F}_{T-1}^{\mathcal{U}}$-measurable and $\mathcal{P}$-full measure set. From (3.3.2), we conclude that $\pi_{T-1}(g)(\omega) \leq V_{T-1}^{x,H}(\omega)$. By induction we see that $\pi_0(g) \leq x$ and thus $\pi_0(g) \leq \pi(g)$. Conversely, using (3.6.6) and (3.6.2)

$$V_T^{\pi_0,\hat{H}} = \pi_0(g) + \sum_{t=1}^T \hat{H}_t \Delta S_t \geq \mathcal{E}^T(g) = g \quad \mathcal{P}\text{-q.s.}$$

and therefore $\pi_0(g) \geq \pi(g)$. Thus $\mathcal{E}^0(g) = \pi_0(g) = \pi(g)$ by (3.6.2) and we obtain (recall (3.6.6) and the definition of $\hat{C}$) that

$$V_t^{\pi(g),\hat{H},\hat{C}} = \mathcal{E}^t(g) = \pi_t(g) \quad \mathcal{P}\text{-q.s.}$$

Since $V_T^{\pi(g),\hat{H},\hat{C}} = \mathcal{E}^T(g) = g \quad \mathcal{P}$-q.s., $(\pi(g), \hat{H}, \hat{C})$ is a superhedging strategy and it is also minimal. Indeed let $(x, H, C) \in \mathcal{A}(g)$ then $V_T^{x,H,C} \geq g$ $\mathcal{P}$-q.s. From (3.6.3), $V_t^{x,H,C} \geq \pi_t(g) = V_t^{\pi(g),\hat{H},\hat{C}}$ $\mathcal{P}$-q.s. This concludes the proof. □

3.7 Proofs of Theorem 3.4.3 and Theorem 3.4.5

3.7.1 Proof of Theorem 3.4.3: the one-period case

We now prove Theorem 3.4.3 in the case $T = 1$, where we follow arguments given in Nutz [2016]. Let $g : \Omega^T \to \mathbb{R}$ be Borel. In preparation for the multi-period case we define the set

$$\mathcal{A}_{0,x} = \{(H, c) \in \mathbb{R}^d \times \mathbb{R}_+ \mid x - c + H\Delta S_1 \geq \pi_1(g)\ \mathcal{P}\text{-q.s.}\}.$$

Recall definition $\pi_t(g)$ given in (3.3.2) for $t = 0, 1$ and note that if $(H, c) \in \mathcal{A}_{0,x}$ then also $(H, 0) \in \mathcal{A}_{0,x}$. We thus often write $H \in \mathcal{A}_{0,x}$ instead of $(H, c) \in \mathcal{A}_{0,x}$. Let $U(1,\cdot,\cdot) : \Omega \times [0,\infty) \to \mathbb{R}$ be bounded from above and $\mathcal{F}_1^{\mathcal{U}}$-measurable. Besides let us assume that $x \mapsto U(1,\omega,x)$ is non-decreasing, concave and continuous for each $\omega \in \Omega$. Furthermore let the deterministic function $U(0,\cdot) : [0,\infty) \to \mathbb{R}$ be non-decreasing and continuous. As usual we set $U(t,\omega,x) = -\infty$ for $x < 0$ and $t = 0, 1$. Let us now state the main theorem for $T = 1$:

Proposition 3.7.1. *Let NA($\mathcal{P}$) hold and $x \geq \pi_0(g)$. Then*

$$u(x) := \sup_{(H,c)\in\mathcal{A}_{0,x}} \left(\inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x - c + H\Delta S_1 - \pi_1(g))] + U(0,c) \right) < \infty$$

and there exists $(H^*, c^*) \in \mathcal{A}_{0,x}$ such that

$$\inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x - c^* + H^*\Delta S_1 - \pi_1(g))] + U(0,c^*) = u(x).$$

We prove the result via a lemma. Here we denote

$$L = \text{span}\ (\{\text{supp}(\mathbb{P}\circ(\Delta S_1)^{-1} \mid \mathbb{P}\in\mathcal{P}\}) \subseteq \mathbb{R}^d$$

and the orthogonal complement

$$L^\perp = \{H \in \mathbb{R}^d \mid HV = 0 \text{ for all } V \in L\}.$$

Lemma 3.7.2. *Assume $x \geq \pi_0(g)$. Under NA($\mathcal{P}$) the set $K_x = \mathcal{A}_{0,x} \cap (L \times \mathbb{R}_+) \subseteq \mathbb{R}^{d+1}$ is non-empty, convex and compact.*

Proof. Clearly K_x is convex and closed. It remains to show that K_x is bounded: as by definition of $\pi_0(g)$ clearly $c \in [0, x - \pi_0(g)]$ for all c such that $(H,c) \in \mathcal{A}_{0,x}$ we only need to show that $(H,0) \in K_x$ is bounded. Note that after a translation by $(H_0, 0) \in K_x$ we have $0 \in \tilde{K}_x := K_x - (H_0, 0)$. Now we assume towards a contradiction that there exist $(H_n, 0) \in \tilde{K}_x$ such that $|H_n| \to \infty$. We define $\delta = |H_0| + 1$. We can extract a subsequence $\delta H_n/|H_n|$ that converges to a limit $H \in \mathbb{R}^d$, so $|H| = \delta$. As $\tilde{K}_x$ is convex and contains the origin we have for n large enough $(\delta H_n/|H_n|, 0) \in \tilde{K}_x$. It follows $(H,0) \in \tilde{K}_x$, since $\tilde{K}_x$ is closed. Furthermore

$$H\Delta S_1 \geq \liminf_{n\to\infty} \frac{\pi_1(g) - x - H_0\Delta S_1}{|H_n|/\delta} = 0 \qquad \mathcal{P}\text{-q.s.}$$

By NA($\mathcal{P}$) this implies $H\Delta S_1 = 0$ $\mathcal{P}$-q.s. and thus $H \in L^\perp$ by use of [Nutz, 2016, Lemma 2.6]. As $(H,0) \in \tilde{K}_x$ this implies $H_0 + H \in L$, which means $|H|^2 = -H_0 H$. This contradicts $|H| = \delta$ by Cauchy-Schwarz inequality. □

Proof of Proposition 3.7.1. Fatou's lemma implies that for all $\mathbb{P} \in \mathcal{P}^u$ the function $(H,c) \mapsto \mathbb{E}_{\mathbb{P}}[U(1, x - c + H\Delta S_1 - \pi_1(g))] + U(0,c)$ is upper semicontinuous on $\mathcal{A}_{0,x}$. It follows that $(H,c) \mapsto \inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x - c + H\Delta S_1 - \pi_1(g))] + U(0,c)$ is upper semicontinuous and thus attains its supremum on the compact set K_x. Finally again using [Nutz, 2016, Lemma 2.6] and recalling that $\mathcal{P}^u \subseteq \mathcal{P}$

$$\begin{aligned} &\sup_{(H,c)\in\mathcal{A}_{0,x}} \left(\inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x - c + H\Delta S_1 - \pi_1(g))] + U(0,c) \right) \\ =& \sup_{(H,c)\in K_x} \left(\inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x - c + H\Delta S_1 - \pi_1(g))] + U(0,c) \right). \end{aligned}$$

We conclude using [Aliprantis and Border, 2006, Theorem 2.43, p.44]. □

Corollary 3.7.3. *Under the conditions of Proposition 3.7.1 we have*

$$\sup_{(H,c)\in\mathcal{A}_{0,x}} \left(\inf_{\mathbb{P}\in\mathcal{P}^u} \mathbb{E}_{\mathbb{P}}[U(1, x-c+H\Delta S_1 - \pi_1(g))] + U(0,c) \right)$$
$$= \inf_{\mathbb{P}\in\mathcal{P}^u} \left(\sup_{(H,c)\in\mathcal{A}_{0,x}} \left(\mathbb{E}_{\mathbb{P}}[U(1, x-c+H\Delta S_1 - \pi_1(g))] + U(0,c) \right) \right).$$

Proof. Note that K_x is compact, convex and $\mathcal{P}^u$ is convex. Define

$$f : K_x \times \mathfrak{P}(\Omega) \to \mathbb{R} \quad (H,c,\mathbb{P}) \mapsto \mathbb{E}_{\mathbb{P}}[U(1, x-c+H\Delta S_1 - \pi_1(g))] + U(0,c)$$

and note that $(H,c) \mapsto f(H,c,\mathbb{P})$ is upper semicontinuous and concave. Furthermore $\mathbb{P} \mapsto f(H,c,\mathbb{P})$ is convex on $\mathcal{P}^u$. The claim follows from Corollary 2 in Terkelsen [1973]. □

Remark 3.7.4. The assumption of boundedness from above of $U(1,\cdot,\cdot)$ in Proposition 3.7.1 can be replaced by a weaker condition: Indeed it is sufficient to assume there exists a constant $a > 0$ such that $\omega \mapsto U(1,\omega,a/2)$ is bounded from below and

$$\mathbb{E}_{\mathbb{P}}[U^+(1, x+H\Delta S_1 - \pi_1(g))] < \infty \quad \text{for all } H \in \mathcal{A}_{0,x} \text{ and } \mathbb{P} \in \mathcal{P}^u$$

as well as

$$\mathbb{E}_{\mathbb{P}}[U^+(1,a)] < \infty \quad \text{for all } \mathbb{P} \in \mathcal{P}^u.$$

The proof of Proposition 3.7.1 then follows along the lines of [Rásonyi and Stettner, 2006, Lemma 1] and [Nutz, 2016, Lemma 2.8] after a translation by some $H_0 \in \mathrm{ri}(K_x)$.

3.7.2 Proof of Theorem 3.4.3: The multi-period case

For the rest of this section we assume NA($\mathcal{P}$) and that g is Borel measurable. Furthermore we often abbreviate $\pi_t(g)$ by π_t. To simplify notation we assume $U(0,\cdot,0) = 0$. We give the following definition:

Definition 3.7.5. *We define $U_T(\omega,x) = U(T,\omega,x)$ and for $0 \le t \le T-1$*

$$U_t(\omega,x) := \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \left(\inf_{\mathbb{P}\in\mathcal{P}^u_t(\omega)} \mathbb{E}_{\mathbb{P}}[U_{t+1}((\omega,\cdot), x+H\Delta S_{t+1}(\omega,\cdot) - c - \mathbb{1}_{\{t=T-1\}} g(\omega,\cdot))] \right.$$
$$\left. + U(t,\omega,c) \right), \quad x \ge \pi_t(\omega)$$

and $U_t(\omega,x) = -\infty$ otherwise, where for $x \in \mathbb{R}$ we set

$$\mathcal{A}_{0,x}(\omega) := \{(H,c) \in \mathbb{R}^d \times \{0\} \mid x + H\Delta S_1(\omega,\cdot) \ge \pi_1(\omega,\cdot)\ \mathcal{P}_0(\omega)\text{-}q.s.\}$$
$$\mathcal{A}_{t,x}(\omega) := \{(H,c) \in \mathbb{R}^d \times \mathbb{R}_+ \mid x + H\Delta S_{t+1}(\omega,\cdot) - c \ge \pi_{t+1}(\omega,\cdot)\ \mathcal{P}_t(\omega)\text{-}q.s.\}, \quad t \ge 1.$$

We recall from Lemma 3.6.3 that $\pi_t(g)$ is upper semianalytic. This means in particular that

$$\{(\omega, x) \mid x < \pi_t(g)(\omega)\} = \bigcup_{q\in\mathbb{Q}} \pi_t^{-1}((q,\infty)) \times (-\infty, q)$$

is analytic. Next we show by backwards induction, that if Assumption 3.4.2 is satisfied, then U_t has $\mathcal{P}^u$-q.s. the following properties:

Condition 3.7.6. *Let $0 \le t \le T-1$. The function $U_t : \Omega^t \times \mathbb{R} \to [-\infty, \infty)$ is lower semianalytic and bounded from above. Furthermore the following properties hold:*

(i) $\omega \mapsto U_t(\omega, x(\omega))$ is bounded from below for $x(\omega) := \pi_t(\omega) + \varepsilon$ and each $\varepsilon > 0$.

(ii) $x \mapsto U_t(\omega, x)$ is non-decreasing, concave and continuous on $[\pi_t(\omega), \infty)$ for each $\omega \in \Omega^t$.

Lemma 3.7.7. *Let $NA(\mathcal{P})$ and Assumptions 3.2.1, 3.4.1 and 3.4.2 hold for $U(t,\cdot,\cdot)$, $0 \le t \le T$. Then there exist functions $\tilde{U}_t : \Omega^t \times (-\infty, \infty) \to [-\infty, \infty)$, which satisfy Condition 3.7.6, such that $\tilde{U}_t = U_t$ $\mathcal{P}^u$-q.s.*

Proof. We prove the claim by induction. Recall that U_T satisfies Assumption 3.4.2. We now show the induction step from $t+1$ to t and therefore first fix $\omega \in \Omega^t$. For simplicity of presentation we assume $t \le T-2$.
We first state some results regarding lower semianalyticity, which lead to the definition of $\tilde{U}_t$: using [Bertsekas and Shreve, 1978, Lemma 7.30, p.177, Prop. 7.47, p.179, Prop. 7.48, p.180], Assumption 3.4.2 and the analytic graph of $\mathcal{P}_t^u$ we see that $\phi : \Omega^t \times (-\infty,\infty) \times \mathbb{R}^d \times \mathbb{R} \to \overline{\mathbb{R}}$

$$\phi(\omega, x, H, c) = \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot), x + H\Delta S_{t+1}(\omega,\cdot) - c)] + U(t,\omega,c)$$

is lower semianalytic as $\Delta S_{t+1}(\omega,\cdot)$ is a Borel measurable functions (and also $g(\omega,\cdot)$ for $t = T-1$). Now we define the function $\tilde{\phi} : \Omega^t \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \to \overline{\mathbb{R}}$

$$\widetilde{\phi}(\omega, x, H, c) = \begin{cases} -\infty & \text{if } (H,c) \notin \mathcal{A}_{t,x} \text{ or } x < \pi_t(g)(\omega) \\ \phi(\omega, x, H, c) & \text{otherwise.} \end{cases}$$

We show that $\tilde{\phi}$ is lower semianalytic. Fix $a \in \mathbb{R}$. Then

$$\begin{aligned}\left\{\widetilde{\phi} < a\right\} &= \{(\omega, x, H, c) \mid \phi(\omega, x, H, c) < a,\ (H,c) \in \mathcal{A}_{t,x}(\omega), x \ge \pi_t(g)(\omega)\} \\ &\cup \{(\omega, x, H, c) \mid (H,c) \notin \mathcal{A}_{t,x}(\omega) \text{ or } x < \pi_t(g)(\omega)\} \\ &= \{\phi < a\} \cup \{(\omega, x, H, c) \mid (H,c) \notin \mathcal{A}_{t,x}(\omega)\} \\ &\cup \{(\omega, x, H, c) \mid x < \pi_t(g)(\omega)\}.\end{aligned}$$

By the same arguments as for the lower seminanalyticity of ϕ we see that

$$\begin{aligned}&\{(\omega, x, H, c) \mid (H, c) \notin \mathcal{A}_{t,x}(\omega)\}\\ = &\left\{(\omega, x, H, c) \;\middle|\; \sup_{\mathbb{P} \in \mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}}[x + H\Delta S_{t+1}(\omega, \cdot) - c - \pi_{t+1}(\omega, \cdot)]^- > 0\right\}\end{aligned}$$

is analytic and the sets

$$\{\phi < a\} \qquad \text{and} \qquad \{(\omega, x, H, c) \in \Omega^t \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \mid x < \pi_t(g)(\omega)\}$$

are analytic, so $\widetilde{\phi}$ is lower semianalytic. Similarly to [Blanchard and Carassus, 2017, Proposition 3.27] we define

$$\tilde{U}_t(\omega, x) = \lim_{n\to\infty} \sup_{(H,c)\in\mathbb{Q}^d\times\mathbb{Q}_+} \widetilde{\phi}\left(\omega, x + \frac{1}{n}, H, c\right). \tag{3.7.1}$$

As the limits and countable supremum of lower semianalytic functions is lower semianalytic, we conclude that $\tilde{U}_t$ is lower semianalytic.
From the definition it is clear that $\tilde{U}_t(\omega, \cdot)$ is non-decreasing and bounded from above. Next we argue that $\tilde{U}_t(\omega, \cdot)$ is concave. As the infimum of concave functions is concave, it is enough to argue that $x \mapsto \sup_{(H,c)\in\mathbb{Q}^d\times\mathbb{Q}_+} \widetilde{\phi}(\omega, x, H, c)$ is concave. This follows very similarly to Rásonyi and Stettner [2006][proof of Prop. 2, p.5]: Indeed, it is enough to show midpoint-concavity of $\sup_{(H,c)\in\mathbb{Q}^d\times\mathbb{Q}_+} \widetilde{\phi}(\omega, \cdot, H, c)$, which is immediate by use of triangle inequality. Concavity implies that $\tilde{U}_t(\omega, \cdot)$ is continuous on $(\pi_t(\omega), \infty)$. By the definition of $\tilde{U}_t$ concavity and continuity extend to $[\pi_t(\omega), \infty)$.
By definition we clearly have

$$\sup_{(H,c)\in\mathbb{Q}^d\times\mathbb{Q}_+} \widetilde{\phi}(\omega, x, H, c) \leq \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \phi(\omega, x, H, c).$$

We now show equality of $U_t(\omega, x)$ and $\tilde{U}_t(\omega, x)$ for $\mathcal{P}^u$-q.e. $\omega \in \Omega^t$. Let us therefore fix $x > \pi_t(\omega)$ and $\omega \in \Omega^t_{\mathrm{NA}}$. Using [Bouchard and Nutz, 2015, Theorem 3.4] and $\mathcal{P}_t^u(\omega) \subseteq \mathcal{P}_t(\omega)$ there exists $\tilde{H} \in \mathbb{R}^d$ such that

$$\pi_t(\omega) + \tilde{H}\Delta S_{t+1}(\omega, \omega') \geq \pi_{t+1}(\omega, \omega') \quad \text{ for } \mathcal{P}_t^u(\omega)\text{-q.e. } \omega' \in \Omega.$$

Take $c < x - \pi_t(\omega)$ and $H \in [0, \infty)^d$ such that

$$H^1 + \cdots + H^d \leq \frac{x - \pi_t(\omega) - c}{\max_{1\leq i\leq d} S_t^i(\omega)}.$$

It follows for $\mathcal{P}_t^u(\omega)$-q.e. $\omega' \in \Omega$ that

$$\begin{aligned} x + (H + \tilde{H})\Delta S_{t+1}(\omega, \omega') - c &= x - \pi_t(\omega) + H\Delta S_{t+1}(\omega, \omega') + \pi_t(\omega) + \tilde{H}\Delta S_{t+1}(\omega, \omega') - c\\ &\geq x - \pi_t(\omega) - HS_t(\omega) + \pi_{t+1}(\omega, \omega') - c\\ &\geq \pi_{t+1}(\omega, \omega'). \end{aligned}$$

Thus the affine hull of $\mathcal{A}_{t,x}(\omega)$ is $\mathbb{R}^{d+1}$ and consequently $\mathrm{Ri}(\mathcal{A}_{t,x}(\omega))$ is an open set in $\mathbb{R}^{d+1}$. This implies

$$\sup_{(H,c)\in\mathbb{Q}^d\times\mathbb{Q}_+} \widetilde{\phi}(\omega,x,H,c) = \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \phi(\omega,x,H,c).$$

for $x>\pi_t(\omega)$. Equality in $x=\pi_t(\omega)$ follows by right-continuity of U_t and $\tilde{U}_t$. Indeed, right-continuity of $U_t(x,\omega)$ in $x=\pi_t(\omega)$ follows by compactness of $\mathcal{A}_{t,\pi_t(\omega)+1}(\omega)\cap(\mathrm{span}(\mathrm{supp}(\{\mathbb{P}\circ(\Delta S_{t+1}(\omega,\cdot)^{-1} \mid \mathbb{P}\in\mathcal{P}_t(\omega)\}))) \times \mathbb{R}$ and Fatou's Lemma.
Lastly we show boundedness of $\tilde{U}_t$ from below: let $x(\omega)=\pi_t(\omega)+\varepsilon$ for some $\varepsilon>0$. By the above arguments there exists $\hat{H}\in\mathbb{Q}^d$ such that $\pi_t(\omega)+\varepsilon/3+\hat{H}\Delta S_{t+1}(\omega,\omega')\geq\pi_{t+1}(\omega,\omega')$ $\mathcal{P}_t^u(\omega)$-a.s. Thus

$$\begin{aligned}\tilde{U}_t(\omega,x(\omega)) &\geq \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_\mathbb{P}[\tilde{U}_{t+1}((\omega,\cdot),x(\omega)+\hat{H}\Delta S_{t+1}(\omega,\cdot)-\varepsilon/3)]+U(t,\omega,\varepsilon/3)\\ &\geq \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_\mathbb{P}[\tilde{U}_{t+1}((\omega,\cdot),\pi_{t+1}(\omega,\cdot)+\varepsilon/3)]+U(t,\omega,\varepsilon/3)\end{aligned}$$

is bounded from below by the induction hypothesis and Assumption 3.4.2. This shows the claim.

□

Lemma 3.7.8. *Let $NA(\mathcal{P})$ and Assumptions 3.2.1, 3.4.1 and 3.4.2 hold for $U(t,\cdot,\cdot)$, $0\leq t\leq T$. Let $t\in\{0,\dots,T-1\}$ and $(H,C)\in\mathcal{A}_{\pi_0}$. There exist universally measurable mappings H^*_{t+1}, c^*_t such that c^*_t is non-negative,*

$$V_{t-1}^{\pi_0,H,C}(\omega)+H_t(\omega)\Delta S_t(\omega)+H^*_{t+1}(\omega)\Delta S_{t+1}(\omega,\cdot)-c^*_t(\omega)\geq\pi_{t+1}(\omega,\cdot)\ \mathcal{P}_t(\omega)\text{-q.s.}$$

and

$$\begin{aligned}&\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_\mathbb{P}\Big[U_{t+1}\Big((\omega,\cdot),V_{t-1}^{\pi_0,H,C}(\omega)+H_t(\omega)\Delta S_t(\omega)+H^*_{t+1}(\omega)\Delta S_{t+1}(\omega,\cdot)-c^*_t(\omega)\\ &\qquad -\mathbb{1}_{\{t=T-1\}}g(\omega,\cdot)\Big)\Big]+U(t,\omega,c^*_t(\omega))\\ &=U_t\Big(\omega,V_{t-1}^{\pi_0,H,C}(\omega)+H_t(\omega)\Delta S_t(\omega)\Big)\end{aligned}$$

for $\mathcal{P}^u$-a.e. $\omega\in\Omega^t$.

Proof. We show that $\tilde{U}_t$ is $\mathcal{F}_t^{\mathcal{U}}\otimes\mathcal{B}(\mathbb{R})$-measurable using Lemma 3.7.7: indeed, we know that $\omega\mapsto\tilde{U}_t(\omega,x)$ is lower seminanalytic and in particular universally measurable. Also $x\mapsto\tilde{U}_t(\omega,x)$ is continuous on $[\pi_t(\omega),\infty)$, bounded from above and $\tilde{U}_t(\omega,x)=-\infty$ for $x<\pi_t(\omega)$. Thus is it concave and upper semicontinuous on $\mathbb{R}$ and the claim follows from [Blanchard and Carassus, 2017, Lemma A.35, p. 1889]. Next we show that the function

$$\phi(\omega,x,H,c)=\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)}\mathbb{E}_\mathbb{P}[\tilde{U}_{t+1}((\omega,\cdot),x+H\Delta S_{t+1}(\omega,\cdot)-c)]+U(t,\omega,c)$$

is $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$-measurable: As we have argued in Lemma 3.7.7 $\omega \mapsto \phi(\omega, x, H, c)$ is lower semianalytic and in particular universally measurable. On the other hand, $x \mapsto \tilde{U}_{t+1}(\omega, x)$ is upper semicontinuous and concave for any $\omega \in \Omega^t$. Since $\tilde{U}_{t+1}$ is bounded from above, an application of Fatou's lemma yields that $(x, H, c) \mapsto \phi(\omega, x, H, c)$ is upper semicontinuous and concave for each $\omega \in \Omega^t$. Again by [Blanchard and Carassus, 2017, Lemma A.35, page 1889] it follows that ϕ is $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$-measurable. Now we define the correspondence

$$\begin{aligned}\Phi(\omega) := \{(H', c') \in \mathbb{R}^d \times \mathbb{R}_+ \mid \phi(\omega, V_{t-1}^{\pi_0, H, C}(\omega) + H_t(\omega)\Delta S_t(\omega) - \mathbb{1}_{\{t=T-1\}} g(\omega, \cdot), H', c') \\ = \tilde{U}_t(\omega, V_{t-1}^{\pi_0, H, C}(\omega) + H_t(\omega)\Delta S_t(\omega))\}, \quad \omega \in \Omega^t.\end{aligned}$$

Then its graph is in $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$. Next we define the function

$$\Upsilon : \omega \mapsto \mathcal{A}_{t, V_{t-1}^{\pi_0, H, C}(\omega) + H_t(\omega)\Delta S_t(\omega)}(\omega).$$

By a slight variation of the arguments given in Bouchard and Nutz [2015][proof of Lemma 4.10, pp.846-848] the graph of Υ is $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$-measurable and thus $\text{graph}(\Upsilon) \cap ((\Omega_{NA}^t \cap \Omega_g^t) \times \mathbb{R}^d \times \mathbb{R}) \in \mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$. Then also the graph of

$$\tilde{\Phi}(\omega) = \begin{cases} \mathcal{A}_{t, V_{t-1}^{\pi_0, H, C}(\omega) + H_t(\omega)\Delta S_t(\omega)}(\omega) \cap \Phi(\omega) & \omega \in \Omega_{NA}^t \cap \Omega_g^t \\ \emptyset & \text{otherwise} \end{cases}$$

is in $\mathcal{F}_t^{\mathcal{U}} \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R})$ and $\tilde{\Phi}$ admits an $\mathcal{F}_t^{\mathcal{U}}$-measurable selector (H_{t+1}^*, c_t^*) on the universally measurable set $\{\tilde{\Phi} \neq \emptyset\} \in \mathcal{F}_t^{\mathcal{U}}$ by the Neumann-Aumann theorem ([Sainte-Beuve, 1974, Cor.1, p.120]). We extend (H_{t+1}^*, c_t^*) by setting $H_{t+1}^* = c_t^* = 0$ on $\{\tilde{\Phi} \neq \emptyset\}$. Moreover, as $(H, C) \in \mathcal{A}_{\pi_0}$, (3.6.3) implies that $V_t^{\pi_0, H, C} \geq \pi_t(g)$ $\mathcal{P}$-q.s. and the one-period case given in Proposition 3.7.1 applied with $x = V_{t-1}^{\pi_0, H, C}(\omega) + H_t(\omega)\Delta S_t(\omega)$, Lemma 3.6.1 as well as existence of superhedging strategies as stated in [Bouchard and Nutz, 2015, Theorem 3.4] show that $\tilde{\Phi}(\omega) \neq \emptyset$ for $\mathcal{P}^u$-q.e. $\omega \in \Omega^t$. This shows the claim as $U_t = \tilde{U}_t$ $\mathcal{P}^u$-q.s. □

Proof of Theorem 3.4.3. Let $(\hat{H}_1, 0)$ be an optimal strategy for

$$\inf_{\mathbb{P} \in \mathcal{P}_0^u} \mathbb{E}_{\mathbb{P}}(U_1[\pi_0 + H_1 \Delta S_1)]$$

as in Lemma 3.7.8. Proceeding recursively, we use Lemma 3.7.8 to define the strategy $\omega \mapsto (H_{t+1}^*, c_t^*)(\omega)$ for

$$\begin{aligned}\inf_{\mathbb{P} \in \mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}}[U_{t+1}((\omega, \cdot), V_{t-1}^{\pi_0, H^*, C^*}(\omega) + H_t^*(\omega)\Delta S_t(\omega) + H_{t+1}(\omega)\Delta S_{t+1}(\omega, \cdot) - c_t(\omega) \\ - \mathbb{1}_{\{t=T-1\}} g(\omega, \cdot))] + U(t, c_t(\omega))\end{aligned}$$

where $1 \le t \le T-1$ and define $C_t^* = \sum_{s=1}^t c_s^*$ as well as $\Delta C_T^* = V_{T-1}^{\pi_0,H^*,C^*} + H_T^* \Delta S_T - g$. By construction we then have $(H^*, C^*) \in \mathcal{A}_{\pi_0}$. To establish that (H^*, C^*) is optimal we first show that

$$\inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{s=1}^T U(s, \Delta C_s^*) \right] \ge U_0(\pi_0). \tag{3.7.2}$$

Let $0 \le t \le T-1$. By definition of (H^*, C^*) we have

$$\begin{aligned} &\inf_{\mathbb{P}' \in \mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}'}[U_{t+1}((\omega, \cdot), V_t^{\pi_0,H^*,C^*}(\omega) + H_{t+1}^*(\omega)\Delta S_{t+1}(\omega, \cdot) - \mathbb{1}_{\{t=T-1\}} g(\omega, \cdot))] \\ &+ U(t, \omega, \Delta C_t^*(\omega)) = U_t(\omega, V_{t-1}^{\pi_0,H^*,C^*}(\omega) + H_t^*(\omega)\Delta S_t(\omega)) \end{aligned}$$

for all $\omega \in \Omega^t$ outside a $\mathcal{P}^u$-polar set. Let $\mathbb{P} \in \mathfrak{P}$, then $\mathbb{P} = \mathbb{P}_0 \otimes \cdots \otimes \mathbb{P}_{T-1}$ for some selectors $\mathbb{P}_t$ of $\mathcal{P}_t^u$, $0 \le t \le T-1$ and we conclude via Fubini's theorem that

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}} \left[U_{t+1} \left(V_t^{\pi_0,H^*,C^*} + H_{t+1}^* \Delta S_{t+1} - \mathbb{1}_{\{t=T-1\}} g \right) + \sum_{s=1}^t U(s, \Delta C_s^*) \right] \\ &= \mathbb{E}_{(\mathbb{P}_0 \otimes \cdots \mathbb{P}_{t-1})(d\omega)} \Bigg(\mathbb{E}_{\mathbb{P}_t(\omega)} \Big[U_{t+1} \Big((\omega, \cdot), V_t^{\pi_0,H^*,C^*}(\omega) + H_{t+1}^*(\omega) \Delta S_{t+1}(\omega, \cdot) \\ &\qquad - \mathbb{1}_{\{t=T-1\}} g(\omega, \cdot) \Big) \Big] + \sum_{s=1}^t U\left(s, \omega, \Delta C_s^*(\omega)\right) \Bigg) \\ &\ge \mathbb{E}_{\mathbb{P}_0 \otimes \cdots \otimes \mathbb{P}_{t-1}} \left[U_t \left(V_{t-1}^{\pi_0,H^*,C^*} + H_t^* \Delta S_t \right) + \sum_{s=1}^{t-1} U\left(s, \Delta C_s^*\right) \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[U_t \left(V_{t-1}^{\pi_0,H^*,C^*} + H_t^* \Delta S_t \right) + \sum_{s=1}^{t-1} U\left(s, \Delta C_s^*\right) \right]. \end{aligned}$$

A repeated application of this inequality shows (3.7.2). To conclude that (H^*, C^*) is optimal, it remains to prove that

$$U_0(\pi_0) \ge \sup_{(H,C) \in \mathcal{A}_{\pi_0}} \inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[\sum_{s=1}^T U(s, \Delta C_s) \right] =: v(\pi_0).$$

To this end we fix an arbitrary $(H, C) \in \mathcal{A}_{\pi_0}$ and first show that

$$\begin{aligned} &\inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[U_t \left(V_{t-1}^{\pi_0,H,C} + H_t \Delta S_t \right) + \sum_{s=1}^{t-1} U(s, \Delta C_s) \right] \\ &\ge \inf_{\mathbb{P} \in \mathcal{P}^u} \mathbb{E}_{\mathbb{P}} \left[U_{t+1} \left(V_t^{\pi_0,H,C} + H_{t+1} \Delta S_{t+1} - \mathbb{1}_{\{t=T-1\}} g \right) + \sum_{s=1}^t U(s, \Delta C_s) \right] \end{aligned} \tag{3.7.3}$$

for $1 \le t \le T-1$. Let $\varepsilon > 0$. As in the proof of Lemma 3.7.7

$$\begin{aligned} (\omega, \mathbb{P}) \mapsto \mathbb{E}_{\mathbb{P}} &\left[U_{t+1}((\omega, \cdot), V_t^{\pi_0,H,C}(\omega) + H_{t+1} \Delta S_{t+1} - \mathbb{1}_{\{t=T-1\}} g(\omega, \cdot)) \right] \\ &+ \sum_{s=1}^t U(s, \omega, \Delta C_s(\omega)), \end{aligned}$$

is lower semianalytic. Using [Bertsekas and Shreve, 1978, Prop. 7.50, p. 184 & Prop. 7.44, p.172] for $\omega \in \Omega^t$ outside a $\mathcal{P}^u$-polar set we have for some universally measurable ε-optimal selector $\mathbb{P}_t^\varepsilon$ that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_t^\varepsilon(\omega)}\Big[U_{t+1}\Big((\omega,\cdot),V_t^{\pi_0,H,C}(\omega)+H_{t+1}(\omega)\Delta S_{t+1}(\omega,\cdot)-\mathbb{1}_{\{t=T-1\}}g(\omega,\cdot)\Big)\Big]\\
&\qquad\quad+\sum_{s=1}^{t}U(s,\omega,\Delta C_s(\omega))-\varepsilon\\
&\le\ (-\varepsilon)^{-1}\vee\bigg(\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)}\mathbb{E}_{\mathbb{P}}\bigg[U_{t+1}\bigg((\omega,\cdot),V_t^{\pi_0,H,C}(\omega)+H_{t+1}(\omega)\Delta S_{t+1}(\omega,\cdot)\\
&\qquad\qquad\qquad\qquad-\mathbb{1}_{\{t=T-1\}}g(\omega,\cdot)\bigg)\bigg]+\sum_{s=1}^{t}U(s,\omega,\Delta C_s(\omega))\bigg)\\
&\le\ (-\varepsilon)^{-1}\vee\bigg(\sup_{(H',c')\in\mathcal{A}_{t,V_{t-1}^{\pi_0,H,C}(\omega)+H_t(\omega)\Delta S_t(\omega)}(\omega)}\ \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)}\mathbb{E}_{\mathbb{P}}\Big[U_{t+1}\big((\omega,\cdot),V_{t-1}^{\pi_0,H,C}(\omega)\\
&+H_t(\omega)\Delta S_t(\omega)-c'+H'\Delta S_{t+1}(\omega,\cdot)-\mathbb{1}_{\{t=T-1\}}g(\omega,\cdot)\big)\Big]\\
&+\sum_{s=1}^{t-1}U(s,\omega,\Delta C_s(\omega))+U(t,\omega,c')\bigg)\\
&=\ (-\varepsilon)^{-1}\vee\bigg(U_t(\omega,V_{t-1}^{\pi_0,H,C}(\omega)+H_t(\omega)\Delta S_t(\omega))+\sum_{s=1}^{t-1}U(s,\omega,\Delta C_s(\omega))\bigg).
\end{aligned}$$

Given $\mathbb{P}\in\mathcal{P}^u$ we thus have

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\bigg[(-\varepsilon)^{-1}\vee\bigg(U_t(V_{t-1}^{\pi_0,H,C}+H_t\Delta S_t)+\sum_{s=1}^{t-1}U(s,\Delta C_s)\bigg)\bigg]\\
&\ge\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_t^\varepsilon}\bigg[U_{t+1}(V_t^{\pi_0,H,C}+H_{t+1}\Delta S_{t+1}-\mathbb{1}_{\{t=T-1\}}g)+\sum_{s=1}^{t}U(s,\Delta C_s)\bigg]-\varepsilon\\
&\ge\inf_{\mathbb{P}'\in\mathcal{P}^u}\mathbb{E}_{\mathbb{P}'}\bigg[U_{t+1}(V_t^{\pi_0,H,C}+H_{t+1}\Delta S_{t+1}-\mathbb{1}_{\{t=T-1\}}g)+\sum_{s=1}^{t}U(s,\Delta C_s)\bigg]-\varepsilon.
\end{aligned}$$

As $\varepsilon>0$ and $\mathbb{P}\in\mathcal{P}^u$ were arbitrary (3.7.3) follows. Noting that

$$U_0(\pi_0)=\inf_{\mathbb{P}\in\mathcal{P}^u}\mathbb{E}_{\mathbb{P}}[U_0(V_0^{\pi_0,H,C})],$$

a repeated application of (3.7.3) yields

$$\begin{aligned}
U_0(\pi_0)\ge\inf_{\mathbb{P}\in\mathcal{P}^u}\mathbb{E}_{\mathbb{P}}[U_1(\pi_0+H_1\Delta S_1)]\ge\cdots\ge\inf_{\mathbb{P}\in\mathcal{P}^u}\mathbb{E}_{\mathbb{P}}\bigg[U_T(V_{T-1}^{\pi_0,H,C}+H_T\Delta S_T-g)\\
+\sum_{s=1}^{T-1}U(s,\Delta C_s)\bigg]=\inf_{\mathbb{P}\in\mathcal{P}^u}\mathbb{E}_{\mathbb{P}}\bigg[\sum_{s=1}^{T}U(s,\Delta C_s)\bigg].
\end{aligned}$$

As $(H,C)\in\mathcal{A}_{\pi_0}$ was arbitrary, it follows that $U_0(\pi_0)\ge v(\pi_0)$. This concludes the proof, since $\pi_0=\pi(g)$. □

3.7.3 Proof of Theorem 3.4.5

Proof. Existence of an optimal investment consumption strategy follows by Theorem 3.4.3. We now show uniqueness of optimisers. We fix $0 \leq t \leq T-1$ and recall the definition of $\tilde{U}_t$ given in Lemma 3.7.7. Note that one can show that the function

$$(\omega, \mathbb{P}) \mapsto \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot), x + H\Delta S_{t+1}(\omega,\cdot) - c - \mathbb{1}_{\{t=T-1\}}g(\omega,\cdot))] + U(t,c)$$

is lower semianalytic by reducing the above expression to a supremum over a countable set as in the proof of Lemma 3.7.7. Recall that again by Lemma 3.7.7 there exists a set of full $\mathcal{P}^u$ measure on which $\tilde{U}_t = U_t$ for all $0 \leq t \leq T$. For the rest of the proof we take ω in the intersection of this set with Ω^t_{NA}. Using the same Jankov-von-Neumann argument as in the proof of Theorem 3.4.3 and Corollary 3.7.3 we conclude that for each $0 \leq t \leq T-1$ there exists a sequence $\mathbb{P}^n_t : \Omega^t \to \mathfrak{P}(\Omega)$ of universally measurable kernels such that $\mathbb{P}^n_t(\omega) \in \mathcal{P}^u_t(\omega)$ and for $x \geq \pi_t(g)(\omega)$

$$\lim_{n\to\infty} \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \mathbb{E}_{\mathbb{P}^n_t(\omega)}[\tilde{U}_{t+1}((\omega,\cdot), x + H\Delta S_{t+1}(\omega,\cdot) - c - \mathbb{1}_{\{t=T-1\}}g(\omega,\cdot))] + U(t,c)$$

equals $\tilde{U}_t(\omega,x)$. Since $\mathcal{P}^u_t(\omega)$ is compact, there exists a probability measure $\mathbb{P}^*_t(\omega) \in \mathcal{P}^u_t(\omega)$ and a subsequence $\{n_k(\omega)\}_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} \mathbb{P}^{n_k(\omega)}_t(\omega) = \mathbb{P}^*_t(\omega)$. We now show, that for $\mathcal{P}^u$-q.e. $\omega \in \Omega^t$ and $x \geq \pi_t(\omega)$ the functions

$$\begin{aligned} U_t(\omega,x) = &\sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \inf_{\mathbb{P}\in\mathcal{P}^u_t(\omega)} \mathbb{E}_{\mathbb{P}}[U_{t+1}((\omega,\cdot), x + H\Delta S_{t+1}(\omega,\cdot) - c - \mathbb{1}_{\{t=T-1\}}g(\omega,\cdot))] \\ &+ U(t,\omega,c) \end{aligned}$$

have a unique optimizer $(H,c) \in \mathcal{A}_{t,x}(\omega)$. For notational convenience we assume that $0 \leq t \leq T-2$. We note that by concavity of $\tilde{U}_{t+1}$ and $U(t,\cdot)$ the function

$$(H,c) \mapsto \inf_{\mathbb{P}\in\mathcal{P}^u_t(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), y + H\Delta S_{t+1}(\omega,\cdot) - c\right)\right) + U(t,c)$$

is concave. Now assume that there are (H^1,c^1), $(H^2,c^2) \in \mathcal{A}_{t,x}(\omega)$ such that

$$\begin{aligned} &\inf_{\mathbb{P}\in\mathcal{P}^u_t(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), x + H^1\Delta S_t(\omega,\cdot) - c^1\right)\right) + U(t,c^1) \\ = &\inf_{\mathbb{P}\in\mathcal{P}^u_t(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), x + H^2\Delta S_t(\omega,\cdot) - c^2\right)\right) + U(t,c^2) \\ = &\tilde{U}_t(\omega,x). \end{aligned}$$

Note that for the strategy $(H^3, c^3) := ((H^1 + H^2)/2, (c^1 + c^2)/2) \in \mathcal{A}_{t,x}(\omega)$ we have by concavity

$$\begin{aligned}
&\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), x + H^3\Delta S_t(\omega,\cdot) - c^3\right)\right) + U(t,c^3) \\
&\geq \frac{1}{2}\Bigg(\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}((\omega,\cdot),\left(x + H^1\Delta S_t(\omega,\cdot) - c^1\right)\right) + U(t,c^1) \\
&+ \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), y + H^2\Delta S_{t+1}(\omega,\cdot) - c^2\right)\right) + U(t,c^2)\Bigg) = \tilde{U}_t(\omega,x).
\end{aligned}$$

We thus conclude

$$\inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} E_{\mathbb{P}}\left(\tilde{U}_{t+1}\left((\omega,\cdot), x + H^3\Delta S_t(\omega,\cdot) - c^3\right)\right) + U(t,c^3) = \tilde{U}_t(\omega,x).$$

Furthermore, for any $x \geq \pi_t(\omega)$ and any maximizer $(\tilde{H},\tilde{c}) \in \mathcal{A}_{t,x}(\omega)$ of $\tilde{U}_t(\omega,x)$ we have

$$\begin{aligned}
&\sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \left(\mathbb{E}_{\mathbb{P}_t^{n_k(\omega)}(\omega)}[\tilde{U}_{t+1}((\omega,\cdot), x + H\Delta S_{t+1}(\omega,\cdot) - c)] + U(t,c)\right) \\
&\geq \mathbb{E}_{\mathbb{P}_t^{n_k(\omega)}(\omega)}[\tilde{U}_{t+1}((\omega,\cdot), x + \tilde{H}\Delta S_{t+1}(\omega,\cdot) - \tilde{c})] + U(t,\tilde{c}) \qquad (3.7.4)\\
&\geq \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot), x + \tilde{H}\Delta S_{t+1}(\omega,\cdot) - \tilde{c})] + U(t,\tilde{c}) = \tilde{U}_t(\omega,x),
\end{aligned}$$

so taking limits in (3.7.4) we find

$$\lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_t^{n_k(\omega)}(\omega)}[\tilde{U}_{t+1}((\omega,\cdot), x + \tilde{H}\Delta S_{t+1}(\omega,\cdot) - \tilde{c})] + U(t,\tilde{c}) = \tilde{U}_t(\omega,x).$$

Furthermore we note that by assumption and Lemma 3.7.7 $\tilde{U}_t(\omega,y)$ is bounded by some C on $\{(\omega,x) \in \Omega^t \times \mathbb{R} \mid x \geq \pi_t(g)(\omega)\}$, non-decreasing as well as continuous in y and g is continuous. Furthermore the superhedging prices $\omega \mapsto \pi_t(g)(\omega)$ are continuous by assumption.

We now show by backwards induction that the function $(\omega,x) \mapsto \tilde{U}_t(\omega,x)$ is continuous in every point of the set $\{(\omega,x) \in \Omega^t \times \mathbb{R} \mid x \geq \pi_t(g)(\omega)\}$: let us assume the hypothesis is true for $t+1$ and fix $\omega \in \Omega^t$, $x \geq \pi_t(g)(\omega)$. We note first that by Remark 3.5.1 and the same contradiction argument as in the proof of Proposition 3.3.7 choosing $|\tilde{\omega} - \omega| \leq \tilde{\delta}$ for $\tilde{\delta} > 0$ small enough we can assume that for any superhedging strategy $(H,c) \in \mathcal{A}_{t,\pi_t(g)(\tilde{\omega})}(\tilde{\omega})$ we have $|(H,c)| \leq \tilde{C}$ for some $\tilde{C} > 1/2$ independently of $\tilde{\omega}$.

Next we make the following observation: As $\mathcal{P}_t^u(\omega)$ is weakly compact by assumption, there exists a compact set $[0,K]^d \subseteq \Omega$, such that $\mathbb{P}(([0,K]^d)^c) \leq \varepsilon/(48\tilde{C})$ for all $\mathbb{P} \in \mathcal{P}_t^u(\omega)$. By the induction hypothesis $(v,y) \mapsto \tilde{U}_{t+1}(v,y)$ is continuous in every point of the set $\{(v,y) \in \Omega^{t+1} \times \mathbb{R} \mid y \geq \pi_{t+1}(g)(v)\}$ and thus uniformly continuous on a compact subset. There exists $\delta_0/2 > 0$ such that for $v,\tilde{v} \in B_1(\omega) \times \{u \in \Omega \mid \inf_{\tilde{u}\in[0,K]^d} |u - \tilde{u}| \leq \delta_0\}$, $y \in [\pi_{t+1}(g)(v), 2\tilde{C}K]$ and $|(v,y) - (\tilde{v},\tilde{y})| \leq \delta_0$ we have

$$\left|\tilde{U}_{t+1}(v,y) - \tilde{U}_{t+1}(\tilde{v},\tilde{y})\right| \leq \varepsilon/24. \qquad (3.7.5)$$

By Assumption 3.4.4.(1) we can adjust $\tilde{\delta} > 0$, such that for all $\tilde{\omega} \in \Omega^t$ with $|\omega - \tilde{\omega}| < \tilde{\delta}$ and for all $\mathbb{P} \in \mathcal{P}_t^u(\omega)$, there exists $\tilde{\mathbb{P}} \in \mathcal{P}_t^u(\tilde{\omega})$ such that $d_L(\mathbb{P}, \tilde{\mathbb{P}}) \leq \tilde{\varepsilon} := \delta_0/(2\tilde{C}) \wedge \varepsilon/(48\tilde{C})$. It follows by Strassen's theorem that there exists a measure $\pi \in \mathfrak{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and two random variables $O \sim \mathbb{P} \circ (S_{t+1})^{-1}(\omega, \cdot)$ and $\tilde{O} \sim \tilde{\mathbb{P}} \circ (S_{t+1})^{-1}(\tilde{\omega}, \cdot)$ such that $\pi(|O - \tilde{O}| \geq \tilde{\varepsilon}) \leq \tilde{\varepsilon}$. Thus we conclude that for $y, \tilde{y} : \Omega \to \mathbb{R}$ with $|y(o) - \tilde{y}(\tilde{o})| \leq \delta_0$ whenever $\pi_{t+1}(\tilde{\omega}, o) \leq y(o) \leq 2\tilde{C}K$ and $|o - \tilde{o}| \leq \tilde{\varepsilon}$

$$
\begin{aligned}
&\left|\mathbb{E}_{\mathbb{P}}\left[\tilde{U}_{t+1}(\omega, \cdot), y(\cdot))\right] - \mathbb{E}_{\tilde{\mathbb{P}}}\left[\tilde{U}_{t+1}((\tilde{\omega}, \cdot), \tilde{y}(\cdot))\right]\right| \qquad (3.7.6)\\
= &\left|\mathbb{E}_{\pi}\left[\tilde{U}_{t+1}(\omega, O), y(O)) - \tilde{U}_{t+1}((\tilde{\omega}, \tilde{O}), \tilde{y}(\tilde{O}))\right]\right| \\
\leq &\ \mathbb{E}_{\pi}\left[\left|\tilde{U}_{t+1}(\omega, O), y(O)) - \tilde{U}_{t+1}((\tilde{\omega}, \tilde{O}), \tilde{y}(\tilde{O}))\right| \mathbb{1}_{\{O \in [0,K]^d,\ |O - \tilde{O}| \leq \tilde{\varepsilon}\}}\right] \\
+ &\ \frac{\tilde{C}\varepsilon}{12\tilde{C}} \leq \varepsilon/12 + \varepsilon/12 = \varepsilon/6.
\end{aligned}
$$

Now we modify $\tilde{\delta} > 0$ such that $|\pi_t(g)(\omega) - \pi_t(g)(\tilde{\omega})| \leq \delta_0/2$ if $|\omega - \tilde{\omega}| \leq \tilde{\delta}$. Furthermore applying Proposition 3.7.1 for the function $(\omega, x) \mapsto \tilde{U}_{t+1}(\omega, x)$ there exists a maximiser $(H', c') \in \mathcal{A}_{t,\tilde{x}}(\tilde{\omega})$ of

$$
\sup_{(H,c) \in \mathcal{A}_{t,\tilde{x}}(\tilde{\omega})} \inf_{\mathbb{P} \in \mathcal{P}_t^u(\tilde{\omega})} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\tilde{\omega}, \cdot), \tilde{x} + H\Delta S_{t+1}(\tilde{\omega}, \cdot) - c)] + U(t, c).
$$

Next, if x and $\tilde{x} \geq \pi_t(g)(\tilde{\omega})$ are close there exists a strategy $(H, c) \in \mathcal{A}_{t,x}(\omega)$ for some c with $|c - c'| \leq \delta_0/2$. Furthermore there exists $\mathbb{P} \in \mathcal{P}_t^u(\omega)$ such that

$$
\tilde{U}_t(\omega, x) \geq \mathbb{E}_{\mathbb{P}}\left[\tilde{U}_{t+1}((\omega, \cdot), x + H\Delta S_{t+1}(\omega, \cdot) - c)\right] + U(t, c) - \varepsilon/6.
$$

Note that we can modify $\tilde{\delta} > 0$ such that $|(\omega, HS_t(\omega)) - (\tilde{\omega}, HS_t(\tilde{\omega}))| \leq (\tilde{C} + 2)\tilde{\delta} \leq \delta_0/2$. Now by (3.7.6) with $y(\cdot) = x + H\Delta S_{t+1}(\omega, \cdot) - c$ and $\tilde{y}(\cdot) = \tilde{x} + H\Delta S_{t+1}(\tilde{\omega}, \cdot) - c'$ together with boundedness of $\tilde{U}_{t+1}$ and the fact that $\mathbb{P}(([0, K]^d)^c) \leq \varepsilon/(48\tilde{C})$ and noting that $|H - H'| \leq 2\tilde{C}$

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega, \cdot), x + H\Delta S_{t+1}(\omega, \cdot) - c)] + U(t, c) - \varepsilon/6 \\
\geq &\ \mathbb{E}_{\tilde{\mathbb{P}}}\left[\tilde{U}_{t+1}[(\tilde{\omega}, \cdot), \tilde{x} + H'\Delta S_{t+1}(\tilde{\omega}, \cdot) - c')\right] + U(t, c') - \varepsilon/2 \\
\geq &\ \tilde{U}_t(\tilde{\omega}, \tilde{x}) - \varepsilon/2.
\end{aligned}
$$

Exchanging the roles of ω and $\tilde{\omega}$ concludes the proof of the induction step.
This shows in particular continuity of $\omega' \mapsto \tilde{U}_{t+1}((\omega, \omega'), x + \tilde{H}\Delta S_{t+1}(\omega, \omega') - \tilde{c})$ as $\omega' \mapsto x + \tilde{H}\Delta S_{t+1}(\omega, \omega') - \tilde{c}$ is continuous. As this function is also $\mathcal{P}_t^u(\omega)$-q.s. bounded by Lemma 3.7.7 (recall that $(\tilde{H}, \tilde{c}) \in \mathcal{A}_{t,x}(\omega)$), we conclude by use of the Portmanteau theorem (here

$\tilde{U}_{t+1}$ can be extended continuously by Tietze's theorem) that

$$\begin{aligned}\tilde{U}_t(\omega,x) &= \liminf_{k\to\infty} \mathbb{E}_{\mathbb{P}_t^{n_k(\omega)}(\omega)}[\tilde{U}_{t+1}((\omega,\cdot),x+\tilde{H}\Delta S_{t+1}(\omega,\cdot)-\tilde{c})] + U(t,\tilde{c}) \\ &\geq \mathbb{E}_{\mathbb{P}_t^*(\omega)}[\tilde{U}_{t+1}((\omega,\cdot),x+\tilde{H}\Delta S_{t+1}(\omega,\cdot)-\tilde{c})] + U(t,\tilde{c}) \\ &\geq \inf_{\mathbb{P}\in\mathcal{P}_t^u(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot),x+\tilde{H}\Delta S_{t+1}(\omega,\cdot)-\tilde{c})] + U(t,\tilde{c}),\end{aligned}$$

which yields for $x \geq \pi_t(\omega)$

$$\tilde{U}_t(\omega,x) = \mathbb{E}_{\mathbb{P}_t^*(\omega)}[\tilde{U}_{t+1}((\omega,\cdot),x+\tilde{H}\Delta S_{t+1}(\omega,\cdot)-\tilde{c})] + U(t,\tilde{c}).$$

In particular for $i=1,2$

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}_t^*\omega)}\left[\tilde{U}_{t+1}\left((\omega,\cdot),x+H^3\Delta S_{t+1}(\omega,\cdot)-c^3\right)\right] + U(t,c^3) \\ =\ &\mathbb{E}_{\mathbb{P}_t^*(\omega)}\left[\tilde{U}_{t+1}\left((\omega,\cdot),x+H^i\Delta S_{t+1}(\omega,\cdot)-c^i\right)\right] + U(t,c^i).\end{aligned}$$

Now since

$$(H,c) \mapsto \mathbb{E}_{\mathbb{P}_t^*(\omega)}[\tilde{U}_{t+1}((\omega,\cdot),x+H\Delta S_{t+1}(\omega,\cdot)-c)] + U(t,c)$$

is concave and strictly concave in c, we need to have $c^1 = c^2$ and

$$H^1\Delta S_{t+1}(\omega,\cdot) = H^2\Delta S_{t+1}(\omega,\cdot) \quad \mathbb{P}_t^*(\omega)-\text{a.s}$$

Lastly denote by Ξ_t the correspondence

$$\Xi_t(\omega) = \left\{\mathbb{P}\in\mathcal{P}_t^u(\omega) \;\middle|\; \tilde{U}_t(x,\omega) = \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot),x+H\Delta S_{t+1}(\omega,\cdot)-c)] + U(t,c)\right\}$$

for $x \geq \pi_t(\omega)$ and note that by measurable selection arguments as in Bouchard and Nutz [2015][proof of Lemma 4.10, p. 848] the set

$$\left\{(\omega,\mathbb{P})\in\operatorname{graph}(\mathcal{P}_t^u) \;\middle|\; \sup_{(H,c)\in\mathcal{A}_{t,x}(\omega)} \mathbb{E}_{\mathbb{P}}[\tilde{U}_{t+1}((\omega,\cdot),x+H\Delta S_{t+1}(\omega,\cdot)-c)] + U(t,c) - \tilde{U}_t(x,\omega) \leq 0\right\}$$

is an element of $\mathbf{A}(\mathcal{F}_t^{\mathcal{U}}\otimes\mathcal{B}(\mathfrak{P}(\Omega)))$, where $\mathbf{A}(\mathcal{F}_t^{\mathcal{U}}\otimes\mathcal{B}(\mathfrak{P}(\Omega)))$ is the set of all nuclei of Suslin schemes on $\mathcal{F}_t^{\mathcal{U}}\otimes\mathcal{B}(\mathfrak{P}(\Omega))$. In consequence there exists an $\mathcal{F}_t^{\mathcal{U}}$-measurable function $\mathbb{P}_t^*:\Omega^t\to\mathfrak{P}(\Omega)$ such that $\operatorname{graph}(\mathbb{P}_t^*)\subseteq\operatorname{graph}(\Xi_t)$. This concludes the proof. □

Remark 3.7.9. If we assume that $H^1-H^2\in\operatorname{span}_{\mathbb{P}_t^*(\omega)}(\Delta S_{t+1}(\omega,\cdot))$, then $H^1=H^2$.

Chapter 4

Robust estimation of superhedging prices

4.1 Introduction

Computation of risk associated to a given financial position is one of the fundamental operations market participants have to perform. For institutional players, like banks, it is regulated by Basel Committee on Banking Supervision [2013] which dictates rules and requirements for such risk assessments. A golden standard has long been given by Value-at-Risk (VaR), however more recently this is being replaced by convex risk measures like Average VaR (Expected Shortfall) or more sophisticated approaches which include market modelling. Consequently, there is an abundant literature on VaR estimation and some more recent works related to statistical estimation of law-invariant risk measures, see Cont et al. [2010], Krätschmer et al. [2012, 2014], Kou et al. [2013], Pichler [2013], Embrechts et al. [2015, 2018]. All these works consider a static situation with no trading involved.

In contrast, in this chapter we consider estimation of risk for an agent who can trade in the market to offset her risk exposure. To put in evidence the novelty and relevance of our setting, we concentrate on one, simple but canonical, way to assess risk: the *superhedging price*. Consider a one-period frictionless market with prices (S_t, S_{t+1}) denominated in units of a fixed numeraire. The current stock prices S_t are known and the future prices S_{t+1} are modelled as random variables, say with return $r := S_{t+1}/S_t$ drawn from a distribution $\mathbb{P}$ on $\mathbb{R}^d_+$. For a payoff $g : \mathbb{R}^d_+ \to \mathbb{R}$, its superhedging price is given by:

$$\pi^{\mathbb{P}}(g) := \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H(r-1) \geq g(r) \ \mathbb{P}\text{-a.s.}\}. \tag{4.1.1}$$

In this simple setting, an arbitrage strategy is $H \in \mathbb{R}^d$ such that $\mathbb{P}(H(r-1) \geq 0) = 1$ and $\mathbb{P}(H(r-1) > 0) > 0$ and if no such strategy exists we say that no-arbitrage $\mathrm{NA}(\mathbb{P})$ holds. By the Fundamental Theorem of Asset Pricing, absence of arbitrage is equivalent to existence of a probability measure $\mathbb{Q}$, equivalent to $\mathbb{P}$, under which S is a martingale, i.e., $\mathbb{E}_{\mathbb{Q}}[r] = 1$. There might be more than one such measure and they can all be used for pricing. Taking the supremum over $\mathbb{E}_{\mathbb{Q}}[g]$ enables to compute the maximal feasible price for g and this, by the fundamental pricing-hedging duality, is the same as the superhedging price of g:

$$\pi^{\mathbb{P}}(g) = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{E}_{\mathbb{Q}}[r]=1} \mathbb{E}_{\mathbb{Q}}[g], \tag{4.1.2}$$

for all Borel g, cf. [Föllmer and Schied, 2004, Thm. 1.31]. Despite its theoretical importance and practical relevance, to the best of our knowledge, there has been no attempt to study statistical estimation of the superhedging price. The results in this chapter fill this important gap. Instead of postulating a measure $\mathbb{P}$, we build estimators of $\pi^{\mathbb{P}}(g)$ directly from historical observations of returns $r_1, \ldots, r_N$ and study their properties. Furthermore, we extend the estimators to take into account also the option price data. This is practically relevant and methodologically novel in that it allows a coherent and simultaneous use of historical time-series data with current option price data or, in mathematical finance jargon, the *physical measure* data and the *risk neutral measure* data.

In contrast, in existing approaches historical returns are, if at all, only used indirectly to compute $\pi^{\mathbb{P}}(g)$. In classical mathematical finance, one first postulates a family of plausible models $\{\mathbb{P}_\theta : \theta \in \Theta\}$. Such choice may be influenced by stylised features of historical returns, see Gatheral et al. [2018] for a recent example, as well as by other considerations, e.g., of computational tractability. Thereon, historical returns are not used and only the "future facing" options price data is exploited to select a candidate pricing measure $\mathbb{Q}_\theta$. More recently, pioneered by Mykland [2000, 2003b,a] in a continuous-time setting and pursued within the so-called robust approach to pricing and hedging, it was suggested to use historical returns to select a *prediction set*, i.e., the set of paths on which the superhedging property is required, and then to compute the resulting cheapest superhedge which trades in stocks and options, see Hou and Obłój [2018], Burzoni et al. [2019]. Our approach inherits from that perspective but takes a statistical viewpoint and evolves it into a dynamic and asymptotically consistent methodology.

To describe our approach, suppose we observe d-dimensional historical returns $r_1 = S_1/S_0, \ldots, r_N = S_N/S_{N-1}$ and for simplicity assume that these are non-negative i.i.d. realisations of a distribution $\mathbb{P}$ which satisfies the no-arbitrage condition. We can equivalently represent the observations through their associated empirical measures

$$\hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{r_i},$$

which are well known to converge weakly to $\mathbb{P}$ as $N \to \infty$, see [Van der Vaart, 1998, Theorem 19.1, p. 266]. This suggests a very natural way to approximate the superhedging price by simply using $\hat{\mathbb{P}}_N$ in place of $\mathbb{P}$. We show in Theorem 4.2.1 below that the resulting *plugin estimator* $\hat{\pi}_N(g) := \pi^{\hat{\mathbb{P}}_N}(g)$ is asymptotically consistent:

$$\lim_{N\to\infty} \hat{\pi}_N(g) = \pi^{\mathbb{P}}(g), \qquad \mathbb{P}^\infty\text{-a.s.},$$

where $\mathbb{P}^\infty$ denotes the law of the process $(r_N)_{N\geq 1}$. However, we also show that $\hat{\pi}_N$ has serious shortcomings. First, it is not (statistically) robust: small perturbations of $\mathbb{P}$ can lead to large changes in the distribution of $\hat{\pi}_N$. We argue that the Lévy-Prokhorov metric used in the classical definition of statistical robustness, Definition 4.2.6, is not appropriate when looking at the financial context of derivatives pricing. We propose and study alternative metrics and ensuing notions of statistical robustness in Section 4.4.

Second, the plugin estimator also lacks robustness from the financial point of view of risk management. In fact, $\hat{\pi}_N$ is monotone in N and converges from below so it is always a lower estimate of the risk: $\hat{\pi}_N \leq \pi^{\mathbb{P}}$. In Theorem 4.2.11, and in more detail in Section 4.9, we study the convergence rates for the plugin estimators. This, in the one-dimensional case $d = 1$, could be exploited to build conservative estimates for the superhedging price $\pi^{\mathbb{P}}$.

A first intuition to improve the plugin estimator could be to turn to estimators of the support of $\mathbb{P}$. Indeed, the superhedging price $\pi^{\mathbb{P}}(g)$, say for a continuous g, only depends on $\mathbb{P}$ via its support. We could thus replace the $\hat{\mathbb{P}}_N$-a.s. inequality in the plugin estimator by an inequality on an estimator of the support of $\mathbb{P}$. Such estimators are well studied in statistics, going back to Geffroy [1964], Chevalier [1976], Devroye and Wise [1980], Grenander [1981], see also Cuevas [1990], Korostelëv et al. [1995], Mammen and Tsybakov [1995], Hardle et al. [1995], Polonik [1995], Tsybakov [1997], Cuevas and Rodríguez-Casal [2004], Casal [2007]. Unfortunately, this approach does not seem to hold any ground. First, convergence of support estimators usually imposes strong conditions on $\mathbb{P}$, e.g., compactness and convexity of the support and/or existence of a density. Second, for the convergence of the superhedging prices we would also need to impose some uniform continuity assumptions on g. However, under such conditions on $\mathbb{P}$ and g, we could directly improve the plugin estimator and consider a suitable $\hat{\pi}_N + a_N$, see Section 4.2.4.

Instead, to address the shortcomings of the plugin estimator, we propose novel estimators, which we introduce in Section 4.3. They exploit the dual formulation of the superhedging price in (4.1.2). In order to achieve financial robustness and to increase our point estimates we need to consider a larger class of martingale measures. Thus we consider

$$\pi_{\mathcal{Q}_N}(g) = \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}_{\mathbb{Q}}[g],$$

where $\mathcal{Q}_N$ is a subset of all martingale measures $\mathcal{M}$. The plugin estimator corresponds to taking $\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \hat{\mathbb{P}}_N\}$ and it is natural to replace it with

$$\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} : \exists \tilde{\mathbb{P}} \in B_N(\hat{\mathbb{P}}_N) \text{ s.t. } \mathbb{Q} \sim \tilde{\mathbb{P}}\},$$

where $B_N(\hat{\mathbb{P}}_N)$ is some "ball" in the space of probability measures around the empirical measure $\hat{\mathbb{P}}_N$. We show that this can lead to a consistent estimator if we use a sufficiently strong metric, e.g., the Wasserstein infinity metric $\mathcal{W}^\infty$. In general however such $\mathcal{Q}_N$ is too large. Instead, our main insight is to consider a tradeoff between the radius of the balls and the behaviour of martingale densities:

$$\hat{\mathcal{Q}}_N := \{\mathbb{Q} \in \mathcal{M} \mid \|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \leq k_N \text{ for some } \tilde{\mathbb{P}} \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)\},$$

where $B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)$ denotes the p-Wasserstein ball of radius ε_N around $\hat{\mathbb{P}}_N$ and $\varepsilon_N \to 0$ as well as $k_N \to \infty$. With a suitable choice of ε_N, k_N, we establish consistency of $\pi_{\hat{\mathcal{Q}}_N}(g)$ for a regular g, see Theorem 4.3.6, and also their financial robustness, see Corollary 4.3.7. This also allows us to study the cases when the estimator naturally extends to the setting of superhedging under model uncertainty about $\mathbb{P}$, see Corollary 4.3.8. The statistical robustness of $\pi_{\hat{\mathcal{Q}}_N}(g)$ is shown in Section 4.4, see Theorem 4.4.2. In Section 4.5 we extend our analysis to the case when risk is assessed not using the superhedging capital but rather via a generic risk measure ρ admitting a Kusuoka representation (Kusuoka [2001], see (4.5.1) for a definition). We stress that this is substantially different to all the works recalled at the beginning of this introduction since we consider an agent who can trade and optimises her position to offset the risk. We propose an estimator, inspired by $\pi_{\hat{\mathcal{Q}}_N}(g)$, and show its consistency.

Finally, we also propose another estimator:

$$\sup_{\mathbb{Q} \in \mathcal{M}} \left(\mathbb{E}_{\mathbb{Q}}[g] - C_N \left(\inf_{\hat{\mathbb{Q}} \sim \hat{\mathbb{P}}_N,\ \hat{\mathbb{Q}} \in \mathcal{M}} \left\| \frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}} \right\|_\infty - 1 \right) \right),$$

which is inspired by penalty methods used in risk measures and their representations as non-linear expectations. Asymptotic consistency of this estimator is shown in Thereom 4.3.12 and holds for an arbitrary measurable bounded g.

The rest of the chapter is organised as follows. In Section 4.2 we study the plugin estimator $\hat{\pi}_N$: its consistency, convergence rates and robustness, both statistical and financial. In Section 4.3 we propose improved estimators and establish consistency for all of them, under different sets of assumptions. Subsequently, in Section 4.4, we discuss statistical robustness of all the estimators. We show in particular that no estimator can be robust in the classical sense of Tukey-Huber-Hampel, and suggest ways to amend the classical definition to make it more appropriate to the superhedging price estimation. We then detail further applications of the main results. In particular we discuss convergence of superhedging strategies in Section 4.6 and partially extend the results to a multiperiod setting in Section 4.7. Section 4.8 discusses estimators $\pi_{\mathcal{Q}_N}$ for generic sets of martingale measures $\mathcal{Q}_N$ and derives necessary and sufficient conditions for asymptotic consistency of estimators. In particular, it motivates the estimators studied in Section 4.3. Finally, Section 4.9 studies in more detail convergence rates of the plugin estimator $\hat{\pi}_N$ when $d = 1$ and is auxiliary to Section 4.2.4. The last sections contain the remaining proofs, along with auxiliary results.

Notation. We write $\mathcal{P}(A)$ for the set of probability measures on $A \subset \mathbb{R}^d$. $\mathbb{P}_n \Rightarrow \mathbb{P}$ denotes weak convergence of measures. $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ is a generic distribution for returns r so that $\mathbb{E}_{\mathbb{P}}[r] = \int_{\mathbb{R}^d_+} x\mathbb{P}(dx)$. We let $\mathcal{M} = \{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^d_+) : \mathbb{E}_{\mathbb{Q}}[r] = 1\}$ denote the set of martingale measures for the stock prices (S_t, S_{t+1}), where $S_t > 0$ is fixed and $S_{t+1} = rS_t$. We write $\mathcal{M}_A$ for the set of martingale measures supported on A. We say that $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ does not admit arbitrage, or that NA($\mathbb{P}$) holds, if $\{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \mathbb{P}\} \neq \emptyset$. Above, and throughout, H is a row vector, r is a column vector and 1 denotes either a scalar or a column vector $(1, \dots, 1)^{\mathrm{T}}$.

4.2 The plugin estimator

Recall that we want to build an estimator for the superhedging price $\pi^{\mathbb{P}}(g)$. The easiest and possibly most natural way to do this is simply to replace the measure $\mathbb{P}$ with the empirical measures $\hat{\mathbb{P}}_N$. This yields the *plugin estimator*:

$$\hat{\pi}_N(g) := \pi^{\hat{\mathbb{P}}_N}(g). \tag{4.2.1}$$

In this section we develop the necessary tools to show asymptotic consistency of this estimator and understand its properties. The proofs are reported in Section 4.10.

4.2.1 Consistency

We now state the main result of this section:

Theorem 4.2.1. *Let $\mathbb{P}_1, \mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ and $g : \mathbb{R}^d_+ \to \mathbb{R}$ be Borel measurable. Assume that $r_1, r_2, \dots$ are realisations of a time-homogeneous ergodic Markov chain with initial distribution $\mathbb{P}_1$ and unique invariant distribution $\mathbb{P}$ such that $\mathbb{P}_1 \ll \mathbb{P}$. Then*

$$\lim_{N\to\infty} \hat{\pi}_N(g) = \pi^{\mathbb{P}}(g) \qquad \mathbb{P}^\infty\text{-}a.s., \tag{4.2.2}$$

where $\mathbb{P}^\infty$ denotes the law of the Markov process started from $\mathbb{P}_1$.

Remark 4.2.2. The assumptions in the above theorem are standard in econometric theory and cover a variety of models frequently used for modelling of financial returns data. We refer to Corollary 4.10.4 for sufficient conditions for stationarity (with exponential decay rates) for various random coefficient autoregressive models, e.g., linear and power GARCH and stochastic autoregressive volatility models, which are frequently used for option pricing. Nevertheless we remark that this assumption rules out deterministic trends, structural breaks and seasonalities, which need to be treated separately.

The proof for a general g follows by Lusin's theorem from the case of a continuous claim g which in turn depends on the characterisation of the superhedging price using concave envelopes, which we now recall.

Definition 4.2.3. *Let $g : \mathbb{R}_+^d \to \mathbb{R}$ be Borel. For $A \subseteq \mathbb{R}_+^d$ and $x \in A$ we define the pointwise concave envelope*

$$\hat{g}_A(x) = \inf\{u(x) \mid u : \mathbb{R}_+^d \to \mathbb{R} \text{ concave, } u \geq g \text{ on } A\}.$$

We define the $\mathbb{P}$-a.s. concave envelope as

$$\hat{g}_{\mathbb{P}}(x) = \inf\{u(x) \mid u : \mathbb{R}_+^d \to \mathbb{R} \text{ concave, } u \geq g \ \mathbb{P}\text{-a.s.}\}.$$

It is well known that in the definition of concave envelopes above we could take infimum over affine functions instead of concave functions. It follows from the definition of the superhedging price in (4.1.1) that we have

$$\pi^{\mathbb{P}}(g) = \hat{g}_{\mathbb{P}}(1) \quad \text{and} \quad \hat{\pi}_N(g) = \hat{g}_{\hat{\mathbb{P}}_N}(1) = \hat{g}_{\{r_1,\dots,r_N\}}(1). \tag{4.2.3}$$

Properties and computational methods for concave envelopes, or more generally for convex hulls of a set of discrete points, have been studied in many applied sciences and there are a number of efficient numerical routines available for their calculation. Naturally computational complexity increases with higher dimensions. Nevertheless there exist algorithms determining approximative convex hulls, whose complexity is independent of the dimension, see for instance Sartipizadeh and Vincent [2016].

To establish a dual formulation for the plugin estimator, assume now that $\mathbb{P}_1 \ll \mathbb{P}$ as well as no-$\mathbb{P}$-arbitrage, NA($\mathbb{P}$), holds and recall this implies the pricing-hedging duality, cf. (4.1.2). It turns out that since $\mathrm{supp}(\hat{\mathbb{P}}_N) \subseteq \mathrm{supp}(\mathbb{P})$ this already implies that NA($\hat{\mathbb{P}}_N$) holds for N large enough. More generally we have:

Proposition 4.2.4. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+^d)$ and $(\mathbb{P}^N)_{N\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}_+^d$ such that $\mathbb{P}^N \Rightarrow \mathbb{P}$ and supp($\mathbb{P}^N$) $\subseteq$ supp($\mathbb{P}$). Then*

$$NA(\mathbb{P}) \quad \Leftrightarrow \quad \exists N_0 \in \mathbb{N} \text{ s.t. } NA(\mathbb{P}^N) \text{ for all } N \geq N_0.$$

In particular, if NA($\mathbb{P}$) holds then in the setup of Theorem 4.2.1 we also have

$$\lim_{N\to\infty} \hat{\pi}_N(g) = \lim_{N\to\infty} \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}^N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = \pi^{\mathbb{P}}(g) \qquad \mathbb{P}^\infty\text{-a.s.} \tag{4.2.4}$$

We close this section considering an extended setup where in addition to the traded assets S, whose historical prices we observe, there also exist options in the market, which can be used for hedging g. If the market enlarged with those options does not allow for an arbitrage, the superhedging price of g in this market is again approximated by the plugin estimator, which now also allows for trading in the options. More precisely, we have the following:

Corollary 4.2.5. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ and $g : \mathbb{R}^d_+ \to \mathbb{R}$ be Borel-measurable. In addition to the assets S, assume that there are $\tilde{d}$ traded options with continuous payoffs $f_1(r)$ and prices f_0 in the market. Define the evaluation map*

$$e(r) = \left(r^1, \dots, r^d, f_1^1(r)/f_0^1, \dots, f_1^{\tilde{d}}(r)/f_0^{\tilde{d}}\right)^{\mathrm{T}}$$

and $\tilde{\mathbb{P}} := \mathbb{P} \circ e^{-1}$. Finally assume no arbitrage, NA($\tilde{\mathbb{P}}$), holds. Then, under the assumptions of Theorem 4.2.1, we have $\mathbb{P}^\infty$-a.s.

$$\begin{aligned}
&\lim_{N\to\infty} \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^{d+\tilde{d}} \text{ s.t. } x + H(e(r) - 1) \geq g(r)\ \forall r \in \{r_1, \dots, r_N\}\} \\
=\ & \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^{d+\tilde{d}} \text{ s.t. } x + H(e(r) - 1) \geq g(r)\ \mathbb{P}\text{-a.s.}\} \\
=\ & \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M},\ \mathbb{E}_{\mathbb{Q}}[f_1]=f_0} \mathbb{E}_{\mathbb{Q}}[g].
\end{aligned}$$

It is worth stressing that in the classical approach to pricing and hedging, the historical returns are seen as *physical measure* inputs and might be used, e.g., for extracting stylised features which models should exhibit. In contrast, option prices f_0 are *risk-neutral measure* inputs and would be used to calibrate the pricing measures. To the best of our knowledge consistent use of both in one estimator has not been achieved before.

4.2.2 Statistical robustness

Robustness of estimators is concerned with their sensitivity to perturbation of the sampling measure $\mathbb{P}$. To formalise this, suppose we have a sequence of estimators T_N which can be expressed as a fixed functional $T : \mathcal{P}(\mathbb{R}^d_+) \to \mathbb{R}$ evaluated on the sequence of empirical measures, i.e., $T_N = T(\hat{\mathbb{P}}_N)$. This is clearly the case with the plugin estimator of the superhedging price in (4.2.1). Hampel [1971] proposed the following definition of statistical robustness:

Definition 4.2.6 (Huber and Ronchetti [2009], p. 42)**.** *Let $r_1, r_2, \dots$ be i.i.d. from $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$. The sequence of estimators $T_N = T(\hat{\mathbb{P}}_N)$ is said to be robust at $\mathbb{P}$ if for every $\varepsilon > 0$ there is $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ and $N \geq N_0$ we have*

$$d_L(\mathbb{P}, \tilde{\mathbb{P}}) \leq \delta \quad \Longrightarrow \quad d_L(\mathcal{L}_{\mathbb{P}}(T_N), \mathcal{L}_{\tilde{\mathbb{P}}}(T_N)) \leq \varepsilon,$$

where d_L is the Lévy-Prokhorov metric

$$d_L(\mathbb{P}, \tilde{\mathbb{P}}) := \inf\{\delta > 0 \mid \mathbb{P}(B) \leq \tilde{\mathbb{P}}(B^\delta) + \delta \text{ for all } B \in \mathcal{B}(\mathbb{R}^d_+)\}. \tag{4.2.5}$$

We sometimes say that T_N is robust with respect to d_L to stress the dependency on the particular choice of the metric. A classical result of Hampel, see [Huber and Ronchetti, 2009, Thm. 2.21], states that if T is asymptotically consistent, i.e.

$$T_N = T(\hat{\mathbb{P}}_N) \longrightarrow T(\mathbb{P}), \quad \text{for all } \mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$$

then T_N is robust at $\mathbb{P}$ if and only if $T(\cdot)$ is continuous at $\mathbb{P}$. The following theorem characterises weak continuity of the superhedging price and hence also robustness of its estimators. In particular, it implies that even for i.i.d. returns $\hat{\pi}_N$ is robust only for special combinations of g and $\mathbb{P}$.

Theorem 4.2.7. *Let g be continuous and $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$. Then the functional $\tilde{\mathbb{P}} \mapsto \pi^{\tilde{\mathbb{P}}}(g)$ is lower semicontinuous at $\mathbb{P}$. It is continuous if and only if*

$$\pi^{\mathbb{P}}(g) = \sup_{\mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g]. \tag{4.2.6}$$

In consequence, any asymptotically consistent estimator T_N is robust at $\mathbb{P}$ only if the above equality holds true.

In particular we see that, in general, the plugin estimator $\hat{\pi}_N(g)$ is not robust w.r.t. d_L. The fact that this holds for any asymptotically consistent estimator suggests strongly that the classical definition of robustness is not adequate in the present context. The superhedging price $\pi^{\mathbb{P}}(g)$ is concerned with the support of $\mathbb{P}$ in the sense that for $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}(\mathbb{R}^d_+)$ with equal supports, and for a continuous g, we have $\pi^{\mathbb{P}_1}(g) = \pi^{\mathbb{P}_2}(g)$. In contrast, any δ-perturbation in the Lévy-Prokhorov sense allows for arbitrary changes to the support, see Lemma 4.10.1. In particular, even if $\mathbb{P}$ satisfies no-arbitrage, measures in its neighbourhood may not and one may not employ (4.1.2) for these. To control the support, we can consider $d_H(\text{supp}(\mathbb{P}), \text{supp}(\tilde{\mathbb{P}}))$, for $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ and where d_H denote the Hausdorff metric on closed subsets of $\mathbb{R}^d_+$.

Proposition 4.2.8. *Let $g : \mathbb{R}^d_+ \to \mathbb{R}$ be uniformly continuous and let $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ such that $NA(\mathbb{P})$ holds. Then the functional $\mathcal{P}(\mathbb{R}^d_+) \ni \tilde{\mathbb{P}} \to \pi^{\tilde{\mathbb{P}}}(g)$ is continuous w.r.t. the pseudometric $d_H(supp(\mathbb{P}), supp(\tilde{\mathbb{P}}))$.*

Alas, this does not allow us to recover statistical robustness of the plugin estimator as the pseudometric above does not admit control over the tails of $\mathbb{P}$. Instead, in Section 4.4.2, we consider a stronger $\mathcal{W}^\infty$ metric which allows to obtain an analogue to Hampel's robustness result.

4.2.3 Financial robustness

The plugin estimator $\hat{\pi}_N$ not only lacks statistical robustness, as seen above, but is also not a financially robust estimate of risk. In fact, if $\mathbb{P}_1 \ll \mathbb{P}$, it converges to the superhedging price from below, i.e., $\hat{\pi}_N \nearrow \pi^{\mathbb{P}}$. From a risk-management perspective one would like to find a consistent estimator for the $\mathbb{P}$-a.s. superhedging price converging from above. However, as we now show, this is not possible in general. As a direct consequence of the discontinuity of the superhedging functional with respect to the Lévy-Prokhorov metric d_L, the convergence from above at some confidence level cannot be achieved in practical applications.

Proposition 4.2.9. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ satisfy $NA(\mathbb{P})$ and g be bounded and Lipschitz continuous. Then, there exists no consistent estimator T_N of $\pi^{\mathbb{P}}(g)$ such that for a confidence level $\alpha \in [0,1]$ there exists $N_0 \in \mathbb{N}$ and*

$$\mathbb{P}^\infty\left(T_N \geq \sup_{\mathbb{Q}\in\mathcal{M},\ \mathbb{Q}\sim\mathbb{P}} \mathbb{E}_{\mathbb{Q}}[g] \text{ for all } N \geq N_0\right) \geq \alpha \tag{4.2.7}$$

for all $\mathbb{P} \in \mathbb{P}(\mathbb{R}^d_+)$.

Thus, in order to achieve the above property (4.2.7) it is necessary to make additional regularity assumptions on $\mathbb{P}$ and g. We show that this is possible for suitably conservative estimators, see Section 4.3.2 below. In the case of the plugin estimator, we can never achieve convergence from above but we can develop an understanding of the order of magnitude of the difference $\pi^{\mathbb{P}}(g) - \hat{\pi}_N(g)$. We first do this by studying the convergence rates, see also Section 4.9. Secondly, we achieve this via notions of statistical robustness suited for the plugin estimator, see Section 4.4.2.

4.2.4 Convergence rates

We now investigate the convergence rate in (4.2.4). While motivated by financial considerations, the question is of independent interest. We focus on the one-dimensional case. We let $F_{\mathbb{P}}$ be the cumulative distribution function of $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+)$ and $d_N = \sup_{r\in\mathbb{R}_+} |F_{\hat{\mathbb{P}}_N}(r) - F_{\mathbb{P}}(r)|$ denote the Kolmogorov-Smirnov distance between $\hat{\mathbb{P}}_N$ and $\mathbb{P}$.

Definition 4.2.10. *For $N \in \mathbb{N}$ and $k = 1, \dots, \lfloor 1/(3d_N) \rfloor$ we define the interquantile distance*

$$\kappa_k^N = F_{\mathbb{P}}^{-1}(3kd_N) - F_{\mathbb{P}}^{-1}(3(k-1)d_N \vee 0+) \text{ for } k = 1, \dots \lfloor 1/(3d_N) \rfloor.$$

Furthermore we set

$$\kappa_0^N = \begin{cases} F_{\mathbb{P}}^{-1}(1) - F_{\mathbb{P}}^{-1}(1-d_N) & \text{if } \mathbb{P} \text{ has bounded support,} \\ 0 & \text{otherwise.} \end{cases}$$

and let $\kappa^N = \sup_{k \in \{0, \dots, \lfloor 1/(3d_N) \rfloor\}} \kappa_k^N$.

We can now establish the speed of convergence for the plugin estimator.

Theorem 4.2.11. *In the setup of Theorem 4.2.1 assume that $NA(\mathbb{P})$ holds and g is bounded and uniformly continuous with $|g(r) - g(\tilde{r})| \leq \delta(|r - \tilde{r}|)$ for some $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\delta(r) \to 0$ for $r \to 0$. Then, as $N \to \infty$,*

$$\begin{aligned} \pi^{\mathbb{P}}(g) - \hat{\pi}_N(g) &= \sup_{\mathbb{Q} \sim \mathbb{P},\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] - \sup_{\mathbb{Q} \sim \hat{\mathbb{P}}_N,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \\ &= \begin{cases} \mathcal{O}(\delta(\kappa^N)) & \text{if } \mathbb{P} \text{ has bounded support,} \\ \mathcal{O}\left(\delta(\kappa^N) + \frac{1}{F_{\mathbb{P}}^{-1}(1-d_N)}\right) & \text{otherwise.} \end{cases} \end{aligned}$$

Remark 4.2.12. When the support of $\mathbb{P}$ is bounded, the above result holds for all continuous g. Furthermore κ^N tends to 0 as $N \to \infty$.

Lemma 4.2.13 (Dvoretzky-Kiefer-Wolfowitz, cf. [Kosorok, 2008, Thm. 11.6])**.** *Suppose the returns $r_1, r_2, \dots$ are i.i.d. samples from $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+)$. Then for every $\varepsilon > 0$*

$$\mathbb{P}^\infty(d_N > \varepsilon) \leq 2e^{-2N\varepsilon^2}.$$

Theorem 4.2.11 and Lemma 4.2.13 yield probabilistic bounds on the distribution of $\pi^{\mathbb{P}}(g) - \hat{\pi}_N(g)$. This is explored in Section 4.9, where we also prove Theorem 4.2.11 and provide extensions of Lemma 4.2.13.

4.3 Improved estimators for the $\mathbb{P}$-a.s. superhedging price

In the last section we have seen that the plugin estimator is asymptotically consistent but has important shortcomings from a statistical and financial point of view. To address these, we propose now new estimators and investigate their asymptotic behaviour as well as their robustness. To construct these, we consider "balls" around the empirical measure $\hat{\mathbb{P}}_N$ and we rely on recent convergence rate results of $\hat{\mathbb{P}}_N$ to $\mathbb{P}$ for the choice of the radii.

We start by considering balls in the Wasserstein$-\infty$ metric, which offers a very good control over the support but where we need to make strong assumptions on $\mathbb{P}$ to control the rate of convergence for $\hat{\mathbb{P}}_N$. Subsequently, in Section 4.3.2, we consider Wasserstein$-p$ metrics, $p \geq 1$. The use of weaker metrics allows us to treat all measures admitting suitable finite moments but requires a penalisation over the dual (pricing) measures. In fact, our estimators rely on a suitable combination of results on convergence of empirical measures with insights into pricing and control over martingale densities. Similarly to the spirit of Corollary 4.2.5 above, we combine the *physical measure-* and the *risk neutral measure-*arguments, see (4.3.4). We note that using Wasserstein metrics, as opposed to weaker metrics, allows us to control the first moment which is important for no-arbitrage reasons, see Section 4.8. Finally, in Section 4.3.3, we consider much larger balls, indeed all of $\mathcal{M}$, and let penalisation select the appropriate measures. Short proofs are given here, the proofs are reported in Appendix 4.11.

4.3.1 Wasserstein $\mathcal{W}^\infty$ balls

When considering robustness of the plugin estimator we saw that to consider measures in a ball around $\hat{\mathbb{P}}_N$ we have to consider a notion of distance which, unlike the Lévy-Prokhorov metric, controls the supports. This is achieved by the Wasserstein-∞ distance

$$\begin{aligned} \mathcal{W}^\infty(\mathbb{P}, \tilde{\mathbb{P}}) &:= \inf_{\gamma \in \Pi(\mathbb{P}, \tilde{\mathbb{P}})} \gamma\text{-ess-sup } |x - y| \\ &= \inf \left\{ \varepsilon > 0 \;\middle|\; \mathbb{P}(B) \leq \tilde{\mathbb{P}}(B^\varepsilon),\ \tilde{\mathbb{P}}(B) \leq \mathbb{P}(B^\varepsilon)\ \forall B \in \mathcal{B}(\mathbb{R}^d_+) \right\}, \end{aligned} \tag{4.3.1}$$

where $\Pi(\mathbb{P}, \tilde{\mathbb{P}})$ denotes the set of all probability measures γ with marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$ and where the equality between the definition and the second representation is a consequence of the Skorokhod representation theorem. A direct comparison of (4.3.1) with (4.2.5) reveals that $\mathcal{W}^\infty$ controls the support in a way that d_L does not. However, one immediate issue with considering $\mathcal{W}^\infty$ is that if $\mathbb{P}$ has unbounded support then $\mathcal{W}^\infty(\mathbb{P}, \hat{\mathbb{P}}_N) = \infty$ for all $N \in \mathbb{N}$ since $\hat{\mathbb{P}}_N$ are finitely supported. For this reason, and also to obtain appropriate confidence intervals, in order to build a good estimator using $\mathcal{W}^\infty$-balls we have to impose relatively strong assumptions on $\mathbb{P}$:

Assumption 4.3.1. *The measure $\mathbb{P}$ is an element of $\mathcal{P}(A)$ for a connected, open and bounded set $A \in \mathcal{B}(\mathbb{R}^d_+)$ with a Lipschitz boundary. Furthermore $\mathbb{P}$ admits a density $\rho : A \to (0, \infty)$ such that there exists $\alpha \geq 1$ for which $1/\alpha \leq \rho(r) \leq \alpha$ on A.*

Under the above assumption, we have explicit bounds on $\mathcal{W}^\infty(\mathbb{P}, \hat{\mathbb{P}}_N)$. The case $d = 1$ follows from Kiefer-Wolfowitz bounds while the case $d \geq 2$ was established in [Trillos and Slepčev, 2015, Thm. 1.1].

Lemma 4.3.2. *Assume that $\mathbb{P}$ fulfils Assumption 4.3.1 and NA($\mathbb{P}$) holds. Furthermore let $r_1, r_2, \dots$ be i.i.d. samples from $\mathbb{P}$. If $d = 1$, then except on a set with probability $O(\exp(-2\sqrt{N}))$, $\mathcal{W}^\infty(\mathbb{P}, \hat{\mathbb{P}}_N) \leq l_N(1, \alpha, A) := \alpha N^{-1/4}$. If $d \geq 2$, then except on a set with probability $O(N^{-2})$,*

$$\mathcal{W}^\infty(\mathbb{P}, \hat{\mathbb{P}}_N) \leq l_N(d, \alpha, A) := C(\alpha, A) \begin{cases} \frac{\log(N)^{3/4}}{N^{1/2}} & \text{if } d = 2, \\ \frac{\log(N)^{1/d}}{N^{1/d}} & \text{if } d \geq 3. \end{cases}$$

We let $B_\varepsilon^\infty(\mathbb{P})$ denote a $\mathcal{W}^\infty$-ball of radius ε around $\mathbb{P}$. The above lemma allows to deduce consistency of the estimator based on such $\mathcal{W}^\infty$ balls:

Proposition 4.3.3. *Consider $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+^d)$ satisfying NA($\mathbb{P}$) and Assumption 4.3.1, and let α, $l_N := l_N(d, \alpha, A)$ be as in Lemma 4.3.2. Let g be a continuous function and $r_1, r_2, \dots$ be i.i.d. samples from $\mathbb{P}$. Then*

$$\hat{\pi}_N^\infty(g) := \sup_{\tilde{\mathbb{P}} \in B_{l_N}^\infty(\hat{\mathbb{P}}_N)} \pi^{\tilde{\mathbb{P}}}(g) \searrow \pi^{\mathbb{P}}(g), \quad \text{as } N \to \infty, \quad \mathbb{P}^\infty - a.s.$$

Remark 4.3.4. The practical use of $\mathcal{W}^\infty$ estimators requires a good handle on l_N. Its dependence on the set A is mild – from Trillos and Slepčev [2015] we see that if a bi-Lipschitz homeomorphism $\phi : \overline{A} \to [0, 1]^d$ exists, then the constant C depends on the domain A only via the Lipschitz constant of ϕ. However, the knowledge of α requires uniform *a priori* estimates on the density of $\mathbb{P}$ on A. This should be contrasted with $\mathcal{W}^p$ estimators below, which only require finiteness of certain moments.

Proof of Proposition 4.3.3. From Lemma 4.3.2, an application of Borel-Cantelli shows that, $\mathbb{P}^\infty$-a.s., for N large enough $\mathbb{P} \in B_{l_N}^\infty(\hat{\mathbb{P}}_N)$ and, in particular, $\hat{\pi}_N^\infty \geq \pi^{\mathbb{P}}$. Further, as $d_H(\text{supp}(\mathbb{P}), \text{supp}(\tilde{\mathbb{P}})) \leq \mathcal{W}^\infty(\mathbb{P}, \tilde{\mathbb{P}})$ by (4.3.1), for all $\tilde{\mathbb{P}} \in B_{l_N}^\infty(\hat{\mathbb{P}}_N)$ we have $\text{supp}(\tilde{\mathbb{P}}) \subseteq \overline{\text{supp}(\mathbb{P})^{2l_N}}$, a compact on which g is uniformly continuous. For measures supported on this compact, Proposition 4.2.8 yields continuity of $\tilde{\mathbb{P}} \to \pi^{\tilde{\mathbb{P}}}$ with respect to $\mathcal{W}^\infty$, which in turn implies consistency of $\hat{\pi}_N^\infty$ and concludes the proof. □

Thus $\hat{\pi}_N^\infty$ is not only consistent but also financially robust. We shall see in Corollary 4.4.6 below, that it is also statistically robust with respect to $\mathcal{W}^\infty$. However, these results only hold for measures $\mathbb{P}$ which satisfy Assumption 4.3.1. In the next section we introduce a family of estimators which exhibit similar desirable properties for a much larger class of measures $\mathbb{P}$.

4.3.2 Wasserstein $\mathcal{W}^p$ balls and martingale densities

We assume no-arbitrage NA($\mathbb{P}$) holds and exploit (4.1.2) to consider estimators of the form

$$\pi_{\mathcal{Q}_N}(g) := \sup_{\mathbb{Q}\in\mathcal{Q}_N} \mathbb{E}_{\mathbb{Q}}[g] \tag{4.3.2}$$

for different specifications of the sets of martingale measures $\mathcal{Q}_N$ based on "balls" around $\hat{\mathbb{P}}_N$. In order to guarantee asymptotic consistency we have to ascertain that the true measure $\mathbb{P}$ is contained in these balls and that we have some control over the tails of the martingale measures in $\mathcal{Q}_N$. Our crucial insight, following recent work of Esfahani and Kuhn [2018], is to work with Wasserstein metrics defined, for $p \geq 1$ and $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ with a finite p^{th} moment, by

$$\mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}}) = \left(\inf\left\{\int_{\mathbb{R}^d_+\times\mathbb{R}^d_+} |r-s|^p \gamma(dr, ds) \;\middle|\; \gamma \in \Pi(\mathbb{P}, \tilde{\mathbb{P}})\right\}\right)^{1/p}$$

where $\Pi(\mathbb{P}, \tilde{\mathbb{P}})$ is the set of probability measures on $\mathbb{R}^d_+ \times \mathbb{R}^d_+$ with marginals $\mathbb{P}$ and $\tilde{\mathbb{P}}$. In case $p = 1$, Kellerer [1982] showed that Kantorovitch-Rubinstein duality (see [Dudley, 2004, Theorem 11.8.2, p.421]) has a particularly nice expression:

$$\mathcal{W}^1(\mathbb{P}, \tilde{\mathbb{P}}) = \sup_{f\in\mathcal{L}_1} \left(\int_{\mathbb{R}^d_+} f(y)d\mu(y) - \int_{\mathbb{R}^d_+} f(y)d\nu(y)\right), \tag{4.3.3}$$

where $\mathcal{L}_1$ denotes the 1-Lipschitz continuous functions $f : \mathbb{R}^d_+ \to \mathbb{R}$.[1] A Wasserstein ball around $\mathbb{P}$ is denoted

$$B^p_\varepsilon(\mathbb{P}) = \{\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+) | \mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}}) \leq \varepsilon\}.$$

For a given $\varepsilon \geq 0$ and $k \in (0, \infty]$, let

$$D^p_{\varepsilon,k}(\mathbb{P}) := \left\{\mathbb{Q} \in \mathcal{M} \;\middle|\; \left\|\frac{d\mathbb{Q}}{d\tilde{\mathbb{P}}}\right\|_\infty \leq k \text{ for some } \tilde{\mathbb{P}} \in B^p_\varepsilon(\mathbb{P})\right\}. \tag{4.3.4}$$

One's first intuition might be to use $\mathcal{Q}_N = D^p_{\varepsilon_N,\infty}(\hat{\mathbb{P}}_N)$ in (4.3.2). Interestingly, this does not work as the balls are too large. Indeed, Wasserstein distance metrises weak convergence and Lemma 4.10.1 shows that any ball around $\hat{\mathbb{P}}_N$ includes measures with full support. As it turns out, to obtain a consistent estimator a subtle interplay is required between ε and k in (4.3.4).

[1] We develop the theory for all $p \geq 1$. In practice, the choice of p has to be made by the statistician. From Theorem 4.4.2 and the equation above Corollary 4.3.7 it is apparent that, for robustness, one wants to take p as large as possible. This however makes moment assumptions more restrictive. The cases $p = 1$ and $p = 2$ are the most popular in literature, given in particular the nice duality for $p = 1$ and the fact that L^2 is a Hilbert space.

Assumption 4.3.5. *1.* $r_1, r_2, \dots$ *are realisations of a time-homogeneous ergodic Markov chain with initial distribution* $\mathbb{P}_1$ *and unique invariant distribution* $\mathbb{P}$ *such that* $\mathbb{P}_1 \ll \mathbb{P}$ *and* $\|d\mathbb{P}_1/d\mathbb{P}\|_{L^s(\mathbb{P})} < \infty$ *for some* $s > 3$*. Furthermore* $\mathbb{E}_{\mathbb{P}}[|r|^q] < \infty$ *for some* $q > 2ps/(s-2)$ *and there exists a sequence* $(\rho_N)_{N\in\mathbb{N}}$ *with* $\sum_{N\in\mathbb{N}} \rho_N < \infty$ *such that if* $r_1 \sim \mathbb{P}$

$$\mathbb{E}\left[\mathbb{E}[f(r_N) - m(f)|r_1]^2\right] \le \rho_N^2, \tag{4.3.5}$$

holds for all $\|f\|_\infty \le 1$*, all* $N \in \mathbb{N}$*, where* $m(f) = \mathbb{E}[f(r_1)]$*.*

2. $r_1, r_2, \dots$ *are i.i.d. samples of* $\mathbb{P}$ *and there exist* $a, c > 0$ *such that* $\mathbb{E}_{\mathbb{P}}[\exp(c|r|^a)] < \infty$*.*

We again refer to Corollary 4.10.4 for examples of processes, which satisfy Assumption 4.3.5. Clearly Assumption 4.3.5.2 implies 4.3.5.1. Under this assumption Fournier and Guillin [2015], see also Esfahani and Kuhn [2018], used concentration of measure techniques to obtain rates of the decay for $\mathbb{P}^\infty(\mathcal{W}(\mathbb{P}, \hat{\mathbb{P}}_N) \ge \varepsilon)$, see Lemma 4.11.2 and Lemma 4.11.5. This allows to compute explicitly a function $\varepsilon_N : (0,1) \to \mathbb{R}_+$ with $\varepsilon_N(\beta) \searrow 0$ as $N \to \infty$, such that

$$\mathbb{P}^\infty(\mathcal{W}^p(\mathbb{P}, \hat{\mathbb{P}}_N) \ge \varepsilon_N(\beta)) \le \beta, \quad N \ge 1.$$

We say that Assumption 4.3.5 holds if either Assumption 4.3.5.1 holds and then ε_N is given in (4.11.8) or Assumption 4.3.5.2 holds and ε_N is then given in (4.11.1). We state now the main result in this section.

Theorem 4.3.6. *Let* g *be either Lipschitz continuous and bounded from below or continuous and bounded,* $p \ge 1$ *and* $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ *satisfying NA(*$\mathbb{P}$*). Suppose Assumption 4.3.5 holds and* $\beta_N \in (0,1)$ *satisfy* $\lim_{N\to\infty} \beta_N = 0$ *and* $\lim_{N\to\infty} \varepsilon_N(\beta_N) = 0$*. Pick a sequence* $k_N = o(1/\varepsilon_N(\beta_N))$*. Then, the limit in* $\mathbb{P}^\infty$*-probability*

$$\lim_{N\to\infty} \pi_{\hat{\mathcal{Q}}_N}(g) = \pi^{\mathbb{P}}(g), \tag{4.3.6}$$

holds, where $\hat{\mathcal{Q}}_N := D^p_{\varepsilon_N(\beta_N), k_N}(\hat{\mathbb{P}}_N)$*. Furthermore, if* $(\beta_N)_{N\in\mathbb{N}}$ *satisfies* $\sum_{N=1}^\infty \beta_N < \infty$ *then the limit* (4.3.6) *also holds* $\mathbb{P}^\infty$*-almost surely.*

The above result shows that $\pi_{\hat{\mathcal{Q}}_N}$ is an asymptotically consistent estimator of $\pi^{\mathbb{P}}$. Note that we assume no arbitrage NA($\mathbb{P}$) so that, using (4.1.2), the convergence above is equivalent to

$$\lim_{N\to\infty} \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g].$$

We write $\hat{\mathcal{Q}}^p_N = \hat{\mathcal{Q}}_N$ when we want to stress the dependence on p. As mentioned above, the consistency depends crucially on the choice of $\hat{\mathcal{Q}}_N$. We discuss this further and motivate the above choice in Section 4.8. For $p > 1$, $D^p_{\varepsilon,k}(\mathbb{P})$ are weakly compact but $D^1_{\varepsilon,k}(\mathbb{P})$ is not even weakly closed in general, see Lemma 4.11.3. In case of $p = 1$, taking weak closure of $\hat{\mathcal{Q}}^1_N$ could destroy the asymptotic consistency of the estimator, e.g., taking $g(r) = (r - 1)$ in the example in the proof of Lemma 4.11.3.

For the particular choice of $\beta_N = \exp(-\sqrt{N})$ under Assumption 4.3.5.2 an explicit computation yields that for N large enough we have

$$\varepsilon_N(\beta_N) = \left(\frac{\log(c_1 \exp(\sqrt{N}))}{c_2 N}\right)^{1/\min(\max(d/p,2),a/(2p))} \sim \frac{1}{N^{1/\min(\max(2d/p,4),a/p)}}.$$

However many other choices of β_N are feasible. The essential point is that for summable (β_N), a Borel-Cantelli argument implies that for N large enough the true distribution $\mathbb{P}$ is within an $\varepsilon_N(\beta_N)$-ball around $\hat{\mathbb{P}}_N$. This allows us to deduce a sufficient condition for financial robustness of our estimator:

Corollary 4.3.7. *In the setup of Theorem 4.3.6 with $\sum_{N=1}^{\infty} \beta_N < \infty$, let g be such that*

$$\exists C \in \mathbb{R}_+ \quad s.t. \quad \sup_{\mathbb{Q}\in\mathcal{M},\ \|\frac{d\mathbb{Q}}{d\mathbb{P}}\|_\infty \le C} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = \pi^{\mathbb{P}}(g). \tag{4.3.7}$$

Then $\pi_{\hat{\mathcal{Q}}_N}(g) \ge \pi^{\mathbb{P}}(g)$ for N large enough so that the estimator is asymptotically consistent and converges from above.

The condition (4.3.7) is motivated by an approximation result, see Lemma 4.11.4. It allows us also to consider the case when we are unsure about the *true* measure $\mathbb{P}$ and instead prefer to superhedge under all measures in its small neighbourhood.

Corollary 4.3.8. *In the setup of Theorem 4.3.6, fix $C > 0$ and assume there exists $\mathbb{Q} \in \mathcal{M}$ such that $\|d\mathbb{Q}/d\mathbb{P}\|_\infty < C$. Consider $C_N \to C$ and a fixed $\varepsilon > 0$. Then*

$$\lim_{N\to\infty} \sup_{\nu\in B^p_{\varepsilon+\varepsilon_N}(\hat{\mathbb{P}}_N),\, \|d\mathbb{Q}/d\nu\|_\infty < C_N} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\nu\in B^p_{\varepsilon}(\mathbb{P}),\, \|d\mathbb{Q}/d\nu\|_\infty < C} \mathbb{E}_{\mathbb{Q}}[g]$$

holds in $\mathbb{P}^\infty$-probability and $\mathbb{P}^\infty$-a.s. whenever $\sum_{N=1}^{\infty} \beta_N < \infty$.

We close this section with two examples illustrating that the assumptions on regularity of g in Theorem 4.3.6 can not be easily relaxed.

Example 4.3.9 (g unbounded, not Lipschitz). Set $g(r) = (r-1)^2$ and consider $(r_N)_{N\ge1}$ i.i.d. from $\mathbb{P} = \delta_1$. For $r_N \ge 2$ consider the measures

$$\nu_N = \frac{\varepsilon_N(\beta_N)}{2}\ \delta_0 + \left(1 - \frac{r_N\ \varepsilon_N(\beta_N)}{2(r_N - 1)}\right)\ \delta_1 + \frac{\varepsilon_N(\beta_N)}{2(r_N - 1)}\ \delta_{r_N}$$

and

$$\mathbb{Q}_N = \frac{\varepsilon_N(\beta_N)}{2\sqrt{\varepsilon_N(\beta_N)}}\,\delta_0 + \left(1 - \frac{r_N\varepsilon_N(\beta_N)}{2(r_N-1)\sqrt{\varepsilon_N(\beta_N)}}\right)\,\delta_1 + \frac{\varepsilon_N(\beta_N)}{2(r_N-1)\sqrt{\varepsilon_N(\beta_N)}}\,\delta_{r_N}.$$

Then $\mathcal{W}^1(\nu_N, \delta_1) \le \varepsilon_N(\beta_N)$,

$$\left\| \frac{d\mathbb{Q}_N}{d\nu_N} \right\|_\infty \le \frac{1}{\sqrt{\varepsilon_N(\beta_N)}}$$

and choosing $r_N = 1/\varepsilon_N(\beta_N)$ we find

$$\mathbb{E}_{\mathbb{Q}_N}[g] \ge \frac{\sqrt{\varepsilon_N(\beta_N)}}{2(1/\varepsilon_N(\beta_N)-1)}(1/\varepsilon_N(\beta_N)-1)^2 \to \infty, \quad \text{ as } N \to \infty.$$

Example 4.3.10 (g bounded, discontinuous). Set $g(r) = \mathbb{1}_{\{r\neq 1\}}$ and consider $(r_N)_{N\ge 1}$ i.i.d. from $\mathbb{P} = \delta_1$. Let

$$\nu_N = \frac{1}{2}\delta_{1-\varepsilon_N(\beta_N)/2} + \frac{1}{2}\delta_{1+\varepsilon_N(\beta_N)/2},$$

then $\mathcal{W}^1(\nu_N, \delta_1) = \varepsilon_N(\beta_N)/2$. We conclude

$$\lim_{N\to\infty} \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_{\mathbb{Q}}[g] \ge \lim_{N\to\infty} \sup_{\mathbb{Q}\ll\nu_N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = 1 \neq 0 = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g].$$

Let us remark that $\pi_{\hat{\mathcal{Q}}_N}(g)$ acting on infinite dimensional spaces is bounded by a more sophisticated version of the plugin estimator. To see this define the Average Value at risk of g at level $1/k$, for $k \ge 1$, by

$$AV@R^{\mathbb{P}}_{1/k}(g) = \max_{\tilde{\mathbb{P}}\sim\mathbb{P},\ \|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \le k} \mathbb{E}_{\tilde{\mathbb{P}}}[g].$$

In dimension one, $d = 1$, it can be re-expressed, see [Föllmer and Schied, 2004, Thm. 4.47], as

$$AV@R^{\mathbb{P}}_{1/k}(g) := k\int_{1-1/k}^{1} F^{-1}_{\mathbb{P}\circ g^{-1}}(x)dx,$$

which makes the link with the classical Value-at-Risk apparent. If we now include the ability to trade and optimise the final position, by the translation-invariance of $AV@R^{\mathbb{P}}_{1/k}(\cdot)$ we can write

$$\begin{aligned}&\inf_{H\in\mathbb{R}^d} AV@R^{\mathbb{P}}_{1/k}(g(r) - H(r-1)) \\ &\quad = \inf\left\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } AV@R^{\mathbb{P}}_{1/k}(g(r) - H(r-1) - x) \le 0\right\},\end{aligned}$$

which is a superhedging price, where the acceptance cone is given by an $AV@R$ constraint. An analogous representation and bounds for $\pi_{\hat{\mathcal{Q}}_N}$ follow:

Corollary 4.3.11. *In the setup of Theorem 4.3.6, let g be 1-Lipschitz and $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ satisfying $NA(\mathbb{P})$. Then there exists $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$*

$$\begin{aligned}
&\inf_{H\in\mathbb{R}^d} AV@R^{\hat{\mathbb{P}}_N}_{1/k_N}(g(r)-H(r-1)) \\
&\leq \inf_{H\in\mathbb{R}^d} \sup_{\tilde{\mathbb{P}}\in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} AV@R^{\tilde{\mathbb{P}}}_{1/k_N}(g(r)-H(r-1)) = \pi_{\hat{\mathcal{Q}}_N}(g) \\
&= \inf \Bigg\{ x\in\mathbb{R} \mid \exists H \in \mathbb{R}^d \ s.t. \qquad\qquad (4.3.8) \\
&\qquad \sup_{\tilde{\mathbb{P}}\in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} AV@R^{\tilde{\mathbb{P}}}_{1/k_N}(g(r)-H(r-1)-x) \leq 0 \Bigg\} \\
&\leq \inf_{|H|\leq 1} AV@R^{\hat{\mathbb{P}}_N}_{1/k_N}(g(r)-H(r-1)) + 2k_N\varepsilon_N(\beta_N)
\end{aligned}$$

on a set of probability greater or equal than $1-\beta_N$.

Note that

$$\pi_{\hat{\mathcal{Q}}_N}(g) = \sup_{\tilde{\mathbb{P}}\in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} \sup_{\|d\nu/d\tilde{\mathbb{P}}\|_\infty \leq k_N} \inf_{H\in\mathbb{R}^d} \mathbb{E}_\nu \left[g - H(r-1)\right].$$

The proof proceeds by using a min-max argument to interchange the two suprema and the infimum above and uses continuity of $\tilde{\mathbb{P}} \mapsto AV@R^{\tilde{\mathbb{P}}}_{1/k_N}(g-H(r-1))$ w.r.t. to $\mathcal{W}^1$, see Pichler [2013] and Appendix 4.11 for details.

We provide a method for the direct calculation of the Wasserstein estimator implemented in TensorFlow[2], which is based on recent duality results obtained in Eckstein et al. [2018]. As this approximation is computationally quite costly when a large sample size is used, we opt to compute the upper bound of $\pi_{\hat{\mathcal{Q}}_N}(g)$ given in (4.3.8) instead. This is shown in Figure 4.1.

4.3.3 A penalty approach: estimator for discontinuous payoffs

In the previous section we introduced the estimator $\pi_{\hat{\mathcal{Q}}_N}$ where $\hat{\mathcal{Q}}_N$ were based on Wasserstein balls around $\hat{\mathbb{P}}_N$. This estimator allowed us to address fundamental shortcomings of the plugin estimator but, as the counterexamples demonstrated, it is only asymptotically consistent under suitable regularity assumptions on g and/or further assumptions on $\mathbb{P}$. To construct an estimator which would be consistent also for discontinuous payoffs while preserving some of the desirable robustness properties of $\pi_{\hat{\mathcal{Q}}_N}$, it is natural to turn to penalty methods used in risk measures and their representations as non-linear expectations. Namely

[2]Our Python implementation for all of the numerical examples in the chapter can be found at `https://github.com/johanneswiesel/Stats-for-superhedging`.

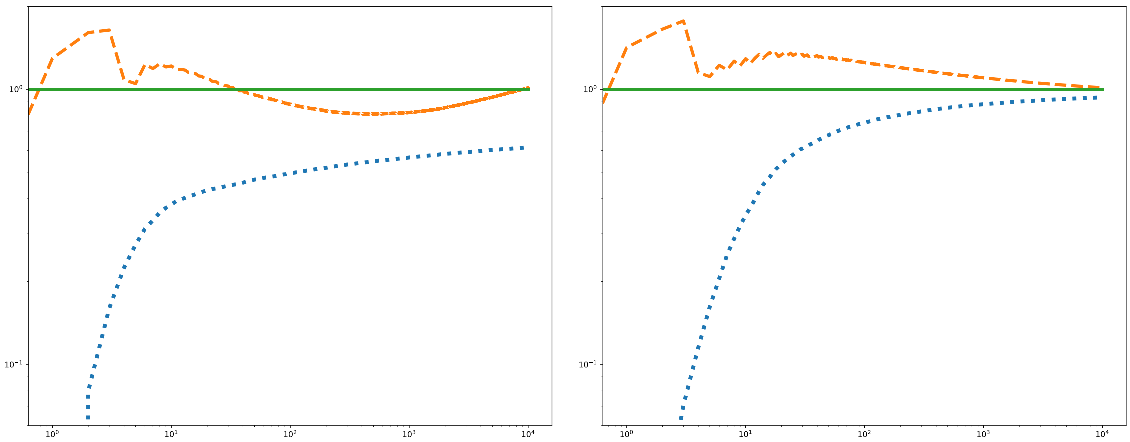

Figure 4.1: Convergence of the plugin estimator $\hat{\pi}_N$ (dotted) and the Wasserstein estimator $\pi_{\hat{\mathbb{Q}}_N}$ (dashed) to the true value (solid) as $N \to \infty$ for $g(r) = (1-r)\mathbb{1}_{\{r \le 1\}} - \sqrt{r-1}\mathbb{1}_{\{r>1\}}$, $\mathbb{P} = \mathrm{Exp}(1)$ (left) and $g(r) = (r-2)^+$, $\mathbb{P} = \exp(\mathcal{N}(0,1))$ (right). Results averaged over 10^3 runs.

we use the maximum norm of the Radon-Nikodym derivative, rather than the Wasserstein distance, in the penalisation term.

Theorem 4.3.12. *In the setting of Theorem 4.2.1, let $NA(\mathbb{P})$ hold and let $g : \mathbb{R}^d_+ \to \mathbb{R}_+$ be Borel-measurable and bounded by some constant $C > 0$. Then for any $C_N \overset{N\to\infty}{\longrightarrow} C$ we have*

$$\lim_{N\to\infty} \sup_{\mathbb{Q}\in\mathcal{M}} \left(\mathbb{E}_{\mathbb{Q}}[g] - C_N \left(\inf_{\hat{\mathbb{Q}}\sim\hat{\mathbb{P}}_N,\ \hat{\mathbb{Q}}\in\mathcal{M}} \left\| \frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}} \right\|_\infty - 1 \right) \right) = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \tag{4.3.9}$$

$\mathbb{P}^\infty$−a.s., where for two probability measures $\mathbb{Q}, \hat{\mathbb{Q}}$ the expression $\left\| \frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}} \right\|_\infty = \infty$ if $\hat{\mathbb{Q}}$ is not absolutely continuous w.r.t. $\mathbb{Q}$.

The direct implementation of (4.3.9) proves numerically expensive and unstable due to the fraction $\|d\hat{\mathbb{Q}}/d\mathbb{Q}\|_\infty$ appearing in the penalisation term. Thus, in Figure 4.2, we show an upper bound on the penalty estimator derived in the proof of Theorem 4.3.12 in Appendix 4.11. We focus on more tractable properties of the plugin and Wasserstein estimator for the rest of the chapter.

4.4 Statistical robustness of superhedging price estimators

Recall that in Theorem 4.2.7 we showed that classical robustness in the sense of Hampel can not hold unless the superhedging price is trivial in that (4.2.6) holds. This is closely related to properties of Lévy-Prokhorov metric, see Lemma 4.10.1, and we are naturally led to consider stronger metrics than d_L, which offer a better control on the support. Below, we investigate the use of Wasserstein distances. First we consider $\mathcal{W}^p$ for $p \ge 1$ which is

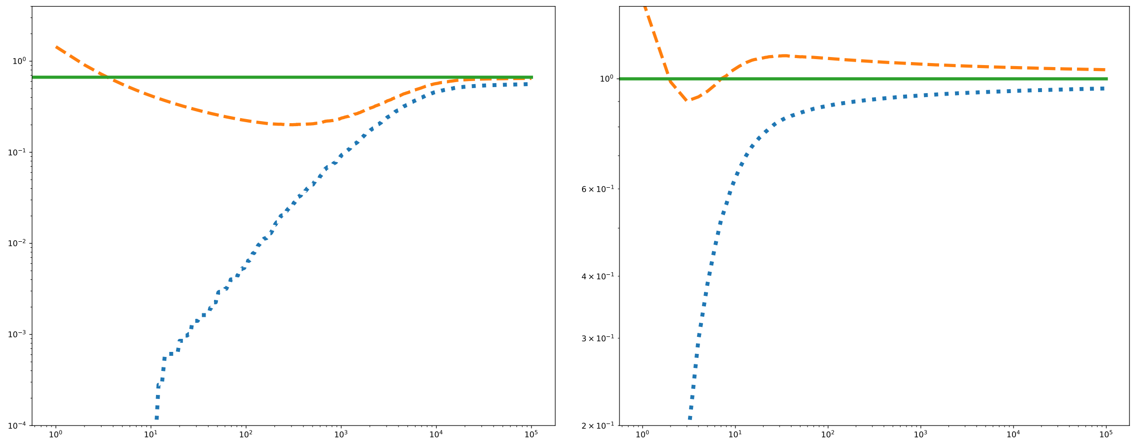

Figure 4.2: Convergence of the penalty estimator (dashed) in Theorem 4.3.12 and the plugin estimator $\hat{\pi}_N$ (dotted) to the true value (solid) as $N \to \infty$ for $g(r) = \mathbb{1}_{\{r \leq 0.5\}}$, $\mathbb{P} = \mathbb{P}^{10}$ from [Obłój and Wiesel, 2020, Example 4.9.1] (left) and $g(r) = \mathbb{1}_{\{r \leq 0.5\}}$, $\mathbb{P} = \mathrm{Exp}(1)$ (right). Results averaged over 10^3 runs.

sufficient to establish robustness of the estimator $\pi_{\hat{\mathcal{Q}}_N}$ from Section 4.3.2. Then we turn to an even stronger metric $\mathcal{W}^\infty$ which is needed to study the plugin estimator.

4.4.1 Robustness with respect to the Wasserstein-Hausdorff metric

Following Li and Lin [2017] we consider Wasserstein-Hausdorff distance, i.e., a Hausdorff distance between subsets of $\mathcal{P}(\mathbb{R}^d_+)$ equipped with $\mathcal{W}^p$:

Definition 4.4.1. *Let $\mathfrak{P}, \tilde{\mathfrak{P}} \subseteq \mathcal{P}(\mathbb{R}^d_+)$. The p-Wasserstein-Hausdorff metric between sets $\mathfrak{P}$ and $\tilde{\mathfrak{P}}$ is given by*

$$\mathcal{W}^p(\mathfrak{P}, \tilde{\mathfrak{P}}) := \max\left(\sup_{\mathbb{P}\in\mathfrak{P}} \inf_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} \mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}}), \sup_{\tilde{\mathbb{P}}\in\tilde{\mathfrak{P}}} \inf_{\mathbb{P}\in\mathfrak{P}} \mathcal{W}^p(\mathbb{P}, \tilde{\mathbb{P}}) \right).$$

In this generality $\mathcal{W}^p(\mathfrak{P}, \tilde{\mathfrak{P}})$ can take the value infinity. Properties of this quantity are discussed in Li and Lin [2017] assuming compactness and uniform integrability of $\mathfrak{P}$ and $\tilde{\mathfrak{P}}$. We apply this distance to the sets of the form $\hat{\mathcal{Q}}_N = D^p_{\varepsilon_N(\beta_N),k_N}(\hat{\mathbb{P}}_N)$, see (4.3.4), and we note that $D^1_{\varepsilon_N(\beta_N),k_N}(\hat{\mathbb{P}}_N)$ is neither compact nor uniformly integrable, see Lemma 4.11.3. We used these sets in Section 4.3.2 to define consistent estimators $\pi_{\hat{\mathcal{Q}}_N}$, see (4.3.2) and Theorem 4.3.6. The following establishes their robustness:

Theorem 4.4.2. *Fix $p \geq 1$. The estimator $\pi_{\hat{\mathcal{Q}}_N}$ studied in Theorem 4.3.6 is robust with respect to the p-Wasserstein-Hausdorff metric in the sense that*

$$\sup_{g\in\mathcal{L}_1} \left| \pi_{\hat{\mathcal{Q}}^1_N}(g) - \pi_{\hat{\mathcal{Q}}^2_N}(g) \right| \leq \mathcal{W}^p(\hat{\mathcal{Q}}^1_N, \hat{\mathcal{Q}}^2_N),$$

where $\hat{\mathcal{Q}}^i_N = D^p_{\varepsilon_N(\beta_N),k_N}(\hat{\mathbb{P}}^i_N)$ for $\mathbb{P}^i \in \mathcal{P}(\mathbb{R}^d_+)$, $i = 1, 2$.

Proof. Note that for all $g \in \mathcal{L}_1$ and $\mathbb{Q}^i \in \hat{\mathcal{Q}}_N^i$, $i = 1, 2$, we have

$$\mathbb{E}_{\mathbb{Q}^1}[g] - \mathbb{E}_{\mathbb{Q}^2}[g] = \int_{\mathbb{R}_+^d \times \mathbb{R}_+^d} g(r) - g(s) d\gamma(r,s) \leq \int_{\mathbb{R}_+^d \times \mathbb{R}_+^d} |r - s| d\gamma(r,s),$$

where $\gamma \in \Pi(\mathbb{Q}^1, \mathbb{Q}^2)$ is a probability measure on $\mathbb{R}_+^d \times \mathbb{R}_+^d$ with marginals $\mathbb{Q}^1$ and $\mathbb{Q}^2$. Taking the infimum over all these probability measures γ yields

$$\mathbb{E}_{\mathbb{Q}^1}[g] - \mathbb{E}_{\mathbb{Q}^2}[g] \leq \mathcal{W}^p(\mathbb{Q}^1, \mathbb{Q}^2)$$

for all $p \geq 1$. The claim follows. $\square$

Remark 4.4.3. It follows in particular that if $\mathbb{P}^1, \mathbb{P}^2$ admit no arbitrage then $\lim_{N\to\infty} \mathcal{W}^p(\hat{\mathcal{Q}}_N^1, \hat{\mathcal{Q}}_N^2) = 0$ implies $\text{supp}(\mathbb{P}^1) = \text{supp}(\mathbb{P}^2)$. Indeed, otherwise there exists a Lipschitz continuous function g such that

$$\sup_{\mathbb{Q}\sim\mathbb{P}^1,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \neq \sup_{\mathbb{Q}\sim\mathbb{P}^2,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g],$$

so, by consistency, $\lim_{N\to\infty} \mathcal{W}^p(\hat{\mathcal{Q}}_N^1, \hat{\mathcal{Q}}_N^2) > 0$, $\mathbb{P}^\infty$-a.s.

4.4.2 Robustness with respect to $\mathcal{W}^\infty$ and perturbations of the support

We reconsider now robustness of the plugin estimator from Section 4.2. In analogy to the previous section, it seems natural to simply consider the Hausdorff distance between the supports of $\hat{\mathbb{P}}_N^1$ and $\hat{\mathbb{P}}_N^2$. In Proposition 4.2.8 we established continuity of $\mathcal{P}(\mathbb{R}_+^d) \ni \tilde{\mathbb{P}} \to \pi^{\tilde{\mathbb{P}}}(g)$ in the pseudometric $d_H(\text{supp}(\mathbb{P}), \text{supp}(\tilde{\mathbb{P}}))$ but noted that it was not sufficient for a robustness result. Recalling the Lévy metric in (4.2.5), if $d_L(\mathbb{P}, \tilde{\mathbb{P}}) \leq \varepsilon$ then $\tilde{\mathbb{P}}$ can be obtained from $\mathbb{P}$ by redistributing ε mass to arbitrary points on $\mathbb{R}_+^d$, while $(1-\varepsilon)$ mass can only be moved in an ε-neighbourhood (in the Euclidean distance) of where $\mathbb{P}$ allocated mass. As we have observed before, the former operation causes problems, as it changes the null sets of the measure uncontrollably. This is no longer possible under our pseudometric. However, to obtain robustness, we have to restrict redistribution of mass to an ε-neighbourhood for all sets and not only for the whole support. This is achieved by the $\mathcal{W}^\infty$ metric as is clear from the second representation in (4.3.1). This leads to the following extended notion of robustness:

Definition 4.4.4. *Let $\mathfrak{P} \subseteq \mathcal{P}(\mathbb{R}_+^d)$ and $r_1, r_2, \dots$ be i.i.d. with distribution $\mathbb{P} \in \mathfrak{P}$. The sequence of estimators $T_N = T(\hat{\mathbb{P}}_N)$ is said to be robust at $\mathbb{P} \in \mathfrak{P}$ w.r.t. $\mathcal{W}^\infty$ on $\mathfrak{P}$, if for all $\varepsilon > 0$ there exist $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and all $\tilde{\mathbb{P}} \in \mathfrak{P}$*

$$\mathcal{W}^\infty(\tilde{\mathbb{P}}, \mathbb{P}) \leq \delta \quad \Longrightarrow \quad d_L(\mathcal{L}_{\tilde{\mathbb{P}}}(T_N), \mathcal{L}_{\mathbb{P}}(T_N)) \leq \varepsilon.$$

The following asserts robustness of the plugin estimator in the above sense and is the main result in this section.

Theorem 4.4.5. *Let* $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ *such that* $NA(\mathbb{P})$ *holds. Then, for a uniformly continuous* g*, the plugin estimator* $\hat{\pi}_N(g)$ *is robust at* $\mathbb{P}$ *w.r.t.* $\mathcal{W}^\infty$ *on* $\mathcal{P}(\mathbb{R}^d_+)$*.*

The proof of Theorem 4.4.5 is reported in [Obłój and Wiesel, 2020, Section 4.12]. There are ways to weaken the continuity assumption on g and obtain robustness on some $\mathfrak{P} \subseteq \mathcal{P}(\mathbb{R}^d_+)$, see Corollary 4.12.2. We close this section with a result on robustness of $\hat{\pi}^\infty_N(g)$ from Proposition 4.3.3.

Corollary 4.4.6. *Let* g *be a continuous function and* $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ *satisfying* $NA(\mathbb{P})$ *and Assumption 4.3.1. Then, the estimator* $\hat{\pi}^\infty_N(g)$ *from Proposition 4.3.3 is robust at* $\mathbb{P}$ *w.r.t.* $\mathcal{W}^\infty$ *on* $\mathcal{P}(\mathbb{R}^d_+)$*.*

4.5 Risk measurement estimation

The $\mathbb{P}$-a.s. superhedging price $\pi^{\mathbb{P}}(g)$ is a very conservative assessment of risk of a short position in a liability with payoff g. Instead, we could use a risk measure $\rho_{\mathbb{P}}$ for such an assessment, as proposed by Cheridito et al. [2017], leading to

$$\pi^{\rho_{\mathbb{P}}}(g) := \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ such that } \rho_{\mathbb{P}}(g - x - H(r-1)) \leq 0\}.$$

Note that we include above the ability to trade in the market in order to (optimally) reduce the risk of g. We consider $\rho_{\mathbb{P}}$ with Kusuoka's representation

$$\rho_{\mathbb{P}}(g) = \sup_{\mu \in \mathfrak{P}} \int_0^1 \mathrm{AV@R}^{\mathbb{P}}_\alpha(g) d\mu(\alpha), \tag{4.5.1}$$

for a set $\mathfrak{P}$ of probability measures on $[0,1]$. This is not very restrictive since this representation, first obtained in Kusuoka [2001], holds for any law invariant coherent risk measure, see Jouini et al. [2006]. Importantly, it enables us to think of $\rho_{\mathbb{P}}(g)$ as a function of the underlying measure $\mathbb{P}$. Much like we did for $\pi^{\mathbb{P}}(g)$, we would like to estimate $\pi^{\rho_{\mathbb{P}}}(g)$ directly from the observed stock returns. To this end we introduce the following estimator

$$\pi^\rho_{B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)}(g) := \inf\Big\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ such that } \sup_{\tilde{\mathbb{P}} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} \rho_{\tilde{\mathbb{P}}}(g - x - H(r-1)) \leq 0\Big\}$$

as the natural equivalent to $\pi_{\hat{\mathcal{Q}}_N}(g)$. In particular, if $\mathfrak{P} = \{\delta_\alpha\}$ where $\alpha \in [0,1]$, we simply have $\rho_{\mathbb{P}}(g) = \mathrm{AV@R}^{\mathbb{P}}_\alpha(g)$ and the corresponding estimator $\pi^\rho_{B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)}(g)$ resembles the Wasserstein $\mathcal{W}^p$ estimator for fixed level $1/k_N := \alpha$. We have the following consistency result:

Proposition 4.5.1. *Assume g satisfies $|g(r) - g(\tilde{r})| \le L|r - \tilde{r}|$ for some $L \in \mathbb{R}$ and that $\sup_{\mu \in \mathfrak{P}} \int_0^1 \mu(d\alpha)/\alpha^{1/p} < \infty$. Then for any $\mathbb{P}$ satisfying NA($\mathbb{P}$) and Assumption 4.3.5 the limit in $\mathbb{P}^\infty$-probability*

$$\lim_{n\to\infty} \pi^{\rho}_{B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)}(g) = \pi^{\rho_{\mathbb{P}}}(g)$$

holds. If Assumption 4.3.5.2 is satisfied, then the limit also holds $\mathbb{P}^\infty$-a.s.

Proof. The "$\ge$"-inequality follows in the proof of Theorem 4.3.6. We now prove the opposite inequality using Pichler [2013][Corollary 11, p.538]. Fix $\varepsilon > 0$. Note that there exists $H \in \mathbb{R}^d$ such that $\rho_{\mathbb{P}}(g - \pi^{\rho_{\mathbb{P}}}(g) - \varepsilon - H(r-1)) \le 0$. Then for all $\tilde{\mathbb{P}} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)$

$$\begin{aligned}
\pi^{\rho_{\tilde{\mathbb{P}}}}(g) &\le \rho_{\tilde{\mathbb{P}}}(g - H(r-1)) = \rho_{\tilde{\mathbb{P}}}(g - \pi^{\rho_{\mathbb{P}}}(g) - \varepsilon - H(r-1)) + \pi^{\rho_{\mathbb{P}}}(g) + \varepsilon \\
&= [\rho_{\tilde{\mathbb{P}}}(g - \pi^{\rho_{\mathbb{P}}}(g) - \varepsilon - H(r-1)) - \rho_{\mathbb{P}}(g - \pi^{\rho_{\mathbb{P}}}(g) - \varepsilon - H(r-1))] \\
&+ \rho_{\mathbb{P}}(g - \pi^{\rho_{\mathbb{P}}}(g) - \varepsilon - H(r-1)) + \pi^{\rho_{\mathbb{P}}}(g) + \varepsilon \\
&\le \tilde{L}\mathcal{W}^p(\tilde{\mathbb{P}}, \mathbb{P}) \sup_{\mu \in \mathfrak{P}} \int_0^1 \frac{\mu(d\alpha)}{\alpha^{1/p}} + \pi^{\rho_{\mathbb{P}}}(g) + \varepsilon.
\end{aligned}$$

As $\varepsilon > 0$ was arbitrary the claim follows. □

Finally, we present a simple empirical test of the performance of our estimators. We simulate weekly returns according to a GARCH$(1,1)$ model:

$$r_n = \sqrt{\frac{\mu - 2}{\mu}} \eta_n \sqrt{h_n}, \quad h_n = \omega + \beta h_{n-1} + \alpha r_{n-1}^2,$$

where $\omega = 0.02, \beta = 0.1, \alpha = 0.8$ and η_n is standard student-t distributed with $\mu = 5$ degrees of freedom. We take 1000 samples from the above $\mathbb{P} \sim$ GARCH$(1,1)$ and calculate the plugin estimator $\pi^{\mathrm{AV@R}_{0.95}}_{\hat{\mathbb{P}}_N}((r-1)^+)$ and the Wasserstein estimator $\pi^{\mathrm{AV@R}_{0.95}}_{B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)}((r-1)^+)$. We compare this to a parametric estimator of $\pi^{\mathrm{AV@R}^{\mathbb{P}}_{0.95}}((r-1)^+)$, where we first estimate the parameters of the GARCH(1,1) model and then compute $\pi^{\mathrm{AV@R}^{\tilde{\mathbb{P}}}_{0.95}}((r-1)^+)$ given the estimated model $\tilde{\mathbb{P}}$. For each of these estimates, we use a running window of 50 weeks, which is in line with the Basel III regulations for calculating the 10-day AV@R (see [Basel Committee on Banking Supervision, 2019, MAR33, p.89]), which set the minimum length of the historical observation period to be one year. We also consider the case when the parameters of the model change for observations $330 - 670$. The behaviour of the three estimators is shown in Figure 4.3. Both the Wasserstein and plugin estimator approximate the true value reasonably well – the Wasserstein estimator being the most conservative estimator. The parametric estimator exhibits the most erratic behaviour which is due to the unstable estimation of the GARCH(1,1) parameters with only 50 data points. This shows advantages of our proposed estimators when compared with a parametric approach, even when the model is correctly selected. The last pane in Figure 4.3 uses S&P500 weekly returns data from 01/01/2006- 01/01/2015 with a moving window of 50 weeks. The GARCH(1,1) estimator, for which the model is mis-specified, does not pick up any markets movements, while both the Plugin and Wasserstein estimator detect the financial crisis and its aftermath. For similar plots but with GARCH(1,1) using log-returns, we refer to Section 4.13. While preliminary, we believe that this simple empirical study points to clear advantages of our approach. In particular, it is encouraging to see that in the last pane, the Wasserstein estimators clearly picks up the financial crisis and the Eurozone debt crisis periods. A further in-depth study of comparative performance of different estimators is clearly needed and is left for future research.

4.6 Convergence of Superhedging strategies

Given the consistency results of Sections 4.2 & 4.3 establishing convergence of superhedging prices we now turn to the convergence of the corresponding superhedging strategies. We start with the case of the plugin estimator.

Theorem 4.6.1. *Consider $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ satisfying NA($\mathbb{P}$). Suppose $\overline{\bigcup_{N\geq 1} \mathrm{supp}(\hat{\mathbb{P}}_N)} = \mathrm{supp}(\mathbb{P})$. Then for any sequence of trading strategies $(H_N)_{N\in\mathbb{N}}$ satisfying*

$$\hat{\pi}_N(g) + H_N(r-1) \geq g(r) \quad \hat{\mathbb{P}}_N\text{-a.s.}$$

there exists a subsequence $(H_{N_k})_{k\in\mathbb{N}}$ converging to some $H \in \mathbb{R}^d$ which satisfies

$$\pi^{\mathbb{P}}(g) + H(r-1) \geq g(r) \quad \mathbb{P}\text{-a.s.}$$

Proof. Note that we can assume without loss of generality that $H_N \in \mathrm{lin}(\mathrm{supp}(\mathbb{P}) - 1)$ for all $N \in \mathbb{N}$. NA($\mathbb{P}$) implies that the sequence $(H_N)_{N\in\mathbb{N}}$ is bounded. Indeed, assume towards a contradiction that $H_N \to \infty$. Clearly $H_N/|H_N| \to \tilde{H} \in \mathbb{R}^d$ with $|\tilde{H}| = 1$. Also since

$$\hat{\pi}_N(g) + H_N(r-1) - g(r) \geq 0 \quad \hat{\mathbb{P}}_N\text{-a.s.}$$

we conclude $\tilde{H}(r-1) \geq 0$ on supp($\mathbb{P}$). By NA($\mathbb{P}$) $\tilde{H}(r-1) = 0$ follows $\mathbb{P}$-a.s. and as $\tilde{H} \in \mathrm{lin}(\mathrm{supp}(\mathbb{P}) - 1))$ we have $\tilde{H} = 0$, which contradicts $|\tilde{H}| = 1$. This shows that $(H_N)_{N\in\mathbb{N}}$ is bounded. Thus there exists a subsequence $(H_{N_k})_{k\in\mathbb{N}}$ of $(H_N)_{N\in\mathbb{N}}$ such that $H_{N_k} \to H$. Lastly we note that for each $r \in \{r_1, \dots, \}$ we have

$$\pi^{\mathbb{P}}(g) + H(r-1) = \lim_{k\to\infty} (\hat{\pi}_{N_k}(g) + H_{N_k}(r-1)) \geq g(r).$$

Thus the claim follows for continuous g. As in the proof of Theorem 4.2.1 we can then extend the result to general g using Lusin's theorem. This concludes the proof. □

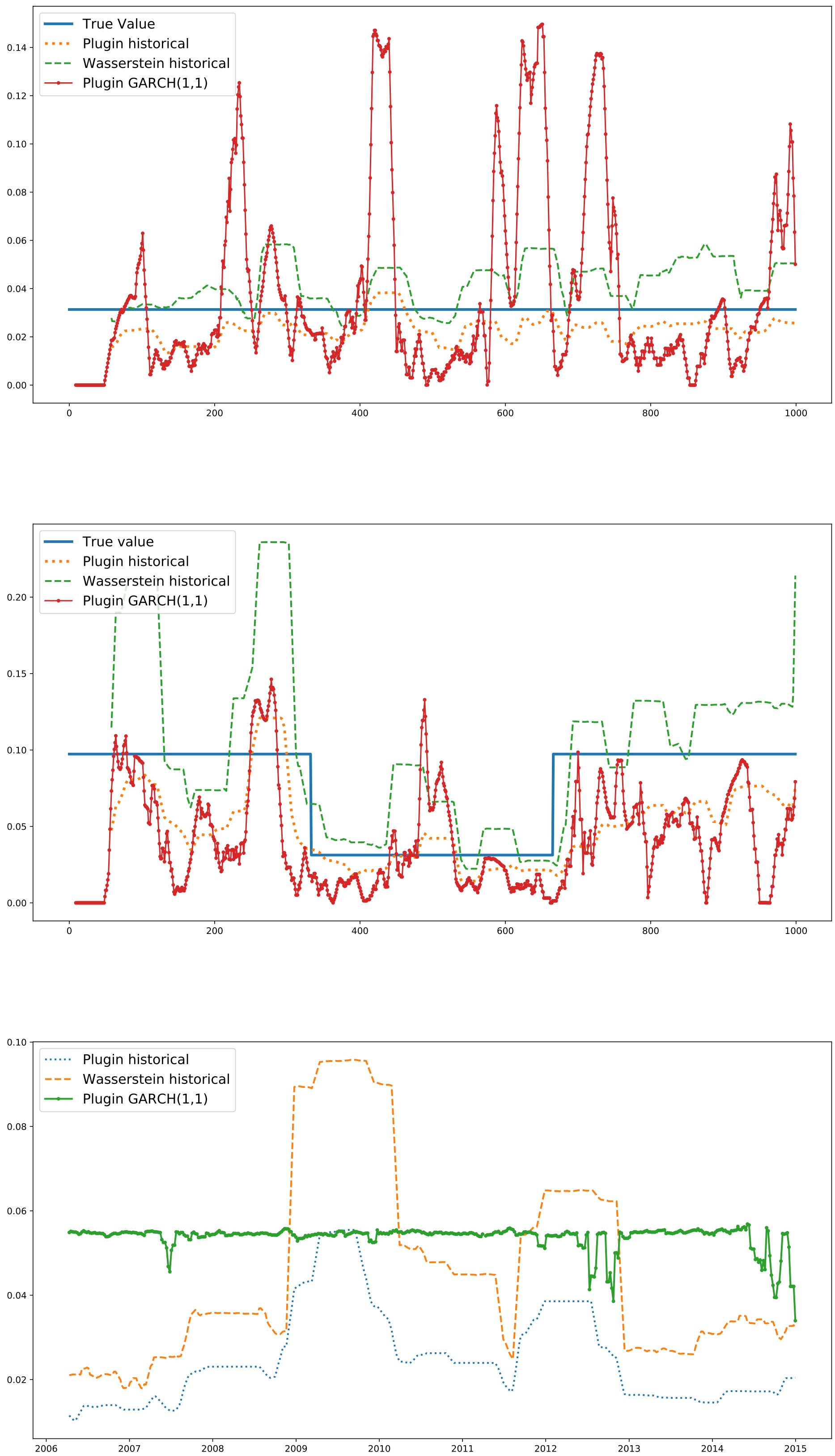

Figure 4.3: Comparison of estimates for $\pi^{\mathrm{AV@R}^{\tilde{\mathbb{P}}}_{0.95}}((r-1)^+)$. Estimates use a rolling window of 50 data points and we plot the average of the last 10 (first two panes) or 5 (last pane) estimates. The data is from $\mathbb{P} \sim \mathrm{GARCH}(1,1)$ (first pane) and its variant with a change in the parameters for the middle third of the time series. The last pane uses S&P500 returns.

Remark 4.6.2. The above claim remains true if we replace $\hat{\pi}_N$ by any consistent estimator which dominates $\hat{\pi}_N$. In particular it is valid for the $\mathcal{W}^\infty$-estimator $\hat{\pi}_N^\infty$ introduced in Proposition 4.3.3 and the penalty estimator introduced in Theorem 4.3.12.

In general one cannot replace the claim $H_{N_k} \to H$ in Theorem 4.6.1 by $H_N \to H \in \mathbb{R}^d$ as the following example shows.

Example 4.6.3. Take $d = 1$ and

$$\hat{\mathbb{P}}_N = \frac{1}{1-2^N} \sum_{k=1}^N 2^{-k} \delta_{1-(-2)^{-k+1}}.$$

Let $g(r) = 1 - |r-1|$. Note that $\mathbb{P}^N \Rightarrow \mathbb{P} := \sum_{k=1}^\infty 2^{-k}\delta_{1-(-2)^{-k+1}}$ and

$$\pi^{\mathbb{P}_N}(g) = 1 - \frac{2}{3} 2^{-N+2} \uparrow \pi^{\mathbb{P}}(g) = 1.$$

Then the sequence of superhedging strategies $(H_N)_{N\in\mathbb{N}}$ is uniquely determined as the slope of the line through the points $(1-(-2)^{-N+2}, g(1-(-2)^{-N+2}))$, $(1-(-2)^{-N+1}, g(1-(-2)^{-N+1}))$, i.e.

$$\begin{aligned} H_N &= \frac{|1-(-2)^{-N+2}-1| - |1-(-2)^{-N+1}-1|}{1-(-2)^{-N+1}-1+(-2)^{-N+2}} = \frac{2^{-N+2}-2^{-N+1}}{-(-2)^{-N+1}+(-2)^{-N+2}} \\ &= \frac{1}{3(-1)^{N+2}} = \frac{(-1)^N}{3}, \end{aligned}$$

which diverges.

Next we establish a corresponding result for the Wasserstein estimator $\pi_{\hat{\mathcal{Q}}_N}(g)$. Recall the definition of AV@R given in Section 4.3.

Theorem 4.6.4. *Let g be Lipschitz continuous and bounded from below. Also let $\mathcal{W}^1(\hat{\mathbb{P}}_N, \mathbb{P}) \le \varepsilon_N(\beta_N)$ for large $N \in \mathbb{N}$, $k_N = o(1/\varepsilon_N(\beta_N))$ and $\lim_{N\to\infty} \pi_{\hat{\mathcal{Q}}_N}(g) = \pi^{\mathbb{P}}(g)$. Then for every sequence of trading strategies $(H_N)_{N\in\mathbb{N}}$ satisfying*

$$AV@R_{1/k_N}^{\hat{\mathbb{P}}_N}\left(g(r) - H_N(r-1) - \pi_{\hat{\mathcal{Q}}_N}(g) - 1/N\right) \le 0 \tag{4.6.1}$$

there exists a subsequence $(H_{N_k})_{k\in\mathbb{N}}$ converging to some $H \in \mathbb{R}^d$ which satisfies

$$\pi^{\mathbb{P}}(g) + H(r-1) \ge g(r) \quad \mathbb{P}\text{-}a.s.$$

Proof. As in Corollary 4.3.11, we see that

$$\pi_{\hat{\mathcal{Q}}_N}(g) \ge \inf\left\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } AV@R_{1/k_N}^{\hat{\mathbb{P}}_N}(g - H(r-1) - x) \le 0\right\}.$$

Recall $\lim_{N\to\infty} \pi_{\hat{\mathbb{Q}}_N}(g) = \pi^{\mathbb{P}}(g)$. As g is Lipschitz continuous Lemma 4.11.1 shows that

$$\begin{aligned}&\Big| AV@R^{\hat{\mathbb{P}}_N}_{1/k_N}\left(g(r) - H(r-1) - \pi_{\hat{\mathbb{Q}}_N}(g)\right) - AV@R^{\mathbb{P}}_{1/k_N}\left(g(r) - H(r-1) - \pi_{\hat{\mathbb{Q}}_N}(g)\right)\Big| \\ &\leq \tilde{C} k_N \mathcal{W}^1(\hat{\mathbb{P}}_N, \mathbb{P}) \leq \tilde{C} k_N \varepsilon_N(\beta_N)\end{aligned} \tag{4.6.2}$$

for some $\tilde{C} > 0$. Note that similarly to the arguments in the proof of Theorem 4.6.1 we can assume that $(H_N)_{N\in\mathbb{N}}$ is bounded, so a subsequence of $(H_N)_{N\in\mathbb{N}}$ (which we also denote by $(H_N)_{N\in\mathbb{N}}$ for convenience) converges to some $H \in \mathbb{R}^d$. Fix $N \in \mathbb{N}$. Then for all $M \geq N$ we have

$$\begin{aligned}&\sup_{\|d\tilde{\mathbb{P}}/d\hat{\mathbb{P}}_N\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M] \\ &\leq 2\tilde{C} k_N \varepsilon_N(\beta_N) + \sup_{\|d\tilde{\mathbb{P}}/d\hat{\mathbb{P}}_M\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M] \\ &\leq 2\tilde{C} k_N \varepsilon_N(\beta_N) + \sup_{\|d\tilde{\mathbb{P}}/d\hat{\mathbb{P}}_M\|_\infty \leq k_M} \mathbb{E}_{\tilde{\mathbb{P}}}[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M] \\ &\leq 2k_N \tilde{C} \varepsilon_N(\beta_N).\end{aligned}$$

Let $\varepsilon > 0$. Then

$$\begin{aligned}&\Big| \sup_{\|d\tilde{\mathbb{P}}/d\hat{\mathbb{P}}_N\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M\right] \\ &\quad - \sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H(r-1) - \pi^{\mathbb{P}}(g)\right] \Big| \\ &\leq \tilde{C} k_N \mathcal{W}^1(\hat{\mathbb{P}}_N, \mathbb{P}) + \Big| \sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M\right] \\ &\quad - \sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \leq k_N} \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H(r-1) - \pi^{\mathbb{P}}(g)\right] \Big| \leq \varepsilon\end{aligned} \tag{4.6.3}$$

by (4.6.2) and $N \leq M$ large enough since we have

$$\begin{aligned}&\sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \leq k_N} \Big| \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H_M(r-1) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M\right] - \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H(r-1) - \pi^{\mathbb{P}}(g)\right] \Big| \\ \leq &\sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty \leq k_N} |H_M - H||\mathbb{E}_{\tilde{\mathbb{P}}}[r-1]| + |\pi^{\mathbb{P}}(g) - \pi_{\hat{\mathbb{Q}}_M}(g) - 1/M| \to 0 \quad (M \to \infty).\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, this implies

$$\sup_{\|d\tilde{\mathbb{P}}/d\mathbb{P}\|_\infty < \infty} \mathbb{E}_{\tilde{\mathbb{P}}}\left[g(r) - H(r-1) - \pi^{\mathbb{P}}(g)\right] \leq 0.$$

This concludes the proof. □

4.7 Multiperiod results

In this section we partly extend results from Sections 4.2 and 4.3 to the case $T > 1$. As before we assume $g : (\mathbb{R}^d_+)^T \to \mathbb{R}$ is Borel. Let $\mathbb{F}$ be generated by the coordinate mappings $\mathbf{r}_t(x) = x_t$, $x \in (\mathbb{R}^d_+)^T$. We write $\mathbf{r}_{i:j}$ for the vector $(\mathbf{r}_i, \dots, \mathbf{r}_j)$, $1 \leq i < j \leq T$ and $\mathbf{r} = \mathbf{r}_{1:T}$. The martingale measures $\mathcal{M}^T$ are now defined via

$$\mathcal{M}^T = \{\mathbb{Q} \in \mathcal{P}((\mathbb{R}^d_+)^T) \mid \mathbb{E}_{\mathbb{Q}}[\mathbf{r}_t | \mathcal{F}_{t-1}] = 1 \text{ for all } t = 1, \dots, T\}.$$

We only consider $\mathbb{R}^d_+$-valued i.i.d. observations $r_1, \dots, r_N$ here[3]. We now concatenate concave envelopes via the following procedure:

Definition 4.7.1. *We define*

$$g(\boldsymbol{r}_{1:t}, \cdot) : \mathbb{R}^d_+ \to \mathbb{R}, \quad \tilde{\boldsymbol{r}} \mapsto g(\boldsymbol{r}_{1:t}, \tilde{\boldsymbol{r}}).$$

Then we recursively set for $\Omega \subseteq (\mathbb{R}^d_+)$

$$\begin{aligned} g^{\Omega}_{T,T}(\boldsymbol{r}) &:= g(\boldsymbol{r}) \\ g^{\Omega}_{t,T}(\boldsymbol{r}_{1:t}) &:= (\widehat{g^{\Omega}_{t+1}(\boldsymbol{r}_{1:t}, \cdot)})_{\Omega}(1), \qquad t = 1, \dots, T-1, \\ g^{\Omega}_{0,T} &:= (\widehat{g^{\Omega}_{1}(\cdot)})_{\Omega}(1). \end{aligned}$$

In analogy with the one-period case we set $\pi^{\hat{\mathbb{P}}_N}_T(g) := g^{\{r_1,\dots,r_N\}}_{0,T}$ *for* $r_i \in \mathbb{R}^d_+$, $i = 1, \dots, N$.

Definition 4.7.2. *For* $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ *we define*

$$\begin{aligned} g^{\mathbb{P}}_{T,T}(\boldsymbol{r}) &:= g(\boldsymbol{r}) \\ g^{\mathbb{P}}_{t,T}(\boldsymbol{r}_{1:t}) &:= (\widehat{g^{\mathbb{P}}_{t+1}(\boldsymbol{r}_{1:t}, \cdot)})_{\mathbb{P}}(1), \quad t = 1, \dots, T-1, \\ g^{\mathbb{P}}_{0,T} &:= (\widehat{g^{\mathbb{P}}_{1}(\cdot)})_{\mathbb{P}}(1) \end{aligned}$$

and define $\mathbb{P}^{\otimes T} = \underbrace{\mathbb{P} \otimes \dots \otimes \mathbb{P}}_{T \text{ times}}$. *Furthermore we set* $\pi^{\mathbb{P}}_T(g) := g^{\mathbb{P}}_{0,T}$.

We quote the following result:

Lemma 4.7.3 (ref. Föllmer and Schied [2004][Theorem 1.31, p.19)**.** *Let* $g : (\mathbb{R}^d_+)^T \to \mathbb{R}$ *be Borel-measurable. Then under* $NA(\mathbb{P})$

$$\begin{aligned} g^{\mathbb{P}}_{0,T} &= \sup_{\mathbb{Q} \sim \mathbb{P}^{\otimes T}} \mathbb{E}_{\mathbb{Q}}[g(\boldsymbol{r})] \\ &= \inf \left\{ x \in \mathbb{R} \;\middle|\; \exists H \in \mathcal{H}(\mathbb{F}) \ s.t. \ x + \sum_{s=1}^{T} H_s(\boldsymbol{r}_s - 1) \geq g(\boldsymbol{r}) \ \mathbb{P}^{\otimes T}\text{-}a.s. \right\}. \end{aligned}$$

[3] The more general Markovian case is likely to require a genuinely novel approach, including using a modified definition of the empirical measure, see Backhoff et al. [2020]. We leave it for future research.

We can now formulate the following multiperiod analogue of Theorem 4.2.1:

Theorem 4.7.4. *Let $\mathbb{P}_1, \mathbb{P}$ be probability measures on $\mathbb{R}^d_+$ and $g : (\mathbb{R}^d_+)^T \to \mathbb{R}$ be Borel-measurable. Assume that the observations $r_1, r_2, \dots$ are i.i.d. samples of $\mathbb{P}$. Then $\mathbb{P}^\infty$-a.s.*

$$\lim_{N\to\infty} \pi_T^{\hat{\mathbb{P}}_N}(g) = \lim_{N\to\infty} g_{0,T}^{\{r_1,\dots,r_N\}} = g_{0,T}^{\mathbb{P}} = \pi_T^{\mathbb{P}}(g).$$

Proof. We prove the claim by induction over $T \in \mathbb{N}$. The case $T = 1$ follows from Theorem 4.2.1. Thus we assume that we have shown that for each $\mathbf{r}_1 \in \mathbb{R}^d_+$

$$\lim_{N\to\infty} g_{1,2}^{\{r_1,\dots,r_N\}}(\mathbf{r}_1) = g_{1,2}^{\mathbb{P}}(\mathbf{r}_1), \quad \mathbb{P}^\infty\text{-a.s.}$$

By Lusin's theorem there exists a sequence of compact sets $K_n \subseteq \text{supp}(\mathbb{P})$ such that $\mathbb{P}(\mathbb{R}^d_+ \setminus K_n) \le 1/n$ and $g_{1,2}^{\mathbb{P}}|_{K_n}$ is continuous. As in the proof of Theorem 4.2.1 we have

$$g_{0,2}^{\mathbb{P}} = \widehat{(g_{1,2}^{\mathbb{P}}(\cdot))}_{\mathbb{P}}(1) = \lim_{n\to\infty} \widehat{(g_{1,2}^{\mathbb{P}}(\cdot))}_{K_n}(1).$$

As the concave envelope of pointwise increasing functions is increasing, we conclude that $g_{1,2}^{\{r_1,\dots,r_N\}}(1)$ is increasing in $N \in \mathbb{N}$. Thus $\mathbb{P}^\infty$-a.s.

$$\widehat{\left(g_{1,2}^{\mathbb{P}}(\cdot)\right)}_{K_n}(1) = \left(\lim_{N\to\infty} \widehat{g_{1,2}^{\{r_1,\dots,r_N\}}(\cdot)}\right)_{K_n}(1) = \lim_{N\to\infty} \widehat{\left(g_{1,2}^{\{r_1,\dots,r_N\}}(\cdot)\right)}_{K_n}(1).$$

Interchanging limits (as they are suprema) yields

$$\begin{aligned}
g_{0,2}^{\mathbb{P}} &= \lim_{N\to\infty} \lim_{n\to\infty} \widehat{\left(g_{1,2}^{\{r_1,\dots,r_N\}}(\cdot)\right)}_{K_n}(1) \\
&= \lim_{N\to\infty} \lim_{n\to\infty} \lim_{M\to\infty} \widehat{\left(g_{1,2}^{\{r_1,\dots,r_N\}}(\cdot)\right)}_{K_n \cap \{r_1,\dots,r_M\}}(1) \\
&= \lim_{N\to\infty} \lim_{M\to\infty} \widehat{\left(g_{1,2}^{\{r_1,\dots,r_N\}}(\cdot)\right)}_{\{r_1,\dots,r_M\}}(1) = \lim_{N\to\infty} g_{0,2}^{\{r_1,\dots,r_N\}}.
\end{aligned}$$

This concludes the proof for $T = 2$. The general induction step follows analogously. $\square$

A similar reasoning can be applied for the estimator $\pi_{\hat{\mathcal{Q}}_N}(g)$ from Theorem 4.3.6 with $\sum_{N=1}^\infty \beta_N < \infty$.

Corollary 4.7.5. *Let $\mathbb{P}$ be a probability measure on $\mathbb{R}^d_+$ and $g : (\mathbb{R}^d_+)^T \to \mathbb{R}$ be 1-Lipschitz and bounded from below. Assume that the observations $r_1, r_2, \dots$ are i.i.d. with respect to $\mathbb{P}$. Then under NA($\mathbb{P}$)*

$$\begin{aligned}
&\lim_{N\to\infty} \sup_{q_1,\dots,q_T} \int_{(\mathbb{R}^d_+)^T} g(\boldsymbol{r}) q_T(\boldsymbol{r}_{1:T-1}; d\boldsymbol{r}_T) \dots q_2(\boldsymbol{r}_1; d\boldsymbol{r}_2) q_1(d\boldsymbol{r}_1) \\
= &\sup_{\mathbb{Q}\in\mathcal{M}^T,\ \mathbb{Q}\sim\mathbb{P}^{\otimes T}} \mathbb{E}_{\mathbb{Q}}[g(\boldsymbol{r})], \qquad \mathbb{P}^\infty\text{-a.s.},
\end{aligned}$$

where $(\boldsymbol{r}_1, \dots, \boldsymbol{r}_t) \mapsto q_{t+1}(\boldsymbol{r}_1, \dots, \boldsymbol{r}_t; \cdot)$ are Borel measurable mappings from $(\mathbb{R}^d_+)^t$ to $\hat{\mathcal{Q}}_N$, $t = 0, \dots, T-1$.

Proof. We show the claim by backwards induction. Fix $\mathbf{r}_{1:T-1} \in (\mathbb{R}^d_+)^{T-1}$. As in the proof of Theorem 4.3.6, we have $\mathbb{P} \in B^p_{\varepsilon_N}(\hat{\mathbb{P}}_N)$ for N large enough $\mathbb{P}^\infty$-a.s. The "$\geq$"-inequality follows as in the proof of Theorem 4.3.6. Indeed,

$$\sup_{q_T \in \hat{\mathcal{Q}}_N} \int_{\mathbb{R}^d_+} g(\mathbf{r}) q_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T) \dots q_2(\mathbf{r}_1; d\mathbf{r}_2) q_1(d\mathbf{r}_1)$$
$$\geq \sup_{q_T \in \mathcal{M},\ \|dq_T/d\mathbb{P}\|_\infty \leq k_N} \int_{\mathbb{R}^d_+} g(\mathbf{r}) q_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T) \dots q_2(\mathbf{r}_1; d\mathbf{r}_2) q_1(d\mathbf{r}_1).$$

As g is bounded from below, passing to the limit with $N \to \infty$ gives the result. Now we show the"$\leq$"-inequality. This follows directly from the fact that $\mathbf{r}_T \mapsto g(\mathbf{r}_{1:T-1}, \mathbf{r}_T)$ is 1-Lipschitz, so by the proof of Theorem 4.3.6

$$\sup_{q_T \in \hat{\mathcal{Q}}_N} \int_{\mathbb{R}^d_+} g(\mathbf{r}) q^N_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T) - \sup_{q_T \in \mathcal{M},\ q_T \sim \mathbb{P}} \int_{\mathbb{R}^d_+} g(\mathbf{r}) dq_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T)$$
$$\leq 2k_N \varepsilon_N(\beta_N).$$

For the induction step we note that for some set $C \subseteq \mathcal{P}(\mathbb{R}^d_+)$, $t \in 1, \dots T$ and some 1-Lipschitz function $f : (\mathbb{R}^d_+)^t \to \mathbb{R}$ we have

$$\left| \sup_{q_t \in C} \int_{\mathbb{R}^d_+} f(\mathbf{r}) dq_t(\mathbf{r}_{1:t-1}, \mathbf{r}_t) - \sup_{q_t \in C} \int_{\mathbb{R}^d_+} f(\tilde{\mathbf{r}}_{1:t-1}, \mathbf{r}_t) dq_t(\tilde{\mathbf{r}}_{1:t-1}, \mathbf{r}_t) \right|$$
$$\leq |\mathbf{r}_{1:t-1} - \tilde{\mathbf{r}}_{1:t-1}| .$$

In particular the functions

$$\sup_{q_T \in \mathcal{M},\ q_T \sim \mathbb{P}} \int_{\mathbb{R}^d_+} g(\mathbf{r}) q_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T) \quad \text{and} \quad \sup_{q_T \in \hat{\mathcal{Q}}_N} \int_{\mathbb{R}^d_+} g(\mathbf{r}) q_T(\mathbf{r}_{1:T-1}; d\mathbf{r}_T)$$

are 1-Lipschitz continuous. The claim now follows using [Bertsekas and Shreve, 1978, Prop. 7.34, p.154]. □

Remark 4.7.6. Similar statements are valid for the penalty estimator and bounded functions g as well as for the $\mathcal{W}^\infty$-estimator with continuous g.

4.8 Asymptotic Consistency, Arbitrage and Contiguity

We discuss now the consistency of $\pi_{\mathcal{Q}_N}(g)$ in (4.3.2) for general sets of martingale measures $\mathcal{Q}_N$. This, in particular, provides a detailed motivation for the construction of our improved estimator in Theorem 4.3.6. Throughout, we assume NA($\mathbb{P}$) to be able to use the dual formulation (4.1.2) and recall that NA($\mathbb{P}$), by the First Fundamental Theorem of Asset pricing, is equivalent to existence of $\mathbb{Q} \in \mathcal{M}$, $\mathbb{Q} \sim \mathbb{P}$.

Clearly, a first necessary condition for asymptotic consistency of $\pi_{\mathcal{Q}_N}(g)$ is that $\mathcal{Q}_N \neq \emptyset$ for N large enough. For the plugin estimator, when $\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \hat{\mathbb{P}}_N\}$, this was established in Proposition 4.2.4 which outlined the relationship between NA($\mathbb{P}$) and NA($\hat{\mathbb{P}}_N$). The proof crucially relied on the fact that $\text{supp}(\hat{\mathbb{P}}_N) \subseteq \text{supp}(\mathbb{P})$. Indeed, for general measures $\nu_N \Rightarrow \mathbb{P}$ the affine hull $\text{aff}(\text{supp}(\nu_N))$ could be of higher dimension than $\text{aff}(\text{supp}(\mathbb{P}))$ and thus there is no relationship between NA($\mathbb{P}$) and NA(ν_N) as the following example shows:

Example 4.8.1. Take $d = 1$ and $\mathbb{P} = \delta_1$. Obviously $\nu_N := \delta_2/N + (1 - 1/N)\delta_1$ converges weakly to $\mathbb{P}$. While NA($\mathbb{P}$) holds, NA(ν_N) is never fulfilled. Conversely $\nu_N := \delta_0/(2N) + \delta_1/(2N) + (1 - 1/N)\delta_2 \Rightarrow \delta_2$, so there exists a $\mathbb{P}$-Arbitrage while there is no ν_N-Arbitrage.

It is thus both natural and necessary to maintain a relationship between $\mathcal{Q}_N$ and $\{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \hat{\mathbb{P}}_N\}$. A minimal property seems to be one of an asymptotic inclusion in the sense that for some sequence $(k_N)_{N\in\mathbb{N}}$ with $\lim_{N\to\infty} k_N = \infty$ we have

$$\mathbb{P}^\infty(\{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \hat{\mathbb{P}}_N \ \& \ \|d\mathbb{Q}/d\hat{\mathbb{P}}_N\|_\infty \leq k_N\} \subseteq \mathcal{Q}_N \text{ for large } N) = 1. \tag{4.8.1}$$

Then the consistency of the plugin estimator in Theorem 4.2.1 implies that

$$\liminf_{N\to\infty} \pi_{\mathcal{Q}_N}(g) \geq \lim_{N\to\infty} \hat{\pi}_N(g) \geq \pi^{\mathbb{P}}(g) \quad \mathbb{P}^\infty\text{-a.s.}$$

The main task now is to identify sequences of sets $\mathcal{Q}_N$ which satisfy (4.8.1) and for which the reverse inequality

$$\limsup_{N\to\infty} \pi_{\mathcal{Q}_N}(g) \leq \pi^{\mathbb{P}}(g) = \sup_{\mathbb{Q}\in\mathcal{M}:\mathbb{Q}\sim\mathbb{P}} \mathbb{E}_{\mathbb{Q}}[g] \quad \mathbb{P}^\infty\text{-a.s.} \tag{4.8.2}$$

holds. For this, we need the $\mathcal{Q}_N$ to asymptotically decrease to the set $\{\mathbb{Q} \in \mathcal{M} : \mathbb{Q} \sim \mathbb{P}\}$. More formally, denoting the ε-neighbourhood of a set A by A^ε, the following can be seen to be a necessary condition for (4.8.2) when g is continuous and bounded:

$$\begin{aligned} &\liminf_{N\to\infty} \nu_N(A^\varepsilon) = 0 \text{ for some sequence } (\nu_N)_{N\in\mathbb{N}} \text{ such that } \nu_N \Rightarrow \mathbb{P} \text{ implies} \\ &\lim_{N\to\infty} \mathbb{Q}_N(A) = 0 \text{ for all } (\mathbb{Q}_N)_{N\in\mathbb{N}}, \text{ such that } \mathbb{Q}_N \in \mathcal{Q}_N \text{ for all } N \in \mathbb{N}, \end{aligned} \tag{4.8.3}$$

for all $A \in \mathcal{B}(\mathbb{R}^d_+)$ and all $\varepsilon > 0$. This condition, which is trivially satisfied if $\mathcal{Q}_N = \{\mathbb{Q} \in \mathcal{M} \mid \mathbb{Q} \sim \hat{\mathbb{P}}_N\}$, can be construed as a contiguity condition on $(\nu_N)_{N\in\mathbb{N}}$ and $(\mathbb{Q}_N)_{N\in\mathbb{N}}$. The notion of contiguity goes back to Le Cam ([Le Cam and Yang, 1990, Chp. 5]) and was used by Kabanov and Kramkov [1995] in a continuous-time setting to describe large financial markets. We introduce a weaker version here, which is sufficient for our setting:

Definition 4.8.2. *A sequence of probability measures* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *is o-contiguous wrt. a sequence* $(\nu_N)_{N\in\mathbb{N}}$*, if for all* $\varepsilon > 0$ *and for all open sets* $O \in \mathcal{B}(\mathbb{R}^d_+)$ $\lim_{N\to\infty} \nu_N(O^\varepsilon) = 0$ *implies* $\lim_{N\to\infty} \mathbb{Q}_N(O) = 0$*. Sequences* $(\nu_N)_{N\in\mathbb{N}}$ *and* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *are mutually o-contiguous if* $(\nu_N)_{N\in\mathbb{N}}$ *is o-contiguous wrt.* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *and* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *is o-contiguous wrt.* $(\nu_N)_{N\in\mathbb{N}}$*. Given two sequences of sets of probability measures* $(\mathfrak{P}_N)_{N\in\mathbb{N}}$ *and* $(\mathcal{Q}_N)_{N\in\mathbb{N}}$ *we say that* $(\mathcal{Q}_N)_{N\in\mathbb{N}}$ *is o-contiguous wrt.* $(\mathfrak{P}_N)_{N\in\mathbb{N}}$ *if for every sequence* $\mathbb{Q}_N \in \mathcal{Q}_N$ *there exists a sequence* $\nu_N \in \mathfrak{P}_N$ *such that* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *is o-contiguous wrt.* $(\nu_N)_{N\in\mathbb{N}}$*.*

Though (4.8.3) is a necessary condition for asymptotic consistency of $\pi_{\mathcal{Q}_N}(g)$ it is not sufficient as the following example shows:

Example 4.8.3. Let $\mathbb{P} = 1/3(\delta_0 + \delta_1 + \delta_2)$, then $\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}((1-r)^+) = 1/2$. Note that for $\nu_N = 1/3(1-1/N)(\delta_0 + \delta_1 + \delta_2) + \delta_{N-4}/N$ we have $\nu_N \Rightarrow \mathbb{P}$ and $\mathbb{Q}_N = (1 - 3/N)\delta_0 + 1/N(\delta_1 + \delta_2 + \delta_{N-4}) \sim \nu_N$. Furthermore $\mathbb{Q}_N$ is o-contiguous with respect to ν_N. Nevertheless $\mathbb{Q}_N \Rightarrow \delta_0$ and thus $E_{\mathbb{Q}_N}((1-r)^+) > 3/4$ for $N > 12$, so asymptotic consistency is not satisfied.

The above example shows that even though $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ is o-contiguous wrt. $(\nu_N)_{N\in\mathbb{N}}$, $\mathbb{Q}_N$ converges weakly to a measure $\mathbb{Q}$, which is not a martingale measure. A way to resolve this problem, is to demand uniform integrability of $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ as was done in Hubalek and Schachermayer [1998] in a continuous-time setting. This allows to establish the following:

Lemma 4.8.4. *Assume NA*$(\mathbb{P})$ *holds. Let* $\mathfrak{P}_N$ *be such that for every sequence* $(\nu_N)_{N\in\mathbb{N}}$ *with* $\nu_N \in \mathfrak{P}_N$ *for all* $N \in \mathbb{N}$ *we have* $\nu_N \Rightarrow \mathbb{P}$*. If* $k_N \to \infty$ *and*

1. $\{\mathbb{Q} \sim \hat{\mathbb{P}}_N,\ \mathbb{Q} \in \mathcal{M},\ \|d\mathbb{Q}/d\mathbb{P}\|_\infty \le k_N\} \subseteq \mathcal{Q}_N$ *for large* N*,*
2. $\mathcal{Q}_N$ *is o-contiguous wrt.* $\mathfrak{P}_N$*,*
3. $\mathcal{Q}_N$ *is such that every sequence* $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ *with* $\mathbb{Q}_N \in \mathcal{Q}_N$ *for all* $N \in \mathbb{N}$ *is uniformly integrable,*

then $\pi_{\mathcal{Q}_N}(g)$ *is asymptotically consistent for all bounded and continuous* g*.*

We omit the proof since relying on uniform integrability is not possible in general: if supp$(\mathbb{P})$ is unbounded (2) and (3) of Lemma 4.8.4 cannot be fulfilled at the same time. The following example illustrates this:

Example 4.8.5. Consider $\mathbb{P}$ with supp$(\mathbb{P}) = \mathbb{R}_+$ and $g(r) = (1-r)^+$. Then $\lim_{N\to\infty} \mathbb{Q}_N(\{0\}) = 1$ holds for every sequence $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ such that

$$\lim_{N\to\infty} \mathbb{E}_{\mathbb{Q}_N}[(1-r)^+] = \sup_{\mathbb{Q}\in\mathcal{M}_{\mathbb{R}_+}} \mathbb{E}_\mathbb{Q}[(1-r)^+] = 1.$$

Obviously δ_0 is not a martingale measure, so $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ cannot be uniformly integrable.

Thus uniform integrability of $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ is in general too strict a requirement. In order to resolve this problem we strengthen the assumption, that ν_N converge weakly to the true measure $\mathbb{P}$. Instead we look at convergence in Wasserstein distance, which is known to metrize the weak convergence (cf. [Villani, 2008, Theorem 6.9, p. 96]).

4.9 Convergence rates for the plugin estimator

We study now in more detail the convergence rates for the plugin estimator $\hat{\pi}_N$ in the one-dimensional case. In particular, we prove Theorem 4.2.11 and also state some extensions.

Following the approach of Hampel, it would be natural to consider the *influence curve* of the superhedging functional, which is simply its Gateaux derivative at $\mathbb{P}$ in the direction of δ_r and represents the marginal influence of an additional observation with value r when the sample size goes to infinity. This however produces trivial results: the function $\varepsilon \mapsto \pi^{(1-\varepsilon)\mathbb{P}+\varepsilon\delta_r}(g)$ is constant on $(0,1)$ so the influence curve at $(r,\mathbb{P})$ will be equal to zero or infinity. This happens because $(1-\varepsilon)\mathbb{P}+\varepsilon\delta_r$ have the same support for all $\varepsilon \in (0,1)$. Clearly, to assess sensitivity of $\pi^{\mathbb{P}}$ to changes in $\mathbb{P}$ we have to vary the support. To this end we consider

$$\pi^{\mathbb{P}}(g) - \inf_{A\subseteq\mathbb{R}^d_+:\mathbb{P}(A)\geq 1-\varepsilon} \pi^{\mathbb{P}|_A}(g)$$

as $\varepsilon \to 0$ and study its natural normalisation. The following examples shows that ε is not the correct normalisation.

Example 4.9.1. Take again $g(r) = |r-1|\wedge 1$ and note that instead of considering $\mathbb{P} = \lambda_{[0,2]}/2$ we can thin out the tails of $\mathbb{P}$ by setting

$$\left.\frac{d\mathbb{P}^n}{d\mathbb{P}}\right|_{[0,1]} = \frac{r^n}{n+1} \quad \text{and} \quad \left.\frac{d\mathbb{P}^n}{d\mathbb{P}}\right|_{[1,2]} = \frac{(2-r)^n}{n+1}$$

Naturally, $\pi^{\mathbb{P}} = \pi^{\mathbb{P}_n}$ but with increasing n the probability mass is less well spread over the support of $\mathbb{P}^n$. We calculate for $n \geq 2$ that $F_{\mathbb{P}^n}(r) = r^{n+1}/2$ for $r \leq 1$ and thus $F^{-1}_{\mathbb{P}^n}(p) = \sqrt[n+1]{2p}$ for $p < 1/2$. This readily implies

$$\pi^{\mathbb{P}}(g) - \inf_{A\subseteq\mathbb{R}^d_+:\mathbb{P}(A)\geq 1-\varepsilon} \pi^{\mathbb{P}|_A}(g) = 1 - g(\sqrt[n+1]{\varepsilon}) = \sqrt[n+1]{\varepsilon}.$$

The above examples motivates using quantile functions for normalisation as stated in Theorem 4.2.11 which we now prove.

Proof of Theorem 4.2.11. As we have noted before $\hat{\pi}_N(g) \leq \pi^{\mathbb{P}}(g)$ and $\hat{\pi}_N(g)$ is non-decreasing in N. Let us first consider $\mathbb{P}$ with bounded support. It suffices to show that

$$\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \leq \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}_N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}_N}[g] + \mathcal{O}(\delta(\kappa^N)).$$

Without loss of generality we assume $r_1 \leq r_2 \leq \cdots \leq r_N$ for the rest of the proof. By the definition of d_N

$$F_{\mathbb{P}}^{-1}((p-d_N) \vee 0+) \leq F_{\hat{\mathbb{P}}_N}^{-1}(p) \leq F_{\mathbb{P}}^{-1}(p+d_N)$$

holds for all $p \in [0,1]$. Thus we note that for all $i = 2, \dots, N$

$$\begin{aligned} r_i - r_{i-1} &= F_{\hat{\mathbb{P}}_N}^{-1}\left(\frac{i}{N}\right) - F_{\hat{\mathbb{P}}_N}^{-1}\left(\frac{i-1}{N}\right) \\ &\leq \sup_{k=1,\dots\lfloor 1/(3d_N)\rfloor} F_{\mathbb{P}}^{-1}(3kd_N) - F_{\mathbb{P}}^{-1}(3(k-1)d_N \vee 0+) \leq \kappa_N. \end{aligned}$$

Next we remark that by definition of $\pi^{\mathbb{P}}(g)$ and Proposition 4.2.8 there exists $C > 0$ such that

$$\pi^{\mathbb{P}}(g) - \pi^{\mathbb{P}(\cdot|[r_1,r_N])}(g) \leq C\delta(F_{\mathbb{P}}^{-1}(d_N) - F_{\mathbb{P}}^{-1}(0+)) + C\delta(F_{\mathbb{P}}^{-1}(1) - F_{\mathbb{P}}^{-1}(1-d_N)\rfloor)).$$

Take $\mathbb{Q} \sim \mathbb{P}(\cdot|[r_1, r_N])$, $\mathbb{Q} \in \mathcal{M}$. We want to apply a simple Balayage construction to redistribute the mass of $\mathbb{Q}$ on the support of $\hat{\mathbb{P}}_N$. Thus we set

$$\begin{aligned} \mathbb{Q}_N(\{r_1\}) &= \int_{[r_1,r_2)} \frac{r-r_2}{r_1-r_2}\mathbb{Q}(dr), \\ \mathbb{Q}_N(\{r_i\}) &= \int_{[r_{i-1},r_i)} \frac{r-r_{i-1}}{r_i-r_{i-1}}\mathbb{Q}(dr) + \int_{[r_i,r_{i+1})} \left(1 - \frac{r-r_i}{r_{i+1}-r_i}\right)\mathbb{Q}(dr) \\ &\qquad \text{for } i = 2, \dots, N-1. \\ \mathbb{Q}_N(\{r_N\}) &= \int_{[r_{N-1},r_N]} \frac{r-r_N}{r_{N-1}-r_N}. \end{aligned}$$

A straightforward calculation shows $\mathbb{Q}_N(\{r_1, \dots, r_N\}) = 1$ and $E_{\mathbb{Q}_N}[r] = 1$. We have

$$\begin{aligned} &\left|\int_{r_1}^{r_N} gd\mathbb{Q}_N - \int_{r_1}^{r_N} gd\mathbb{Q}\right| \\ &= \sum_{k=2}^{N} \int_{r_{k-1}}^{r_k} \left|\frac{g(r_k)(r-r_{k-1}) - g(r_{k-1})(r-r_k)}{r_k - r_{k-1}} - g(r)\right| \mathbb{Q}(dr) \\ &\leq \delta(\kappa^N). \end{aligned} \tag{4.9.1}$$

For $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+)$ with unbounded support and g bounded by $D > 0$ we note that

$$\pi^{\mathbb{P}}(g) - \pi^{\mathbb{P}(\cdot|[0,r_N])}(g) \leq \frac{2D}{r_N} \leq \frac{2D}{F_{\mathbb{P}}^{-1}(1-d_N)}.$$

This concludes the proof. □

In order to improve upon, and further specify the results in Theorem 4.2.11, we recall Lemma 4.11.4 and make the following easy observation:

Lemma 4.9.2. *Let $C > 0$ and $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+^d)$. Then the set $\{\mathbb{Q} \in \mathcal{M} \mid \|d\mathbb{Q}/d\mathbb{P}\| \leq C\}$ is weakly compact.*

Proof. As

$$\limsup_{K\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{Q}_n}[|r|\mathbb{1}_{\{|r|\geq K\}}] \leq \limsup_{K\to\infty} C\mathbb{E}_{\mathbb{P}}[|r|\mathbb{1}_{\{|r|\geq K\}}] = 0,$$

the claim follows. □

Corollary 4.9.3. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+)$ have bounded support and let $g : supp(\mathbb{P}) \to \mathbb{R}$ be continuous such that $|g(r) - g(\tilde{r})| \leq \delta(|r - \tilde{r}|)$ for some monotone $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\delta(r) \to 0$ for $r \to 0$. If* (4.3.7) *holds then*

$$\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] - \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}_N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = \mathcal{O}(d_N^{1/2} + \delta(d_N^{1/2})) \qquad \mathbb{P}^\infty\text{-a.s.}$$

Proof. Note that by assumption it is sufficient to consider martingale measures $\mathbb{Q} \sim \mathbb{P}$ such that $\|d\mathbb{Q}/d\mathbb{P}\|_\infty \leq C$ for some $C > 0$. Similarly to the proof of Theorem 4.2.11 we set

$$\begin{aligned}
\mathbb{Q}_N(\{r_1\}) &= \int_{[r_1,r_2)} \frac{r - r_2}{r_1 - r_2}\mathbb{Q}(dr) + \mathbb{Q}([0, r_1)),\\
\mathbb{Q}_N(\{r_i\}) &= \int_{[r_{i-1},r_i)} \frac{r - r_{i-1}}{r_i - r_{i-1}}\mathbb{Q}(dr) + \int_{[r_i,r_{i+1})} \left(1 - \frac{r - r_i}{r_{i+1} - r_i}\right)\mathbb{Q}(dr)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } i = 2, \dots, N-1.\\
\mathbb{Q}_N(\{r_N\}) &= \int_{[r_{N-1},r_N]} \frac{r - r_N}{r_{N-1} - r_N} + \mathbb{Q}((r_N, \infty)).
\end{aligned}$$

and note that $\mathbb{Q}_N(\{r_1, \dots, r_N\}) = 1$ and

$$|E_{\mathbb{Q}_N}[r] - 1| = |E_{\mathbb{Q}_N}[r] - \mathbb{E}_{\mathbb{Q}}[r]| = \int_0^{r_1} |r_1 - r|\, d\mathbb{Q}(r) + \int_{r_n}^\infty |r_n - r| d\mathbb{Q}(r) \leq K d_N$$

for some $K > 0$. Also

$$\pi^{\mathbb{P}}(g) - \hat{\pi}_N(g) \leq \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] - \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}_N,\ |\mathbb{E}_{\mathbb{Q}}(r-1)|\leq K d_N} \mathbb{E}_{\mathbb{Q}}[g] + 2\delta(K d_N)$$

We further assume that g is bounded by D. Then (4.9.1) becomes

$$\begin{aligned}
\left|\int_{r_1}^{r_N} g d\mathbb{Q}_N - \int_{r_1}^{r_N} g d\mathbb{Q}\right| &= \sum_{k=2}^N \int_{r_{k-1}}^{r_k} \left|\frac{g(r_k)(r - r_{k-1}) - g(r_{k-1})(r - r_k)}{r_k - r_{k-1}} - g(r)\right| \mathbb{Q}(dr)\\
&\leq \delta(d_N^{1/2}) + 2CD\left|\left\{k \in \{1, \dots, \lfloor 1/(3d_N)\rfloor\} \;\middle|\; |\kappa_k^N| \geq d_N^{1/2}\right\}\right| 3d_N\\
&\leq \delta(d_N^{1/2}) + \frac{6CDF_{\mathbb{P}}^{-1}(1)d_N}{d_N^{1/2}} = \mathcal{O}(d_N^{1/2} + \delta(d_N^{1/2})).
\end{aligned}$$

This concludes the proof. □

Remark 4.9.4. We note that the above asymptotic result can be used to set up a utility based hedging problem to approximate $\pi^{\mathbb{P}}$. If we let

$$\alpha_N := U\left(C(d_N^{1/2} + \delta(d_N^{1/2}))\right)$$

for a concave and strictly increasing U and some $C \in \mathbb{R}_+$ then

$$\pi^{\mathbb{P}}(g) \approx \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}_N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] + U^{-1}(\alpha_N)$$

is the value of the utility based hedging problem under $\hat{\mathbb{P}}_N$

$$\inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R} \text{ s.t. } U(x + H(r_1) - g(r)) \geq \alpha_N \text{ for } r = r_1, \dots, r_N\}.$$

Let us now state a generalisation of Lemma 4.2.13 to Markov chains, where we recall Definition 4.10.2 and notation from Section 4.2.4:

Lemma 4.9.5 (cf. [Kosorok, 2008, Theorem 11.24, p.228])**.** *Assume that $r_1, r_2, \dots$ are realisations of a stationary β-mixing Markov chain with exponential decay and invariant measure $\mathbb{P}$ as its initial distribution. Then, as $N \to \infty$*

$$\sqrt{N} \sup_{x\in\mathbb{R}_+} \left|F_{\hat{\mathbb{P}}_N}(x) - F_{\mathbb{P}}(x)\right| \Rightarrow \sup_{x\in\mathbb{R}_+} |G(F_{\mathbb{P}}(x))|,$$

where G is a standard Brownian bridge on $[0, 1]$.

Proof. Setting $\mathcal{F} = \{\mathbb{1}_{(-\infty,x]} \ : \ x \in \mathbb{R}_+\}$, we can choose $p = 4$ in [Kosorok, 2008, Theorem 11.24, p.228] and immediately check that for $\mathbb{P}$ with geometric mixing rate ρ we have

$$\sum_{k=1}^{\infty} k\rho^k < \infty$$

as well as $J_{[]}(\infty, \mathcal{F}, L_4(\mathbb{P})) < \infty$, where $J_{[]}(\infty, \mathcal{F}, L_4(\mathbb{P}))$ denotes the bracketing integral for $\mathcal{F}$ and $p = 4$ (see e.g. [Kosorok, 2008, p. 17]). This concludes the proof. □

For examples of stationary β-mixing Markov chains with exponential decay we refer to Lemma 4.10.3 and Corollary 4.10.4. The above lemma can be used to obtain asymptotic confidence bounds for Corollary 4.9.3. More precisely, for $N \to \infty$, we obtain

$$\begin{aligned}
\mathbb{P}^{\infty}(|\pi^{\mathbb{P}}(g) - \hat{\pi}_N(g)| \geq \varepsilon) &\leq \mathbb{P}^{\infty}(d_N^{1/2} + \delta(d_N^{1/2}) \geq \varepsilon/C) \\
&\leq \mathbb{P}^{\infty}(d_N \geq f(\varepsilon/C)) \\
&\lesssim \mathbb{P}^{\infty}\left(\sup_{x\in\mathbb{R}_+} |G(F_{\mathbb{P}}(x))| \geq f(\varepsilon/C)\sqrt{N}\right)
\end{aligned}$$

for some constant $C > 0$ and $f : \mathbb{R}_+ \to \mathbb{R}_+$ being the square of the inverse of $x \to x + \delta(x)$. Similarly, under some regularity assumptions on $F_{\mathbb{P}}^{-1}$, an analogous but non-asymptotic estimate can be given for Theorem 4.2.11 using Lemma 4.2.13.

To close this section we consider the convergence rate in some cases when Theorem 4.2.11 does not apply. Note that Theorem 4.2.11 applies for a continuous g whenever $\mathbb{P}$ has bounded support or if there exists $K > 0$ such that

$$\sup_{\mathbb{Q}\sim\mathbb{P}|_{[0,K]},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g]. \tag{4.9.2}$$

Suppose now that no such K exists. Clearly if $g(r)/r \to \infty$ as $r \to \infty$ then $\pi^{\mathbb{P}}(g) = \infty$ so consider g with linear growth: $g(r)/r \to c \in \mathbb{R}$ as $r \to \infty$. As $\mathbb{P}$ necessarily has unbounded support we can take a sequence $K_n \to \infty$ and, by the above condition, some $(H_n)_{n\in\mathbb{N}}$ such that

$$\pi^{\mathbb{P}(\cdot|[0,K_n])}(g) + H_n(K_n - 1) = g(K_n).$$

As $\pi^{\mathbb{P}(\cdot|[0,K_n])}(g) \to \pi^{\mathbb{P}}(g)$ we conclude $H_n \uparrow c$. Thus $\pi^{\mathbb{P}}(g) = \max_{\lambda\in[0,1]\cap\mathrm{supp}(\mathbb{P})}(g(\lambda) - c(\lambda - 1)) = g(\tilde{r}) - c(\tilde{r} - 1)$ for some $\tilde{r} \in [0,1] \cap \mathrm{supp}(\mathbb{P})$. In particular

$$\pi^{\mathbb{P}(\cdot|[0,K_n])}(g) \geq \frac{K_n - 1}{K_n - \tilde{r}} g(\tilde{r}) + \frac{1 - \tilde{r}}{K_n - \tilde{r}} g(K_n).$$

Clearly $\hat{\pi}_N(g) \leq \pi^{\mathbb{P}(\cdot|[0,r_n])}(g)$ so the above, using that g is bounded on $[0,1]$, implies that the convergence rate is at most of the order of

$$\frac{1}{\max_{i=1,\ldots,N} r_i} + \left(c - \frac{g(\max_{i=1,\ldots,N} r_i)}{\max_{i=1,\ldots,N} r_i}\right) \tag{4.9.3}$$

but could be slower. Typically, e.g., if $\mathbb{P}$ has a density bounded from below in the neighbourhood of $\tilde{r}$, the tails of the distribution $\mathbb{P}$ are the decisive feature for the convergence rate for $\hat{\pi}_N$ and (4.9.3) holds. We discuss this in more detail in the examples below.

Example 4.9.6. We provide now some examples to illustrate different convergence rates which might be observed in the context of Theorem 4.2.11 or (4.9.3) above. We use $N = 10^5$ realisations of 1000 runs for our numerical illustrations. We start with some examples when Theorem 4.2.11 applies either directly or because (4.9.2) holds.

- $g(r) = |r - 1|$, $\mathbb{P} = \lambda|_{[0,2]}/2$. Note that $\mathbb{P}(\max_{i=1,\ldots,N} r_i \leq x) = (x/2)^N$. Thus
$$\mathbb{E}\left[\max_{i=1,\ldots,N} r_i - 1\right] = \int_0^2 \frac{Nx^{N-1}(x-1)}{2^N} dx = \frac{2N}{N+1} - 1 = 1 - \frac{1}{N+1}.$$
Convergence rate $\mathcal{O}(1/N)$ (Figure 4.4).
- $g(r) = |r - 1|$, $\mathbb{P} = \mathbb{P}^{29}$ from Example 4.9.1. Convergence rate $\mathcal{O}(1/N^{1/30})$ (Figure 4.5).

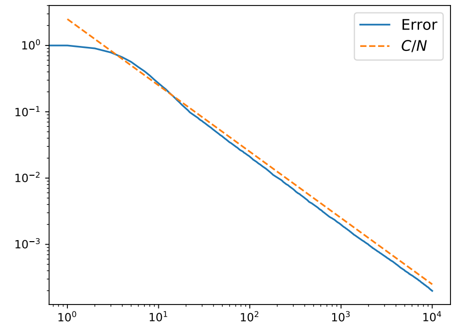

Figure 4.4: $g(r) = |r-1|$, $\mathbb{P} = \lambda|_{[0,2]}/2$

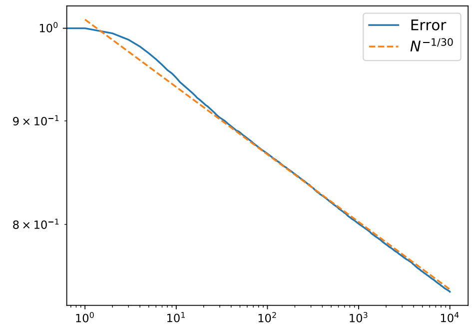

Figure 4.5: $g(r) = |r-1|$, $\mathbb{P} = \mathbb{P}^{29}$

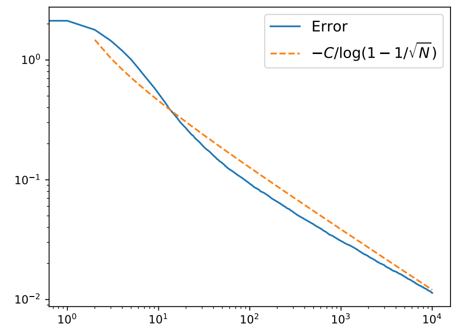

Figure 4.6: $g(r) = (2-r)\mathbb{1}_{\{r\leq 1\}} + \sqrt{r}\mathbb{1}_{\{r\geq 1\}}$, $\mathbb{P} = \text{Exp}(1)$

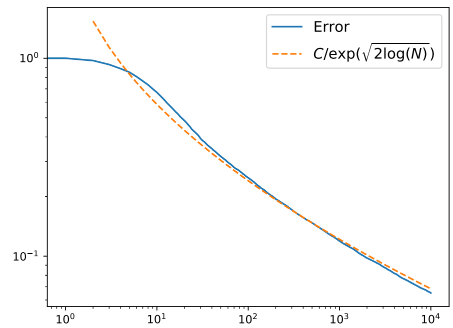

Figure 4.7: $g(r) = (r-2)^+$, $\mathbb{P} = \exp(\mathcal{N}(0,1))$

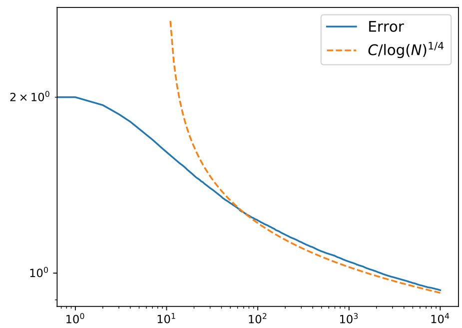

Figure 4.8: $g(r) = |r-1| + \sqrt{r-1}\mathbb{1}_{\{r\geq 1\}}$, $\mathbb{P} = |\mathcal{N}(0,1)|$

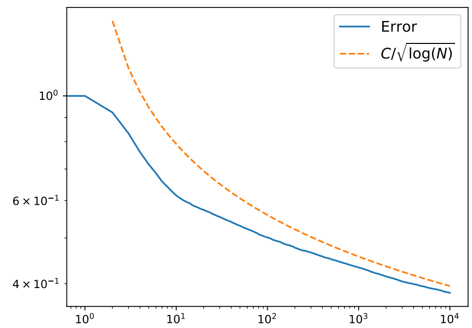

Figure 4.9: $g(r) = (1-r)\mathbb{1}_{\{r\leq 1\}} - \sqrt{r-1}\mathbb{1}_{\{r>1\}}$, $\mathbb{P} = \text{Exp}(1)$

- $g(r) = (2-r)\mathbb{1}_{\{r\leq 1\}} + \sqrt{r}\mathbb{1}_{\{r\geq 1\}}$, $\mathbb{P} = \text{Exp}(1)$. Note that $2 + x/8$ is tangential to $\sqrt{x}$ in $x = 16$. In particular $\pi^{\mathbb{P}}(g) = 2.125 = \pi^{\mathbb{P}(\cdot|[0,16])}(g)$. Convergence rate $\mathcal{O}(F_{\mathbb{P}}^{-1}(d_N)) = \mathcal{O}(-\log(1-1/\sqrt{N}))$ (Figure 4.6).

We move now to examples where we can not rely on Theorem 4.2.11 but use (4.9.3) instead and show the bound may be sharp. The asymptotic distribution of $\max_{i=1,\ldots,N} r_i$ can be determined using classical results from extreme value theory. In particular, the scaled maximum of exponential/normal/lognormal random variables converges weakly to a Gumbel distributed random variable Y (see [Takahashi, 1987, Examples 2,3, p.199]).

- $g(r) = (r-2)^+$, $\mathbb{P} = \exp(\mathcal{N}(0,1))$. Here
$$\max_{i=1,\dots,N} r_i \stackrel{d}{=} \frac{Y \exp(\sqrt{2\log N})}{\sqrt{2\log N}} + \exp(\sqrt{2\log N}) \sim \exp(\sqrt{2\log N}).$$
Convergence rate $\mathcal{O}(1/\max_{i=1,\dots,N} r_i) = \mathcal{O}(1/\exp(\sqrt{2\log N}))$ (Figure 4.7).

- $g(r) = |r-1| - \sqrt{r-1}\mathbb{1}_{\{r\geq 1\}}$, $\mathbb{P} = |\mathcal{N}(0,1)|$. Here
$$\max_{i=1,\dots,N} r_i \stackrel{d}{\sim} \frac{Y}{\sqrt{2\log N}} + \left(\sqrt{2\log N} - \frac{\log(4\pi\log N)}{2\sqrt{2\log N}}\right) \sim \sqrt{\log N}.$$
Convergence rate $\mathcal{O}(1/(\max_{i=1,\dots,N} r_i)^{1/2}) = \mathcal{O}(1/\sqrt[4]{\log N})$ (Figure 4.8).

- $g(r) = (1-r)\mathbb{1}_{\{r\leq 1\}} - \sqrt{r-1}\mathbb{1}_{\{r>1\}}$, $\mathbb{P} = \mathrm{Exp}(1)$. Here
$$\max_{i=1,\dots,N} r_i \stackrel{d}{=} Y + \log N \sim \log N.$$
Convergence rate $\mathcal{O}(1/(\max_{i=1,\dots,N} r_i)^{1/2}) = \mathcal{O}(1/\sqrt{\log N})$ (Figure 4.9).

4.10 Additional results and proofs for Section 4.2

Proof of Theorem 4.2.1. First note that if A_n is a non-decreasing sequence of sets with $A = \lim_n A_n = \cup_n A_n$ then $\hat{g}_A = \lim_n \hat{g}_{A_n}$. The "$\geq$" inequality is obvious and the reverse follows since $\hat{g}_{A_n}$ is a non-decreasing sequence of concave functions thus its limit is a concave function dominating g on A.

Using Lusin's theorem (see [Cohn, 1980, Theorem 7.4.3, page 227]) we can find an increasing sequence K_n of compact subsets of $\mathrm{supp}(\mathbb{P})$ such that $\mathbb{P}(\mathbb{R}^d_+ \setminus K_n) \leq 1/n$ and $g|_{K_n}$ is continuous. Continuity of g on K_n implies that $\hat{g}_{K_n} = \hat{g}_{\mathbb{P}_{|K_n}} \leq \hat{g}_{\mathbb{P}}$. On the other hand, by the argument above, $\lim_n \hat{g}_{K_n} = \hat{g}_{\cup_n K_n} \geq \hat{g}_{\mathbb{P}}$ since $\mathbb{P}(\cup_n K_n) = 1$. We conclude that $\lim_n \hat{g}_{K_n} = \hat{g}_{\mathbb{P}}$. Further, by Birkhoff's ergodic theorem (see [Kallenberg, 2002, Theorem 9.6, p.159]) and $\mathbb{P}_1 \ll \mathbb{P}$ we have
$$\bigcup_N \mathrm{supp}(\hat{\mathbb{P}}^N) = \{r_1, r_2, \dots\} \quad \text{is a.s. dense in } \mathrm{supp}(\mathbb{P})$$
and hence $\hat{g}_{K_n \cap \{r_1, r_2, \dots\}} = \hat{g}_{K_n}$ a.s. By the argument above, we thus have
$$\lim_{N\to\infty} \hat{g}_{\hat{\mathbb{P}}_N} = \hat{g}_{\{r_1,\dots\}} = \hat{g}_{\cup_n K_n \cap \{r_1,\dots\}} = \lim_{n\to\infty} \hat{g}_{K_n} = \hat{g}_{\mathbb{P}}, \quad \mathbb{P}^\infty\text{-a.s.}$$
where the second equality follows since the inclusion $\{r_1, r_2, \dots\} \subset \cup_n K_n$ holds $\mathbb{P}^\infty$-a.s. We conclude using (4.2.3). □

Proof of Proposition 4.2.4. Assume first that NA($\mathbb{P}$) holds. Denote the relative interior of the convex hull of a set $A \subseteq \mathbb{R}^d$ by ri(A). Furthermore write lin(A) for the linear hull of A and aff(A) for the affine hull. We recall that $1 \in \text{ri}(\text{supp}(\mathbb{P}))$ if and only if NA($\mathbb{P}$) holds. Consequently there exist $\varepsilon > 0$, $0 \leq k \leq d$ and $r_i^{\pm} \in \text{conv}(\text{supp}(\mathbb{P}))$ with $r_i^{\pm} = 1 \pm \varepsilon e_i$ for $i = 1, \dots, k$, where $e_1, \dots e_k$ are an orthonormal basis of the space $\text{lin}(\text{supp}(\mathbb{P}) - 1)$. Fix some $r_i^{\pm}$. Then there exists $n \in \mathbb{N}$ and $\tilde{r}_1, \dots, \tilde{r}_n \in \text{supp}(\mathbb{P})$ such that $r_i^{\pm}$ can be written as a convex combination of $\tilde{r}_1, \dots, \tilde{r}_n$. Denote by A the finite collection of $\tilde{r}$ for all $r_i^{\pm}$, $i = 1, \dots, k$. Choosing $0 < \delta < \varepsilon$ sufficiently small we have $1 \in \text{ri}(\{\tilde{r}_i^{\pm} \mid i = 1, \dots, k\})$ for each choice of $\tilde{r}_i^{\pm} \in B_\delta(r_i^{\pm}) \cap \text{aff}(\text{supp}(\mathbb{P}))$. As $\text{supp}(\mathbb{P}^N) \subseteq \text{supp}(\mathbb{P})$ and $\mathbb{P}^N \Rightarrow \mathbb{P}$ there exists $N_0 \in \mathbb{N}$ such that

$$
\begin{aligned}
\mathbb{P}^N(B_\delta(\tilde{r}) \cap \text{aff}(\text{supp}(\mathbb{P}))) &\geq \mathbb{P}^N(B_\delta(\tilde{r}) \cap \text{aff}(\text{supp}(\mathbb{P}^N))) \\
&= \mathbb{P}^N(B_\delta(\tilde{r})) > 0
\end{aligned}
$$

for all $\tilde{r} \in A$ and $N \geq N_0$. Thus $1 \in \text{ri}(\text{supp}(\mathbb{P}^N))$ and NA($\mathbb{P}^N$) holds.
Conversely if there exists $H \in \mathbb{R}^d$ such that $\mathbb{P}(H(r-1) > 0) > 0$ and $\mathbb{P}(H(r-1) \geq 0) = 1$ then again by continuity $H(r-1) \geq 0$ on $\text{supp}(\mathbb{P}) \supseteq \text{supp}(\mathbb{P}^N)$ and the set $\{r \mid H(r-1) > 0\}$ is open. Thus there exists $N_0 \in \mathbb{N}$ such that $\mathbb{P}^N(H(r-1) > 0) > 0$ for all $N \geq N_0$. □

Proof of Corollary 4.2.5. By NA($\tilde{\mathbb{P}}$) we have

$$
\begin{aligned}
&\inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^{d+\tilde{d}} \text{ s.t. } x + H(e(r) - 1) \geq g(r) \ \mathbb{P}\text{-a.s.}\} \\
&= \sup_{\mathbb{Q} \sim \mathbb{P},\ \mathbb{Q} \in \mathcal{M},\ \mathbb{E}_\mathbb{Q}(f_1) = f_0} \mathbb{E}_\mathbb{Q}[g].
\end{aligned}
$$

As e is continuous $\hat{\mathbb{P}}_N \Rightarrow \mathbb{P}$ implies $\widehat{(\tilde{\mathbb{P}})}_N \Rightarrow \tilde{\mathbb{P}}$. Next assume $\text{supp}(\hat{\mathbb{P}}_N) \subseteq \text{supp}(\mathbb{P})$. Again by continuity of f_1 we conclude that

$$
e(\text{supp}(\mathbb{P})) = \text{graph}(f_1) \cap (\text{supp}(\mathbb{P}) \times \mathbb{R}^{\tilde{d}})
$$

is closed. Thus we clearly have $\text{supp}(\tilde{\mathbb{P}}) \subseteq e(\text{supp}(\mathbb{P}))$. We show $e(\text{supp}(\mathbb{P})) \subseteq \text{supp}(\tilde{\mathbb{P}})$. Assume towards a contradiction there exists $e(r) \in e(\text{supp}(\mathbb{P})) \setminus \text{supp}(\tilde{\mathbb{P}})$. Note that for any sequence $(r_n)_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} r_n = r$ we have $\lim_{n \to \infty} e(r_n) = e(r)$. Thus there exists $\varepsilon > 0$ such that $e(B_\varepsilon(r)) \cap \text{supp}(\tilde{\mathbb{P}}) = \emptyset$. But $\tilde{\mathbb{P}}(e(B_\varepsilon(r))) = \mathbb{P}(B_\varepsilon(r)) > 0$, a contradiction. Thus $\text{supp}(\widehat{(\tilde{\mathbb{P}})}_N) \subseteq \text{supp}(\tilde{\mathbb{P}})$ so that $\{e(r_1), e(r_2) \dots\}$ are dense in $\text{supp}(\tilde{\mathbb{P}})$ and Theorem 4.2.1 is still applicable for the enlarged market. The martingale constraint $\mathbb{E}_\mathbb{Q}[\tilde{r}^{d+i} - 1] = 0$ for $i = 1, \dots, \tilde{d}$ is then equivalent to $\mathbb{E}_\mathbb{Q}[f_1^i(S)] = f_0^i(S_0)$. □

Lemma 4.10.1. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+^d)$ with finite first moment and $\varepsilon > 0$. Then for all $x \in \mathbb{R}_+^d$ the ball $B_\varepsilon^1(\mathbb{P})$ in the 1-Wasserstein metric contains $\lambda\delta_x + (1-\lambda)\mathbb{P}$ for some $\lambda \in (0,1)$. In* *particular, any* $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ *can be written as a weak limit of probability measures* $\mathbb{P}^N$ *with* *supp*$(\mathbb{P}^N) = \mathbb{R}^d_+$.

Proof of Lemma 4.10.1. Let $\varepsilon > 0$ be given. We define $\nu = \lambda\delta_x + (1-\lambda)\mu$ and recall that $\mathcal{L}_1$ denotes the 1-Lipschitz functions $f : \mathbb{R}^d_+ \to \mathbb{R}$. Then

$$\begin{aligned}\mathcal{W}^1(\mu,\nu) &= \sup_{f\in\mathcal{L}_1}\left|\int_{\mathbb{R}^d_+} f(y)d\mu(y) - \int_{\mathbb{R}^d_+} f(y)d\nu(y)\right| = \lambda \sup_{f\in\mathcal{L}_1}\left|\int_{\mathbb{R}^d_+} f(y) - f(x)d\mu(y)\right| \\ &= \lambda\int_{\mathbb{R}^d_+}|x-y|d\mu(y) < \varepsilon\end{aligned}$$

for $\lambda > 0$ sufficiently small. □

Proof of Theorem 4.2.7. Fix $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ and a sequence $\mathbb{P}^N$ converging to $\mathbb{P}$. Let $\{r_1, r_2, \dots\}$ be dense in supp$(\mathbb{P})$. Fix $n \geq 1$ and note that, for any $i \geq 1$, weak convergence implies that $\mathbb{P}^N(B_{1/n}(r_i)) > 0$ for all N large enough. In particular there exists $r_i^n \in B_{1/n}(r_i)$ such that $\hat{g}_{\mathbb{P}^N}(r_i^n) \geq g(r_i^n)$. Thus, by the same reasoning as in the proof of Theorem 4.2.1 above,

$$\liminf_{\mathbb{P}^N\Rightarrow\mathbb{P}} \pi^{\mathbb{P}^N}(g) = \liminf_{N\to\infty} \hat{g}_{\mathbb{P}^N}(1) \geq \lim_{n\to\infty}\lim_{k\to\infty} \hat{g}_{\{r_1^n, r_2^n, \dots, r_k^n\}}(1) = \pi^{\mathbb{P}}(g).$$

We conclude using Lemma 4.10.1 since for a sequence with supp$(\mathbb{P}^N) = \mathbb{R}^d_+$, by continuity of g, we have, for all $N \geq 1$,

$$\pi^{\mathbb{P}^N}(g) = \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H(r-1) \geq g \text{ on } \mathbb{R}^d_+\} = \sup_{\mathbb{Q}\in\mathcal{M}_{\mathbb{R}^d_+}} \mathbb{E}_{\mathbb{Q}}[g].$$

For the second part of the theorem, assume $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ is such that

$$\pi^{\mathbb{P}}(g) < \sup_{\mathbb{Q}\in\mathcal{M}_{\mathbb{R}^d_+}} \mathbb{E}_{\mathbb{Q}}[g] = \inf\{x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t. } x + H(r-1) \geq g \text{ on } \mathbb{R}^d_+\}.$$

Take a sequence $(\mathbb{P}^N)_{N\in\mathbb{N}}$, as above, with supp$(\mathbb{P}^N) = \mathbb{R}^d_+$ and $\mathbb{P}^N \Rightarrow \mathbb{P}$. Fix $\varepsilon > 0$ such that

$$2\varepsilon < \pi^{\mathbb{P}^N}(g) - \pi^{\mathbb{P}}(g).$$

For every $\delta > 0$ there exists $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ we have $d_L(\mathbb{P}^N, \mathbb{P}) \leq \delta$. Let T_N be an asymptotically consistent estimator of $\pi^{\mathbb{P}}(g)$. Then, for all N large enough

$$\begin{aligned}d_L(\mathcal{L}_{\mathbb{P}^{N_0}}(T_N), \mathcal{L}_{\mathbb{P}}(T_N)) \geq{}& d_L(\delta_{\pi^{\mathbb{P}}(g)}, \delta_{\pi^{\mathbb{P}^{N_0}}(g)}) - d_L(\mathcal{L}_{\mathbb{P}^{N_0}}(T_N), \delta_{\pi^{\mathbb{P}^{N_0}}(g)}) \\ &- d_L(\delta_{\pi^{\mathbb{P}}(g)}, \mathcal{L}_{\mathbb{P}}(T_N)) \geq \varepsilon.\end{aligned}$$

Thus T_N is not robust at $\mathbb{P}$, which shows the claim. □

Proof of Proposition 4.2.8. Let us fix $\varepsilon > 0$. As g is uniformly continuous, there exists $\delta > 0$, such that for $|r - \tilde{r}| \le \delta$ we have $|g(r) - g(\tilde{r})| \le \varepsilon/3$. Let us now consider $\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ such that $d_H(\mathrm{supp}(\mathbb{P}), \mathrm{supp}(\tilde{\mathbb{P}})) \le \delta$. Then $\pi^{\tilde{\mathbb{P}}}(g) < \infty$ and there exists $H_{\tilde{\mathbb{P}}} \in \mathbb{R}^d$ such that

$$\pi^{\tilde{\mathbb{P}}}(g) + \varepsilon/3 + H_{\tilde{\mathbb{P}}}(\tilde{r} - 1) \ge g(\tilde{r}) \quad \text{for all } \tilde{r} \in \mathrm{supp}(\tilde{\mathbb{P}}).$$

Next we note that by no-arbitrage arguments detailed in [Carassus et al., 2019, Proof of Prop. 3.5] there exists $\delta_0 > 0$ such that for all $\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ with $d_H(\mathrm{supp}(\mathbb{P}), \mathrm{supp}(\tilde{\mathbb{P}})) \le \delta_0$ the strategy $|H_{\tilde{\mathbb{P}}}|$ can be chosen to be bounded by a constant $C > 0$ which depends on g but does not depend on $\tilde{\mathbb{P}}$.
In conclusion consider $\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ such that

$$d_H(\mathrm{supp}(\mathbb{P}), \mathrm{supp}(\tilde{\mathbb{P}})) < \min\{\varepsilon/(3C), \delta, \delta_0\}.$$

Then for every $\tilde{r} \in \mathrm{supp}(\tilde{\mathbb{P}})$ there exists $r \in \mathrm{supp}(\mathbb{P})$ such that $|H_{\tilde{\mathbb{P}}}||r - \tilde{r}| \le \varepsilon/3$ and $|g(r) - g(\tilde{r})| \le \varepsilon/3$. Thus

$$\pi^{\tilde{P}}(g) + \varepsilon + H_{\tilde{\mathbb{P}}}(r - 1) \ge \pi^{\tilde{\mathbb{P}}}(g) + 2\varepsilon/3 + H_{\tilde{\mathbb{P}}}(\tilde{r} - 1) \ge g(\tilde{r}) + \varepsilon/3 \ge g(r).$$

Thus $\pi^{\mathbb{P}}(g) \le \pi^{\tilde{\mathbb{P}}}(g) + \varepsilon$. Exchanging the roles of $\mathbb{P}$ and $\tilde{\mathbb{P}}$ yields the claim. □

Proof of Proposition 4.2.9. We give the following counterexample: Define $\omega_1 = (1, 1, ...)$, $\mathbb{P}_n = (1-1/n)\delta_1 + (1/n)\lambda_{[0,2]}$ and $g(r) = |1-r| \wedge 1$. Obviously $\mathbb{P}_n \Rightarrow \delta_1$ and $\sup_{\mathbb{Q}\sim\delta_1} \mathbb{E}_{\mathbb{Q}}[g] = 0$ as well as $\sup_{\mathbb{Q}\sim\mathbb{P}_n,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = 1$ for all $n \in \mathbb{N}$. By consistency we must have $T_N(\omega_1) \to 0$ as $N \to \infty$, in particular we can assume $T_{N_0}(\omega_1) < 1$. Thus

$$\begin{aligned}\mathbb{P}_n^\infty(T_N \ge 1 \text{ for all } N \ge N_0) &= 1 - \mathbb{P}_n^\infty(T_N < 1 \text{ for some } N \ge N_0)\\ &\le 1 - \mathbb{P}_n^{N_0}(\{\omega_1\}) = 1 - (1 - 1/n)^{N_0} \to 0\\ &\qquad (n \to \infty).\end{aligned}$$

□

4.10.1 Proof of Remark 4.2.2

With regards to Remark 4.2.2 let us now define the following concepts:

Definition 4.10.2. *Suppose that $\{X_n \mid n \in \mathbb{N}\}$ is a time-homogeneous Markov chain with initial distribution $\mathbb{P}$ as its invariant measure and transition kernel K. Then $\{X_n\}$ is called stationary β-mixing with exponential decay if there exist $0 < \rho < 1$ and $c > 0$ such that*

$$\int \|K^n(x, \cdot) - \mathbb{P}(\cdot)\|_{\mathrm{TV}}\, \mathbb{P}(dx) \le c\rho^n \quad \forall n \in \mathbb{N},$$

where $\|\cdot\|_{TV}$ *defines the total variation norm.*
The Markov chain $\{X_n \mid n \in \mathbb{N}\}$ *is called geometrically ergodic if there exists a probability measure* $\mathbb{P} \in \mathcal{P}(\mathbb{R}_+^d)$*, a constant* $0 < \rho < 1$ *and a* $\mathbb{P}$*-integrable non-negative measurable function* g *such that*

$$\|K^n(x,\cdot) - \mathbb{P}(\cdot)\|_{\mathrm{TV}} \leq \rho^n g(x) \quad \forall n \in \mathbb{N},\ \forall x \in \mathbb{R}_+^d.$$

The following lemma lists stationarity and ergodicity properties for some common time series models following Mokkadem [1990], Boussama [1998], He and Teräsvirta [1999], Carrasco and Chen [2002], Basrak et al. [2002], Lindner [2009], Francq and Zakoian [2019], where for simplicity we only consider the simplest cases and refer to the references above for a more comprehensive picture:

Lemma 4.10.3. *The following conditions are sufficient for the* β*-mixing property with exponential decay (if started from the invariant distribution) for the processes* $\{r_n \mid n \in \mathbb{N}\}$ *and* $\{h_n \mid n \in \mathbb{N}\}$ *defined below, as well as for geometric ergodicity of the process* $\{h_n \mid n \in \mathbb{N}\}$*:*

- *Augmented GARCH(1,1) models, where*

$$r_n = \sqrt{h_n}\eta_n, \quad n = 0, 1, \ldots$$
$$\Gamma(h_{n+1}) = c(e_n)\Gamma(h_n) + g(e_{n+1}),$$

  and $\{\eta_n\}$ *is an i.i.d. sequence independent of* h_0 *with* $\mathbb{E}[\eta_n] = 0$*,* $\mathbb{E}[\eta_n^2] = 1$ *and continuous positive density with respect to the Lebesgue measure,* Γ *is an increasing function and* e_{n+1} *is a measurable function of* η_n*.*

  - *LGARCH:* $\beta + \alpha < 1$*, where*

$$h_n = \omega + \beta h_{n-1} + \alpha \eta_{n-1}^2 h_{n-1}$$

    and $\omega > 0$*,* $\beta \geq 0$*,* $\alpha \geq 0$*. Furthermore, if there exists an integer* $s \geq 1$ *such that* $(\beta + \alpha)^s < 1$ *or* $\mathbb{E}[|\eta_n|^{2s}] < \infty$ *and* $\beta + \alpha < 1/(\mathbb{E}[|\eta_n|^{2s}])^{1/s}$*, then* $\mathbb{E}[|r_n|^{2s}] < \infty$*.*

  - *MGARCH:* $|\beta| < 1$*, where*

$$\log(h_n) = \omega + \beta \log(h_{n-1}) + \alpha \log(h_{n-1}^2)$$

    and $\omega > 0$*,* $\beta \geq 0$*,* $\alpha \geq 0$*.*

  - *EGARCH:* $|\beta| < 1$*, where*

$$\log(h_n) = \omega + \beta \log(h_{n-1}) + \alpha(|\eta_{n-1}| + \gamma \eta_{n-1})$$

    and $\gamma \neq 0$*.*

– *NGARCH:* $\beta + \alpha(1 + c^2) < 1$, *where*

$$h_n = \omega + \beta h_{n-1} + \alpha(\eta_{n-1} - c)^2 h_{n-1}$$

and $\omega > 0$, $\beta \geq 0$, $\alpha \geq 0$. *Furthermore, if there exists an integer* $s \geq 1$ *such that* $\mathbb{E}[(\beta + \alpha(\eta_n - c)^2)^s] < 1$ *or* $\mathbb{E}[|\eta_n|^{2s}] < \infty$ *and* $\beta + \alpha < 1/(\mathbb{E}[|\eta_n - c|^{2s}])^{1/s}$, *then* $\mathbb{E}[|r_n|^{2s}] < \infty$.

– *VGARCH:* $\beta < 1$, *where*

$$h_n = \omega + \beta h_{n-1} + \alpha(\eta_{n-1} - c)^2$$

and $\omega > 0$, $\beta, \alpha \geq 0$. *Furthermore, if there exists an integer* $s \geq 1$ *such that* $\beta < 1$, $\mathbb{E}[|\eta_n|^{2s}] < \infty$, *then* $\mathbb{E}[|r_n|^{2s}] < \infty$.

– *TSGARCH:* $\beta + \alpha\mathbb{E}|\eta_t| < 1$, *where*

$$\sqrt{h_n} = \omega + \beta\sqrt{h_{n-1}} + \alpha_1|\eta_{n-1}|\sqrt{h_{n-1}}$$

and $\omega > 0$, $\beta \geq 0$, $\alpha_1 > 0$ *and* $\alpha_1 + \alpha_2 \geq 0$. *Furthermore, if there exists an integer* $s \geq 1$ *such that* $\mathbb{E}[(\beta + \alpha_1|\eta_n|)^s] < \infty$, *then* $\mathbb{E}[|r_n|^s] < \infty$.

– *GJR-GARCH:* $\beta + \alpha_1 + \alpha_2\mathbb{E}\max(0, -\eta_n)^2 < 1$, *where*

$$h_n = \omega + \beta h_{n-1} + \alpha_1\eta_{n-1}^2 h_{n-1} + \alpha_2\max(0, -\eta)^2 h_{n-1}$$

and $\omega > 0$, $\beta \geq 0$, $\alpha_1 > 0$ *and* $\alpha_1 + \alpha_2 \geq 0$. *Furthermore, if there exists an integer* $s \geq 1$ *such that* $\mathbb{E}[(\beta + \alpha_1\eta_n^2 + \alpha_2\max(0, -\eta_n)^2)^s] < \infty$, *then* $\mathbb{E}[|r_n|^{2s}] < \infty$.

– *TGARCH:* $\beta + \alpha_1|\eta_n| + \alpha_2\mathbb{E}\max(0, -\eta_n) < 1$, *where*

$$\sqrt{h_n} = \omega + \beta\sqrt{h_{n-1}} + \alpha_1|\eta_{n-1}|\sqrt{h_{n-1}} + \alpha_2\max(0, -\eta)\sqrt{h_{n-1}}$$

and $\omega > 0$, $\beta \geq 0$, $\alpha_1 > 0$ *and* $\alpha_1 + \alpha_2 \geq 0$. *Furthermore, if there exists an integer* $s \geq 1$ *such that* $\mathbb{E}[(\beta + \alpha_1|\eta_n| + \alpha_2\max(0, -\eta_n))^s] < \infty$, *then* $\mathbb{E}[|r_n|^s] < \infty$.

- *PGARCH*(a, b) *models:* $\sum_{i=1}^{b}\alpha_i + \sum_{j=1}^{a}\beta_j < 1$ *and* $\mathbb{E}(|\eta_n|^{2\delta}) < \infty$, *where*

$$r_n = \sqrt{h_n}\eta_n, \quad n = 0, 1, \ldots$$
$$h_{n+1}^{\delta} = \omega + \sum_{i=1}^{b}\alpha_i h_{n+1-i}^{\delta}|\eta_{n+1-i}|^{2\delta} + \sum_{j=1}^{a}\beta_j h_{n+1-j}^{\delta}$$

and $a, b \geq 1$, $\delta > 0$, $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, b$, $\beta_j \geq 0$, $j = 1, \ldots, a$. *This includes GARCH*(a, b) *for* $\delta = 1$ *and TSGARCH*(a, b) *for* $\delta = 1/2$. *Furthermore* $\mathbb{E}[|r_n|^{2\delta}] < \infty$.

- *Stochastic autoregressive volatility:* $\mathbb{E}|u_n| < \infty$, $\mathbb{E}|\beta + \alpha u_n| < \infty$, *where*

$$\begin{aligned} r_n &= \sigma_n z_n, \\ \log \sigma_n &= \omega + \beta \log \sigma_{n-1} + (\gamma + \alpha \log \sigma_{n-1}) u_n \end{aligned}$$

  and $\{z_n\}_{n\in\mathbb{N}}$, $\{u_n\}_{n\in\mathbb{N}}$ *are mutually independent i.i.d. variables with zero means and unit variances,* $\alpha + \beta > 0$, $\alpha + \gamma > 0$.

As a direct consequence we obtain the following:

Corollary 4.10.4. *Let the distribution of* $\{r_n\}_{n\in\mathbb{N}}$ *be given by one of the models satisfying the corresponding conditions in Lemma 4.10.3 for stationarity and let* $\mathbb{P}_1 = \mathbb{P}$ [4]. *Then the conditions of Theorem 4.2.1 are satisfied. Furthermore Assumption 4.3.5.1 is satisfied in the following cases:*

- *LGARCH: There exists an integer* $q > 3p$ *such that* $(\beta + \alpha)^q < 1$ *or* $\mathbb{E}[|\eta_n|^{2q}] < \infty$ *and* $\beta + \alpha < 1/(\mathbb{E}[|\eta_n|^{2q}])^{1/q}$.
- *NGARCH: There exists an integer* $q > 3p$ *such that* $\mathbb{E}[(\beta + \alpha(\eta_n - c)^2)^q] < 1$ *or* $\mathbb{E}[|\eta_n|^{2q}] < \infty$ *and* $\beta + \alpha < 1/(\mathbb{E}[|\eta_n - c|^{2q}])^{1/q}$.
- *VGARCH: There exists an integer* $q > 3p$ *such that* $\beta < 1$, $\mathbb{E}[|\eta_n|^{2q}] < \infty$.
- *TS-GARCH: There exists an integer* $q > 6p$ *such that* $\mathbb{E}[(\beta + \alpha_1|\eta_n|)^q] < \infty$.
- *GJR-GARCH: There exists an integer* $q > 3p$ *such that* $\mathbb{E}[(\beta+\alpha_1\eta_n^2+\alpha_2 \max(0, -\eta_n)^2)^q] < \infty$.
- *TGARCH: There exists an integer* $q > 6p$ *such that* $\mathbb{E}[(\beta+\alpha_1|\eta_n|+\alpha_2 \max(0, -\eta_n))^q] < \infty$.
- *PGARCH:* $\delta > 3p$.

Proof. The first claim follows directly from Lemma 4.10.3. To show the second claim it is sufficient to check that there exists $q > 6p$ with $\mathbb{E}[|r_n|^q] < \infty$ and (4.3.5) holds. Using again Lemma 4.10.3, existence of $q > 6p$ with $\mathbb{E}[|r_n|^q] < \infty$ is satisfied under the conditions listed in the statement of the corollary. Lastly we note that

$$\|K^n(r_1, \cdot) - \mathbb{P}(\cdot)\|_{\mathrm{TV}} = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} |\mathbb{E}[f(r_n)|r_1] - \mathbb{E}[f(r_1)]|$$

[4] We note that GARCH is traditionally used for logarithmic returns $\log(S_n/S_{n-1})$ and not raw returns $r_n = S_n/S_{n-1}$ in the econometric literature. As $r \mapsto \log(r)$ is continuous this is no restriction for Theorem 4.2.1, however imposing Assumption 4.3.5 for $r_n = \log(S_n/S_{n-1})$ would mean that we had to demand exponential moments of the GARCH models. Even in a parametric setting, statistical inference in a heavy-tailed GARCH environment is notoriously difficult, see e.g. Hall and Yao [2003]. On the other hand, for short time scales we find $r_n \sim 1$ and the approximation $\log(r_n) \sim r_n - 1$ is quite accurate.

and thus conclude that for $0 < \rho < 1$ in the β-mixing with exponential decay property of (r_n), as given by Lemma 4.10.3, $r_1 \sim \mathbb{P}$ and all $\|f\|_\infty \leq 1$ we have

$$\begin{aligned}(\mathbb{E}[(\mathbb{E}[f(r_n)|r_1] - \mathbb{E}[f(r_1)])^2])^{1/2} &\leq (\mathbb{E}[2|\mathbb{E}[f(r_n)|r_1] - \mathbb{E}[f(r_1)]|])^{1/2} \\ &\leq \left(\mathbb{E}\left[2 \sup_{\|f\|_\infty \leq 1} |\mathbb{E}[f(r_n)|r_1] - \mathbb{E}[f(r_1)]|\right]\right)^{1/2} \\ &\leq (4\mathbb{E}\left[\|K^n(r_1, \cdot) - \mathbb{P}(\cdot)\|_{\mathrm{TV}}\right])^{1/2} \\ &\leq 2c\rho^{n/2},\end{aligned}$$

which is summable. □

A test for (strict) stationarity of GARCH models was developed in Francq and Zakoïan [2012]. Applying their test to daily returns of CAC, DAX, DJA, DJI, DJT, DJU, FTSE, Nasdaq, Nikkei, SMI, and SP500, from January 2, 1990 to January 22, 2009 the authors conclude that non-stationarity of the time series is not plausible.

4.11 Additional results and proofs for Section 4.3

Proof of Lemma 4.3.2. It suffices to argue the case $d = 1$ since the for $d \geq 2$ the result is shown in [Trillos and Slepčev, 2015, Thm. 1.1]. The Kiefer-Wolfowitz bounds (see Lemma 4.2.13) yield

$$\mathbb{P}^\infty\left(\sup_{r \in \mathbb{R}_+} |F_{\mathbb{P}}(r) - F_{\hat{\mathbb{P}}_N}(r)| \geq N^{-1/4}\right) \leq \exp(-2\sqrt{N}).$$

Recall also that A is connected, open and bounded and $\mathbb{P}$ satisfies Assumption 4.3.1. It follows that $F_{\mathbb{P}}(x) + \varepsilon \leq F_{\mathbb{P}}(x + \alpha\varepsilon)$ for $x \in \mathbb{R}$. We thus conclude that with $\mathbb{P}^\infty$-probability greater than $(1 - \exp(-2\sqrt{N}))$ we have, for $x \in \mathbb{R}$,

$$\begin{aligned}F_{\mathbb{P}}(x - \alpha N^{-1/4}) \leq F_{\mathbb{P}}(x) - N^{-1/4} &\leq F_{\hat{\mathbb{P}}_N}(x) \leq F_{\mathbb{P}}(x) + N^{-1/4} \\ &\leq F_{\mathbb{P}}(x + \alpha N^{-1/4})\end{aligned}$$

and in particular $\mathcal{W}^\infty(\mathbb{P}, \hat{\mathbb{P}}_N) \leq \alpha N^{-1/4}$. □

Proof of Corollary 4.3.8. For the "$\leq$"-inequality take $\mathbb{Q}_N \in \mathcal{M}$ and $\nu_N \in B^p_{\varepsilon+\varepsilon_N}(\hat{\mathbb{P}}_N)$ such that $\|d\mathbb{Q}_N/d\nu_N\|_\infty < C_N$. By assumption there exists a martingale measure $\mathbb{Q} \sim \mathbb{P}$ such that $\|d\mathbb{Q}/d\mathbb{P}\|_\infty \leq C - \delta$ for some $\delta > 0$. Define

$$\delta_N := \frac{2|C_N - C|(C - \delta)}{C_N - C + \delta} \vee \left(\frac{2\varepsilon_N C_N}{\varepsilon + 2\varepsilon_N} - |C_N - C|\right)$$

for all $N \in \mathbb{N}$. For $N \in \mathbb{N}$ large enough choose

$$\lambda_N \in \left(\frac{|C_N - C| + \delta_N}{C_N}, \frac{\delta_N}{C - \delta}\right).$$

Then $\lambda_N \in (0,1)$, $\lim_{N\to\infty} \lambda_N = 0$ and $(1-\lambda_N)\nu_N + \lambda_N \mathbb{P} \in B_\varepsilon^p(\mathbb{P})$ on a set of probability at least $1-\beta_N$. Furthermore $(1-\lambda_N)C_N \le C - \delta_N$ and thus

$$(1-\lambda_N)C_N + \lambda_N(C-\delta) \le C - \delta_N + \lambda_N(C-\delta) < C - \delta_N + \delta_N = C$$

for all $N \in \mathbb{N}$. Then for $\tilde{\mathbb{Q}}_N := (1-\lambda_N)\mathbb{Q}_N + \lambda_N \mathbb{Q}$ we have

$$\frac{d\tilde{\mathbb{Q}}_N}{d((1-\lambda_N)\nu_N + \lambda_N \mathbb{P})} = \frac{d((1-\lambda_N)\mathbb{Q}_N + \lambda_N \mathbb{Q})}{d((1-\lambda_N)\nu_N + \lambda_N \mathbb{P})} < C$$

and, as $\mathbb{Q}_N, \mathbb{Q} \in \mathcal{M}$ and for any $\tilde{\mathbb{Q}} \in \mathcal{M}$

$$\mathbb{E}_{\tilde{\mathbb{Q}}}[|r|] \le \sqrt{d}\,\mathbb{E}_{\tilde{\mathbb{Q}}}[|\max_{1\le i\le d} r_i|] \le \sqrt{d} \sum_{i=1}^d \mathbb{E}_{\tilde{\mathbb{Q}}}[|r_i|] = \sqrt{d} \sum_{i=1}^d \mathbb{E}_{\tilde{\mathbb{Q}}}[r_i] = d^{3/2},$$

there exists $M > 0$ such that

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_N}[g] &\le (1-\lambda_N)\mathbb{E}_{\mathbb{Q}_N}[g] + \lambda_N \mathbb{E}_{\mathbb{Q}}[g] + \lambda_N |\mathbb{E}_{\mathbb{Q}_N}[g] - \mathbb{E}_{\mathbb{Q}}[g]| \\ &\le \sup_{\nu \in B_\varepsilon^p(\mathbb{P}),\ \|d\mathbb{Q}/d\nu\|_\infty < C} \mathbb{E}_{\mathbb{Q}}[g] + \lambda_N M,\end{aligned}$$

where the last inequality holds with $\mathbb{P}^\infty$-probability at least $1-\beta_N$.
For the "$\ge$"-inequality we proceed similarly to the proof of Theorem 4.3.6 and, if needed, also use the above ideas. Specifically, we fix M and take $\mathbb{Q}_M \in \mathcal{M}$ such that $\|d\mathbb{Q}_M/d\nu\|_\infty \le C$ for some $\nu \in B_\varepsilon^p(\mathbb{P})$ and

$$\sup_{\nu \in B_\varepsilon^p(\mathbb{P}),\ \|d\mathbb{Q}/d\nu\|_\infty \le C} \mathbb{E}_{\mathbb{Q}}[g] \le \mathbb{E}_{\mathbb{Q}_M}[g] + \frac{1}{M}.$$

We then define $\tilde{\mathbb{Q}}_N := (1-\lambda_N)\mathbb{Q}_M + \lambda_N \mathbb{Q}$, where $\lambda_N := |C - C_N|/\delta$. Then $(1-\lambda_N)\nu + \lambda_N \mathbb{P} \in B^p_{\varepsilon+\varepsilon_N}(\hat{\mathbb{P}}_N)$ on a set of $\mathbb{P}^\infty$-probability at least $1-\beta_N$ and $d\tilde{\mathbb{Q}}/d((1-\lambda_N)\nu + \lambda_N \mathbb{P}) \le C_N$. Taking $N \to \infty$ and then $M \to \infty$ concludes the proof. □

We recall here the continuity property of $AV@R^{\mathbb{P}}_{1/k_N}(g)$:

Lemma 4.11.1 (Pichler [2013], Cor. 11, p.538)**.** *For $g \in \mathcal{L}^1$ and $\mathbb{P}, \tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ we have*

$$\left| AV@R^{\mathbb{P}}_{1/k_N}(g) - AV@R^{\tilde{\mathbb{P}}}_{1/k_N}(g) \right| \le k_N \mathcal{W}^1(\mathbb{P}, \tilde{\mathbb{P}}).$$

This is used for the subsequent proof.

Proof of Corollary 4.3.11. Recall that

$$\pi_{\hat{\mathcal{Q}}_N}(g) = \sup_{\tilde{\mathbb{P}} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} \sup_{\|d\nu/d\tilde{\mathbb{P}}\|_\infty \le k_N} \inf_{H \in \mathbb{R}^d} \mathbb{E}_\nu \left[g - H(r-1) \right].$$

We now want to interchange the two suprema and the infimum above, which can be done by the same arguments as in [Bartl, 2019, Proofs of Theorem 3.1 and Lemma 3.2]. Indeed, as NA($\mathbb{P}$) holds and as $\mathbb{P} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)$ with high probability we conclude by [Föllmer and Schied, 2004, Theorem 1.48, p.29] that there exists $N_0 \in \mathbb{N}$ such that

$$1 \in \mathrm{ri}(\{\mathbb{E}_\nu = [r] \mid \exists \tilde{\mathbb{P}} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N) \text{ s.t. } \|d\nu/d\tilde{\mathbb{P}}\|_\infty \le k_N\})$$

for all $N \ge N_0$, where $\mathrm{ri}(A)$ denotes the relative interior of the convex hull of a set $A \in \mathcal{B}(\mathbb{R}^d_+)$. This implies

$$\begin{aligned}\pi_{\hat{\mathcal{Q}}_N}(g) &= \inf_{H\in\mathbb{R}^d} \sup_{\tilde{\mathbb{P}}\in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} \sup_{\|d\nu/d\tilde{\mathbb{P}}\|_\infty \le k_N} \mathbb{E}_\nu\left[g - H(r-1)\right] \\ &= \inf_{H\in\mathbb{R}^d} \sup_{\tilde{\mathbb{P}}\in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)} AV@R^{\tilde{\mathbb{P}}}_{1/k_N}\left[g - H(r-1)\right].\end{aligned}$$

The first inequality in (4.3.8) is trivial, while the second inequality follows from 2-Lipschitz-continuity of $g(r) - H(r-1)$ for $|H| \le 1$ and continuity of $\tilde{\mathbb{P}} \mapsto AV@R^{\tilde{\mathbb{P}}}_{1/k_N}(g - H(r-1))$ w.r.t. to $\mathcal{W}^1$, see Pichler [2013] or Lemma 4.11.1. □

Proof of Theorem 4.3.12. Clearly

$$\begin{aligned}&\lim_{N\to\infty} \sup_{\mathbb{Q}\in\mathcal{M}} \left(\mathbb{E}_\mathbb{Q}[g] - C_N\left(\inf_{\hat{\mathbb{Q}}\sim\hat{\mathbb{P}}_N,\ \hat{\mathbb{Q}}\in\mathcal{M}} \left\|\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right\|_\infty - 1\right)\right) \\ &\ge \lim_{N\to\infty} \sup_{\mathbb{Q}\sim\hat{\mathbb{P}}_N,\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}[g] = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}[g] \quad \mathbb{P}^\infty\text{-a.s.}\end{aligned}$$

It remains to establish the reverse inequality. Let us first assume $C_N = C$. Fix $\mathbb{Q} \in \mathcal{M}$ and consider $\hat{\mathbb{Q}} \sim \hat{\mathbb{P}}_N$, $\hat{\mathbb{Q}} \in \mathcal{M}$ such that $\hat{\mathbb{Q}} \ll \mathbb{Q}$. Define $\varepsilon \ge 0$ through

$$\mathbb{Q}(r \notin \mathrm{supp}(\hat{\mathbb{P}}_N)) + \sum_{i=1}^N (\mathbb{Q}(r_i) - \hat{\mathbb{Q}}(r_i))^+ = \varepsilon.$$

Note that we can write $\mathbb{Q} = f(r)\hat{\mathbb{Q}} + \nu$ for some Radon-Nykodym derivative f and some measure ν singular to $\hat{\mathbb{Q}}$, where $\nu(\mathbb{R}^d_+) \le \varepsilon$. Let $J := \{i \in \{0, \dots, N\} \mid \hat{\mathbb{Q}}(r_i) \ge \mathbb{Q}(r_i)\}$. Then we must have

$$\sum_{i\in J} \hat{\mathbb{Q}}(r_i) - \mathbb{Q}(r_i) = \varepsilon.$$

In particular

$$\left\|\frac{1}{f}\right\|_\infty = \left\|\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right\|_\infty \ge \frac{1}{1-\varepsilon},$$

because otherwise

$$\begin{aligned}\sum_{i\in J}\hat{\mathbb{Q}}(r_i)-\mathbb{Q}(r_i) &< \frac{\varepsilon}{1-\varepsilon}\sum_{i\in J}\mathbb{Q}(r_i)\\ &= \frac{\varepsilon}{1-\varepsilon}\left(1-\mathbb{Q}(r\notin \operatorname{supp}(\hat{\mathbb{P}}_N))-\sum_{i=1}^{N}(\mathbb{Q}(r_i)-\hat{\mathbb{Q}}(r_i))^+-\sum_{i\in\{1,\dots,N\}\setminus J}\hat{\mathbb{Q}}(r_i)\right)\\ &\leq \varepsilon,\end{aligned}$$

a contradiction. Furthermore, by the definition of ε, $\mathbb{E}_{\mathbb{Q}}[g]-\mathbb{E}_{\hat{\mathbb{Q}}}[g]\leq C\varepsilon$. Thus

$$\left(\mathbb{E}_{\mathbb{Q}}[g]-C\left(\left\|\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right\|_\infty-1\right)\right)-\mathbb{E}_{\hat{\mathbb{Q}}}[g]\leq C\left(\varepsilon-\frac{1}{1-\varepsilon}+1\right)=\frac{-C\varepsilon^2}{1-\varepsilon}\leq 0,$$

so there is no gain from shifting mass and in particular

$$\sup_{\hat{\mathbb{Q}}\sim\hat{\mathbb{P}}_N,\ \hat{\mathbb{Q}}\in\mathcal{M}}\left(\mathbb{E}_{\mathbb{Q}}[g]-C\left(\left\|\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right\|_\infty-1\right)\right)\leq\sup_{\hat{\mathbb{Q}}\sim\hat{\mathbb{P}}_N,\ \hat{\mathbb{Q}}\in\mathcal{M}}\mathbb{E}_{\hat{\mathbb{Q}}}[g].$$

For $C_N\to C$ we have

$$\begin{aligned}\left(\mathbb{E}_{\mathbb{Q}}[g]-C_N\left(\left\|\frac{d\hat{\mathbb{Q}}}{d\mathbb{Q}}\right\|_\infty-1\right)\right)-\mathbb{E}_{\hat{\mathbb{Q}}}[g] &\leq C\varepsilon-C_N\left(\frac{1}{1-\varepsilon}-1\right)\\ &= C\varepsilon-C_N\left(\frac{\varepsilon}{1-\varepsilon}\right),\end{aligned}$$

which is non-negative for $\varepsilon\leq 1-C_N/C$. Finally, the case when $\hat{\mathbb{Q}}$ is not absolutely continuous with respect to $\mathbb{Q}$ is trivial since then $\|d\hat{\mathbb{Q}}/d\mathbb{Q}\|_\infty=\infty$. □

4.11.1 Proof of Theorem 4.3.6

For simplicity of exposition we first prove Theorem 4.3.6 under Assumption 4.3.5.2. We adopt the notation of Section 4.3.

Lemma 4.11.2 (Fournier and Guillin [2015] Theorem 2)**.** *Under Assumption 4.3.5.2 we have*

$$\mathbb{P}^\infty(\mathcal{W}^p(\mathbb{P},\hat{\mathbb{P}}_N)\geq\varepsilon)\leq\begin{cases}c_1\exp(-c_2N\varepsilon^{\min(\max(d/p,2),a/(2p))}) & \textit{if } \varepsilon\leq 1,\\ c_1\exp(-c_2N\varepsilon^{a/(2p)}) & \textit{if } \varepsilon>1\end{cases}$$

for $N\geq 1$, $d\neq 2p$ and $\varepsilon>0$, where c_1,c_2 are positive constants that only depend on p, d, a, c and $\mathbb{E}_{\mathbb{P}}[\exp(c|r|^a)]$. Thus for some confidence level $\beta\in(0,1)$ we can choose

$$\varepsilon_N(\beta):=\begin{cases}\left(\frac{\log(c_1\beta^{-1})}{c_2N}\right)^{1/\min(\max(d/p,2),a/(2p))} & \textit{if } N\geq\frac{\log(c_1\beta^{-1})}{c_2},\\ \left(\frac{\log(c_1\beta^{-1})}{c_2N}\right)^{(2p)/a} & \textit{if } N<\frac{\log(c_1\beta^{-1})}{c_2}\end{cases}\tag{4.11.1}$$

which yields $\mathbb{P}^\infty(\mathcal{W}^p(\mathbb{P},\hat{\mathbb{P}}_N)\geq\varepsilon_N(\beta))\leq\beta$.

Lemma 4.11.3. *Fix $N \in \mathbb{N}$ and $p \geq 1$. Let $\mathbb{Q}_n \in D^p_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$ such that $\mathbb{Q}_n \Rightarrow \mathbb{Q} \in \mathcal{P}(\mathbb{R}^d_+)$ for $n \to \infty$. Then $|E_{\mathbb{Q}}[r-1]| \leq K k_N \varepsilon_N(\beta_N)$ for some $K > 0$. $D^p_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$ is weakly compact for $p > 1$. In general $D^1_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$ is not weakly closed.*

Proof. Take a sequence $\mathbb{Q}_n \in D^p_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$ such that $\mathbb{Q}_n \Rightarrow \mathbb{Q}$. For every $n \in \mathbb{N}$ there exists $\nu_n \in B^p_{\varepsilon_N(\beta_N)}(\mathbb{P})$ such that $\|d\mathbb{Q}_n/d\nu_n\|_\infty \leq k_N$. First observe that for any n we have

$$\mathbb{E}_{\mathbb{Q}_n}[r \wedge K] \leq k_N \left(\mathbb{E}_{\mathbb{P}}[r \wedge K] + \varepsilon_N(\beta_N)\right) \leq k_N \left(\mathbb{E}_{\mathbb{P}}[r] + \varepsilon_N(\beta_N)\right) < \infty. \tag{4.11.2}$$

It follows, by weak convergence and monotone convergence theorem, that

$$\mathbb{E}_{\mathbb{Q}}[r] \leq k_N \left(\mathbb{E}_{\mathbb{P}}[r] + \varepsilon_N(\beta_N)\right) < \infty.$$

Next, we show $|\mathbb{E}_{\mathbb{Q}}[r-1]| \leq \text{const} \cdot k_N \varepsilon_N(\beta_N)$. As

$$|\mathbb{E}_{\mathbb{Q}}[r-1]| \leq \sqrt{d} \max_{1 \leq i \leq d} |\mathbb{E}_{\mathbb{Q}}[r_i - 1]| \leq \sqrt{d} \sum_{i=1}^{d} |\mathbb{E}_{\mathbb{Q}}[r_i - 1]|$$

it is enough to consider the case $d = 1$. Then

$$\begin{aligned} |\mathbb{E}_{\mathbb{Q}}[r-1]| \leq & \left|\mathbb{E}_{\mathbb{Q}}[(r-1) \wedge K] - \mathbb{E}_{\mathbb{Q}_n}[(r-1) \wedge K]\right| \\ & + \left|\mathbb{E}_{\mathbb{Q}}\left[(r-1-K)\mathbb{1}_{\{r \geq K+1\}}\right] - \mathbb{E}_{\mathbb{Q}_n}\left[(r-1-K)\mathbb{1}_{\{r \geq K+1\}}\right]\right| \end{aligned} \tag{4.11.3}$$

Consider now the terms on the RHS. The third term can be made arbitrarily small by taking large K since $\mathbb{Q}$ admits first moment. The fourth term can be bounded, in analogy to (4.11.2), as follows:

$$\mathbb{E}_{\mathbb{Q}_n}[(r-K)\mathbb{1}_{\{r \geq K\}}] \leq k_N \mathbb{E}_{\mathbb{P}}[(r-K)\mathbb{1}_{\{r \geq K\}}] + k_N \varepsilon_N(\beta_N) \leq 2k_N \varepsilon_N(\beta_N),$$

where we took K large enough. Finally, for a fixed K, the difference between the first two terms can be made small by taking n large due to weak convergence of measures. The bound $|\mathbb{E}_{\mathbb{Q}}[r-1]| \leq \text{const} \cdot k_N \varepsilon_N(\beta_N)$ follows.

If $p > 1$, the ball $B^p_{\varepsilon_N}(\mathbb{P})$ is weakly compact (see [Villani, 2008, Def. 6.8., p.96]), in particular uniformly integrable, and it follows that also the fourth term on the RHS of (4.11.3) converges to zero, uniformly in n, as $K \to \infty$

$$\mathbb{E}_{\mathbb{Q}_n}[(r-1-K)\mathbb{1}_{\{r \geq K+1\}}] \leq k_N \limsup_{K \to \infty} \sup_n \mathbb{E}_{\nu_n}[r \mathbb{1}_{r \geq K}] = 0.$$

Further, possibly on a subsequence, $(\nu_n)_{n \in \mathbb{N}}$ converges weakly to a limit $\nu \in B^p_{\varepsilon_N}(\mathbb{P})$. By regularity of probability measures it is sufficient to test $d\mathbb{Q}/d\nu$ against bounded continuous functions and we easily conclude that $\|d\mathbb{Q}/d\nu\|_\infty \leq k_N$. This shows compactness of $D^p_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$.

Consider now $p = 1$. We give the following counterexample: take $\mathbb{P} = \delta_1$ and set for $r_n \geq 2$

$$\nu_n = \frac{\varepsilon_N(\beta_N)}{2}\ \delta_0 + \left(1 - \frac{r_n\ \varepsilon_N(\beta_N)}{2(r_n-1)}\right)\ \delta_1 + \frac{\varepsilon_N(\beta_N)}{2(r_n-1)}\ \delta_{r_n}.$$

Furthermore let

$$\mathbb{Q}_n = \frac{\varepsilon_N(\beta_N)}{2\sqrt{\varepsilon_N(\beta_N)}}\ \delta_0 + \left(1 - \frac{r_n \varepsilon_N(\beta_N)}{2(r_n-1)\sqrt{\varepsilon_N(\beta_N)}}\right)\ \delta_1 + \frac{\varepsilon_N(\beta_N)}{2(r_n-1)\sqrt{\varepsilon_N(\beta_N)}}\ \delta_{r_n}.$$

Then $\mathcal{W}^1(\nu_n, \delta_1) \leq \varepsilon_N(\beta_N)$ and

$$\left\| \frac{d\mathbb{Q}_N}{d\nu_N} \right\|_\infty \leq \frac{1}{\sqrt{\varepsilon_N(\beta_N)}}.$$

Assume $\lim_{n\to\infty} r_n = \infty$. Then

$$\mathbb{Q}_n \Rightarrow \frac{\sqrt{\varepsilon_N(\beta_N)}}{2}\ \delta_0 + \left(1 - \frac{\sqrt{\varepsilon_N(\beta_N)}}{2}\right)\ \delta_1 \notin \mathcal{M}.$$

□

Taking the closure of $D^1_{\varepsilon_N(\beta_N),k_N}(\mathbb{P})$ would ensure compactness, but consistency of the estimator $\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_\mathbb{Q}[g]$ in Theorem 4.3.6 would be lost in general since the closure might include non-martingale measures. To see this, take for instance $g(r) = (r-1)$ in the above example.

Proof of Theorem 4.3.6 under Assumption 4.3.5.2. Let us assume that Assumption 4.3.5.2 is satisfied. Note that the "$\geq$"-inequality follows from Lemma 4.11.4. Indeed, Lemma 4.11.4 implies that for all $N \in \mathbb{N}$ there exists a martingale measure $\mathbb{Q}_N \sim \mathbb{P}$ with $\|d\mathbb{Q}_N/d\mathbb{P}\|_\infty \leq k_N$ such that

$$\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}[g] \leq \mathbb{E}_{\mathbb{Q}_N}[g] + a_N,$$

with $a_N \to 0$ as $N \to \infty$. With $\mathbb{P}^\infty$-probability $(1-\beta_N)$ we have $\mathbb{P} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)$ and hence $\mathbb{Q}_N \in \hat{\mathcal{Q}}_N$. This gives

$$\mathbb{P}^\infty\left(\sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}[g] - a_N \leq \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{E}_\mathbb{Q}[g]\right) \geq 1 - \beta_N, \quad N \geq 1.$$

We recall that NA($\mathbb{P}$) gives $\pi^\mathbb{P}(g) = \sup_{\mathbb{Q}\sim\mathbb{P},\ \mathbb{Q}\in\mathcal{M}} \mathbb{E}_\mathbb{Q}[g]$ and hence, for any $\varepsilon > 0$ and N large enough, $\mathbb{P}^\infty(\pi_{\hat{\mathcal{Q}}_N}(g) - \pi^\mathbb{P}(g) \leq -\varepsilon) \leq \beta_N \to 0$ as $N \to \infty$.

For the "$\leq$"-inequality, we assume g is Lipschitz continuous bounded from below and we take $H \in \mathbb{R}^d$ such that

$$\pi^\mathbb{P}(g) + H(r-1) \geq g \quad \mathbb{P}\text{-a.s.}$$

Take a sequence $\mathbb{Q}_N \in \hat{\mathcal{Q}}_N$ with $\mathbb{E}_{\mathbb{Q}_N}[g] \geq \pi_{\hat{\mathcal{Q}}_N}(g) - a_N$. By definition, there exist $\nu_N \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)$ such that $\|d\mathbb{Q}_N/d\nu_N\|_\infty \leq k_N$. In particular, with $\mathbb{P}^\infty$-probability $(1-\beta_N)$ we have $\nu_N \in B^p_{2\varepsilon_N(\beta_N)}(\mathbb{P})$. Let us define

$$A = \{\pi^{\mathbb{P}}(g) + H(r-1) - g \geq 0\}.$$

Then

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}_N}[g] &\leq \mathbb{E}_{\mathbb{Q}_N}[(\pi^{\mathbb{P}}(g) + H(r-1))\mathbb{1}_A + g\mathbb{1}_{A^c}] \\ &= \pi^{\mathbb{P}}(g) + \mathbb{E}_{\mathbb{Q}_N}[H(r-1)] + \mathbb{E}_{\mathbb{Q}_N}\left[(g - H(r-1) - \pi^{\mathbb{P}}(g))\mathbb{1}_{A^c}\right] \end{aligned} \tag{4.11.4}$$

and the second term on the RHS vanishes since $\mathbb{Q}_N$ is a martingale measure. To treat the last term on the RHS consider the function

$$\tilde{g} := (g - H(r-1) - \pi^{\mathbb{P}}(g)) \vee 0$$

which is non-negative, C-Lipschitz for some $C > 0$ and $\{\tilde{g} > 0\} = A^c$. Since $\mathbb{P}(A^c) = 0$ we have in particular

$$\left|\int_{A^c} \tilde{g} d\nu_N\right| = \left|\int_{A^c} \tilde{g} d\nu_N - \int_{A^c} \tilde{g} d\mathbb{P}\right| = \left|\int \tilde{g} d\nu_N - \int \tilde{g} d\mathbb{P}\right|$$

which by the Kantorovich-Rubinstein duality (4.3.3) is dominated by $C\mathcal{W}^1(\nu_N, \mathbb{P}) \leq C\mathcal{W}^p(\nu_N, \mathbb{P})$. We conclude that, for any $\varepsilon > 0$,

$$\mathbb{P}^\infty\left(\pi_{\hat{\mathcal{Q}}_N}(g) - \pi^{\mathbb{P}}(g) \geq \varepsilon\right) \leq \mathbb{P}^\infty\left(a_N + Ck_N\mathcal{W}^p(\nu_N, \mathbb{P}) \geq \varepsilon\right) \leq \beta_N$$

for N large enough since $\varepsilon_N k_N \to 0$. This establishes the convergence of $\pi_{\hat{\mathcal{Q}}_N}(g)$ to $\pi^{\mathbb{P}}(g)$ in $\mathbb{P}^\infty$-probability. Further, whenever $\sum_{N=1}^\infty \beta_N < \infty$, a simple application of Borel-Cantelli lemma, similarly as in [Esfahani and Kuhn, 2018, Lemma 3.7], shows that the convergence holds $\mathbb{P}^\infty$-a.s. This concludes the proof in the case of Lipschitz continuous g bounded from below and under Assumption 4.3.5.2.

It remains to argue the "$\leq$"-inequality when g is bounded and continuous. We fix a small $\delta > 0$ and define

$$\tilde{\mathcal{Q}}_N := \left\{\mathbb{Q} \in \mathcal{P}(\mathbb{R}^d_+) \mid \exists \tilde{\mathbb{P}} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N) \text{ such that } \|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \leq k_N/(1-\delta)\right\}.$$

and set

$$\pi_{\tilde{\mathcal{Q}}_N, [0,K]^d}(g) := \sup_{\mathbb{Q} \in \tilde{\mathcal{Q}}_N, \mathbb{Q} \in \mathcal{M},\ \mathrm{supp}(\mathbb{Q}) \subseteq [0,K]^d} \mathbb{E}_{\mathbb{Q}}(g).$$

Similarly to Corollary 4.3.11 we see that for $K > 1$ large enough

$$\pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) = \inf \left\{ x \in \mathbb{R} \mid \exists H \in \mathbb{R}^d \text{ s.t.} \sup_{\mathbb{Q} \in \tilde{\mathcal{Q}}_N,\ \mathrm{supp}(\mathbb{Q}) \subseteq [0,K]^d} \mathbb{E}_{\mathbb{Q}}[g(r) - H(r-1) - x] \le 0 \right\}.$$

Thus there exists a sequence $H_N \in \mathbb{R}^d$ such that

$$\sup_{\mathbb{Q} \in \tilde{\mathcal{Q}}_N,\ \mathrm{supp}(\mathbb{Q}) \subseteq [0,K]^d} \mathbb{E}_{\mathbb{Q}}[g(r) - H_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g)] \le 1/N \tag{4.11.5}$$

for all $N \in \mathbb{N}$. Take K large enough so that $d/K < \delta$. For notational simplicity we assume that $\mathbb{P}$ has full support. Recall that g is bounded, $|g| \le C$. We now show that H_N^i is bounded. Let us first look at the lower bound: for this, we take $i \in \{1, \dots, d\}$ and suppose that $H_N^i \le 0$ (otherwise we trivially bound H_N^i from below by 0). We work on the set $\{\mathbb{P} \in B^p_{\varepsilon_N(\beta_N)}(\hat{\mathbb{P}}_N)\}$, which has $\mathbb{P}^\infty$-probability at least $1 - \beta_N$. Now we take N large enough such that

$$k_N \cdot \mathbb{P}\left(s_j r_j \ge 1 \text{ for all } j \ne i,\ K/2 < r_i < K\right) \ge 1$$

for all $(s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_d) \in \{-1,1\}^{d-1}$. Then defining $s_j = -\mathrm{sign}(H_N^j)$ for all $j \ne i$ and

$$\frac{d\mathbb{Q}^{s_1,\dots,s_{i-1},s_{i+1},s_d}}{d\mathbb{P}} := \frac{\mathbb{1}_{\{s_j r_j \ge 1 \text{ for all } j \ne i,\ K/2 < r_i < K\}}}{\mathbb{P}(s_j r_j \ge 1 \text{ for all } j \ne i,\ K/2 < r_i < K)}$$

we have by (4.11.5)

$$-H_N^i(K/2 - 1) \le -H_N \mathbb{E}_{\mathbb{Q}}[r-1] \le \frac{1}{N} - \mathbb{E}_{\mathbb{Q}}[g(r)] + \pi_{\hat{\mathcal{Q}}_N}(g) \le 2C + 1$$

and thus $H_N^i \ge -(2(C+1))/(K/2-1)$. On the other hand, assuming $H_N^i \ge 0$ and setting

$$\frac{d\mathbb{Q}^{s_1,\dots,s_{i-1},s_{i+1},s_d}}{d\mathbb{P}} := \frac{\mathbb{1}_{\{s_j r_j \ge 1 \text{ for all } j \ne i,\ 0 < r_i < 1/2\}}}{\mathbb{P}(s_j r_j \ge 1 \text{ for all } j \ne i,\ 0 < r_i < 1/2)}$$

we got by the same arguments as above

$$H_N^i/2 \le -H_N \mathbb{E}_{\mathbb{Q}}[r-1] \le \frac{1}{N} - \mathbb{E}_{\mathbb{Q}}[g(r)] + \pi_{\hat{\mathcal{Q}}_N}(g) \le 2C + 1$$

for a possibly larger N. In conclusion thus $-(2(C+1))/(K/2-1) \le H_N^i \le 4C+2$. Now we set

$$\tilde{H}_N^i := (C - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g))/(K-1) \vee H_N^i, \qquad i \in \{1, \dots, d\}.$$

Obviously $\tilde{H}_N^i \ge 0$ and

$$\pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) + \tilde{H}_N^i(r^i - 1) \ge C \tag{4.11.6}$$

for $r^i \geq K$. Furthermore as $(C - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g))/(K-1) \leq 2C/(K-1)$ and $H_N^i \geq -2(C+1)/(K/2-1)$ we have for $r^i \leq 1$

$$0 \geq (\tilde{H}_N^i - H_N^i)(r^i - 1) \geq -(\tilde{H}_N^i - H_N^i) \geq -4(C+1)/(K/2-1). \tag{4.11.7}$$

Consider now $\mathbb{Q}_N \in \hat{\mathcal{Q}}_N$. Note that by Markov's inequality we have

$$\mathbb{Q}_N([0,K]^d) \geq 1 - \frac{d}{K} \geq 1 - \delta.$$

Then

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_N}\left[g(r) - \tilde{H}_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) - 4d(C+1)/(K/2-1) - d^2(4C+2)/K\right] \\
&\leq \mathbb{E}_{\mathbb{Q}_N}\left[\left(g(r) - \tilde{H}_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) - 4d(C+1)/(K/2-1)\right)\mathbb{1}_{[0,K]^d}\right] \\
&\qquad + \mathbb{E}_{\mathbb{Q}_N}\left[\left(g(r) - \tilde{H}_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) - d^2(4C+2)/K\right)\mathbb{1}_{([0,K]^d)^c}\right] \\
&\leq \mathbb{E}_{\mathbb{Q}_N}\left[\left(g(r) - H_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g)\right)\mathbb{1}_{[0,K]^d}\right] \\
&\leq \sup_{\mathbb{Q}\in\tilde{\mathcal{Q}}_N,\ \mathrm{supp}(\mathbb{Q})\subseteq[0,K]^d} \mathbb{E}_{\mathbb{Q}}\left[g(r) - H_N(r-1) - \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g)\right] \leq \frac{1}{N}.
\end{aligned}$$

where we used (4.11.6) and (4.11.7) as well as $|H_N^j|\mathbb{Q}_N(([0,K]^d)^c) \leq d(4C+2)/K$ for $j \neq i$ in the second inequality. In the last inequality we used (4.11.5). By definition of $\pi_{\hat{\mathcal{Q}}_N}(g)$, and since $\mathbb{E}_{\mathbb{Q}_N}[\tilde{H}_N(r-1)] = 0$, this shows

$$\pi_{\hat{\mathcal{Q}}_N}(g) \leq \frac{4d(C+1)}{K/2-1} + \frac{d^2(4C+2)}{K} + \frac{1}{N} + \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g).$$

We obtain the same result in the case of $\mathbb{P}$ without full support with a possible addition of the term constant times $\varepsilon_N k_N$ on the RHS. Note that we either have $r_i = 1$ $\mathbb{P}$-a.s., or else by NA(P), $\mathbb{P}$ puts mass on r_i on either side of 1. If r_i is bounded under $\mathbb{P}$, we can take $\tilde{H}_N^i = H_N^i$ and the arguments remain the same but we have the additional error term as $\mathbb{Q}_N$ can put small mass on unbounded r_i.

Lastly we fix $\varepsilon > 0$, $K > 1$ such that

$$\frac{4d(C+1)}{K/2-1} + \frac{d^2(4C+2)}{K} \leq \varepsilon$$

and take a sequence $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ which satisfies $\pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) \leq \varepsilon + \mathbb{E}_{\mathbb{Q}_N}[g]$ for all $N \in \mathbb{N}$. Note that all $(\mathbb{Q}_N)_{N\in\mathbb{N}}$ are martingale measures supported on $[0,K]^d$. As g is uniformly continuous on $[0,K+1]^d$, there exists a constant $L > 0$ such that $|g(r) - g(\tilde{r})| \leq \varepsilon + L|r - \tilde{r}|$ for all $r, \tilde{r} \in [0,K+1]^d$. Define

$$\hat{g}(r) := \sup_{u\in[0,K]^d} (g(u) - (L \vee 2C)|u - r| - \varepsilon)$$

and note that $g(r)-\varepsilon \le \hat{g}(r)$ on $[0,K]^d$ as well as $\hat{g}(r) \le g(r)$ on $[0,K+1]^d$ and $\hat{g}(r) \le -C$ on $([0,K+1]^d)^c$. In conclusion $\hat{g}(r) \le g(r)$. For fixed $u \in [0,K]^d$ we write

$$g(u) - (L \vee 2C)|u-r| - \varepsilon \le g(u) - (L \vee 2C)|u-r'| - \varepsilon + (L \vee 2C)|r-r'|$$

Taking suprema over $u \in [0,K]^d$ on both sides we conclude

$$\hat{g}(r) \le \hat{g}(r') + (L \vee 2C)|r-r'|.$$

Exchanging the roles of r and r' finally yields $|\hat{g}(r) - \hat{g}(r')| \le (L \vee 2C)|r-r'|$ for $r, r' \in \mathbb{R}^d_+$. Thus we can argue as in the proof of Theorem 4.3.6 for Lipschitz continuous claims (and using the same notation) to obtain

$$\begin{aligned}
\pi_{\hat{\mathcal{Q}}_N}(g) - \pi^{\mathbb{P}}(g) &\le \pi_{\hat{\mathcal{Q}}_N,[0,K]^d}(g) - \pi^{\mathbb{P}}(g) + \varepsilon + \frac{1}{N} \\
&\le \mathbb{E}_{\mathbb{Q}_N}[g] - \pi^{\mathbb{P}}(g) + 2\varepsilon + \frac{1}{N} \\
&\le \mathbb{E}_{\mathbb{Q}_N}[\hat{g}] - \pi^{\mathbb{P}}(g) + 3\varepsilon + \frac{1}{N} \\
&\le \pi^{\mathbb{P}}(\hat{g}) - \pi^{\mathbb{P}}(g) + 3\varepsilon + \hat{C}k_N \mathcal{W}^p(\hat{\mathbb{P}}_N, \mathbb{P}) + \frac{1}{N} \\
&\le \pi^{\mathbb{P}}(g) - \pi^{\mathbb{P}}(g) + 3\varepsilon + \hat{C}k_N \mathcal{W}^p(\hat{\mathbb{P}}_N, \mathbb{P}) + \frac{1}{N}.
\end{aligned}$$

for some $\hat{C} > 0$. In particular

$$\mathbb{P}^\infty\left(\pi_{\hat{\mathcal{Q}}_N}(g) - \pi^{\mathbb{P}}(g) \ge \varepsilon\right) \le \mathbb{P}^\infty\left(3\varepsilon + \hat{C}k_N \mathcal{W}^p(\hat{\mathbb{P}}_N, \mathbb{P}) + \frac{1}{N} \ge 4\varepsilon\right) \le \beta_N.$$

As ε was arbitrary and $\varepsilon_N k_N \to 0$ the claim follows. □

We have used the following lemma:

Lemma 4.11.4 (Rásonyi [2002], Cor. 3.3)**.** *For a measurable function g bounded from below*

$$\sup_{\|\frac{d\mathbb{Q}}{d\mathbb{P}}\|_\infty < \infty,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] = \sup_{\mathbb{Q} \sim \mathbb{P},\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g].$$

Before we finish the proof of Theorem 4.3.6 we first give the convergence rates for $\hat{\mathbb{P}}_N$ under Assumption 4.3.5.1.[5] This is a slight modification of the result in Fournier and Guillin [2015] but the proof is essentially the same as that of Theorem 15 therein and is hence omitted.

[5] According to [Fournier and Guillin, 2015, Comments after Theorems 14 & 15] the L^2-L^2-decay property stated in [Fournier and Guillin, 2015, Theorem 15] is actually too strong as only functions bounded by one are considered, as given in our assumptions.

Lemma 4.11.5 (Fournier and Guillin [2015], Proof of Theorem 15)**.** *Under Assumption 4.3.5.1 there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mathbb{P}}[\mathcal{W}^p(\mathbb{P}, \hat{\mathbb{P}}_N)^p] \leq \kappa_N$$
$$:= C \begin{cases} N^{-1/2} + N^{-(q_s-p)/q_s} & \text{if } p > d_s/(2s) \text{ and } q_s \neq 2p, \\ N^{-1/2}\log(1+N) + N^{-(q_s-p)/q_s} & \text{if } p = d_s/(2s) \text{ and } q_s \neq 2p, \\ N^{-p/d} + N^{-(q_s-p)/q_r} & \text{if } p \in (0, d_s/2) \text{ and } q_s \neq d_s/(d_s-p) \end{cases}$$

and $q_s := q(s-2)/(2s)$ and $d_s := d(3s+2)/(2s)$.

Proof of Theorem 4.3.6 under Assumption 4.3.5.1. By Assumption 4.3.5.1 we have for $\varepsilon \geq 0$

$$\mathbb{P}(\mathcal{W}^p(\mathbb{P}, \hat{\mathbb{P}}_N) \geq \varepsilon) \leq \kappa_N/\varepsilon^p$$

using Markov's inequality, so in particular we can choose

$$\varepsilon_N(\beta) := \left(\frac{\kappa_N}{\beta}\right)^{1/p}. \tag{4.11.8}$$

The rest of the proof now follows as under Assumption 4.3.5.2. □

4.12 Additional results and proofs for Section 4.4

Before we prove Theorem 4.4.5, we recall the following result:

Lemma 4.12.1 (Shorack and Wellner [2009] Theorem 26.1, p. 828 & Ex. 26.2, p.833)**.** *Let $\mathcal{C}$ denote all closed balls in $\mathbb{R}^d_+$. Then for all $\varepsilon > 0$*

$$\lim_{N\to\infty} \sup_{\mathbb{P}\in\mathcal{P}(\mathbb{R}^d_+)} \mathbb{P}^\infty\left(\sup_{M\geq N}\sup_{\overline{B}\in\mathcal{C}} |\hat{\mathbb{P}}_M(\overline{B}) - \mathbb{P}(\overline{B})| \geq \varepsilon\right) = 0.$$

Proof of Theorem 4.4.5. Let $\varepsilon > 0$ and fix $\mathbb{P}^0 \in \mathcal{P}(\mathbb{R}^d_+)$ such that $\mathrm{NA}(\mathbb{P}^0)$ holds. As g is uniformly continuous, its $\mathbb{P}$-concave envelope operator is continuous at $\mathbb{P}_0$, see Proposition 4.2.8, and we can take δ small enough so that for all $\bar{\mathbb{P}}, \tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ with $d_H(\mathrm{supp}(\bar{\mathbb{P}}), \mathrm{supp}(\tilde{\mathbb{P}})) \leq 2\delta$ we have $|\pi^{\bar{\mathbb{P}}}(g) - \pi^{\tilde{\mathbb{P}}}(g)| \leq \varepsilon/9$.

We first argue that we can restrict to a compact set. Indeed we have

$$|\pi^{\hat{\mathbb{P}}^0_N}(g) - \pi^{\hat{\mathbb{P}}^1_N}(g)| \leq |\pi^{\hat{\mathbb{P}}^0_N}(g) - \pi^{\mathbb{P}^0}(g)| + |\pi^{\mathbb{P}^0}(g) - \pi^{\mathbb{P}^1}(g)| \\ + |\pi^{\mathbb{P}^1}(g) - \pi^{\hat{\mathbb{P}}^1_N}(g)|. \tag{4.12.1}$$

We choose $\mathbb{P}^1$ such that $\mathcal{W}^\infty(\mathbb{P}^1, \mathbb{P}^0) \leq \delta/4$, i.e. $\mathbb{P}^1(B) \leq \mathbb{P}^0(B^{\delta/4})$ and $\mathbb{P}^0(B) \leq \mathbb{P}^1(B^{\delta/4})$ for all $B \in \mathcal{B}(\mathbb{R}^d_+)$. Thus in particular

$$\mathrm{supp}(\mathbb{P}^0) \subseteq \mathrm{supp}(\mathbb{P}^1)^{\delta/2} \quad \text{and} \quad \mathrm{supp}(\mathbb{P}^1) \subseteq \mathrm{supp}(\mathbb{P}^0)^{\delta/2}.$$

Note also that, by the choice of δ, we have

$$|\pi^{\mathbb{P}^0(\cdot|[0,L]^d)}(g) - \pi^{\mathbb{P}^1(\cdot|[0,L]^d)}(g)| \le \varepsilon/9 \quad \text{and} \quad |\pi^{\mathbb{P}^0}(g) - \pi^{\mathbb{P}^1}(g)| \le \varepsilon/9 \tag{4.12.2}$$

By a monotone limit argument, see the proof of Theorem 4.2.1, we have

$$\sup_{L\in\mathbb{N}} \pi^{\mathbb{P}^0(\cdot|[0,L]^d)}(g) = \pi^{\mathbb{P}^0}(g),$$

so there for all L large enough

$$|\pi^{\mathbb{P}^0(\cdot|[0,L]^d)}(g) - \pi^{\mathbb{P}^0}(g)| \le \varepsilon/9.$$

Fix such L and set $K := [0,L]^d$. Note that by (4.12.2) we now have

$$\begin{aligned}\pi^{\mathbb{P}^1}(g) - \pi^{\mathbb{P}^1(\cdot|K)}(g) &\le \pi^{\mathbb{P}^1}(g) - \pi^{\mathbb{P}^0}(g) + \pi^{\mathbb{P}^0}(g) - \pi^{\mathbb{P}^0(\cdot|K)}(g)\\ &\quad + \pi^{\mathbb{P}^0(\cdot|K)}(g) - \pi^{\mathbb{P}^1(\cdot|K)}(g) \le \varepsilon/9 + \varepsilon/9 + \varepsilon/9 = \varepsilon/3.\end{aligned}$$

In particular for $i = 0, 1$

$$\begin{aligned}0 \le \pi^{\mathbb{P}^i}(g) - \pi^{\hat{\mathbb{P}}^i_N}(g) &= \pi^{\mathbb{P}^i}(g) - \pi^{\mathbb{P}^i(\cdot|K)}(g) + \pi^{\mathbb{P}^i(\cdot|K)}(g) - \pi^{\hat{\mathbb{P}}^i_N}(g)\\ &\le \varepsilon/3 + \pi^{\mathbb{P}^i(\cdot|K)}(g) - \pi^{\hat{\mathbb{P}}^i_N}(g)\\ &\le \varepsilon/3 + \pi^{\mathbb{P}^i(\cdot|K)}(g) - \pi^{\hat{\mathbb{P}}^i_N(\cdot|K)}(g).\end{aligned} \tag{4.12.3}$$

We proceed to bound the difference of the last two terms on the RHS. Let us write $\mathrm{supp}(\hat{\mathbb{P}}^i_N) = \{r^i_1, \dots, r^i_N\}$, $i = 0, 1$. We now argue that

$$\mathbb{P}^\infty(d_H(\{r^i_1, \dots, r^i_N\} \cap K, \mathrm{supp}(\mathbb{P}^i) \cap K) > 2\delta) \le \varepsilon/2, \quad i = 0, 1, \tag{4.12.4}$$

for all $N \ge N_*$ for some N_* independent of $\mathbb{P}^1$. To this end, take $M \in \mathbb{N}$ and deterministic points $\tilde{r}_1, \dots, \tilde{r}_M \in \mathrm{supp}(\mathbb{P}^0) \cap K$ such that

$$\cup_{k=1}^M B_{\delta/2}(\tilde{r}_k) \supseteq \mathrm{supp}(\mathbb{P}^0)^{\delta/2} \cap K \supseteq \mathrm{supp}(\mathbb{P}^1) \cap K,$$

where $B_\delta(\tilde{r}) = \{r \in \mathbb{R}^d : |r - \tilde{r}| < \delta\}$. Then as M is finite, there exists $N_0 \in \mathbb{N}$ such that for all $N \ge N_0$

$$\mathbb{P}^\infty(\{\forall k \in \{1, \dots, M\}\ \exists j \in \{1, \dots, N\} \text{ s.t. } |\tilde{r}_k - r^0_j| \le \delta\}) \ge 1 - \varepsilon/2.$$

Set

$$\alpha := \min_{k=1,\dots,M} \mathbb{P}^0(B_{\delta/4}(\tilde{r}_k)) > 0.$$

Then $\mathbb{P}^1(B_{\delta/2}(\tilde{r}_k)) \geq \mathbb{P}^0(B_{\delta/4}(\tilde{r}_k)) \geq \alpha$ for all $k = 1, \ldots, M$. By Lemma 4.12.1 there exists $N_* \geq N_0$ such that for all $N \geq N_*$ and all $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$

$$\mathbb{P}^\infty \left(\sup_{\overline{B} \in \mathcal{C}} |\hat{\mathbb{P}}_N(\overline{B}) - \mathbb{P}(\overline{B})| \geq \alpha \right) < \varepsilon/2,$$

If there exists $k \in \{1, \ldots, M\}$ such that for all $j \in \{1, \ldots, N\}$ $|\tilde{r}_k - r^1_j| \geq \delta$ then

$$|\hat{\mathbb{P}}^1_N(\overline{B}_{\delta/2}(\tilde{r}_k)) - \mathbb{P}^1(\overline{B}_{\delta/2}(\tilde{r}_k))| = \mathbb{P}^1(\overline{B}_{\delta/2}(\tilde{r}_k) \geq \alpha,$$

in particular $\sup_{\overline{B} \in \mathcal{C}} |\hat{\mathbb{P}}^1_N(\overline{B}) - \mathbb{P}^1(\overline{B})| \geq \alpha$. Thus

$$\mathbb{P}^\infty \left(\forall k \in \{1, \ldots, M\} \ \exists j \in \{1, \ldots, N\} \text{ s.t. } |\tilde{r}_k - r^1_j| \leq \delta \} \right) \geq 1 - \varepsilon/2.$$

On the other hand, by the choice of $\{\tilde{r}_1, \ldots \tilde{r}_M\}$ for any $i \in \{0, 1\}$ and any $j \in \{1, \ldots, N\}$ with $r^i_j \in K$ there exists $k \in \{1, \ldots, M\}$ such that $|r^i_j - \tilde{r}_k| \leq \delta$. Note that $\{\tilde{r}_1, \ldots \tilde{r}_M\} \subseteq \operatorname{supp}(\mathbb{P}^0) \cap K$ and we conclude that (4.12.4) holds. It then follows from our choice of δ that for all $N \geq N_*$ we have

$$\mathbb{P}^\infty \left(\pi^{\mathbb{P}^0(\cdot|K)}(g) - \pi^{\hat{\mathbb{P}}^0_N(\cdot|K)}(g) \leq \varepsilon/9, \ \pi^{\mathbb{P}^1(\cdot|K)}(g) - \pi^{\hat{\mathbb{P}}^1_N(\cdot|K)}(g) \leq \varepsilon/9 \right) > 1 - \varepsilon.$$

Hence, by (4.12.3), we deduce that

$$\mathbb{P}^\infty \left(\pi^{\mathbb{P}^0}(g) - \pi^{\hat{\mathbb{P}}^0_N}(g) \leq 4\varepsilon/9, \ \pi^{\mathbb{P}^1}(g) - \pi^{\hat{\mathbb{P}}^1_N}(g) \leq 4\varepsilon/9 \right) > 1 - \varepsilon.$$

Combining the above with (4.12.1) and (4.12.2) we have

$$\mathbb{P}^\infty \left(\left| \pi^{\hat{\mathbb{P}}^0_N}(g) - \pi^{\hat{\mathbb{P}}^1_N}(g) \right| > \varepsilon \right) < \varepsilon.$$

Using Strassen's theorem ([Huber, 1996, Theorem 2.13, p. 30]) we deduce that

$$d_L(\mathcal{L}_{\mathbb{P}^1}(\hat{\pi}_N), \mathcal{L}_{\mathbb{P}^0}(\hat{\pi}_N)) \leq \varepsilon.$$

This concludes the proof. □

Corollary 4.12.2. *Let $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d_+)$ such that $NA(\mathbb{P})$ holds.*

(i) Let g be continuous and $\mathfrak{P} \subseteq \mathcal{P}(\mathbb{R}^d_+)$. If $\mathbb{P} \in \mathfrak{P}$ and for all $\delta > 0$, there exists a compact set $K \subseteq \mathbb{R}^d_+$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \left(\pi^{\tilde{\mathbb{P}}}(g) - \pi^{\tilde{\mathbb{P}}(\cdot|K)}(g) \right) \leq \delta,$$

then $\hat{\pi}_N(g)$ is robust at $\mathbb{P}$ wrt. $\mathcal{W}^\infty$ on $\mathfrak{P}$.

(ii) Let g be a continuous function of linear growth and $\mathfrak{P} \subseteq \mathcal{P}(\mathbb{R}_+)$. If $\mathbb{P} \in \mathfrak{P}$, $\mathfrak{P}$ is uniformly integrable and for all $\delta > 0$ there exists $C > 0$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \left(\pi^{\tilde{\mathbb{P}}}(g) - \sup_{\|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \right) \le \delta, \tag{4.12.5}$$

then $\hat{\pi}_N$ is robust at $\mathbb{P}$ wrt. $\mathcal{W}^\infty$ on $\mathfrak{P}$.

Proof. *(i)*: Let $\varepsilon > 0$. By assumption we can find a compact set $K \subseteq \mathbb{R}^d_+$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \left(\pi^{\tilde{\mathbb{P}}}(g) - \pi^{\tilde{\mathbb{P}}(\cdot|K)}(g) \right) \le \varepsilon/3.$$

The rest of the proof follows as in the proof of Theorem 4.4.5 above using uniform continuity of g on K.

(ii): Let $\varepsilon > 0$ and choose $C > 0$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \left(\pi^{\tilde{\mathbb{P}}}(g) - \sup_{\|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] \right) \le \varepsilon/3,$$

By assumption there exists a constant $D > 0$ such that $g(r) \le D(1 + |r|)$ for $|r|$ large enough. As $\mathfrak{P}$ is uniformly integrable, there exists $L > 0$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \sup_{\mathbb{Q} \in \mathcal{M},\ \|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C} \mathbb{E}_{\mathbb{Q}}[g \mathbb{1}_{\{|r| \ge L\}}] \le \sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} C \mathbb{E}_{\tilde{\mathbb{P}}}[D(1+|r|)\mathbb{1}_{\{|r| \ge L\}}] \le \varepsilon/3.$$

Note that there exists $A > 0$ such that the hedging strategies

$$\{H \in \mathbb{R}^d \mid \pi^{\tilde{\mathbb{P}}(\cdot|[0,L]^d)}(g) + H(r-1) \ge g(r) \quad \tilde{\mathbb{P}}(\cdot|[0,L]^d)\text{-a.s.}\}$$

contain an element bounded by some constant $A > 0$ for all $L > 0$ large enough: Otherwise there exist sequences $(L_n)_{n \in \mathbb{N}}$, $(H_n)_{n \in \mathbb{N}}$ and $(\tilde{\mathbb{P}}_n)_{n \in \mathbb{N}}$ such that $H_n \to \infty$. Note that by uniform integrability of $\tilde{\mathbb{P}}_n$ we have $\tilde{\mathbb{P}}_n \Rightarrow \tilde{\mathbb{P}}$ and also $\tilde{\mathbb{P}}_n(\cdot|[0, L_n]^d) \Rightarrow \tilde{\mathbb{P}}$, where NA($\tilde{\mathbb{P}}$) holds. Take $\tilde{H}_n := H_n/|H_n|$, then after possibly taking a subsequence $\tilde{H}_n \to \tilde{H}$ with $|\tilde{H}| = 1$ and $\tilde{H}(r-1) \ge 0$ $\tilde{\mathbb{P}}$-a.s., which leads to $\tilde{H} = 0$ by NA($\tilde{\mathbb{P}}$), a contradiction. Take $L > 0$ such that

$$\sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} \sup_{\mathbb{Q} \in \mathcal{M},\ \|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C} \mathbb{E}_{\mathbb{Q}}[A|r|\mathbb{1}_{\{|r| \ge L\}}] \le \sup_{\tilde{\mathbb{P}} \in \mathfrak{P}} C \mathbb{E}_{\tilde{\mathbb{P}}}[A|r|\mathbb{1}_{\{|r| \ge L\}}] \le \varepsilon/3.$$

Then for all $\tilde{\mathbb{P}} \in \mathfrak{P}$ and for all hedging strategies H bounded by A

$$\begin{aligned}
\pi^{\tilde{\mathbb{P}}}(g) - \varepsilon &\le \sup_{\|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g] - 2\varepsilon/3 \\
&\le \sup_{\|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g \mathbb{1}_{\{|r| \le L\}}] - \varepsilon/3 \\
&\le \sup_{\|d\mathbb{Q}/d\tilde{\mathbb{P}}\|_\infty \le C,\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g \mathbb{1}_{\{|r| \le L\}} - |H|\,|r|\mathbb{1}_{\{|r| \ge L\}}] \\
&\le \sup_{\mathbb{Q} \sim \tilde{\mathbb{P}},\ \mathbb{Q} \in \mathcal{M}} \mathbb{E}_{\mathbb{Q}}[g \mathbb{1}_{\{|r| \le L\}} - |H|\,|r|\mathbb{1}_{\{|r| \ge L\}}] \\
&= \pi^{\tilde{\mathbb{P}}}(g \mathbb{1}_{\{|r| \le L\}} - H|r|\mathbb{1}_{\{|r| \ge L\}}) \le \pi^{\tilde{\mathbb{P}}(\cdot|[0,L]^d)}(g) \le \pi^{\tilde{\mathbb{P}}}(g).
\end{aligned}$$

Thus again we can restrict to $K = [0, L]^d$ as before and proceed as in the proof of Theorem 4.4.5. □

Proof of Corollary 4.4.6. Recall that $\mathbb{P}$ is compactly supported, say $\text{supp}(\mathbb{P}) \subseteq \overline{B_R(0)}$ for some $R > 0$, so that we can assume that g is uniformly continuous. Note that we also have that $\text{supp}(\hat{\mathbb{P}}_N) \subseteq \overline{B_R(0)}$, $\mathbb{P}^\infty$-a.s. Theorem 4.4.5 then shows robustness of $\hat{\pi}_N$. Thus, for any $\varepsilon > 0$, there exists $\delta > 0$ and $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$ and all $\tilde{\mathbb{P}} \in \mathcal{P}(\mathbb{R}^d_+)$ we have:

$$\mathcal{W}^\infty(\tilde{\mathbb{P}}, \mathbb{P}) \leq \delta \quad \Rightarrow \quad d_L(\mathcal{L}_{\tilde{\mathbb{P}}}(\hat{\pi}_N), \mathcal{L}_{\mathbb{P}}(\hat{\pi}_N)) \leq \varepsilon/3.$$

Observe that $\mathcal{W}^\infty(\tilde{\mathbb{P}}, \mathbb{P}) \leq \delta$ implies $d_H(\text{supp}(\mathbb{P}), \text{supp}(\tilde{\mathbb{P}})) \leq \delta$. Proposition 4.2.8 states that the map $\tilde{\mathbb{P}} \mapsto \pi^{\tilde{\mathbb{P}}}(g)$ is continuous in the pseudo-metric $d_H(\text{supp}(\mathbb{P}), \text{supp}(\tilde{\mathbb{P}}))$. By [Bertsekas and Shreve, 1978, Prop. C.2] the collection of closed subsets of $\overline{B_{R+1}(0)}$ equipped with the Hausdorff-metric is compact, so in particular there exists $N_1 \geq N_0$ such that for all $N \geq N_1$ and for all $\tilde{\mathbb{P}} \in B^\infty_{l_N}(\hat{\mathbb{P}}_N)$ we have $|\pi^{\tilde{\mathbb{P}}}(g) - \pi^{\hat{\mathbb{P}}_N}(g)| \leq \varepsilon/3$. This concludes the proof. □

4.13 Additional results for Section 4.5

In this Section we present some simulations complementing Figure 4.3 in the main article. More specifically we compare estimates for the quantity $\pi^{\text{AV@R}^{\mathbb{P}}_{0.95}}(g)$, where we set $g(r) = (r-1)^+$ (Figure 4.10) for historical gold price (WGC/GOLD DAILY USD) returns and Apple (AAPL) returns, and $g(r) = |r-1|$ (Figure 4.11) for historical S&P500 and DAX30 returns. Finally, Figure 4.12 reproduces the case of S&P500 and DAX30 returns and $g(r) = (r-1)^+$ from Figure 4.3 but this time the GARCH(1,1)-estimator uses log-returns instead of simple returns. While before it ignored market crisis now it overreacts to it. We use 50 or 100 data points to build the estimates and plot the average of the last 5 or 10 running estimates.

Figure 4.10: Comparison of estimates for $\pi^{\mathrm{AV@R}^{\mathbb{P}}_{0.95}}((r-1)^+)$. The first two panes show estimates with a rolling window of 100 data points and we plot the average of the last 10 estimates. The first pane uses Apple returns while the second one uses gold returns. The last pane uses gold returns for estimates with a rolling window of 50 data points where we plot the average of the last 5 estimates.

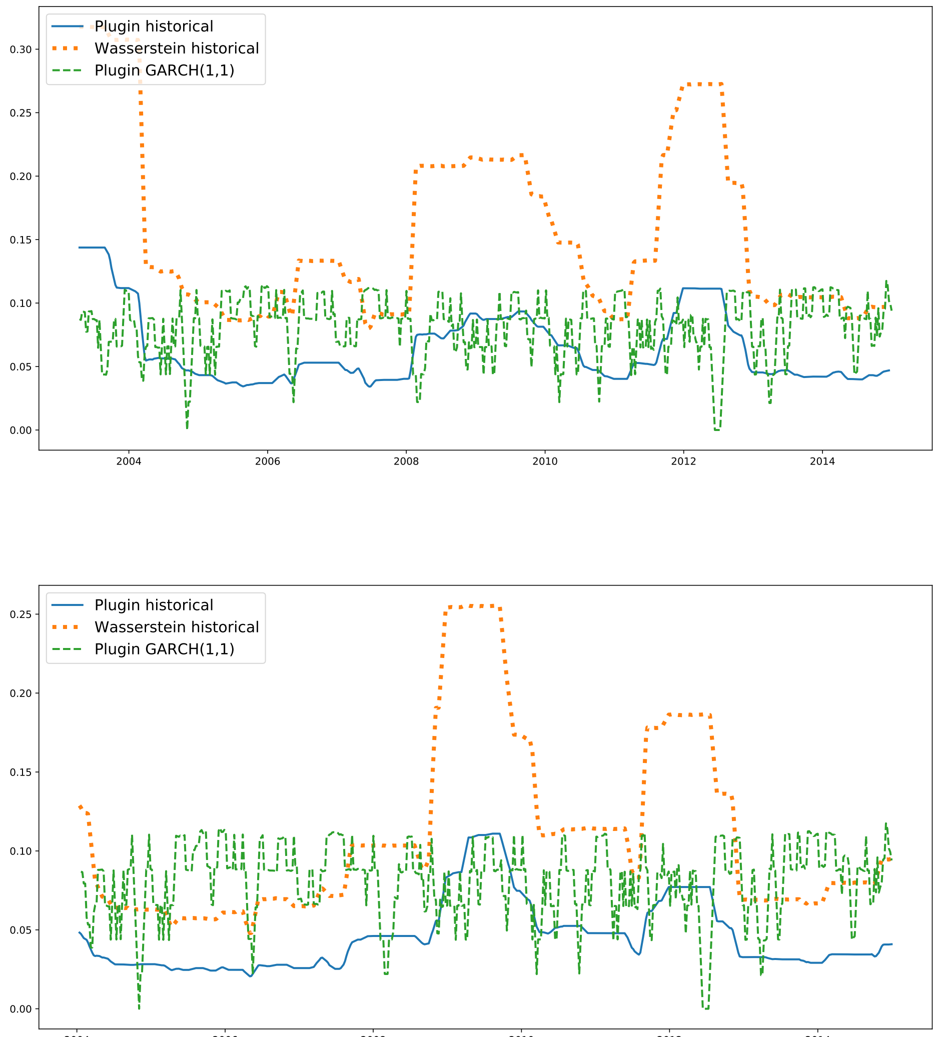

Figure 4.11: Comparison of estimates for $\pi^{\mathrm{AV@R}^{\mathbb{P}}_{0.95}}(|r-1|)$. Estimates use a rolling window of 50 data points and we plot the average of the last 5 estimates. The first pane uses DAX30 returns while the second one uses S&P500 returns.

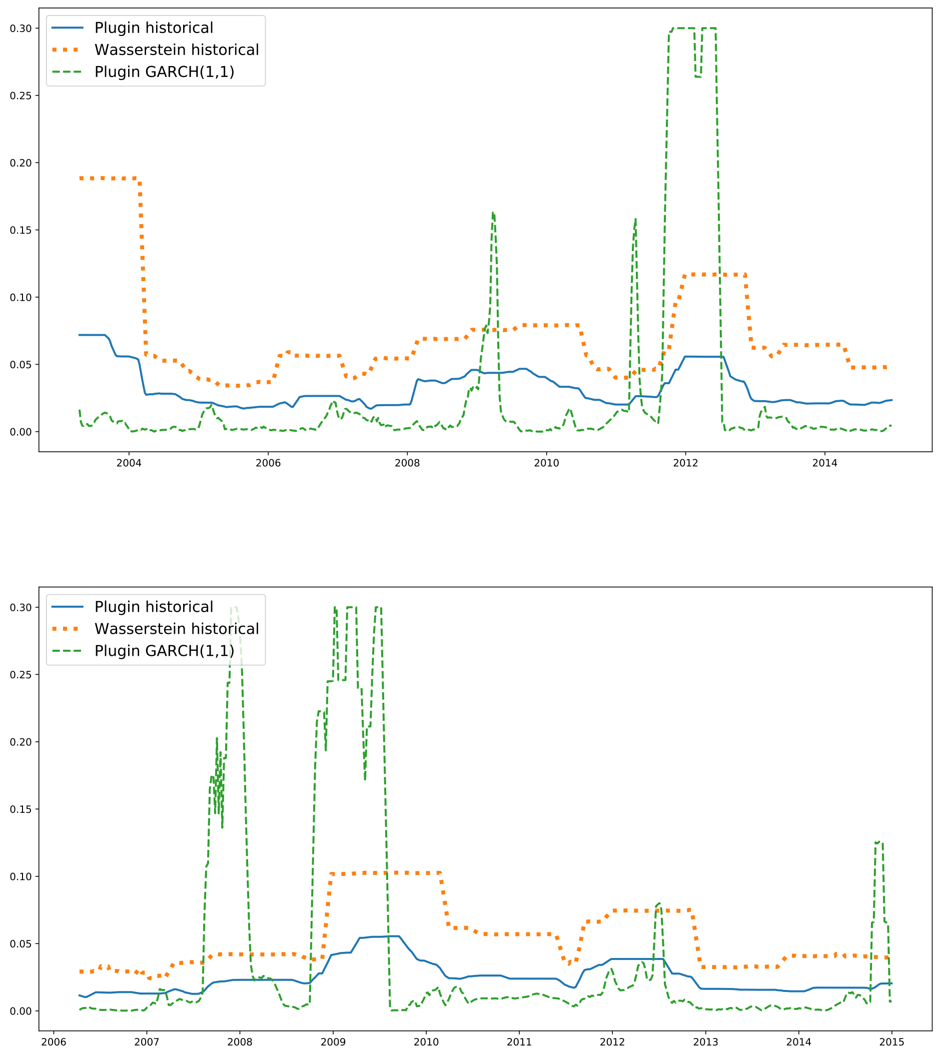

Figure 4.12: Comparison of estimates for $\pi^{\mathrm{AV@R}_{0.95}^{\mathbb{P}}}((r-1)^+)$ estimated on log-returns for the GARCH(1,1) model. Estimates use a rolling window of 50 data points and we plot the average of the last 5 estimates. The first pane uses DAX30 returns while the second one uses S&P500 returns.

Chapter 5

Continuity of the martingale optimal transport problem on the real line

5.1 Introduction: The Martingale optimal transport problem and nested Wasserstein distance

The martingale optimal transport (MOT) problem, which was introduced in Beiglböck et al. [2013] in discrete time and in Galichon et al. [2014] in continuous time, is a version of the optimal transport problem, which was first posed by Gaspard Monge in Monge [1781], with an additional martingale constraint. In recent years it has received considerable attention in the field of robust mathematical finance, as it can be utilised to obtain no-arbitrage pricing bounds. For an overview of recent developments in the field we refer to Beiglböck and Juillet [2016], Beiglböck et al. [2017] and the references therein.
Given two measures μ and ν on the real line, let us denote by $\Pi(\mu, \nu)$ the set of probability measures on $\mathbb{R}^2$ with marginals μ and ν. With this notation at hand the MOT problem reads

$$C(\mu, \nu) := \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c(x_1, x_2)\, \pi(dx_1, dx_2). \tag{5.1.1}$$

Here $\mathcal{M}(\mu, \nu)$ is the set of martingale couplings

$$\mathcal{M}(\mu, \nu) = \left\{ \pi \in \Pi(\mu, \nu) \ : \ \int (x_2 - x_1)\, \pi_{x_1}(dx_2) = 0 \quad \mu\text{-a.s.} \right\},$$

$(\pi_{x_1})_{x_1 \in \mathbb{R}}$ denotes a regular disintegration of the coupling π with respect to its first marginal μ and $c : \mathbb{R}^2 \to \mathbb{R}$ is a Borel measurable function.
In this chapter we establish continuity of the mapping

$$(\mu, \nu) \mapsto C(\mu, \nu) \tag{5.1.2}$$

with respect to the Wasserstein metric and give a new proof of sufficiency of the monotonicity principle for martingale optimal transport, which was introduced in [Beiglböck and Juillet, 2016, Lemma 1.11], as a consequence of this result. Such a continuity property is well known for classical optimal transport (see e.g. [Villani, 2008, Theorem 5.20, p.77]), but has only quite recently been proven for martingale optimal transport in Backhoff-Veraguas and Pammer [2019]. Before, partial results have been obtained in Juillet [2016] and Guo and Obłój [2019]. Establishing continuity of $(\mu, \nu) \mapsto C(\mu, \nu)$ is clearly of paramount importance for any practical applications such as computational methods or statistical estimation, when approximations cannot be avoided or uncertainty in the underlying data is present.

Contrary to Backhoff-Veraguas and Pammer [2019], our main stability result is proved via an estimate of the nested 1-Wasserstein distance $\mathcal{W}^1_{nd}$ between a coupling $\pi \in \Pi(\mu, \nu)$ and the set $\mathcal{M}(\mu, \nu)$. More specifically, we show

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi}) \approx \int \left| \int (x_2 - x_1)\, \pi_{x_1}(dx_2) \right| \mu(dx_1), \tag{5.1.3}$$

where $\mathcal{W}^1_{nd}$ is defined as (see [Backhoff-Veraguas et al., 2017b, Proposition 5.2])

$$\begin{aligned} \mathcal{W}^1_{nd}(\pi, \tilde{\pi}) = \inf_{\gamma^1 \in \Pi(\pi^1, \tilde{\pi}^1)} \Bigg(& \int |x_1 - y_1|\, \gamma^1(dx_1, y_1) \\ & + \int \inf_{\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})} \int |x_2 - y_2|\, \gamma^2(dx_2, dy_2)\, \gamma^1(dx_1, dy_1) \Bigg). \end{aligned} \tag{5.1.4}$$

Here $\pi^1, \tilde{\pi}^1$ denote the first marginals of π and $\tilde{\pi}$ respectively, while $(\pi_{x_1})_{x_1 \in \mathbb{R}}$, $(\tilde{\pi}_{y_1})_{y_1 \in \mathbb{R}}$ denote the disintegrations of π and $\tilde{\pi}$ with respect to the first marginal. On an intuitive level, the nested Wasserstein distance only considers those couplings $\gamma \in \Pi(\pi, \tilde{\pi})$, which respect the information flow formalised by the canonical (i.e. coordinate) filtration $(\mathcal{F}_t)_{t \in \{1,2\}}$: in (5.1.4) this is achieved by first taking an infimum over couplings of $\pi^1, \tilde{\pi}^1$ (i.e. "couplings at time one") and then a second (nested) infimum with respect to the respective disintegrations (i.e. "conditional couplings at time two"). This feature distinguishes $\mathcal{W}^1_{nd}$ from the Wasserstein distance $\mathcal{W}^1$, which also includes "anticipative couplings". We refer to [Backhoff-Veraguas et al., 2019, pp. 2-3] for a well-written introduction to this topic. The nested distance was introduced in Pflug [2009], Pflug and Pichler [2012] in the context of multistage stochastic optimisation and was independently analysed in Lassalle [2018].

Our estimate (5.1.3) complements the results of Backhoff-Veraguas and Pammer [2019], who essentially show continuity of the monotonicity principle for MOT without using the primal formulation (5.1.1) directly. We believe that it is of independent interest as it implies uniform continuity of the mapping $\pi \mapsto \inf_{\tilde{\pi} \in \mathcal{M}(\pi^1, \pi^2)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi})$ under a uniform integrability constraint on the second marginal of π, which we denote by π^2. Furthermore we show that our estimate (5.1.3) is sharp for a class of couplings $\pi \in \Pi(\mu, \nu)$ satisfying a dispersion assumption in the spirit of Hobson and Klimmek [2015], extending results obtained in Jourdain and Margheriti [2018].

The remainder of this chapter is organised as follows: we state our main results in Section 5.2. The proof of Proposition 5.2.4 is given in Section 5.4, while we prove Theorem 5.2.8 in Section 5.5. In Section 5.6 we collect proofs of the remaining results announced in Section 5.2. We also list some generic approximation results for $\mathcal{W}^1_{nd}$ and $\mathcal{W}^p$ in Section 5.3. These will be frequently used in Sections 5.4 and 5.5 and are proved in the appendix.

5.2 Main results

5.2.1 Notation

Let us first outline the notation used in this chapter. We denote by $\mathcal{P}(\mathbb{R}^d)$ the probability measures on $\mathbb{R}^d$ and write $\mathcal{P}_p(\mathbb{R}^d) = \{\mu \in \mathcal{P}(\mathbb{R}^d) \ : \ \int |x|^p \mu(dx) < \infty\}$, where $p \geq 1$ and $|\cdot|$ is the Euclidean distance on $\mathbb{R}^d$. For two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ let $\Pi(\mu, \nu)$ denote the set of couplings $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ and ν. Let $f_*\pi$ denote the push-forward measure of π by the function of $f : \mathbb{R}^d \to \mathbb{R}^d$. For $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ we denote by π^1 and π^2 the push-forward measures of π under the canonical projection to the first coordinate $(x_1, x_2) \mapsto x_1$ and second coordinate $(x_1, x_2) \mapsto x_2$ respectively. Furthermore a disintegration (or regular conditional probability distribution) of π is defined as a family of probability measures $(\pi_{x_1})_{x_1 \in \mathbb{R}^d}$ on $\mathbb{R}^d$ such that for every Borel set $B \subseteq \mathbb{R}^d$ the mapping $x_1 \mapsto \pi_{x_1}(B)$ is Borel measurable and for all Borel sets $A, B \subseteq \mathbb{R}^d$

$$\pi(A \times B) = \int_A \pi_{x_1}(B)\, \pi^1(dx_1).$$

For a general existence result on Polish spaces and fundamental properties of disintegrations we refer to [Stroock and Varadhan, 2007, pp.12-19]. More generally, for a disintegration $(\pi_{x_1})_{x_1 \in \mathbb{R}^d}$ on $\mathbb{R}^d$ and a measure μ on $\mathbb{R}^d$ we denote by $\mu \otimes \pi_{x_1}$ the measure obtained via $\mu \otimes \pi_{x_1}(A \times B) = \int_A \pi_{x_1}(B)\, \mu(dx_1)$ for Borel $A, B \subseteq \mathbb{R}^d$. The product coupling of $\pi \in \mathcal{P}(\mathbb{R}^d)$ and $\tilde{\pi} \in \mathcal{P}(\mathbb{R}^d)$ will be denoted by $\pi \times \tilde{\pi}$. We also write $\mathrm{supp}(\pi)$ for the support of a measure $\pi \in \mathcal{P}(\mathbb{R}^d)$ and often use $\mu(x_1) := \mu(\{x_1\})$ to shorten notation. Given a set $\Gamma \in R^d \times \mathbb{R}^d$ we write $\Gamma^1 := \{x_1 \in \mathbb{R}^d \ : \ \exists x_2 \in \mathbb{R}^d \text{ such that } (x_1, x_2) \in \Gamma\}$ and $\Gamma_{x_1} := \{x_2 \in \mathbb{R}^d \ : \ (x_1, x_2) \in \Gamma\}$.

For $\mu, \nu \in \mathcal{P}(\mathbb{R})$ let $\mathcal{M}(\mu, \nu)$ be the set of martingale couplings

$$\mathcal{M}(\mu, \nu) = \left\{ \pi \in \Pi(\mu, \nu) \ : \ \int (x_2 - x_1)\, \pi_{x_1}(dx_2) = 0 \quad \mu\text{-a.s.} \right\},$$

where $(\pi_{x_1})_{x_1\in\mathbb{R}}$ is a disintegration of π. We denote the convex order of μ and ν by $\mu \preceq_c \nu$. It is well known that $\mu \preceq_c \nu$ is equivalent to $\mathcal{M}(\mu,\nu) \neq \emptyset$ (see Strassen [1965]). We call a set of measures $\mathfrak{P} \subseteq \mathcal{P}(\mathbb{R})$ uniformly integrable if $\lim_{K\to\infty} \sup_{\mu\in\mathfrak{P}} \int_{\{|x|\geq K\}} |x|\, \mu(dx) = 0$. Next we recall the p-Wasserstein distance on $\mathcal{P}(\mathbb{R}^2)$ given by

$$\mathcal{W}^p(\pi,\tilde{\pi}) = \left(\inf_{\gamma\in\Pi(\pi,\tilde{\pi})} \int |x_1 - y_1|^p + |x_2 - y_2|^p \,\gamma(dx,dy) \right)^{1/p},$$

where $\Pi(\pi,\tilde{\pi}) \subseteq \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2)$ is the set of couplings with first marginal $\pi \in \mathcal{P}(\mathbb{R}^2)$ and second marginal $\tilde{\pi} \in \mathcal{P}(\mathbb{R}^2)$, and the nested p-Wasserstein distance

$$\mathcal{W}^p_{nd}(\pi,\tilde{\pi}) = \Bigg(\inf_{\gamma^1\in\Pi(\pi^1,\tilde{\pi}^1)} \bigg(\int |x_1 - y_1^p| \,\gamma^1(dx_1,y_1) + \int \inf_{\gamma^2\in\Pi(\pi_{x_1},\tilde{\pi}_{y_1})} \int |x_2 - y_2|^p \,\gamma^2(dx_2,dy_2)\,\gamma^1(dx_1,dy_1) \bigg) \Bigg)^{1/p}.$$

For ease of notation we furthermore define

$$\varepsilon_\pi := \int \left| \int (x_2 - x_1)\,\pi_{x_1}(dx_2) \right| \mu(dx_1)$$

for $\pi \in \Pi(\mu,\nu)$.

Fix $\mu,\nu \in \mathcal{P}(\mathbb{R})$. We now investigate the nested distance $\mathcal{W}^1_{nd}$ between a coupling $\pi \in \Pi(\mu,\nu)$ and its projection on to the set $\mathcal{M}(\mu,\nu)$. For the sake of clarity we first give a lower bound on $\inf_{\tilde{\pi}\in\mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi,\tilde{\pi})$ and then derive upper bounds under progressively less restrictive assumptions.

5.2.2 Projection on to $\mathcal{M}(\mu,\nu)$: attainment of lower bound

Let us first derive a lower bound on $\inf_{\tilde{\pi}\in\mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi,\tilde{\pi})$:

Lemma 5.2.1. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$, $\mu \preceq_c \nu$ and $\pi \in \Pi(\mu,\nu)$. Then*

$$\inf_{\tilde{\pi}\in\mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi,\tilde{\pi}) \geq \varepsilon_\pi. \tag{5.2.1}$$

We introduce the following assumption:

Assumption 5.2.2 (Barycentre dispersion assumption)**.** *Let $\mu \preceq_c \nu$ and $\pi \in \Pi(\mu,\nu)$. For all $x \in \mathbb{R}$*

$$\int_{\{x_1\geq x\}} (x_2 - x_1)\,\pi(dx_1,dx_2) \geq 0. \tag{5.2.2}$$

In contrast to Hobson and Klimmek [2015] our dispersion assumption 5.2.2 is formulated for π and not just the marginals μ and ν. In order to motivate it, let us recall the Hoeffding–Fréchet coupling $\pi_{HF} \in \Pi(\mu, \nu)$: it enjoys the property, that it is an optimiser for problems of the form

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c(x_1, x_2)\, \pi(dx_1, dx_2),$$

where $c(x_1, x_2) = h(x_2 - x_1)$ and $h : \mathbb{R} \to \mathbb{R}$ is any convex function. In fact π_{HF} is characterised by the following monotonicity property:

$$\begin{aligned} &\text{There exists a Borel set } \Gamma_{HF} \subseteq \mathbb{R}^2 \text{ such that } \pi_{HF}(\Gamma_{HF}) = 1 \\ &\text{and whenever } (x_1, x_2), (y_1, y_2) \in \Gamma_{HF} \text{ and } x_1 < y_1 \text{ then also } x_2 \le y_2. \end{aligned} \tag{5.2.3}$$

This characterisation directly implies the following lemma:

Lemma 5.2.3. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ with $\mu \preceq_c \nu$ and $\pi_{HF} \in \Pi(\mu, \nu)$ be the Hoeffding–Fréchet coupling. Then π_{HF} satisfies the barycentre dispersion assumption 5.2.2.*

It turns out that our lower bound on $\inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi})$ is tight under the barycentre dispersion assumption:

Proposition 5.2.4. *Let $p \ge 1$, $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ satisfy $\mu \preceq_c \nu$. Let $\pi \in \Pi(\mu, \nu)$ satisfy the barycentre dispersion assumption 5.2.2. Then there exists a martingale measure $\pi_{mr} \in \mathcal{M}(\mu, \nu)$ such that*

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi}) = \mathcal{W}^1_{nd}(\pi, \pi_{mr}) = \varepsilon_\pi. \tag{5.2.4}$$

We call a martingale coupling $\pi_{mr} \in \mathcal{M}(\mu, \nu)$ satisfying (5.2.4) a *($\mathcal{W}^1_{nd}$-minimal) martingale rearrangement coupling* of π. We now discuss some basic properties of π_{mr}. Let us first remark that, as $\mathcal{W}^1(\cdot, \cdot) \le \mathcal{W}^1_{nd}(\cdot, \cdot)$, we have $\mathcal{W}^1(\pi, \pi_{mr}) \le \varepsilon_\pi$ in Proposition 5.2.4 and this inequality is strict in general. Furthermore, while existence of π_{mr} is guaranteed in Proposition 5.2.4, uniqueness is not satisfied in general as the following example shows:

Example 5.2.5. Take

$$\bar{\pi} := \frac{1}{4} \left(\delta_{(-2,-3)} + \delta_{(-1,-2)} + \delta_{(1,2)} + \delta_{(2,3)} \right).$$

As $\varepsilon_\pi = 1$ it remains to check that both

$$\pi_1 = \frac{1}{20} \left(4\delta_{(-2,-3)} + \delta_{(-2,2)} + 4\delta_{(-1,-2)} + \delta_{(-1,3)} + 4\delta_{(1,2)} + \delta_{(1,-3)} + 4\delta_{(2,3)} + \delta_{(2,-2)} \right)$$

and

$$\hat{\pi} := \frac{1}{24} \left(5\delta_{(-2,-3)} + \delta_{(-2,3)} + 5\delta_{(2,3)} + \delta_{(2,-3)} \right) + \frac{1}{16} \left(3\delta_{(-1,-2)} + \delta_{(-1,2)} + 3\delta_{(1,2)} + \delta_{(1,-2)} \right)$$

are $\mathcal{W}^1_{nd}$-minimal rearrangement couplings.

By an application of the triangle inequality the following corollary of Proposition 5.2.4 is immediate:

Corollary 5.2.6. *Let $c : \mathbb{R}^2 \to \mathbb{R}$ be L-Lipschitz-continuous. Then*

$$C(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu),\ \pi \text{ satisfies Ass. 5.2.2}} \left(\int c(x_1,x_2)\, \pi(dx_1,dx_2) + L\varepsilon_\pi \right).$$

Proposition 5.2.4 and Corollary 5.2.6 complement Jourdain and Margheriti [2018], who give a powerful characterisation of the above $\mathcal{W}^1_{nd}$-minimal martingale couplings of the Hoeffding-Fréchet coupling π_{HF} by use of its characterisation via quantile functions. More specifically [Jourdain and Margheriti, 2018, Theorem 2.11] states

$$\inf_{\pi \in \mathcal{M}(\mu,\nu)} \int |x_1 - x_2|\, \pi(dx_1,dx_2) \le 2\mathcal{W}^1(\mu,\nu). \tag{5.2.5}$$

We can recover (5.2.5), noting that

$$\begin{aligned}
\inf_{\pi \in \mathcal{M}(\mu,\nu)} \int |x_1 - x_2|\, \pi(dx_1,dx_2) &\le \int |x_1 - x_2|\, \pi_{HF}(dx_1,dx_2) + \varepsilon_{\pi_{HF}} \\
&= \int |x_1 - x_2|\, \pi_{HF}(dx_1,dx_2) \\
&\quad + \int \left| \int (x_2 - x_1)\, \pi_{HF,x_1}(dx_2) \right| \mu(dx_1) \\
&\le 2 \int |x_1 - x_2|\, \pi_{HF}(dx_1,dx_2) = 2\mathcal{W}^1(\mu,\nu),
\end{aligned}$$

where we used Lemma 5.2.3 and Corollary 5.2.6 in the first inequality and Jensen's inequality for the second inequality.

Let us lastly give the following remark.

Remark 5.2.7. While Assumption 5.2.2 is sufficient for (5.2.4), it is not necessary. Indeed, (5.2.4) also holds for the antitone or decreasing monotone coupling π_{AT}, which satisfies

$$\begin{gathered}
\text{There exists a Borel set } \Gamma_{AT} \subseteq \mathbb{R}^2 \text{ such that } \pi_{AT}(\Gamma_{AT}) = 1 \\
\text{and whenever } (x_1,x_2),(y_1,y_2) \in \Gamma_{AT} \text{ and } x_1 < y_1 \text{ then } x_2 \ge y_2.
\end{gathered} \tag{5.2.6}$$

We leave the question of finding a necessary condition for (5.2.4) for future research.

The proofs of the above results are deferred to Sections 5.4 and 5.6 and rely on the following simple observation: let us assume for the moment that $\pi \in \Pi(\mu,\nu) \setminus \mathcal{M}(\mu,\nu)$ is finitely supported and let us consider the barycentres of the disintegration $(\pi_{x_1})_{x_1 \in \mathbb{R}}$

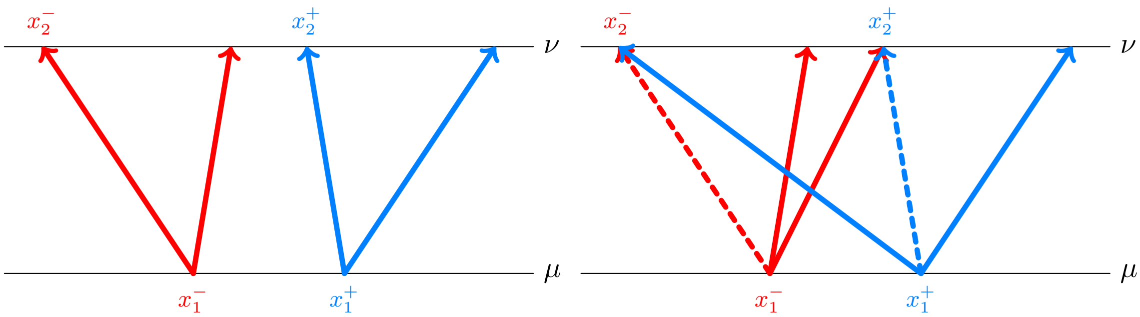

Figure 5.1: Exchange of masses at x_2^- and x_2^+ for the case $x_1^- < x_1^+$.

given by $\left(\int (x_2 - x_1)\, \pi_{x_1}(dx_2)\right)_{x_1 \in \mathrm{supp}(\mu)}$. By the barycentre dispersion assumption 5.2.2 and convex order of μ and ν we can find pairs $x_1^-, x_1^+ \in \mathrm{supp}(\mu)$ such that

$$\int (x_2 - x_1^-)\, \pi_{x_1^-}(dx_2) < 0, \qquad \int (x_2 - x_1^+)\, \pi_{x_1^+}(dx_2) > 0$$

and corresponding points $x_2^- \in \mathrm{supp}\left(\pi_{x_1^-}\right), x_2^+ \in \mathrm{supp}\left(\pi_{x_1^+}\right)$ with $x_2^- < x_2^+$. Assigning a part of the mass at x_2^- and x_2^+ to the disintegrations $\pi_{x_1^+}$ and $\pi_{x_1^-}$ respectively then allows to essentially rectify the barycentres of $\pi_{x_1^-}$ and $\pi_{x_1^+}$ piece by piece without changing the marginal constraints (see Figure 5.1).

5.2.3 Projection on to $\mathcal{M}(\mu,\nu)$: the general case

It turns out that (5.2.4) is not satisfied in general (see Example 5.2.10). Instead we obtain the following relaxation of Proposition 5.2.4 as a main result:

Theorem 5.2.8. *Let $\mathfrak{P} \subseteq \mathcal{P}_1(\mathbb{R})$ be uniformly integrable. Then for every $\delta > 0$ there exists a constant $K = K(\delta, \mathfrak{P})$ such that the following holds: for every measure $\pi \in \Pi(\mu,\nu)$, where $\mu \preceq_c \nu$ and $\nu \in \mathfrak{P}$, we have*

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi}) \leq K\varepsilon_\pi + \delta. \tag{5.2.7}$$

Similarly to Corollary 5.2.6 we obtain:

Corollary 5.2.9. *Let $c : \mathbb{R}^2 \to \mathbb{R}$ be L-Lipschitz-continuous. Then for every $\delta > 0$ there exists a constant $K = K(\delta, \nu)$ such that*

$$\begin{aligned}\inf_{\pi \in \Pi(\mu,\nu)} \int c(x_1,x_2)\, \pi(dx_1,dx_2) + KL\varepsilon_\pi &\leq C(\mu,\nu) \\ &\leq \inf_{\pi \in \Pi(\mu,\nu)} \int c(x_1,x_2)\, \pi(dx_1,dx_2) + KL\varepsilon_\pi + L\delta.\end{aligned}$$

Consequently $C(\mu,\nu)$ can be approximated by an optimal transport problem with cost function $\tilde{c}(x_1,x_2,\pi_{x_1}) = c(x_1,x_2)+K(\delta,\nu)L|\int(y_2-x_1)\,\pi_{x_1}(dy_2)|$ for an L-Lipschitz-continuous cost function c. The function $\tilde{c}$ can be interpreted as a sum of a usual optimal transport cost and a weak optimal transport cost in the spirit of Gozlan et al. [2017]. The penalisation approach of Corollary 5.2.9 is also akin to the numerical approximation results for the MOT problem obtained in Guo and Obłój [2019].

The dependence of K on δ and $\mathfrak{P}$ in Theorem 5.2.8 above is crucial, as the following counterexample shows:

Example 5.2.10. Let us consider

$$\mu^n = \nu^n = \frac{1}{n}\sum_{i=1}^n \delta_i.$$

Then trivially $\mu^n \preceq_c \nu^n$ for all $n \in \mathbb{N}$, $\mathfrak{P} = \{\nu_n \,:\, n \in \mathbb{N}\}$ is not uniformly integrable and the only martingale coupling $\tilde{\pi}^n \in \mathcal{M}(\mu^n,\nu^n)$ is supported on the diagonal $x_1 = x_2$. We take

$$\pi^n = \frac{1}{n}\left(\frac{\delta_{(1,1)}+\delta_{(1,2)}}{2} + \frac{\delta_{(n,n-1)}+\delta_{(n,n)}}{2} + \sum_{i=2}^{n-1}\frac{\delta_{(i,i-1)}+\delta_{(i,i+1)}}{2}\right),$$

which is “almost” a martingale coupling. Then

$$\inf_{\tilde{\pi}\in\mathcal{M}(\mu^n,\nu^n)} \mathcal{W}^1_{nd}(\pi^n,\tilde{\pi}) = \frac{n-1}{n} \qquad \text{and} \qquad \int\left|\int (x_2-x_1)\,\pi^n_{x_1}(dx_2)\right|\,\mu(dx_1) = \frac{1}{n}.$$

Thus for any $0 \le \delta < 1$ there exists no $K > 0$, which fulfils (5.2.7) simultaneously for all $(\pi^n)_{n\in\mathbb{N}}$.

5.2.4 Continuity of MOT

We now turn to our second main result, which establishes continuity of the map $(\mu,\nu) \mapsto C(\mu,\nu)$:

Theorem 5.2.11. *Let $p \ge 1$ and let $(\mu^n)_{n\in\mathbb{N}}$, $(\nu^n)_{n\in\mathbb{N}}$ be two sequences of measures in $\mathcal{P}_p(\mathbb{R})$ with $\mu^n \preceq_c \nu^n$ for all $n \in \mathbb{N}$. Let $\mu,\nu \in \mathcal{P}(\mathbb{R})$ be such that $\lim_{n\to\infty}\mathcal{W}^p(\mu^n,\mu) = 0$ and $\lim_{n\to\infty}\mathcal{W}^p(\nu^n,\nu) = 0$. Furthermore let $c : \mathbb{R}^2 \to \mathbb{R}$ be continuous and such that $|c(x_1,x_2)| \le C(1+|x_1|^p+|x_2|^p)$ for some $C \ge 0$. Then*

$$\lim_{n\to\infty} C(\mu^n,\nu^n) = C(\mu,\nu).$$

This stability result extends the findings of Juillet [2016] and Guo and Obłój [2019]. Juillet [2016] proves continuity of the left-curtain coupling with respect to its marginals in a Wasserstein-type metric. In particular the results obtained only hold for cost functions satisfying the Spence-Mirrlees condition $c_{xyy} < 0$. On the other hand [Guo and Obłój, 2019, Prop. 4.7] assume a Lipschitz-continuous cost function c together with a finite second moment of ν and exploit a duality result for martingale optimal transport. Our result is more general and only considers the primal formulation of $C(\mu, \nu)$ given in (5.1.1). It is akin to a similar stability result in optimal transport with the obvious modifications. The proof of Theorem 5.2.11 extends a natural construction given in [Guo and Obłój, 2019, proof of Proposition 4.2, p. 20], which essentially couples the marginals μ^n, ν^n with the disintegration $(\pi_{x_1})_{x_1 \in \mathbb{R}}$. In a second step one then corrects the new coupling to account for the martingale constraint, which is achieved by an application of Theorem 5.2.8.

5.2.5 An independent proof of the monotonicity principle for MOT

As in classical optimal transport, it is desirable to characterise the sets $\Gamma \subseteq \mathbb{R}^2$, on which optimisers of $C(\mu, \nu)$ live. This has been achieved in the influential work Beiglböck and Juillet [2016] and is known as a monotonicity principle for martingale optimal transport. To set up notation we recall here the notion of a competitor given in Beiglböck and Juillet [2016], which naturally extends the corresponding optimal transport formulation. We recall that α^1 denotes push-forward measure of α under the canonical projection to the first coordinate $x = (x_1, x_2) \mapsto x_1$:

Definition 5.2.12. *Let $\alpha \in \mathcal{P}(\mathbb{R}^2)$. We say that $\alpha' \in \mathcal{P}(\mathbb{R}^2)$ is a competitor of α, if α' has the same marginals as α and*

$$\int y \, \alpha_x(dy) = \int y \, d\alpha'_x(dy) \qquad \alpha^1\text{-a.s.}$$

The following monotonicity principle was first stated in [Beiglböck and Juillet, 2016, Lemma 1.11, p. 49], where necessity and a partial sufficiency result was shown.

Theorem 5.2.13. *Assume that $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ satisfy $\mu \preceq_c \nu$ and that $c : \mathbb{R}^2 \to \mathbb{R}$ is a continuous cost function such that $|c(x_1, x_2)| \leq \tilde{K}(1 + |x_1|^p) + |x_2|^p)$ for some $\tilde{K} \geq 0$ and $p > 1$. Then $\pi \in \mathcal{M}(\mu, \nu)$ is an optimiser of $C(\mu, \nu)$ if and only if there exists a Borel set Γ with $\pi(\Gamma) = 1$ such that the following holds:*
if α is a measure on $\mathbb{R}^2$ with $|supp(\alpha)| < \infty$ and $supp(\alpha) \subseteq \Gamma$, then we have

$$\int c(x_1, x_2) \, \alpha(dx_1, dx_2) \leq \int c(x_1, x_2) \, \alpha'(dx_1, dx_2)$$

for every competitor α' of α.

The proof of necessity was later simplified in Beiglböck and Griessler [2019] and essentially relies on the idea to select competitors in a measurable way. We give here an

independent proof of sufficiency, which uses the stability result stated in Theorem 5.2.11. The idea is to argue by contraposition: take any martingale measure $\pi \in \mathcal{M}(\mu, \nu)$, any set $\Gamma \subseteq \mathbb{R}^2$ such that $\pi(\Gamma) = 1$ and assume π is not optimal for $C(\mu, \nu)$. By an approximation result given in Lemma 5.3.1 it is possible to find martingale measures π^n finitely supported on Γ such that $\lim_{n\to\infty} \mathcal{W}^p(\pi^n, \pi) = 0$. Let us denote the first marginal of π^n by μ^n and the second marginal by ν^n. As π is not optimal and as $(\mu, \nu) \mapsto C(\mu, \nu)$ is continuous, there exists a number $n \in \mathbb{N}$ and a competitor $\pi' \in \mathcal{M}(\mu^n, \nu^n)$ with cost $\int c\, d\pi'$ strictly smaller than $\int c\, d\pi^n$, showing that Γ is not finitely optimal.
In particular this enables us to show sufficiency for continuous functions of polynomial growth similar to Griessler [2016], who uses a splitting property for cyclically monotone sets and the decomposition into irreducible components established in Beiglböck and Juillet [2016].

5.3 Generic approximation results

Let us now list several approximation results for the nested distance $\mathcal{W}^1_{nd}$ and the Wasserstein distance $\mathcal{W}^p$, which we will use throughout the chapter. As these do not immediately follow from the isometric embedding of the space $(\mathcal{P}_p(\mathbb{R}), \mathcal{W}^p_{nd})$ into a Wasserstein space of nested distributions obtained in Backhoff-Veraguas et al. [2017a], we adopt a constructive self-contained approach. The proofs are mainly technical and are thus deferred to the appendix. We also remark that the concrete formulation of the lemmas below is due to the particular settings in Sections 5.4-5.6 below and some of them could be generalised. We have refrained from doing this here for the sake of clarity.

Lemma 5.3.1. *Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$, $\pi \in \Pi(\mu, \nu)$ and $\kappa > 0$. Let $\Gamma \subseteq \mathbb{R}^2$ be a Borel set such that $\pi(\Gamma) = 1$.*

(i) There exists a measure $\hat{\pi}$, which finitely supported on Γ, such that $\mathcal{W}^p_{nd}(\pi, \hat{\pi}) \le \kappa$. Furthermore

$$\int_{\{x_1 \ge x\}} (x_2 - x_1)\, \hat{\pi}(dx_1, dx_2) \ge \int_{\{x_1 \ge x\}} (x_2 - x_1)\, \pi(dx_1, dx_2) - \kappa \tag{5.3.1}$$

for all $x \in \operatorname{supp}(\hat{\pi}^1)$.

(ii) If $\pi \in \mathcal{M}(\mu, \nu)$, then $\hat{\pi}$ can be chosen to be a martingale measure.

Lemma 5.3.2. *Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$, $\mu \preceq_c \nu$, $\pi \in \Pi(\mu, \nu)$ and $\kappa > 0$. Then there exists a finitely supported measure $\bar{\pi} \in \Pi(\bar{\mu}, \bar{\nu})$ such that $\bar{\mu} \preceq_c \bar{\nu}$ and $\mathcal{W}^p_{nd}(\pi, \bar{\pi}) \le \kappa$.*

Lemma 5.3.3. *Let $\mu,\nu \in \mathcal{P}_p(\mathbb{R})$ and $\pi \in \Pi(\mu,\nu)$. Let $\pi^n \in \Pi(\mu^n,\nu^n)$ be a sequence of finitely supported measures and let $(\tilde{\pi}^n)_{n\in\mathbb{N}}$ be another sequence satisfying $\tilde{\pi}^n \in \Pi(\mu^n,\rho^n)$ for all $n \in \mathbb{N}$ and some $(\rho^n)_{n\in\mathbb{N}}$ with $\rho^n \in \mathcal{P}(\mathbb{R})$ for all $n \in \mathbb{N}$. Then*

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i, \frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i\right) \leq \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\pi^i_{x_1}, \tilde{\pi}^i_{x_1})\,\mu^i(dx_1).$$

Lemma 5.3.4. *(i) Fix a sequence $(\kappa_n)_{n\in\mathbb{N}}$. Then there exists a sequence $(\hat{\pi}^n)_{n\in\mathbb{N}}$ of finitely supported measures which satisfy*

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i, \pi\right) \leq \frac{1}{n}\sum_{i=1}^n \kappa_i$$

and for all $n \in \mathbb{N}$. Furthermore for all $n \in \mathbb{N}$ and all $x \in \operatorname{supp}((\hat{\pi}^n)^1)$

$$\int_{\{x_1 \geq x\}} (x_2 - x_1)\,\hat{\pi}^n(dx_1,dx_2) \geq \int_{\{x_1 \geq x\}} (x_2 - x_1)\,\pi(dx_1,dx_2) - \kappa_n$$

holds.

(ii) There exists another sequence $(\bar{\pi}^n)_{n\in\mathbb{N}}$ of finitely supported measures which satisfy $\bar{\pi}^n \in \Pi(\bar{\mu}^n,\bar{\nu}^n)$ such that $\bar{\mu}^n,\bar{\nu}^n \in \mathcal{P}(\mathbb{R})$, $\bar{\mu}^n \preceq_c \bar{\nu}^n$ for all $n \in \mathbb{N}$ and

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i, \pi\right) \leq \frac{1}{n}\sum_{i=1}^n \kappa_i.$$

Lemma 5.3.5. *Let $\mu,\nu,\tilde{\mu},\tilde{\nu}$ be elements of $\mathcal{P}_p(\mathbb{R})$, $\mu \preceq_c \nu$ and let $\pi \in \mathcal{M}(\mu,\nu)$. Then there exists $\tilde{\pi} \in \Pi(\tilde{\mu},\tilde{\nu})$ such that $\mathcal{W}^p(\pi,\tilde{\pi}) \leq \mathcal{W}^p(\mu,\tilde{\mu}) + \mathcal{W}^p(\nu,\tilde{\nu})$ and*

$$\int \left| \int (y_2 - y_1)\,\tilde{\pi}_{y_1}(dy_2)\right| \tilde{\mu}(dy_1) \leq \mathcal{W}^p(\mu,\tilde{\mu}) + \mathcal{W}^p(\nu,\tilde{\nu}). \tag{5.3.2}$$

Lemma 5.3.6. *Let $\mu,\nu \in \mathcal{P}_p(\mathbb{R})$ and $\pi \in \Pi(\mu,\nu)$. Let $(\hat{\pi}^n)_{n\in\mathbb{N}}$ with $\hat{\pi}^n \in \Pi(\mu^n,\nu^n)$ for all $n \in \mathbb{N}$ be a sequence of measures obtained from Lemma 5.3.4. for a sequence $(\kappa_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\kappa_n = 0$ and let $(\breve{\pi}^n)_{n\in\mathbb{N}}$ be another sequence satisfying $\breve{\pi}^n \in \Pi(\mu^n,\nu^n)$ for all $n \in \mathbb{N}$. Then, after extracting a subsequence, there exists a measure $\breve{\pi} \in \Pi(\mu,\nu)$ such that*

$$\lim_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \breve{\pi}^i, \breve{\pi}\right) = 0.$$

Lemma 5.3.7. *Let $\mu,\nu \in \mathcal{P}_p(\mathbb{R})$. Let $(\pi^n)_{n\in\mathbb{N}}$ be a sequence of measures satisfying $\pi^n \in \Pi(\mu^n,\nu^n)$ for all $n \in \mathbb{N}$ and let $(\tilde{\pi}^n)_{n\in\mathbb{N}}$ be another sequence satisfying $\tilde{\pi}^n \in \Pi(\mu^n,\nu^n)$ for all $n \in \mathbb{N}$. Let $\lim_{n\to\infty}\mathcal{W}^p(\mu^n,\mu) = \lim_{n\to\infty}\mathcal{W}^p(\nu^n,\nu) = 0$ and $\lim_{n\to\infty}\mathcal{W}^1(\pi^n,\tilde{\pi}^n) = 0$. Then for any continuous function $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ satisfying $|c(x_1,x_2)| \leq C(1+|x_1|^p+|x_2|^p)$ for some $C \geq 0$ we have*

$$\lim_{n\to\infty} \int c(x_1,x_2)\,\pi^n(dx_1,dx_2) = \lim_{n\to\infty} \int c(x_1,x_2)\,\tilde{\pi}^n(dx_1,dx_2),$$

in particular $\lim_{n\to\infty}\mathcal{W}^p(\pi^n,\tilde{\pi}^n) = 0$.

5.4 Proof of Proposition 5.2.4

5.4.1 Proof of Proposition 5.2.4 for finitely supported measures

Throughout this section we fix two finitely supported measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and $c \geq 0$. To prepare for the general case treated in Section 5.4.2 we introduce the following generalised barycentre assumption for c:

Assumption 5.4.1 (Generalised barycentre dispersion assumption)**.** *Let $\pi \in \Pi(\mu,\nu)$. For all $x \in \mathbb{R}$*

$$\int_{\{x_1 \geq x\}} (x_2 - x_1)\, \pi(dx_1, dx_2) \geq -c. \tag{5.4.1}$$

Definition 5.4.2. *For a sequence of measures $(\pi^n)_{n\in\mathbb{N}}$ we set*

$$\begin{aligned} X_1^{n,+} &:= \left\{ x_1 \in \operatorname{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) > -c \right\}, \\ X_1^{n,0} &:= \left\{ x_1 \in \operatorname{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) = -c \right\}, \\ X_1^{n,-} &:= \left\{ x_1 \in \operatorname{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) < -c \right\}. \end{aligned}$$

for all $n \in \mathbb{N}$. Furthermore

$$\begin{aligned} X_2^{n,+} &:= \bigcup_{x_1^+ \in X_1^{n,+}} \operatorname{supp}(\pi^n_{x_1^+}), \quad X_2^{n,0} := \bigcup_{x_1^0 \in X_1^{n,0}} \operatorname{supp}(\pi_{x_1^0}) \quad \text{and} \\ X_2^{n,-} &:= \bigcup_{x_1^- \in X_1^{n,-}} \operatorname{supp}(\pi^n_{x_1^-}). \end{aligned}$$

Now we fix a measure $\pi \in \Pi(\mu,\nu)$ satisfying Assumption 5.4.1 for $c \geq 0$, set $\pi^{(0)} := \pi$ and assume that $X_1^{0,-} \neq \emptyset$. The general idea formalised in this section will be to iteratively build measures $\pi^{(j)}$ such that $X_1^{j,-}$ is decreasing to $\emptyset$. This is achieved by switching atoms in the support of $\pi^{(0)}_{x_1}$ without changing the marginal constraints. More specifically we will use Algorithm 5.4.3 given below, which is written in a slightly elaborate form in order to prepare for the more complicate case of Algorithm 5.5.4 discussed in Section 5.5. In the definition of the algorithm, we will use Definition 5.4.2 for the sequence of measures $(\pi^{(n)})_{n\in\mathbb{N}}$, which are constructed iteratively.

Algorithm 5.4.3. *Set $j = 0$.*

(i) Define $x_1^-(j) := \max(X_1^{j,-})$ and $x_2^-(j) := \min(\operatorname{supp}(\pi_{x_1^-(j)}))$. Set $x_1^+(j) := \max(X_1^{j,+})$ and $x_2^+(j) := \max(\operatorname{supp}(\pi_{x_1^+(j)}))$.

(ii) Define

$$\lambda^{(j)} := \Bigg[\mu(x_1^-(j)) \min\left\{\tilde\lambda > 0 \,\middle|\, \int (x_2 - x_1^-(j))\, \pi^{(j)}_{x_1^-(j)}(dx_2) + \tilde\lambda(x_2^+(j) - x_2^-(j)) \geq -c\right\}$$
$$\wedge\, \mu(x_1^-(j))\pi^{(j)}_{x_1^-(j)}(x_2^-(j))$$
$$\wedge\, \mu(x_1^+(j)) \min\left\{\tilde\lambda > 0 \,\middle|\, \int (x_2 - x_1^+(j))\, \pi^{(j)}_{x_1^+(j)}(dx_2) + \tilde\lambda(x_2^-(j) - x_2^+(j)) \leq -c\right\}$$
$$\wedge\, \mu(x_1^+(j))\pi^{(j)}_{x_1^+(j)}(x_2^+(j))\Bigg] \cdot (x_2^+(j) - x_2^-(j)).$$

(iii) Define $\rho^{(j)} \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2)$ with via $\rho^{(j)} := ((x_1, x_1)_*\mu) \otimes \rho^{(j)}_{(x_1,x_1)}$, where

$$\rho^{(j)}_{(x_1,x_1)} := (x_2, x_2)_*\pi^{(j)}_{x_1} \quad \textit{for all } x_1 \in \operatorname{supp}(\mu) \setminus \{x_1^-(j), x_1^+(j)\}$$

$$\rho^{(j)}_{(x_1^-(j),x_1^-(j))} := (x_2, x_2)_*\pi^{(j)}_{x_1^-(j)} + \frac{\lambda^{(j)}}{\mu(x_1^-(j))(x_2^+(j) - x_2^-(j))}(\delta_{(x_2^-(j),x_2^+(j))} - \delta_{(x_2^-(j),x_2^-(j))})$$

$$\rho^{(j)}_{(x_1^+(j),x_1^+(j))} := (x_2, x_2)_*\pi^{(j)}_{x_1^+(j)} + \frac{\lambda^{(j)}}{\mu(x_1^+(j))(x_2^+(j) - x_2^-(j))}(\delta_{(x_2^+(j),x_2^-(j))} - \delta_{(x_2^+(j),x_2^+(j))}).$$

Set $\pi^{(j+1)}(dy_1, dy_2) := \int \rho^{(j)}(dx_1, dx_2, dy_1, dy_2)$.

Now set $j = j + 1$ and iterate (i)-(iii). Terminate if $X_1^{j,-} = \emptyset$.

Remark 5.4.4. To simplify notation, we will mostly work with the measure $\pi^{(j+1)}$ in the proofs below. In particular we will use that $\pi^{(j+1)}$ has first marginal μ and

$$\pi^{(j+1)}_{x_1} = \pi^{(j)}_{x_1} \quad \text{for all } x_1 \in \operatorname{supp}(\mu) \setminus \{x_1^-(j), x_1^+(j)\}$$

$$\pi^{(j+1)}_{x_1^-(j)} = \pi^{(j)}_{x_1^-(j)} + \frac{\lambda^{(j)}}{\mu(x_1^-(j))(x_2^+(j) - x_2^-(j))}(\delta_{x_2^+(j)} - \delta_{x_2^-(j)})$$

$$\pi^{(j+1)}_{x_1^+(j)} = \pi^{(j)}_{x_1^+(j)} + \frac{\lambda^{(j)}}{\mu(x_1^+(j))(x_2^+(j) - x_2^-(j))}(\delta_{x_2^-(j)} - \delta_{x_2^+(j)}).$$

Nevertheless the definition of $\rho^{(j)}$ will be crucial for the estimation of $\mathcal{W}^1_{nd}(\pi, \pi^{(j)})$.

Definition 5.4.5. *We denote the number of steps until termination of Algorithm 5.4.3 by $N \in \mathbb{N} \cup \{\infty\}$.*

Lemma 5.4.6. *In every step $0 \leq j \leq N$ of Algorithm 5.4.3 we have $x_1^+(j) > x_1^-(j)$ and $x_2^+(j) > x_2^-(j)$. Furthermore the measures $\pi^{(j)}$ satisfy the generalised barycentre dispersion assumption 5.4.1 for c and $\pi^{(j)} \in \Pi(\mu, \nu)$.*

Proof. By Assumption 5.4.1 there exists $x_1 > x_1^-(0)$ such that $x_1 \in X_1^{0,+}$, in particular $x_1^+(0) > x_1^-(0)$, which also implies $x_2^+(0) > x_1^+(0) - c > x_1^-(0) - c > x_2^-(0)$. We now check that the measure $\pi^{(1)}$ satisfies Assumption 5.4.1. Note that by Algorithm 5.4.3, equation (5.4.1) trivially holds for $\pi^{(1)}$ and $x > x_1^+(0)$. By definition of $\lambda^{(0)}$ we also have

$$\int (x_2 - x_1^+(0))\, \pi^{(1)}_{x_1^+(0)}(dx_2) \geq -c,$$

which implies that (5.4.1) holds for all $x \geq x_1^+(0)$ and by definition of $x_1^-(0)$ then also for all $x > x_1^-(0)$. Next we note that for all $x \leq x_1^-(0)$

$$\begin{aligned}
\int_{\{x_1 \geq x\}} (x_2 - x_1)\, \pi^{(1)}(dx_1, dx_2) &= \int_{\{x_1 \geq x\}} (x_2 - x_1)\, \pi^{(0)}(dx_1, dx_2) \\
&\quad + \mu(x_1^-(0)) \frac{\lambda^{(0)}}{\mu(x_1^-(0))(x_2^+(0) - x_2^-(0))}(x_2^+(0) - x_2^-(0)) \\
&\quad + \mu(x_1^+(0)) \frac{\lambda^{(0)}}{\mu(x_1^+(0))(x_2^+(0) - x_2^-(0))}(x_2^-(0) - x_2^+(0)) \\
&= \int_{\{x_1 \geq x\}} (x_2 - x_1)\, \pi^{(0)}(dx_1, dx_2) \geq -c,
\end{aligned}$$

so the claim follows. Lastly we show that $\pi^{(1)} \in \Pi(\mu, \nu)$. As noted before, the first marginal of $\pi^{(1)}$ is μ, so we only need to check the second marginal. For this we take a Borel set $A \subseteq \mathbb{R}$ and calculate

$$\begin{aligned}
\int_{\mathbb{R} \times A} \pi^{(1)}(dx_1, dx_2) &= \int_{\mathbb{R}} \int_A \pi^{(1)}_{x_1}(dx_2)\, \mu(dx_1) \\
&= \int_{\mathbb{R}} \int_A \pi^{(1)}_{x_1}(dx_2)\, \mu(dx_1) \\
&\quad + \mu(x_1^-(0)) \frac{\lambda^{(0)}}{\mu(x_1^-(0))(x_2^+(0) - x_2^-(0))}(\delta_{x_2^+(0)}(A) - \delta_{x_2^-(0)}(A)) \\
&\quad + \mu(x_1^+(0)) \frac{\lambda^{(0)}}{\mu(x_1^+(0))(x_2^+(0) - x_2^-(0))}(\delta_{x_2^-(0)}(A) - \delta_{x_2^+(0)}(A)) \\
&= \int_{\mathbb{R}} \int_A \pi^{(0)}_{x_1}(dx_2)\, \mu(dx_1) \\
&\quad + \frac{\lambda^{(0)}}{x_2^+(0) - x_2^-(0)}(\delta_{x_2^+(0)}(A) - \delta_{x_2^-(0)}(A) + \delta_{x_2^-(0)}(A) - \delta_{x_2^+(0)}(A)) \\
&= \int_{\mathbb{R}} \int_A \pi^{(0)}_{x_1}(dx_2)\, \mu(dx_1) = \int_{\mathbb{R} \times A} \pi^{(0)}(dx_1, dx_2).
\end{aligned}$$

Applying the above arguments inductively concludes the proof. □

Lemma 5.4.7. *Algorithm 5.4.3 terminates after at most $N \leq |\mathrm{supp}(\mu)|(1 + |\mathrm{supp}(\nu)|)$ steps.*

Proof. For $j \in \mathbb{N}_0$ and all $x_1 \in X_1^{0,+}$ we define the set

$$I^{(j)}(x_1) := \{x_2 \in \mathrm{supp}(\nu) \mid x_2 \leq \max(\mathrm{supp}(\pi^{(j)}_{x_1}))\}.$$

Similarly for all $x_1 \in X_1^{0,-}$ we set

$$I^{(j)}(x_1) := \{x_2 \in \mathrm{supp}(\nu) \mid x_2 \geq \min(\mathrm{supp}(\pi_{x_1}^{(j)}))\}.$$

By the definition of $\lambda^{(j)}$ in Algorithm 5.4.3 we note that in every step j at least one of the following three cases occurs:

(i) $|X_1^{j,+}| - |X_1^{j+1,+}| = 1$ or $|X_1^{j,-}| - |X_1^{j+1,-}| = 1$.

(ii) $|I^{(j)}(x_1^+(j))| - |I^{(j+1)}(x_1^+(j))| \geq 1$.

(iii) $|I^{(j)}(x_1^-(j))| - |I^{(j+1)}(x_1^-(j))| \geq 1$.

Combining this observation with the fact that again by the definition of $\lambda^{(j)}$ we have $X_1^{j,+} \subseteq X_1^{0,+}$ and $X_1^{j,-} \subseteq X_1^{0,-}$ as well as $I^{(j)}(x_1) \subseteq I^{(0)}(x_1)$ for all $x_1 \in X_1^{0,+} \cup X_1^{0,-}$ we conclude that the number of steps N is bounded by

$$\begin{aligned}|X_1^{0,+}| + |X_1^{0,-}| + \sum_{x_1 \in X_1^{0,-}} |I^{(0)}(x_1)| + \sum_{x_1 \in X_1^{0,+}} |I^{(0)}(x_1)| &\leq |\mathrm{supp}(\mu)| + |\mathrm{supp}(\mu)||\mathrm{supp}(\nu)| \\ &= |\mathrm{supp}(\mu)|(1 + |\mathrm{supp}(\nu)|).\end{aligned}$$

This concludes the proof. □

Proof of Proposition 5.2.4 for finitely supported $\pi \in \Pi(\mu, \nu)$. Given Lemmas 5.4.6 and 5.4.7 we only have to show that

$$\mathcal{W}_{nd}^1(\pi^{(N)}, \pi) \leq \int \left| \int (x_2 - x_1)\, \pi_{x_1}^{(0)}(dx_2) \right| \mu(dx_1).$$

Using the triangle inequality we indeed have

$$\begin{aligned}\mathcal{W}_{nd}^1(\pi^{(N)}, \pi) &\leq \int \mathcal{W}^1(\pi_{x_1}^{(N)}, \pi_{x_1})\, \mu(dx_1) \\ &\leq \sum_{j=1}^{N} \int \mathcal{W}^1(\pi_{x_1}^{(j)}, \pi_{x_1}^{(j-1)})\, \mu(dx_1) \\ &\leq \sum_{j=1}^{N} \mu(x_1^-(j-1)) \frac{\lambda^{(j-1)}}{\mu(x_1^-(j-1))(x_2^+(j-1) - x_2^-(j-1))} \\ &\qquad \cdot |x_2^+(j-1) - x_2^-(j-1)| \\ &+ \sum_{j=1}^{N} \mu(x_1^+(j-1)) \frac{\lambda^{(j-1)}}{\mu(x_1^+(j-1))(x_2^+(j-1) - x_2^-(j-1))} \\ &\qquad \cdot |x_2^-(j-1) - x_2^+(j-1)| \\ &= \sum_{j=1}^{N} 2\lambda^{(j-1)}.\end{aligned} \tag{5.4.2}$$

On the other hand, by definition of $\lambda^{(j)}$,

$$\int \left|(x_2 - x_1)\, \pi^{(j-1)}_{x_1}(dx_2)\right| \mu(dx_1) - \int \left|(x_2 - x_1)\, \pi^{(j)}_{x_1}(dx_2)\right| \mu(dx_1) = 2\lambda^{(j-1)}. \tag{5.4.3}$$

Combining (5.4.2) and (5.4.3)

$$\begin{aligned}\mathcal{W}^1_{nd}(\pi^{(N)}, \pi) &\leq \sum_{j=1}^{N} \int \left|(x_2 - x_1)\, \pi^{(j-1)}_{x_1}(dx_2)\right| \mu(dx_1) - \int \left|(x_2 - x_1)\, \pi^{(j)}_{x_1}(dx_2)\right| \mu(dx_1) \\ &\leq \int \left|(x_2 - x_1)\, \pi^{(0)}_{x_1}(dx_2)\right| \mu(dx_1),\end{aligned}$$

which shows the claim.

Lastly using Algorithm 5.4.3 in the special case $c = 0$ the claim now follows for finitely supported measures, as $\mu \preceq_c \nu$ and $X_1^{N,-} = \emptyset$ implies $X_1^{N,+} = \emptyset$. Thus $\pi^{(N)}$ is a martingale. □

5.4.2 Proof of Proposition 5.2.4 for general $\pi \in \Pi(\mu, \nu)$

Throughout this section we fix two measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ such that $\mu \preceq_c \nu$. We now extend the results from Section 5.4.1 to a general coupling $\pi \in \Pi(\mu, \nu)$ satisfying the barycentre dispersion assumption 5.2.2.

Proof of Proposition 5.2.4 for general $\pi \in \Pi(\mu, \nu)$. By Lemma 5.3.4 applied with $\kappa_n = 1/n$ there exists a sequence of finitely supported measures $(\pi^n)_{n \in \mathbb{N}}$ with

$$\lim_{n \to \infty} \mathcal{W}^1_{nd}\left(\frac{1}{n} \sum_{i=1}^{n} \pi^i, \pi\right) = 0.$$

and

$$\int_{\{x_1 \geq x\}} (x_2 - x_1)\, \pi^n(dx_1, dx_2) \geq -1/n$$

for all $x \in \mathbb{R}$ and for all $n \in \mathbb{N}$. Let us denote the marginals of π^n by μ^n and ν^n. In particular π^n satisfies Assumption 5.4.1 with $c_n = 1/n$. Applying Algorithm 5.4.3 and using the proof of Proposition 5.2.4 for finitely supported measures we can find a sequence of measures $(\pi^n_{mr})_{n \in \mathbb{N}}$ such that for all $n \in \mathbb{N}$ we have $\pi^n_{mr} \in \Pi(\mu^n, \nu^n)$,

$$\mathcal{W}^1_{nd}(\pi^n_{mr}, \pi^n) \leq \int \mathcal{W}^1(\pi^n_{mr,x_1}, \pi_{x_1})\, \mu^n(dx_1) \leq \int \left| \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) \right| \mu^n(dx_1) \tag{5.4.4}$$

and

$$\int (x_2 - x_1)\, \pi^n_{mr,x_1}(dx_2) \geq -1/n.$$

for all $x_1 \in \text{supp}(\mu^n)$. We now apply Lemma 5.3.6, which yields a measure $\pi_{mr} \in \Pi(\mu, \nu)$, such that (after taking a subsequence without relabelling)

$$\lim_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{mr}, \pi_{mr}\right) = 0.$$

In particular $\pi_{mr} \in \mathcal{M}(\mu, \nu)$. Using Lemma 5.3.3

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{mr}, \frac{1}{n}\sum_{i=1}^n \pi^i\right) \leq \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\pi^i_{mr,x_1}, \pi^i_{x_1})\, \mu^i(dx_1)$$

and by (5.4.4)

$$\begin{aligned}\mathcal{W}^1_{nd}(\pi_{mr}, \pi) &\leq \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\pi_{mr}, \frac{1}{n}\sum_{i=1}^n \pi^i_{mr}\right) + \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{mr}, \frac{1}{n}\sum_{i=1}^n \pi^i\right)\\ &\quad + \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i, \pi\right)\\ &\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \int \left|\int (x_2 - x_1)\, \pi^i_{x_1}(dx_2)\right| \mu^i(dx_1).\end{aligned}$$

The last expression is equal to

$$\int \left|\int (x_2 - x_1)\, \pi_{x_1}(dx_2)\right| \mu(dx_1)$$

as $\lim_{n\to\infty} \mathcal{W}^1(\pi^n, \pi) = 0$. This proves the claim. □

5.5 Proof of Theorem 5.2.8

Throughout this section we assume $\mu \preceq_c \nu$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$. Furthermore we make use of the notation introduced in Section 5.4 for the case $c = 0$, e.g. we write

$$\begin{aligned}X_1^{n,+} &= \left\{x_1 \in \text{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) > 0\right\},\\ X_1^{n,0} &= \left\{x_1 \in \text{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) = 0\right\},\\ X_1^{n,-} &= \left\{x_1 \in \text{supp}(\mu) \;\middle|\; \int (x_2 - x_1)\, \pi^n_{x_1}(dx_2) < 0\right\}\end{aligned}$$

for a sequence of measures $(\pi^n)_{n\in\mathbb{N}}$.

5.5.1 Proof of Theorem 5.2.8 for finitely supported $\pi \in \Pi(\mu, \nu)$ with common compact support

We prove Theorem 5.2.8 via several lemmas. We first argue for finitely supported $\pi \in \Pi(\mu, \nu)$ and write $\pi^{(0)} = \pi$. To motivate the construction in this section, let us first consider a particular case:

Lemma 5.5.1. *Assume that* $\pi \in \Pi(\mu, \nu)$ *is finitely supported and that* $X_1^0 = \emptyset$. *Then there exist pairs* $(x_1^-, x_1^+) \in X_1^{0,-} \times X_1^{0,+}$ *and* $(x_2^-, x_2^+) \in supp(\pi_{x_1^-}) \times supp(\pi_{x_1^+})$ *such that* $x_2^- < x_2^+$.

Proof. Let us write $\pi^{(0)} = \pi$ and assume towards a contradiction that the claim does not hold. We first note that for all $x_1^+ \in X_1^{0,+}$ there is $x_2^+ \in \text{supp}(\pi^{(0)}_{x_1^+})$ such that $x_2^+ > x_1^+$ and correspondingly for all $x_1^- \in X_1^{0,-}$ there is $x_2^- \in \text{supp}(\pi^{(0)}_{x_1^-})$ such that $x_2^- < x_1^-$. This implies that

$$\begin{aligned} \max\{x_1^+ \ : \ x_1^+ \in X_1^{0,+}\} < \max\{x_2^+ \ : \ x_2^+ \in X_2^{0,+}\} \leq \min\{x_2^- \ : \ x_2^- \in X_2^{0,-}\} \\ < \min\{x_1^- \ : \ x_1^- \in X_1^{0,-}\}. \end{aligned} \tag{5.5.1}$$

We conclude that

$$\{x_1 \in \text{supp}(\mu) \ : \ x_1 \leq \max(X_2^{0,+})\} = X_1^+$$

and

$$\{x_2 \in \text{supp}(\nu) \ : \ x_2 \leq \max(X_2^{0,+})\} = X_2^+.$$

Furthermore

$$\int_{\{x_1 \in X_1^+\}} (x_2 - x_1)\, \pi^{(0)}(dx_1, dx_2) > 0$$

and taking $g(x) := (\max(X_2^{0,+}) - x)^+$

$$\begin{aligned} \int g(x_1)\, \mu(dx_1) &= \int_{\{x_1 \in X_1^+\}} (\max(X_2^{0,+}) - x_1)^+\, \mu(dx_1) \\ &= \int_{\{x_1 \in X_1^+\}} (\max(X_2^{0,+}) - x_1)\, \mu(dx_1) \\ &> \int_{\{x_1 \in X_1^+\}} (\max(X_2^{0,+}) - x_2)\, \pi^{(0)}_{x_1}(dx_2)\mu(dx_1) \\ &= \int_{\{x_2 \in X_2^+\}} (\max(X_2^{0,+}) - x_2)\, \nu(dx_2) = \int g(x_2)\, \nu(dx_2). \end{aligned}$$

Noting that g is convex, this contradicts $\mu \preceq_c \nu$ and shows the claim. □

We are now ready for the general case:

Lemma 5.5.2. *Let* $j \in \mathbb{N}_0$, *assume that* $\pi^{(j)} \in \Pi(\mu, \nu)$ *is finitely supported and there exist no pairs* $(x_1^-, x_1^+) \in X_1^{j,-} \times X_1^{j,+}$ *and* $\left(x_2^-, x_2^+\right) \in supp(\pi^{(j)}_{x_1^-}) \times supp(\pi^{(j)}_{x_1^+})$ *such that* $x_2^- < x_2^+$. *Set* $x_2^+(j) := \max(X_2^{j,+}) \leq \min(X_2^{j,-}) =: x_2^-(j)$. *Then there exist vectors*

$$T_1^j := (x_1^+(j), x_1^{0,1}(j), \ldots, x_1^{0,m_j}(j), x_1^-(j))$$

and

$$T_2^j := \left(x_2^+(j), x_2^{0,1,-}(j), x_2^{0,1,+}(j), \dots, x_2^{0,m_j,+}(j), x_2^-(j)\right)$$

with $1 \leq m_j \leq |supp(\mu)|$ *such that*

$$x_2^{0,1,-}(j) < x_2^+(j) \leq x_2^{0,2,-}(j) < x_2^{0,1,+}(j) \leq \dots \leq x_2^-(j) < x_2^{0,m_j,+}(j), \tag{5.5.2}$$

where $x_2^{0,i,-}(j) = \min(\operatorname{supp}(\pi^{(j)}_{x_1^{0,i}(j)}))$ *and* $x_2^{0,i,+}(j) = \max(\operatorname{supp}(\pi^{(j)}_{x_1^{0,i}(j)}))$ *for* $i = 1, \dots, m_j$ *(see Figure 5.2).*

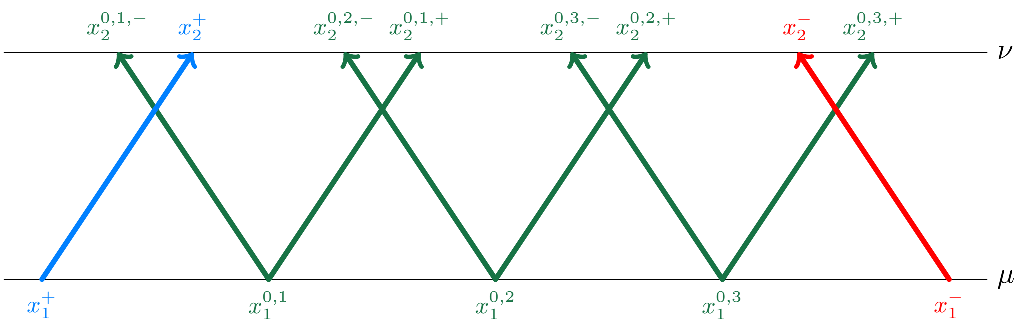

Figure 5.2: T_1^j and T_2^j for $m_j = 3$.

Proof. Note that $x_2^+(j) \leq x_2^-(j)$ follows from the assumption. We now prove the claim inductively. Thus we assume towards a contradiction, that there exists no $x_1^0 \in X_1^{j,0}$ and pair $(x_2^{0,-}, x_2^{0,+})$ with $x_2^{0,-}, x_2^{0,+} \in \operatorname{supp}(\pi^{(j)}_{x_1^0})$ and $x_2^{0,-} < x_2^+(j) < x_2^{0,+}$. This immediately implies

$$\int_{\{x_1 \leq x_2^+(j), x_2 \leq x_2^+(j)\}} (x_2 - x_1)\, \pi^{(j)}(dx_1, dx_2) > 0.$$

Set $g(x) := (x_2^+(j) - x)^+$. Then, as in the proof of Lemma 5.5.1,

$$\begin{aligned}
\int g(x_1)\, \mu(dx_1) &= \int_{\{x_1 \leq x_2^{j,+}(j)\}} (x_2^+(j) - x_1)^+\, \mu(dx_1) \\
&= \int_{\{x_1 \leq x_2^{j,+}(j)\}} (x_2^+(j) - x_1)\, \mu(dx_1) \\
&> \int_{\{x_1 \leq x_2^+(j)\}} (x_2^+(j) - x_2)\, \pi^{(j)}_{x_1}(dx_2) \mu(dx_1) \\
&= \int_{\{x_2 \leq x_2^+(j)\}} (x_2^+(j) - x_2)\, \nu(dx_2) = \int g(x_2)\, \nu(dx_2),
\end{aligned}$$

a contradiction. This shows existence of $x_1^{0,1}(j)$. Let us choose $x_2^{0,1,-}(j) := \min(\operatorname{supp}(\pi^{(j)}_{x_1^{0,1}(j)}))$ and $x_2^{0,1,+}(j) := \max(\operatorname{supp}(\pi^{(j)}_{x_1^{0,1}(j)}))$. Now we iterate the argument until $x_2^{0,k,+}(j) > x_2^-(j)$ for some $k \in \mathbb{N}$ (note that $\operatorname{supp}(\nu)$ is finite). This shows existence of vectors $x_1^{0,1}(j), \dots, x_1^{0,k}(j)$ such that $x_2^{0,i,-}(j) < x_2^{0,i-1,+}(j) < x_2^{0,i,+}(j)$. Note that if $x_2^{0,i,-}(j) <$

$x_2^{0,i-2,+}(j)$ for some $i = 2, \dots, k$ (and $x_2^{0,0}(j) := x_2^+(j)$), then $x_1^{0,i-1}(j)$ can be deleted from $(x_1^{0,1}(j), \dots, x_1^{0,k}(j))$ without changing this property. In conclusion we can assume that $x_2^{0,i,-}(j) \geq x_2^{0,i-2,+}(j)$ for all $i = 2, \dots, m_j$. This shows (5.5.2) and concludes the proof. $\square$

Definition 5.5.3. *We call the tuples T_1^j and T_2^j constructed in Lemma 5.5.2* exchange tuples for $\pi^{(j)}$, *if*

$$|T_1^j| = \min \left\{ |T_1^j| \mid \exists\, T_2^j \text{ such that } (T_1^j, T_2^j) \text{ are as in Lemma } 5.5.2 \right\}.$$

Given Lemmas 5.5.1 and 5.5.2, we now apply the following algorithm:

Algorithm 5.5.4. *Set $j = 0$.*

(i) If there exists some 4-tuple $(x_1^-, x_1^+, x_2^-, x_2^+)$ with $(x_1^-, x_1^+) \in X_1^{j,-} \times X_1^{j,+}$, $x_2^- < x_2^+$ and $\left(x_2^-, x_2^+\right) \in supp(\pi^{(j)}_{x_1^-}) \times supp(\pi^{(j)}_{x_1^+})$, then set $(x_1^-(j), x_1^+(j), x_2^-(j), x_2^+(j)) := (x_1^-, x_1^+, x_2^-, x_2^+)$ and carry out steps (ii)-(iii) of Algorithm 5.4.3. Set $m_j = 0$.

(ii) If there exist no 4-tuples $(x_1^-, x_1^+, x_2^-, x_2^+)$ with $(x_1^-, x_1^+) \in X_1^{j,-} \times X_1^{j,+}$, $x_2^- < x_2^+$ and $\left(x_2^-, x_2^+\right) \in supp(\pi^{(j)}_{x_1^-}) \times supp(\pi^{(j)}_{x_1^+})$, then choose exchange tuples for $\pi^{(j)}$ denoted by T_1^j and T_2^j. Set

$$\lambda^{(j,+)} := \mu(x_1^+(j)) \min \left\{ \tilde{\lambda} > 0 \;\middle|\; \int (x_2 - x_1^+(j))\, \pi^{(j)}_{x_1^+(j)}(dx_2) + \tilde{\lambda}(x_2^{0,1,-}(j) - x_2^{0,+}(j)) \leq 0 \right\}$$
$$\wedge\, \mu(x_1^{0,1}(j)) \pi^{(j)}_{x_1^{0,1}(j)}(x_2^{0,1,-}(j)) \wedge \mu(x_1^+(j)) \pi^{(j)}_{x_1^+(j)}(x_2^+(j))$$

and

$$\lambda^{(j,-)} := \mu(x_1^-(j)) \min \left\{ \tilde{\lambda} > 0 \;\middle|\; \int (x_2 - x_1^-(j))\, \pi^{(j)}_{x_1^-(j)}(dx_2) + \tilde{\lambda}(x_2^{0,m_j,+}(j) - x_2^-(j)) \geq 0 \right\}$$
$$\wedge\, \mu(x_1^{0,m_j}(j)) \pi^{(j)}_{x_1^{0,m_j}(j)}(x_2^{0,m_j,+}(j)) \wedge \mu(x_1^-(j)) \pi^{(j)}_{x_1^-(j)}(x_2^-(j)).$$

Furthermore for $i = 2, \dots, m_j$

$$\lambda^{(j,i)} := \mu(x_1^{0,i}(j)) \pi^{(j)}_{x_1^{0,i}(j)}(x_2^{0,i,-}(j)) \wedge \mu(x_1^{0,i-1}(j)) \pi^{(j)}_{x_1^{0,i-1}(j)}(x_2^{0,i-1,+}(j)).$$

and set

$$\lambda^{(j)} = \lambda^{(j,+)}(x_2^{0,+}(j) - x_2^{0,1,-}(j)) \wedge \min_{i=2,\dots,m_j-1} \lambda^{(j,i)}(x_2^{0,i-1,+}(j) - x_2^{0,i,-}(j))$$
$$\wedge\, \lambda^{(j,-)}(x_2^{0,m_j,+}(j) - x_2^-(j)) > 0.$$

Now define $\rho^{(j)} \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R}^2)$ with via $\rho^{(j)} := ((x_1, x_1)_* \mu) \otimes \rho^{(j)}_{(x_1,x_1)}$, where

$$\begin{aligned}
\rho^{(j)}_{(x_1,x_1)} &:= (x_2, x_2)_* \pi^{(j)}_{x_1} \qquad \text{for all } x_1 \in \operatorname{supp}(\mu) \setminus T_1^j \\
\rho^{(j)}_{(x_1^+(j), x_1^+(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^+(j)} + \frac{\lambda^{(j)}}{\mu(x_1^+(j))(x_2^{0,+}(j) - x_2^{0,1,-}(j))} \\
&\quad \cdot (\delta_{((x_2^+(j), x_2^{0,1,-}(j))} - \delta_{((x_2^+(j), x_2^+(j))}) \\
\rho^{(j)}_{(x_1^{0,1}(j), x_1^{0,1}(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^{0,1}(j)} + \frac{\lambda^{(j)}}{\mu(x_1^{0,1}(j))(x_2^{0,+}(j) - x_2^{0,1,-}(j))} \\
&\quad \cdot (\delta_{(x_2^{0,1,-}(j), x_2^+(j))} - \delta_{(x_2^{0,1,-}(j), x_2^{0,1,-}(j))}) \\
\rho^{(j)}_{(x_1^-(j), x_1^-(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^-(j)} + \frac{\lambda^{(j)}}{\mu(x_1^-(j))(x_2^{0,m_j,+}(j) - x_2^-(j))} \\
&\quad \cdot (\delta_{(x_2^-(j), x_2^{0,m_j,+}(j))} - \delta_{(x_2^-(j), x_2^-(j))}) \\
\rho^{(j)}_{(x_1^{0,m_j}(j), x_1^{0.m_j}(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^{0,m_j}(j)} + \frac{\lambda^{(j)}}{\mu(x_1^{m_j}(j))(x_2^{0,m_j,+}(j) - x_2^-(j))} \\
&\quad \cdot (\delta_{(x_2^{0,m_j,+}(j), x_2^-(j))} - \delta_{(x_2^{0,m_j,+}(j), x_2^{0,m_j,+}(j))})
\end{aligned}$$

and for $i = 2, \dots, m_j$

$$\begin{aligned}
\rho^{(j)}_{(x_1^{0,i}(j), x_1^{0,i}(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^{0,i}(j)} + \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))(x_2^{0,i-1,+}(j) - x_2^{0,i,-}(j)))} \\
&\quad \cdot (\delta_{(x_2^{0,i,-}(j), x_2^{0,i-1,+}(j))} - \delta_{(x_2^{0,i,-}(j), x_2^{0,i,-}(j))}) \\
\rho^{(j)}_{(x_1^{0,i-1}(j), x_1^{0,i-1}(j))} &:= (x_2, x_2)_* \pi^{(j)}_{x_1^{0,i-1}(j)} + \frac{\lambda^{(j)}}{\mu(x_1^{0,i-1}(j))(x_2^{0,i-1,+}(j) - x_2^{0,i,-}(j)))} \\
&\quad \cdot (\delta_{(x_2^{0,i-1,+}(j), x_2^{0,i,-}(j))} - \delta_{(x_2^{0,i-1,+}(j), x_2^{0,i-1,+}(j))}).
\end{aligned}$$

Set $\pi^{(j+1)}(dy_1, dy_2) := \int \rho^{(j)}(dx_1, dx_2, dy_1, dy_2)$.

Set $j = j + 1$. Now iterate (i)-(ii) and terminate if $\pi^{(j)} \in \mathcal{M}(\mu, \nu)$ for some $j \in \mathbb{N}$. In that case, set $\pi_{mr} := \pi^{(j)}$.

Definition 5.5.5. *We denote the number of steps until termination of Algorithm 5.5.4 by $N \in \mathbb{N} \cup \{\infty\}$.*

Remark 5.5.6. As in Section 5.4 we will mostly work with the definition of $\pi^{(j)}$ directly in order to shorten notation. Nevertheless, to make arguments in the proof of Lemma 5.5.13 and Section 5.5.2 precise, we will sometimes use $\rho^{(j)}$ directly. In any case, we conclude that for $j = 0, \dots, N - 1$

$$\mathcal{W}^1_{nd}(\pi^{(j)}, \pi^{(j+1)}) \le \int \int |x_2 - y_2| \, \rho^{(j)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1).$$

Let us now give two more definitions:

Definition 5.5.7. *If $N < \infty$ then we define*

$$\rho(dz^0, dz^1, dz^2, \dots, dz^N) := \rho^{(N-1)}_{z^{N-1}}(dz^N) \dots \rho^{(1)}_{z^1}(dz^2)\rho^{(0)}(dz^0, dz^1),$$

where $z^1, \dots, z^N \in \mathbb{R}^2$. Furthermore, for $0 \le j \le \hat{j} \le N-1$ we set

$$\rho^{(j:\hat{j})}(dx, dy) := \int \rho(dz^0, \dots, dz^{j-1}, dx, dz^{j+1}, \dots, dz^{\hat{j}-1}, dy, dz^{\hat{j}+1}, \dots, dz^N).$$

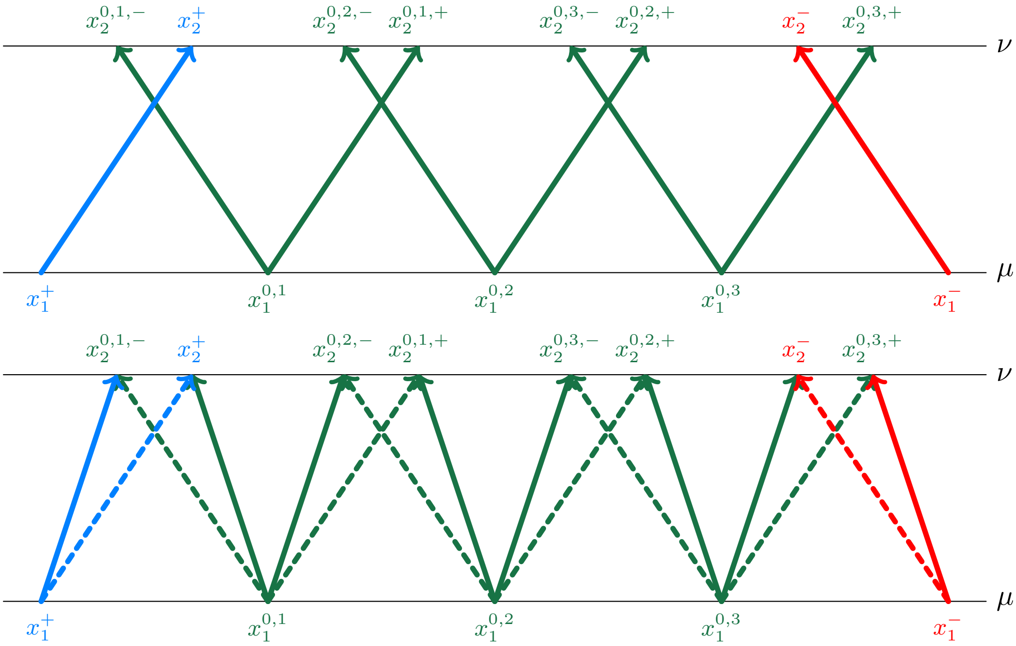

Figure 5.3: Exchange of masses at $x_2^{0,1,-} < x_2^+ \le x_2^{0,2,-} < x_2^{0,1,+} \le x_2^{0,3,-} < x_2^{0,2,+} \le x_2^- < x_2^{0,3,+}$ for the case $m_j = 3$.

Applying Algorithm 5.5.4 recursively, we "rectify" the barycentres of both $\pi^{(j)}_{x_1^+}$ and $\pi^{(j)}_{x_1^-}$, i.e. we shift both of them by the same amount to the left and right respectively. If we are in case *(ii)*, the barycentres of the disintegrations at points $(x_1^{0,1}(j), \dots, x_1^{0,m_j}(j))$ still remain zero (see Figure 5.3). We formally prove this in the following lemmas, which list important properties of Algorithm 5.5.4.

Lemma 5.5.8. *The following properties hold for Algorithm 5.5.4:*

(i) *If $x_1 \in X_1^{\tilde{j},0}$, then $x_1 \in X_1^{j,0}$ for all $j \ge \tilde{j}$.*

(ii) *For any $0 \le j \le N$ we have $\pi^{(j)} \in \Pi(\mu, \nu)$.*

(iii) *For any $0 \le j \le N-1$ we have*

$$\int \left| \int (x_2 - x_1)\, \pi^{(j)}_{x_1}(dx_2) \right| \mu(dx_1) - \int \left| \int (x_2 - x_1)\, \pi^{(j+1)}_{x_1}(dx_2) \right| \mu(dx_1) = 2\lambda^{(j)}.$$

and

$$\mathcal{W}^1_{nd}(\pi^{(j)}, \pi^{(j+1)}) \leq \int \mathcal{W}^1(\pi^{(j)}_{x_1}, \pi^{(j+1)}_{x_1})\, \mu(dx_1) \leq 2(m_j+1)\lambda^{(j)}.$$

Proof. Recalling the observations in Section 5.4.1, we only have to prove the claims for steps j, in which case *(ii)* in Algorithm 5.5.4 is applied.
(i): For any element $x_1^{0,i}(j) \in T_1^j \cap X_1^{j,0}$, $i = 1, \dots, m_j$ we have by Algorithm 5.5.4

$$\begin{aligned}
\int (x_2 - x_1^{0,i}(j))\, \pi^{(j+1)}_{x_1^{0,i}}(j)(dx_2) &= \int (x_2 - x_1^{0,i}(j))\, \pi^{(j)}_{x_1^{0,i}}(j)(dx_2) \\
&+ \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))(x_2^{0,i-1,+}(j) - x_2^{0,i,-}(j))}(x_2^{0,i-1,+}(j) - x_2^{0,i,-}(j)) \\
&+ \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))(x_2^{0,i,+}(j) - x_2^{0,i+1,-}(j))}(x_2^{0,i+1,-}(j) - x_2^{0,i,+}(j)) \\
&= \int (x_2 - x_1^{(0,i)}(j))\, \pi^{(j)}_{x_1^{(0,i)}}(j)(dx_2) + \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))} - \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))} \\
&= 0.
\end{aligned}$$

(ii) We note that it is sufficient to check $\pi^{(j+1)}(\mathbb{R} \times \{x_2\}) = \pi^{(j)}(\mathbb{R} \times \{x_2\})$ for all $x_2 \in T_2^j$. To see this let us consider $x_2^{0,i,+}(j) \in T_2^j$ and calculate

$$\begin{aligned}
\pi^{(j+1)}(\mathbb{R} \times \{x_2^{0,i,+}(j)\}) = \pi^{(j)}(\mathbb{R} \times \{x_2^{0,i,+}(j)\}) &+ \mu(x_1^{0,i+1}(j)) \frac{\lambda^{(j)}}{\mu(x_1^{0,i+1}(j))(x_2^{0,i,+}(j) - x_2^{0,i+1,-})} \\
&- \mu(x_1^{0,i}(j)) \frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))(x_2^{0,i,+}(j) - x_2^{0,i+1,-})} \\
&= \pi^{(j)}(\mathbb{R} \times \{x_2^{0,i,+}(j)\}).
\end{aligned}$$

The cases $x_2^+, x_2^-, x_2^{0,i,-} \in T_2^j$ work analogously.
(iii) The first claim follows from *(i)* and the observation that

$$\begin{aligned}
&\int (x_2 - x_1^+(j))\, \pi^{(j)}_{x_1^+(j)}(dx_2) - \int (x_2 - x_1^+(j))\, \pi^{(j+1)}_{x_1^+(j)}(dx_2) \\
&= \frac{\lambda^{(j)}}{\mu(x_1^+(j))(x_2^{0,+}(j) - x_2^{0,1,-}(j))}(x_2^+(j) - x_2^{0,1,-}(j)) = \frac{\lambda^{(j)}}{\mu(x_1^+)}
\end{aligned}$$

and

$$\begin{aligned}
&\int (x_2 - x_1^-(j))\, \pi^{(j+1)}_{x_1^-(j)}(dx_2) - \int (x_2 - x_1^-(j))\, \pi^{(j)}_{x_1^-(j)}(dx_2) \\
&= \frac{\lambda^{(j)}}{\mu(x_1^-(j))(x_2^{0,m_j,+}(j) - x_2^-(j))}(x_2^{0,m_j,+}(j) - x_2^-(j)) = \frac{\lambda^{(j)}}{\mu(x_1^-)}.
\end{aligned}$$

We now show the second claim. Similarly to *(i)* we conclude that for all $x_1^{0,i}(j) \in T_1^j$

$$\mathcal{W}^1(\pi^{(j)}_{x_1^{0,i}(j)}, \pi^{(j+1)}_{x_1^{0,i}(j)}) \leq 2\frac{\lambda^{(j)}}{\mu(x_1^{0,i}(j))}.$$

Furthermore for $x_1^+(j) \in T_1^j$

$$\mathcal{W}^1(\pi^{(j)}_{x_1^+(j)}, \pi^{(j+1)}_{x_1^+(j)}) \leq \frac{\lambda^{(j)}}{\mu(x_1^+(j))(x_2^+(j) - x_2^{0,1,-}(j))} |x_2^{0,1,-}(j) - x_2^+(j)| = \frac{\lambda^{(j)}}{\mu(x_1^+(j))}$$

and similarly for $x_1^-(j) \in T_1^j$

$$\mathcal{W}^1(\pi^{(j)}_{x_1^-(j)}, \pi^{(j+1)}_{x_1^-(j)}) \leq \frac{\lambda^{(j)}}{\mu(x_1^-(j))(x_2^{0,m_j,+}(j) - x_2^-(j))} |x_2^{0,m_j,+}(j) - x_2^-(j)| = \frac{\lambda^{(j)}}{\mu(x_1^-(j))}.$$

Writing

$$\begin{aligned}\mathcal{W}^1_{nd}(\pi^{(j)}, \pi^{(j+1)}) &\leq \int \mathcal{W}^1(\pi^{(j)}_{x_1}), \pi^{(j+1)}_{x_1})\, \mu(dx_1) \\ &\leq \mu(x_1^+(j))\; \mathcal{W}^1(\pi^{(j)}_{x_1^+(j)}, \pi^{(j+1)}_{x_1^+(j)}) + \sum_{i=1}^{m_j} \mu(x_1^{0,i}(j))\; \mathcal{W}^1(\pi^{(j)}_{x_1^{0,i}}, \pi^{(j+1)}_{x_1^{0,i}}) \\ &+ \mu(x_1^-(j))\; \mathcal{W}^1(\pi^{(j)}_{x_1^-(j)}, \pi^{(j+1)}_{x_1^-(j)}) \leq 2(m_j+1)\lambda^{(j)}\end{aligned}$$

concludes the proof. □

Lemma 5.5.9. *Let us take a finitely supported measure $\pi \in \Pi(\mu, \nu)$, set $\pi^{(0)} = \pi$ and assume that there exist no pairs $(x_1^-, x_1^+) \in X_1^{0,-} \times X_1^{0,+}$ and $\left(x_2^-, x_2^+\right) \in supp(\pi^{(0)}_{x_1^-}) \times supp(\pi^{(0)}_{x_1^+})$ such that $x_2^- < x_2^+$. If we apply Algorithm 5.5.4 to $\pi^{(0)} = \pi$, then the following hold:*

(i) For any steps $j \geq \tilde{j}$ and any $x_1 \in X_1^{\tilde{j},0}$ we have

$$\min(\mathrm{supp}(\pi^{(\tilde{j})}_{x_1})) \leq \min(\mathrm{supp}(\pi^{(j)}_{x_1})) \leq \max(\mathrm{supp}(\pi^{(j)}_{x_1})) \leq \max(\mathrm{supp}(\pi^{(\tilde{j})}_{x_1})).$$

(ii) In every step $0 \leq j \leq N$ there are no pairs $(x_1^-, x_1^+) \in X_1^{j,-} \times X_1^{j,+}$ and $\left(x_2^-, x_2^+\right) \in$ $\mathrm{supp}(\pi^{(j)}_{x_1^-}) \times \mathrm{supp}(\pi^{(j)}_{x_1^+})$ such that $x_2^- < x_2^+$.

(iii) For any $j > \tilde{j}$ and $x_1^0 \in X_1^{j,0} \setminus X_1^{\tilde{j},0}$ we have $\mathrm{supp}(\pi^{(j)}_{x_1}) \cap (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-})) = \emptyset$.

(iv) If $j \geq \tilde{j}$ and either $x_2^{0,i,-}(j) \in T_2^j$ or $x_2^{0,i,+}(j) \in T_2^j$ (or both) is contained in the interval $(\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-}))$ for some $i \in \{1, \dots, m_j\}$ then $x_1^{0,i}(j) \in X_1^{\tilde{j},0} \cap T_1^j$.

Proof. *(i)*: this follows immediately from the definition of $x_2^{0,i,-}(j)$ and $x_2^{0,i,+}(j)$ and (5.5.2) in Lemma 5.5.2 as well as the construction of $\pi^{(j+1)}$ in Algorithm 5.5.4.
(ii): the assertion is true for $j = 0$ by assumption. It then follows for all $j \in \{1, \dots, N\}$ by observing

$$\max(X_2^{j,+}) \leq \max(X_2^{0,+}) \leq \min(X_2^{0,-}) \leq \min(X_2^{j,-}), \tag{5.5.3}$$

which is implied by (5.5.2) in Lemma 5.5.2 and Algorithm 5.5.4.
(iii): by Lemma 5.5.8 $X_1^{\tilde{j},0} \subseteq X_1^{j,0}$ holds. As $x_1^0 \in X_1^{j,0} \setminus X_1^{\tilde{j},0}$ we have $x_1^0 \in X_1^{\tilde{j},+} \cup X_1^{\tilde{j},-}$ and thus clearly $\operatorname{supp}(\pi_{x_1}^{(\tilde{j})}) \cap (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-})) = \emptyset$. Now by (5.5.3) and *(i)* we have $\operatorname{supp}(\pi_{x_1}^{(j)}) \cap (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-})) = \emptyset$.
(iv): let us assume that $x_2^{0,i,+}(j) \in (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-}))$ for some $i \in \{1, \dots, m_j\}$. Furthermore let us assume towards a contradiction that $x_1^{0,i}(j) \in X_1^{j,0} \setminus X_1^{\tilde{j},0}$. Then *(iii)* implies

$$\operatorname{supp}(\pi^{(j)}_{x_1^{0,i}(j)}) \cap (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-})) = \emptyset,$$

a contradiction. The case $x_2^{0,i,-}(j) \in (\max(X_2^{\tilde{j},+}), \min(X_2^{\tilde{j},-}))$ is analogous. □

Having established these basic properties of Algorithm 5.5.4 it is now straightforward to conclude $N < \infty$. More concretely we have the following lemma:

Lemma 5.5.10. *Algorithm 5.5.4 terminates after at most $N \leq |\operatorname{supp}(\mu)|(1 + |\operatorname{supp}(\nu)|)$ steps.*

Proof. For $j \in \mathbb{N}_0$ define the set

$$I^{(j)}(x_1) := \{x_2 \in \operatorname{supp}(\nu) \mid x_2 \leq \max(\operatorname{supp}(\pi_{x_1}^{(j)}))\}$$

for all $x_1 \in X_1^{j,+}$,

$$I^{(j)}(x_1) := \{x_2 \in \operatorname{supp}(\nu) \mid \min(\operatorname{supp}(\pi_{x_1}^{(j)})) \leq x_2 \leq \max(\operatorname{supp}(\pi_{x_1}^{(j)}))\}$$

for all $x_1 \in X_1^{j,0}$ and

$$I^{(j)}(x_1) := \{x_2 \in \operatorname{supp}(\nu) \mid x_2 \geq \min(\operatorname{supp}(\pi_{x_1}^{(j)}))\}$$

for all $x_1 \in X_1^{j,-}$. By the definition of $\lambda^{(j)}$ in Algorithm 5.5.4 we note that in every step j at least one of the following four cases occurs:

(i) $|X_1^{j,+}| - |X_1^{j+1,+}| = 1$ or $|X_1^{j,-}| - |X_1^{j+1,-}| = 1$.

(ii) $|I^{(j)}(x_1^+(j))| - |I^{(j+1)}(x_1^+(j))| \geq 1$.

(iii) $|I^{(j)}(x_1)| - |I^{(j+1)}(x_1)| \geq 1$ for some $x_1 \in X_1^{j,0}$.

(iv) $|I^{(j)}(x_1^-(j))| - |I^{(j+1)}(x_1^-(j))| \geq 1$.

Combining this observation with the fact that again by the definition of $\lambda^{(j)}$ we have $X_1^{j,+} \subseteq X_1^{0,+}$ and $X_1^{j,-} \subseteq X_1^{0,-}$ as well as $I^{(j)}(x_1) \subseteq I^{(0)}(x_1)$ for all $x_1 \in \text{supp}(\mu)$, we conclude that the number of steps N is bounded by

$$\begin{aligned}
&|X_1^{0,+}| + |X_1^{0,-}| + \sum_{x_1 \in X_1^{0,-}} |I^{(0)}(x_1)| + \sum_{x_1 \in X_1^{0,0}} |I^{(0)}(x_1)| + \sum_{x_1 \in X_1^{0,+}} |I^{(0)}(x_1)| \\
&\leq |\text{supp}(\mu)| + |\text{supp}(\mu)||\text{supp}(\nu)| = |\text{supp}(\mu)|(1 + |\text{supp}(\nu)|).
\end{aligned}$$

This concludes the proof. □

We now argue that (5.2.7) holds for finitely supported measures $\pi \in \Pi(\mu, \nu)$. By the triangle inequality

$$\mathcal{W}^1_{nd}(\pi^{(0)}, \pi^{(N)}) \leq \int \mathcal{W}^1(\pi^{(0)}_{x_1}, \pi^{(N)}_{x_1})\, \mu(dx_1) \leq \sum_{j=0}^{N-1} \int \mathcal{W}^1(\pi^{(j)}_{x_1}, \pi^{(j+1)}_{x_1})\, \mu(dx_1). \tag{5.5.4}$$

Thus it is sufficient to consider $\int \mathcal{W}^1(\pi^{(j)}_{x_1}, \pi^{(j+1)}_{x_1})\, \mu(dx_1)$ individually for each $j \in \{1, \dots, N-1\}$. Furthermore, if in step j of Algorithm 5.5.4 case *(i)* is carried out , then by Lemma 5.5.8.*(iii)* we have

$$\begin{aligned}
\int \mathcal{W}^1(\pi^{(j)}_{x_1}, \pi^{(j+1)}_{x_1})\, \mu(dx_1) \leq 2\lambda^{(j)}& \\
&= \int \left| \int (x_2 - x_1)\, \pi^{(j)}(dx_2) \right| \mu(dx_1) \\
&\quad - \int \left| \int (x_2 - x_1)\, \pi^{(j+1)}(dx_2) \right| \mu(dx_1).
\end{aligned}$$

To simplify notation we thus make the following standing assumption for the remainder of this section:

Assumption 5.5.11. *Let $\pi \in \Pi(\mu, \nu)$ be such that there exist no pairs $(x_1^-, x_1^+) \in X_1^- \times X_1^+$ and $\left(x_2^-, x_2^+\right) \in supp(\pi_{x_1^-}) \times supp(\pi_{x_1^+})$ such that $x_2^- < x_2^+$.*

Recalling Lemma 5.5.9.*(ii)*, Assumption 5.5.11 is then satisfied for $\pi^{(j)} \in \Pi(\mu, \nu)$ for all $j \in \{0, \dots, N-1\}$.

For notational convenience we make the following additional conventions for the rest of this section: we set $x_1^{0,0}(j) := x_1^+(j)$ and $x_1^{0,m_j+1}(j) := x_1^-(j)$ as well as $x_2^{0,0,+}(j) := x_2^+(j)$ and $x_2^{0,m_0+1,-}(j) := x_2^-(j)$ for all $j = 1, \dots, N$. This is particularly useful in counting arguments, where we do not need to stress the special role of $x_1^+(j)$ or $x_1^-(j)$ respectively.

In order get some intuition for the general result we now treat the case $N = 1$.

Lemma 5.5.12. *Assume that $N = 1$. Then $\lambda^{(0)}$ is of order $1/m_0^2$.*

Proof. Note that by (5.5.2) the intervals $\left[x_2^{0,i+1,-}(0), x_2^{0,i,+}(0)\right)$, $i = 0, 1, 2, \ldots, m_0$ are disjoint (see Figure 5.2). Furthermore, by definition of $\lambda^{(0)}$

$$\mu\left(x_1^{0,i+1}(0)\right) \geq \pi^{(0)}\left(\left(x_1^{0,i+1}(0), x_2^{0,i+1,-}(0)\right)\right) \geq \frac{\lambda^{(0)}}{x_2^{0,i,+}(0) - x_2^{0,i+1,-}(0)}, \tag{5.5.5}$$

$i = 0, 1, 2, \ldots, m_0$, has to hold. Summing (5.5.5) over $i = 0, 1, 2, \ldots, m_0$ this implies

$$1 \geq \sum_{i=0}^{m_0} \mu\left(x_1^{0,i+1}(0)\right) \geq \sum_{i=0}^{m_0} \frac{\lambda^{(0)}}{x_2^{0,i,+}(0) - x_2^{0,i+1,-}(0)} \geq \frac{\lambda^{(0)}(m_0+1)^2}{\tilde{K}}, \tag{5.5.6}$$

where the last inequality follows from the arithmetic-harmonic mean inequality as $\left[x_2^{0,i+1,-}(0), x_2^{0,i,+}(0)\right)$, $i = 0, 1, 2, \ldots, m_0$ are disjoint and $\text{supp}(\nu) \subseteq [-\tilde{K}, \tilde{K}]$ for some $\tilde{K} > 0$. This shows the desired growth for $\lambda^{(0)}$. □

The general case follows by more involved arguments. In particular we will identify points in the support of μ in each step j of Algorithm 5.5.4, where barycentre mass is only ever shifted in one direction (namely downwards in our case). These yield an upper bound on $\sum_{j=0}^{N-1} \lambda^{(j)}$ by a similar argument as in Lemma 5.5.12 above:

Lemma 5.5.13. *Fix $\delta > 0$ and assume that supp$(\nu) \subseteq [-\tilde{K}, \tilde{K}]$. Let $\delta > 0$. Then there exists a constant $K = K(\delta, \tilde{K})$ such that*

$$\inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}_{nd}^1(\pi, \tilde{\pi}) \leq K\varepsilon_\pi + \delta. \tag{5.5.7}$$

Proof. Let us fix $m \geq 5$ and define $J(m) := \{j \in \{1, \ldots, N-1\} \;:\; m_j \geq m\}$ as well as $\tilde{j} := \min(J(m))$. Next we define the disjoint intervals

$$\begin{aligned}
A_1 &:= \left[x_2^{0,1,-}(\tilde{j}), x_2^{0,5,-}(\tilde{j})\right), \\
A_2 &:= \left[x_2^{0,5,-}(\tilde{j}), x_2^{0,9,-}(\tilde{j})\right), \\
\vdots \quad & \quad \vdots \\
A_{\tilde{m}} &:= \left[x_2^{0,4(\tilde{m}-1)+1,-}(\tilde{j}), x_2^{0,4\tilde{m}+1,-}(\tilde{j})\right),
\end{aligned} \tag{5.5.8}$$

where $\tilde{m} := \lfloor m_{\tilde{j}}/4 \rfloor$. Lemma 5.5.14 states, that for every $j = \tilde{j}, \ldots, N-1$ and every $k = 1, \ldots, \tilde{m}$ there exists $i_{j,k} \in \{1, \ldots, m_j\}$ such that

$$[x_2^{0,i_{j,k},-}(j), x_2^{0,i_{j,k},+}(j)) \subseteq A_k.$$

We make the important convention that $[x_2^{0,i_{j,k},-}(j), x_2^{0,i_{j,k},+}(j))$ is the left-most such interval, i.e.

$$i_{j,k} := \min\{i \in \{1, \ldots, m_j\} \mid [x_2^{0,i,-}(j), x_2^{0,i,+}(j)) \subseteq A_k\}.$$

We denote the corresponding left-neighbouring points $x_1^{0,i_{j,k}-1}(j)$ in T_1^j by $\overleftarrow{x}_1^{j,k}$. Fix now $x_1 \in \operatorname{supp}(\mu)$. We make three important observations:

- First we note that by the definition of $i_{j,k}$ it follows that if $x_1 = \overleftarrow{x}_1^{j,k} = \overleftarrow{x}_1^{\hat{j},\hat{k}}$ for some $j, \hat{j} \geq \tilde{j}$ and some $1 \leq k, \hat{k} \leq \tilde{m}$, then $k = \hat{k}$.

- Secondly, assume that $x_1 = \overleftarrow{x}_1^{j,k}$ for some $\tilde{j} \leq j \leq N-1$ and some $1 \leq k \leq \tilde{m}$. Now if $x_1 = x_1^{0,i}(\hat{j})$ for some $1 \leq \hat{j} \leq N$, $0 \leq i \leq m_{\hat{j}} + 1$ and $x_2^{0,i,-}(\hat{j}) = \min(\operatorname{supp}(\pi^{(\hat{j})}_{x_1^{0,i}(\hat{j})})) \geq \min(A_k)$, then it follows from the definition of $\overleftarrow{x}_1^{j,k}$ as well as Lemma 5.5.9.*(i)* that $\hat{j} > j$. In particular $\hat{j} \geq \max\{\bar{j} \in \{\tilde{j}, \dots, N-1\} \mid x_1 = \overleftarrow{x}_1^{\bar{j},k}\}$.

- Lastly, assume that $x_1 = x_1^i(j) \in T_1^j$ for some $\tilde{j} \leq j \leq N-1$ and some $1 \leq i \leq m_j$. If $x_2^{0,i,+}(j) \geq \min(A_k)$ and $x_2^{0,i+1,-}(j) < \min(A_k)$ then *(i)* of Lemma 5.5.9 yields $\max(\operatorname{supp}(\pi^{(\hat{j})}_{x_1})) \leq x_2^{0,i,+}(j)$ for all $\hat{j} > j$. In particular, if $x_1 = x_1^{\hat{i}}(\hat{j}) \in T_1^{\hat{j}}$ for some $\hat{j} \geq j$ and some $1 \leq \hat{i} \leq m_{\hat{j}}$, then necessarily $x_2^{0,\hat{i},+}(\hat{j}) \leq x_2^{0,i,+}(j)$.

Assume now that for $\hat{j} > j \geq \tilde{j}$ we have $x_1 = \overleftarrow{x}_1^{j,k} = \overleftarrow{x}_1^{\hat{j},k}$, $x_1 = x_1^{0,i}(j) = x_1^{0,\hat{i}}(\hat{j})$ and

$$((x_1, x_2^{0,i+1,-}(j)), (x_1, x_2^{0,\hat{i},+}(\hat{j})) \in \operatorname{supp}(\rho^{(j:\hat{j})}_{(x_1,x_1)}).$$

Combining the three observations above we conclude that $x_2^{0,i+1,-}(j) \geq x_2^{0,\hat{i},+}(\hat{j})$.

Next let us define $J(m,k,x_1) := \{j \in J(m) \ : \ \overleftarrow{x}_1^{j,k} = x_1\}$. Together with the definition of $\lambda^{(j)}$ in Algorithm 5.5.4 and the definition of ρ in Definition 5.5.7 we then conclude that

$$\begin{aligned}
\sum_{j \in J(m,k,x_1)} \lambda^{(j)} &= \sum_{j \in J(m,k,x_1)} \mu(x_1) \int_{A_k \times A_k} (x_2 - y_2)\, \rho^{(j)}_{(x_1,x_1)}(dx_2, dy_2) \\
&= \mu(x_1) \int \sum_{j \in J(m,k,x_1)} (z_2^j - z_2^{j+1})\, \rho_{(x_1,\dots,x_1)}(dz_2^0, \dots, dz_2^N) \\
&\leq \mu(x_1) |A_k|,
\end{aligned}$$

where we used a telescoping argument and the definition of $\overleftarrow{x}_1^{j,k}$ for the last inequality. Summing over $x_1 \in \operatorname{supp}(\mu)$ and $k = 1, \dots, \tilde{m}$ this implies

$$\sum_{k=1}^{\tilde{m}} \sum_{j \in J(m)} \frac{\lambda^{(j)}}{|A_k|} = \sum_{k=1}^{\tilde{m}} \sum_{x_1 \in \operatorname{supp}(\mu)} \sum_{j \in J(m,k,x_1)} \frac{\lambda^{(j)}}{|A_k|} \leq \sum_{x_1 \in \operatorname{supp}(\mu)} \mu(x_1).$$

In particular

$$\begin{aligned}
1 = \sum_{x_1 \in \operatorname{supp}(\mu)} \mu(x_1) &\geq \sum_{j \in J(m)} \sum_{k=1}^{\tilde{m}} \frac{\lambda^{(j)}}{|A_k|} \\
&= \sum_{j \in J(m)} \sum_{k=1}^{\tilde{m}} \frac{\lambda^{(j)}}{x_2^{0,4k+1,-} - x_2^{0,4(k-1)+1,-}} \\
&\geq \sum_{j \in J(m)} \frac{\lambda^{(j)}\, \tilde{m}^2}{\tilde{K}} \geq \sum_{j \in J(m)} \frac{\lambda^{(j)}\, (m-3)^2}{16 \tilde{K}}
\end{aligned}$$

noting that $\tilde{m} \geq (m-3)/4$. This implies

$$\sum_{j \in J(m)} \lambda^{(j)} \leq \frac{16\tilde{K}}{(m-3)^2}$$

and thus

$$\sum_{j \in J(m)} 2(m_j+1)\,\lambda^{(j)} = 2m \sum_{j \in J(m)} \lambda^{(j)} + 2 \sum_{\hat{m} \geq m} \sum_{j \in J(\hat{m})} \lambda^{(j)} \leq \frac{32\tilde{K}m}{(m-3)^2} + \sum_{\hat{m} \geq m} \frac{32\tilde{K}}{(\hat{m}-3)^2}. \tag{5.5.9}$$

Given $\delta > 0$ there exists $m > 0$ such that the last sum on the right hand side of (5.5.9) is less than δ. We take the smallest such m and define $K(\delta, \tilde{K}) := (m+1)$. Using Lemma 5.5.8.*(iii)* and the triangle inequality we conclude that

$$\begin{aligned}
\mathcal{W}^1_{nd}(\pi^{(0)}, \pi^{(N)}) &\leq \int \mathcal{W}^1(\pi^{(0)}_{x_1}, \pi^{(N)}_{x_1})\,\mu(dx_1) \\
&\leq \sum_{j=0}^{N-1} \int \mathcal{W}^1(\pi^{(j)}_{x_1}, \pi^{(j+1)}_{x_1})\,\mu(dx_1) \\
&\leq 2\sum_{j=0}^{N-1} (m_j+1)\lambda^{(j)} \\
&= 2 \sum_{j \in J(K(\delta,\tilde{K}))} (m_j+1)\lambda^{(j)} + 2 \sum_{j \notin J(K(\delta,\tilde{K}))} (m_j+1)\lambda^{(j)} \\
&\leq \delta + K(\delta,\tilde{K}) \sum_{j \notin J(K(\delta,\tilde{K}))} \int \left| \int (x_2 - x_1)\,\pi^{(j)}_{x_1}(dx_2) \right| \mu(dx_1) \\
&\qquad - \int \left| \int (x_2 - x_1)\,\pi^{(j+1)}_{x_1}(dx_2) \right| \mu(dx_1) \\
&\leq \delta + K(\delta,\tilde{K}) \sum_{j=0}^{N-1} \int \left| \int (x_2 - x_1)\,\pi^{(j)}_{x_1}(dx_2) \right| \mu(dx_1) \\
&\qquad - \int \left| \int (x_2 - x_1)\,\pi^{(j+1)}_{x_1}(dx_2) \right| \mu(dx_1) \\
&= \delta + K(\delta,\tilde{K})\varepsilon_\pi.
\end{aligned}$$

In particular $K = K(\delta, \tilde{K})$ depends on δ and the support of ν (via $\tilde{K}$) only. □

We have used the following lemma:

Lemma 5.5.14. *Let $\mu \in \Pi(\mu, \nu)$ be finitely supported, fix $m \in \mathbb{N}$ and recall $J(m) = \{j \in \{1, \ldots, N\} \ : \ m_j \geq m\}$ and $\tilde{j} := \min(J(m))$. Then for every $j = \tilde{j}, \ldots, N-1$ and every $k = 1, \ldots, \tilde{m} = \lfloor m_{\tilde{j}}/4 \rfloor$, there exists $i_{j,k} \in \{1, \ldots, m_j\}$ such that $[x_2^{0,i_{j,k},-}(j), x_2^{0,i_{j,k},+}(j)) \subseteq A_k$.*

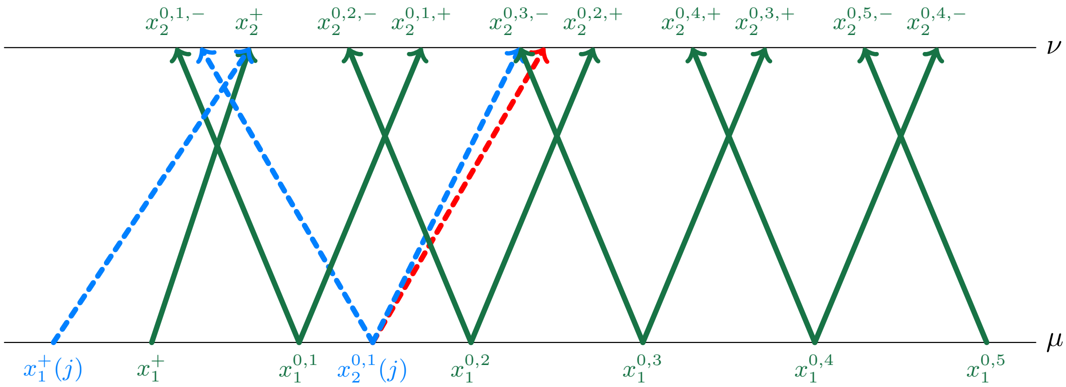

Figure 5.4: Case (a): The first elements of $T_2^{\tilde{j}}$ (green) and of T_2^j (blue dotted). The red dotted arrows show values which lead to a contradiction to Definition 5.5.3.

Proof. We will prove the Lemma by contradiction. In particular we will make use of the properties of Algorithm 5.5.4 established in Lemmas 5.5.8 and 5.5.9 above. Let us fix $j \geq \tilde{j}$. Let us first consider the interval $A_1 = [x_2^{0,1,-}(\tilde{j}), x_2^{0,5,-}(\tilde{j}))$ and note that by equation (5.5.3) we have

$$x_2^+(j) \leq \max(X_2^{\tilde{j},+}) \leq \min(X_2^{\tilde{j},-}) \leq x_2^-(j). \tag{5.5.10}$$

We can assume for notational simplicity that $x_2^+(j) \geq x_2^{0,1,-}(\tilde{j})$: otherwise we consider the smallest x_2 contained in the set $\{x_2^+(j), x_2^{0,1,+}(j), x_2^{0,2,+}(j), \dots, x_2^{0,m_j,+}(j)\}$ satisfying $x_2 \geq x_2^{0,1,-}(\tilde{j})$ and shift the argument to start from this element instead of $x_2^+(j)$ accordingly. Such a smallest x_2 exists by Lemma 5.5.8 and (5.5.10).

We now consider the two cases

(a) $x_2^{0,1,-}(j) \geq x_2^{0,1,-}(\tilde{j})$ and

(b) $x_2^{0,1,-}(j) < x_2^{0,1,-}(\tilde{j})$

separately. Cases (a) and (b) correspond to Figures 5.4 and 5.5, where we have drawn the first elements of T_2^j in blue and $T_2^{\tilde{j}}$ in green. Let us first consider case (a) and note that by definition of $J(m)$ we have $m_{\tilde{j}} \leq m_j$ and $m_{\tilde{j}}$ was chosen minimally according to Definition 5.5.3. We conclude that $x_2^{0,1,+}(j) \leq x_2^{0,3,-}(\tilde{j})$: indeed as $x_2^+(\tilde{j}) \geq x_2^+(j) > x_2^{0,1,-}(j)$ by (5.5.3), we otherwise have $[x_2^+(\tilde{j}), x_2^{0,3,-}(\tilde{j})] \subset [x_2^{0,1,-}(j), x_2^{0,1,+}(j)]$. By Lemma 5.5.9.*(iv)* we have $x_1^{0,1}(j) \in X_1^{0,\tilde{j}}$ and by Lemma 5.5.9.*(i)* there exist $\tilde{x}_2^-, \tilde{x}_2^+ \in \operatorname{supp}(\pi^{(\tilde{j})}_{x_1^{0,1}(j)})$ such that $\tilde{x}_2^- \leq x_2^{0,1,-}(j)$ and $\tilde{x}_2^+ \geq x_2^{0,1,+}(j)$. This leads to a contradiction to minimality of $m_{\tilde{j}}$ at step $\tilde{j}$, as we could replace $x_1^{0,1}(\tilde{j})$ and $x_1^{0,2}(\tilde{j})$ by $x_1^{0,1}(j)$, thus reducing $|T_1^{\tilde{j}}|$ from $m_{\tilde{j}} + 2$ to $m_{\tilde{j}} + 1$. We thus conclude $[x_2^{0,1-}(j), x_2^{0,1,+}(j)) \subseteq A_1$.

The case (b) works similarly, compare Figure 5.5. We conclude $x_2^{0,1,+}(j) \leq x_2^{0,3,-}(\tilde{j})$ (oth-

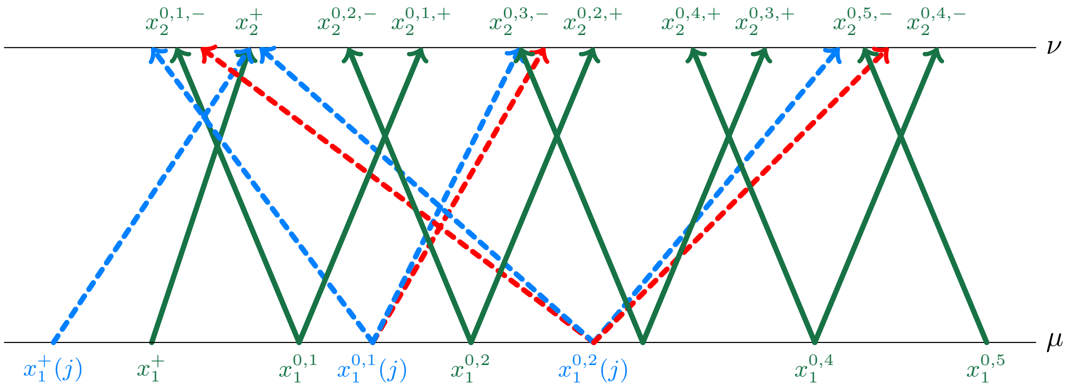

Figure 5.5: Case (b): The first elements of $T_2^{\tilde{j}}$ (green) and of T_2^j (blue dotted). The red dotted arrows show values which lead to a contradiction to Definition 5.5.3.

erwise $x_1^{0,1}(\tilde{j})$ and $x_1^{0,2}(\tilde{j})$ could be replaced by $x_1^{0,1}(j)$ as in case (a), which contradicts minimality) and then $x_2^{0,2,+}(j) \le x_2^{0,5,-}(\tilde{j})$ (otherwise $x_1^{3,0}(\tilde{j})$ and $x_1^{4,0}(\tilde{j})$ could be replaced by $x_1^{2,0}(j)$), and as $x_2^{0,2,-}(j) \ge x_2^+(j) \ge x_2^{0,1,-}(\tilde{j})$ by (5.5.2) and the convention $x_2^+(j) \ge x_2^{0,1,-}(\tilde{j})$, this concludes the proof for A_1. Now we iterate the arguments for $k > 1$, which shows the claim. □

5.5.2 Proof of Theorem 5.2.8 for finitely supported $\pi \in \Pi(\mu, \nu)$ with $\nu \in \mathfrak{P}$

Let $\mathfrak{P} \subseteq \mathcal{P}_1(\mathbb{R})$ be uniformly integrable. We have already proved Theorem 5.2.8 for all finitely supported probability measures $\pi \in \Pi(\mu, \nu)$, where the support of ν is contained in a common compact set $[-\tilde{K}, \tilde{K}]$. We can now extend inequality (5.2.7) to all finitely supported measures $\pi \in \Pi(\mu, \nu)$, for which $\nu \in \mathfrak{P}$.

For this let us recall Definition 5.5.7, specifically

$$\rho(dz^0, dz^1, dz^2, \dots, dz^N) = \rho^{(N-1)}_{z^{N-1}}(dz^N) \dots \rho^{(1)}_{z^1}(dx^2)\rho^{(0)}(dz^0, dz^1)$$

and

$$\rho^{(0:N)}(dx, dy) = \int \rho(dx, dz^1, \dots, , dz^{N-1}, dy).$$

We are now ready to extend the proof of Theorem 5.2.8 to measures $\pi \in \Pi(\mu, \nu)$ where $\nu \in \mathfrak{P}$.

Proof of Theorem 5.2.8 for finitely supported measures with $\nu \in \mathfrak{P}$. Applying Algorithm 5.5.4 as in Section 5.5.1 we obtain a martingale measure $\pi_m := \pi^{(N)} \in \mathcal{M}(\mu, \nu)$ and a coupling $\rho^{(0:N)} \in \Pi(\pi, \pi_m)$. We now show that (5.2.7) holds for π_m: indeed, as $\mathfrak{P}$ is uniformly integrable there exists $\tilde{K} = \tilde{K}(\mathfrak{P}) > 0$ such that $\int_{[-\tilde{K},\tilde{K}]^c} |x_2| \, \nu(dx_2) \le \delta/8$. Next we observe that by the triangle inequality

$$|x_2 - y_2|\mathbb{1}_{\{x_2 \notin [-\tilde{K},\tilde{K}] \vee y_2 \notin [-\tilde{K},\tilde{K}]\}} \le 2(|x_2|\mathbb{1}_{\{x_2 \notin [-\tilde{K},\tilde{K}]\}} + |y_2|\mathbb{1}_{\{y_2 \notin [-\tilde{K},\tilde{K}]\}}) \tag{5.5.11}$$

holds for all $x_2, y_2 \in \mathbb{R}$, so it holds in particular for all (x_2, y_2) such that $(x_2, y_2) \in \text{supp}(\rho^{(0:N)}_{(x_1,x_1)})$ for some $x_1 \in \text{supp}(\mu)$. We next define for all $0 \le j \le N$

$$\tilde{m}_j := \max\Big\{m \in \mathbb{N} : \ \exists i \in \{0, \ldots, m_j - 1\} \text{ s.t. } [x_2^{0,i+k,-}(j), x_2^{0,i+k,+}(j)] \subseteq [-\tilde{K}, \tilde{K}] \\ \text{for all } k = 1, \ldots, m\Big\},$$

where $x_2^{0,i+k,-}(j), x_2^{0,i+k,+}(j)$ denote elements of T_2^j as defined in Lemma 5.5.2. Let us also make the following important observations, which follow immediately from Definition 5.5.3 and Lemma 5.5.9:

- for every $0 \le j \le N$ there are at most four distinct $x_1 \in T_1^j$ for which simultaneously $\text{supp}(\pi^{(j)}_{x_1}) \cap [-\tilde{K}, \tilde{K}] \neq \emptyset$ and $\text{supp}(\pi^{(j)}_{x_1}) \cap [-\tilde{K}, \tilde{K}]^c \neq \emptyset$.
- if $x_1 \in X_1^{j,0}$ for some $0 \le j \le N - 1$ and $\text{supp}(\pi^{(j)}_{x_1}) \subseteq [-\tilde{K}, \tilde{K}]$ then $\text{supp}(\pi^{(j+1)}_{x_1}) \subseteq [-\tilde{K}, \tilde{K}]$.
- if $x_1 \in X_1^{j,0}$ for some $0 \le j \le N - 1$ and $\text{supp}(\pi^{(j)}_{x_1}) \subseteq [-\tilde{K}, \tilde{K}]^c$ then $\text{supp}(\pi^{(j+1)}_{x_1}) \subseteq [-\tilde{K}, \tilde{K}]^c$.

Defining

$$B^j_{\tilde{K}} := \{x_1 \in \text{supp}(\mu) \mid \text{supp}(\pi^{(j)}_{x_1}) \cap [-\tilde{K}, \tilde{K}] \neq \emptyset\}$$

for all $0 \le j \le N - 1$ we thus conclude from the last point that $B^j_{\tilde{K}}$ is non-increasing. In particular writing

$$\begin{aligned}
\mathcal{W}^1_{nd}(\pi, \pi_m) &\le \int\int |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \\
&\le \int\int_{\{x_2 \notin [-\tilde{K},\tilde{K}] \vee y_2 \notin [-\tilde{K},\tilde{K}]\}} |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \\
&+ \int\int_{\{x_2 \in [-\tilde{K},\tilde{K}] \wedge y_2 \in [-\tilde{K},\tilde{K}]\}} |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \\
&\le 2\left(\int_{[-\tilde{K},\tilde{K}]^c} |x_2| \, \nu(dx_2) + \int_{[-\tilde{K},\tilde{K}]^c} |y_2| \, \nu(dy_2)\right) \\
&+ \int_{\{x_1 \in B^1_{\tilde{K}}\}} \int_{[-\tilde{K},\tilde{K}]^2} |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1).
\end{aligned}$$

Next, applying the triangle inequality to $|x_2 - y_2|$ we have

$$\int_{\{x_1 \in B^1_{\tilde{K}}\}} \int_{[-\tilde{K},\tilde{K}]^2} |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1)$$
$$\leq \sum_{j=0}^{N-1} \int_{\{x_1 \in B^j_{\tilde{K}}\}} \int |x_2 - y_2| \, \rho^{(j)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \leq \sum_{j=0}^{N-1} 2(\tilde{m}_j + 4)\lambda^{(j)}$$

by the definitions of $\rho^{(0:N)}$ and $\tilde{m}_j$ combined with a computation similar to the proof of *(iii)* of Lemma 5.5.8. Now we follow exactly the same arguments as in the proof of Lemma 5.5.13 with δ replaced by $\delta/2$ and m_j replaced by $\tilde{m}_j$ to obtain

$$\sum_{j=0}^{N-1} 2(\tilde{m}_j + 4)\lambda^{(j)} \leq \delta/2 + \sum_{j=0}^{N-1} (K(\delta/2, \tilde{K}) + 4)$$
$$\cdot \left(\left| \int (x_2 - x_1) \, \pi^{(j)}_{x_1}(dx_2) \right| \mu(dx_1) - \int \left| \int (x_2 - x_1) \, \pi^{(j+1)}_{x_1}(dx_2) \right| \mu(dx_1) \right)$$
$$\leq \delta/2 + (K(\delta/2, \tilde{K}) + 4)\varepsilon_\pi.$$

Combining the estimates above we finally obtain

$$\begin{aligned}
\mathcal{W}^1_{nd}(\pi, \pi_m) &\leq \int \int |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \\
&\leq 2 \left(\int_{[-\tilde{K},\tilde{K}]^c} |x_2| \, \nu(dx_2) + \int_{[-\tilde{K},\tilde{K}]^c} |y_2| \, \nu(dy_2) \right) \\
&\quad + \int_{\{x_1 \in B^1_{\tilde{K}}\}} \int_{[-\tilde{K},\tilde{K}]^2} |x_2 - y_2| \, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2) \, \mu(dx_1) \\
&\leq \delta/2 + \delta/2 + (K(\delta/2, \tilde{K}) + 4)\varepsilon_\pi \\
&\leq \delta + (K(\delta/2, \tilde{K}) + 4)\varepsilon_\pi.
\end{aligned}$$

The claim follows by setting $K(\delta, \mathfrak{P}) := K(\delta/2, \tilde{K}) + 4$. □

5.5.3 Proof of Theorem 5.2.8 for general $\pi \in \Pi(\mu, \nu)$

Lastly we give the proof of Theorem 5.2.8 for general $\pi \in \Pi(\mu, \nu)$. This follows by arguments similar to the proof of Proposition 5.2.4 in Section 5.4.2.

Proof of Proposition 5.2.4 for general $\pi \in \Pi(\mu, \nu)$. Fix $\delta > 0$. By Lemma 5.3.4.*(ii)* there exists a sequence of finitely supported measures $(\pi^n)_{n \in \mathbb{N}}$ such that $\pi^n \in \Pi(\mu^n, \nu^n)$, where $\mu^n \preceq_c \nu^n$ for all $n \in \mathbb{N}$ and

$$\lim_{n \to \infty} \mathcal{W}^1_{nd} \left(\frac{1}{n} \sum_{i=1}^n \pi^i, \pi \right) = 0.$$

Furthermore, we can assume that for n large enough we have $\sup_{\nu\in\mathfrak{P}}\int_{[-\tilde{K},\tilde{K}]^c}|x_2|\,\nu^n(dx_2)\le\delta/8$. By the finitely supported case we can thus find a sequence of measures $(\pi^n_m)_{n\in\mathbb{N}}$ such that $\pi^n_m\in\mathcal{M}(\mu^n,\nu^n)$ and

$$\mathcal{W}^1_{nd}(\pi^n_m,\pi^n)\le\int\mathcal{W}^1(\pi^n_{m,x_1},\pi^n_{x_1})\,\mu^n(dx_1)\le\delta+K(\delta,\mathfrak{P})\varepsilon_{\pi^n}. \tag{5.5.12}$$

for all $n\in\mathbb{N}$. We now apply Lemma 5.3.6, which yields a measure $\pi_m\in\Pi(\mu,\nu)$, such that (after taking a subsequence)

$$\lim_{n\to\infty}\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n\pi^i_m,\pi_m\right)=0$$

In particular $\pi\in\mathcal{M}(\mu,\nu)$. Using Lemma 5.3.3

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n\pi^i_m,\frac{1}{n}\sum_{i=1}^n\pi^i\right)\le\frac{1}{n}\sum_{i=1}^n\int\mathcal{W}^1(\pi^i_{m,x_1},\pi^i_{x_1})\,\mu^i(dx_1)$$

and by (5.5.12)

$$\begin{aligned}\mathcal{W}^1_{nd}(\pi_m,\pi)&\le\limsup_{n\to\infty}\mathcal{W}^1_{nd}\left(\pi_m,\frac{1}{n}\sum_{i=1}^n\pi^i_m\right)+\limsup_{n\to\infty}\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n\pi^i_m,\frac{1}{n}\sum_{i=1}^n\pi^i\right)\\&\quad+\limsup_{n\to\infty}\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n\pi^i,\pi\right)\\&\le\delta+K(\delta,\mathfrak{P})\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\varepsilon_{\pi^i}.\end{aligned}$$

The last expression is equal to ε_π as $\lim_{n\to\infty}\mathcal{W}^1(\pi^n,\pi)=0$. This proves the claim. □

5.6 Proofs of remaining results in Section 5.2

Proof of Lemma 5.2.1. We observe that for an arbitrary $\tilde{\pi} \in \mathcal{M}(\mu, \nu)$

$$\begin{aligned}
\mathcal{W}^1_{nd}(\pi, \tilde{\pi}) = \inf_{\gamma^1 \in \Pi(\mu,\mu)} \Bigg(\int |x_1 - y_1| \, \gamma^1(dx_1, y_1) \\
&+ \int \inf_{\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})} \int |x_2 - y_2| \, \gamma^2(dx_2, dy_2) \, \gamma^1(dx_1, dy_1) \Bigg) \\
&\geq \inf_{\gamma^1 \in \Pi(\mu,\mu)} \Bigg(\int |x_1 - y_1| \, \gamma^1(dx_1, dy_1) \\
&+ \int \inf_{\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})} \left| \int (x_2 - y_2) \, \gamma^2(dx_2, dy_2) \right| \gamma^1(dx_1, dy_1) \Bigg) \\
&= \inf_{\gamma^1 \in \Pi(\mu,\mu)} \Bigg(\int |x_1 - y_1| \, \gamma^1(dx_1, dy_1) \\
&+ \int \inf_{\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})} \left| \int (x_2 - y_1) \, \gamma^2(dx_2, dy_2) \right| \gamma^1(dx_1, dy_1) \Bigg) \\
&= \inf_{\gamma^1 \in \Pi(\mu,\mu)} \Bigg(\int |x_1 - y_1| \, \gamma^1(dx_1, dy_1) + \int \inf_{\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})} \Bigg| \int x_2 \, \gamma^2(dx_2, dy_2) \\
&- x_1 + x_1 - y_1 \Bigg| \gamma^1(dx_1, dy_1) \Bigg) \\
&\geq \int \left| \int (x_2 - x_1) \, \pi_{x_1}(dx_2) \right| \mu(dx_1) = \varepsilon_\pi
\end{aligned}$$

holds by an application of Jensen's inequality and reverse triangle inequality. This shows the claim. □

Proof of Lemma 5.2.3. Assume towards a contradiction that there exists $x_0 \in \mathbb{R}$ such that

$$\int_{\{x_1 \geq x_0\}} (x_2 - x_1) \, \pi_{HF}(dx_1, dx_2) < 0.$$

Noting that

$$x \mapsto \int_{\{x_1 \geq x\}} (x_2 - x_1) \, \pi_{HF}(dx_1, dx_2)$$

is left-continuous and non-increasing on $\{x \in \mathbb{R} \; : \; \inf(\mathrm{supp}(\pi_{HF,x})) \geq x\}$ we conclude that

$$\mu \left(\left\{ x \in \mathbb{R} \; : \; \int_{\{x_1 \geq x\}} (x_2 - x_1) \, \pi_{HF}(dx_1, dx_2) < 0, \inf(\mathrm{supp}(\pi_{HF,x})) < x \right\} \right) > 0.$$

Consequently we can choose

$$\begin{aligned}
x^* \in \Big\{ x \in \Gamma^1 \; : \; \int_{\{x_1 \geq x\}} (x_2 - x_1) \, \pi_{HF}(dx_1, dx_2) < 0, \\
\inf(\mathrm{supp}(\pi_{HF,x})) < x, \pi_{HF,x}(\Gamma_x) = 1 \Big\} \neq \emptyset,
\end{aligned}$$

where Γ^1 is the projection of Γ onto the first coordinate and Γ_{x_1} is the x_1-section of Γ. We set $\hat{x} = \inf(\mathrm{supp}(\pi_{HF,x^*}))$. For the convex function $x \mapsto (x-\hat{x})^+$ we then have

$$\begin{aligned}\int (x_1-\hat{x})^+ \,\mu(dx_1) &\geq \int_{\{x_1\geq x^*\}} (x_1-\hat{x})\,\mu(dx_1) > \int_{\{x_1\geq x^*\}} (x_2-\hat{x})\,\pi_{HF}(dx_1,dx_2)\\ &= \int_{\{x_1\geq x^*\}\cap\Gamma} (x_2-\hat{x})\,\pi_{HF}(dx_1,dx_2) = \int_{\{x_2\geq \hat{x}\}\cap\Gamma} (x_2-\hat{x})\,\pi_{HF}(dx_1,dx_2)\\ &= \int_{\{x_2\geq\hat{x}\}} (x_2-\hat{x})\,\nu(dx_2) = \int (x_2-\hat{x})^+\,\nu(dx_2),\end{aligned}$$

where we used the definition of x^* for the first and second inequality and (5.2.3) for the equality in the second line. This is a contradiction to $\mu \preceq_c \nu$ and thus proves the claim. □

Proof of Corollary 5.2.6. As every $\pi \in \mathcal{M}(\mu,\nu)$ satisfies $\int (x_2-x_1)\,\pi_{x_1}(dx_2) = 0$ μ-a.s., it clearly fulfils the barycentre dispersion assumption 5.2.2 and we have

$$\inf_{\pi\in\Pi(\mu,\nu),\ \pi \text{ satisfies Ass. 5.2.2}} \Big(\int c(x_1,x_2)\,\pi(dx_1,dx_2) + L\int \Big|\int (x_2-x_1)\,\pi_{x_1}(dx_2)\Big|\,\mu(dx_1)\Big) \leq C(\mu,\nu).$$

Now take any $\pi \in \Pi(\mu,\nu)$ satisfying the barycentre dispersion assumption 5.2.2 and any $\tilde{\pi} \in \mathcal{M}(\mu,\nu)$. Then

$$\begin{aligned}\int c(x_1,x_2)\,\tilde{\pi}(dx_1,dx_2) &\leq \int c(x_1,x_2)\,\pi(dx_1,dx_2)\\ &\quad + \Big(\int c(x_1,x_2)\,\tilde{\pi}(dx_1,dx_2) - \int c(x_1,x_2)\,\pi(dx_1,dx_2)\Big)\\ &\leq \int c(x_1,x_2)\,\pi(dx_1,dx_2) + L\mathcal{W}^1_{nd}(\pi,\tilde{\pi}) \qquad (5.6.1)\end{aligned}$$

as c is L-Lipschitz-continuous. Taking the infimum over $\tilde{\pi} \in \mathcal{M}(\mu,\nu)$ in (5.6.1) and using Proposition 5.2.4 we conclude that

$$\begin{aligned}C(\mu,\nu) &\leq \int c(x_1,x_2)\,\pi(dx_1,dx_2) + L\inf_{\tilde{\pi}\in\mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi,\tilde{\pi})\\ &= \int c(x_1,x_2)\,\pi(dx_1,dx_2) + L\int\Big|\int (x_2-x_1)\,\pi_{x_1}(dx_2)\Big|\,\mu(dx_1).\end{aligned}$$

Taking the infimum over $\pi \in \Pi(\mu,\nu)$ satisfying Assumption 5.2.2 concludes the proof. □

Proof of Remark 5.2.7. Let us first assume that $\pi_{AT} \in \Pi(\mu,\nu)$ is finitely supported. We note that the barycentre assumption is not satisfied in general for π_{AT}, but we may apply Algorithm 5.5.4. Particular care has to be taken if there exists $x_1^* \in \mathrm{supp}(\mu)$ such that $x_1^* \in \mathrm{supp}(\pi_{x_1^*})$. By (5.2.6), there exists at most one such x_1^*. In this case we first apply Algorithm 5.4.3 to x_1^* until $x_1^* \in X_1^{j,0}$ for some $j \in \mathbb{N}$, in such a way that $(\min(X_2^{j,0}), \max(X_2^{j,0})) \cap$

$\mathrm{supp}(\nu) = \bigcup_{x_1 \in X_1^{j,0}} \mathrm{supp}(\pi^{(j)}_{x_1})$. This can always be achieved by exchanging mass in the direct (left or right) neighbourhood of $\pi_{x_1^*}$, as (5.2.6) holds for π. For the rest of the iterations we now leave $\{\pi^{(j)}_{x_1^*} : x_1 \in X_1^{j,0}\}$ unchanged. Formally this can be achieved by following [Beiglböck and Juillet, 2016, proof of Lemma 2.8]: let us define the sub-probability measure π^* via $\pi^*(A) := \pi^{(j)}(A) - \pi^{(j)}(A \cap (\{X_1^{j,0}\} \times \mathbb{R}))$ for all Borel sets A. We call its marginals μ^* and ν^*. It remains to check that these are still in convex order: take any convex function $\varphi : \mathbb{R} \to \mathbb{R}$. Then φ is dominated by a convex function ψ which is linear on $(\min(X_2^{j,0}), \max(X_2^{j,0}))$ and agrees with φ on $\mathbb{R} \setminus (\min(X_2^{j,0}), \max(X_2^{j,0}))$. Then $\int_{X_1^{j,0}} \int \psi(x_2)\pi^{(j)}_{x_1}(dx_2)\, \mu(dx_1) = \int_{X_1^{j,0}} \psi(x_1)\, \mu(dx_1)$ and

$$\begin{aligned}
\int \varphi(x_1)\, \mu^*(dx_1) &\leq \int \psi(x_1)\, \mu^*(dx_1) \\
&= \int \psi(x_1)\, \mu(dx_1) - \int_{X_1^{j,0}} \psi(x_1)\, \mu(dx_1) \\
&\leq \int \psi(x_2)\, \nu(dx_2) - \int_{X_1^{j,0}} \psi(x_1)\, \mu(dx_1) \\
&= \int \psi(x_2)\, \nu(dx_2) - \int_{X_1^{j,0}} \int \psi(x_2)\, \pi^{(j)}_{x_1}(dx_2)\, \mu(dx_1) \\
&= \int \psi(x_2)\, \nu^*(dx_2) = \int \varphi(x_2)\, \nu^*(dx_2),
\end{aligned}$$

thus $\mu^* \preceq_c \nu^*$ follows. We now apply Algorithm 5.5.4 to π^* to obtain a sub-probability measure $\pi^{(*,N)}$ with marginals μ^* and ν^* satisfying the martingale condition $\int (x_2 - x_1)\, \pi^{(*,N)}_{x_1} = 0$ for all $x_1 \in \mathrm{supp}(\mu) \setminus X_1^{j,0}$. We still denote the (bicausal) coupling between π and $\pi^{(N)}$ defined via $\pi^{(N)}(A) := \pi^{(*,N)}(A) + \pi^{(j)}(A \cap (\{X_1^{j,0}\} \times \mathbb{R}))$ for all Borel sets A by $\rho^{(0:N)}$. By construction of $\pi^{(j)}$ and the properties of Algorithm 5.5.4, in particular item (i) of Lemma 5.5.9, we note that for $(x, y) \in \mathrm{supp}(\rho^{(0:N)})$ we have

$$\begin{aligned}
&x_2 \geq y_2 \text{ if } x_1 \in X_1^{0,+}, \\
&x_2 \leq y_2 \text{ if } x_1 \in X_1^{0,-}, \\
&x_2 = y_2 \text{ if } x_1 \in X_1^{0,0}.
\end{aligned}$$

Thus in particular

$$\begin{aligned}
\mathcal{W}^1_{nd}(\pi^{(0)}, \pi^{(N)}) &\leq \int \int |x_2 - y_2|\, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2)\, \mu(dx_1) \\
&= \int_{x_1 \in X_1^{0,+}} \int (x_2 - y_2)\, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2)\, \mu(dx_1) \\
&\quad + \int_{x_1 \in X_1^{0,-}} \int (y_2 - x_2)\, \rho^{(0:N)}_{(x_1,x_1)}(dx_2, dy_2)\, \mu(dx_1) \\
&= \int \left| \int (x_2 - x_1)\, \pi_{x_1}(dx_2) \right| \mu(dx_1).
\end{aligned}$$

This finishes the proof for the finitely supported case.

For the general case we first use Lemma 5.3.4.*(ii)* to find sequences $(\mu^n)_{n\in\mathbb{N}}$ and $(\nu^n)_{n\in\mathbb{N}}$ such that $\mu^n \preceq_c \nu^n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} \mathcal{W}^p(\mu^n,\mu) = 0 = \lim_{n\to\infty} \mathcal{W}^p(\nu^n,\nu)$. We denote by π^n_{AT} the antitone coupling between μ^n and ν^n and apply the finitely supported case above to find a sequence of martingale measures $\pi^n_{mr} \in \mathcal{M}(\mu^n,\nu^n)$ such that $\mathcal{W}^1_{nd}(\pi^n_{AT},\pi^n_{mr}) \le \varepsilon_{\pi^n_{AT}}$. Next we apply Lemma 5.3.6 twice to find measures $\tilde{\pi} \in \Pi(\mu,\nu)$ and $\pi_{mr} \in \mathcal{M}(\mu,\nu)$ such that (after subsequences)

$$\lim_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{AT},\tilde{\pi}\right) = 0 = \lim_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{mr},\pi_{mr}\right).$$

Arguing exactly as in the proof of Proposition 5.2.4 and the proof of Theorem 5.2.8

$$\begin{aligned}\mathcal{W}^1_{nd}(\tilde{\pi},\pi) &\le \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\tilde{\pi},\frac{1}{n}\sum_{i=1}^n \pi^i_{AT}\right) + \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{AT},\frac{1}{n}\sum_{i=1}^n \pi^i\right)\\ &\quad + \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i_{mr},\pi_{mr}\right)\\ &\le \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \int \left| \int (x_2-x_1)\,\pi^i_{AT,x_1}(dx_2)\right| \mu^i(dx_1).\end{aligned}$$

The last expression is equal to

$$\int \left| \int (x_2-x_1)\,\tilde{\pi}_{x_1}(dx_2)\right| \mu(dx_1)$$

as $\lim_{n\to\infty} \mathcal{W}^1(\pi^n_{AT},\tilde{\pi}) = 0$. It remains to notice that $\tilde{\pi} = \pi_{AT}$, which follows from the stability of Optimal Transport (see e.g. [Villani, 2008, Proof of Theorem 5.10, Step 2, pp. 64-65]) and the fact that the antitone coupling is the unique optimiser for $\inf_{\pi\in\Pi(\mu,\nu)} \int h(x_2 - x_1)\,\pi(dx_1,dx_2)$ for any function $h:\mathbb{R}\to\mathbb{R}$, which is strictly concave. □

Proof of Corollary 5.2.9. Fix $\delta > 0$ and $\nu \in \mathcal{P}_1(\mathbb{R})$. As $\mathfrak{P} = \{\nu\}$ is uniformly integrable, we can apply Theorem 5.2.8 to obtain a constant $K(\delta,\nu)$ such that we have

$$\inf_{\tilde{\pi}\in\mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi,\tilde{\pi}) \le K(\delta,\nu)\varepsilon_\pi + \delta \tag{5.6.2}$$

for all $\pi \in \Pi(\mu,\nu)$ where $\mu \in \mathcal{P}(\mathbb{R})$ with $\mu \preceq_c \nu$. As $\varepsilon_\pi = 0$ for all $\pi \in \mathcal{M}(\mu,\nu)$ the first inequality in Corollary 5.2.9 is trivial. Now take any $\pi \in \Pi(\mu,\nu)$ and any $\tilde{\pi} \in \mathcal{M}(\mu,\nu)$. Then as in the proof of Corollary 5.2.6

$$\begin{aligned}\int c(x_1,x_2)\,\tilde{\pi}(dx_1,dx_2) &\le \int c(x_1,x_2)\,\pi(dx_1,dx_2)\\ &\quad + \left(\int c(x_1,x_2)\,\tilde{\pi}(dx_1,dx_2) - \int c(x_1,x_2)\,\pi(dx_1,dx_2)\right)\\ &\le \int c(x_1,x_2)\,\pi(dx_1,dx_2) + L\mathcal{W}^1_{nd}(\pi,\tilde{\pi})\end{aligned} \tag{5.6.3}$$

as c is L-Lipschitz-continuous. Taking the infimum over $\tilde{\pi} \in \mathcal{M}(\mu, \nu)$ in (5.6.3) and using (5.6.2) we conclude that

$$\begin{aligned} C(\mu, \nu) &\leq \int c(x_1, x_2)\, \pi(dx_1, dx_2) + L \inf_{\tilde{\pi} \in \mathcal{M}(\mu,\nu)} \mathcal{W}^1_{nd}(\pi, \tilde{\pi}) \\ &= \int c(x_1, x_2)\, \pi(dx_1, dx_2) + LK(\delta, \nu)\varepsilon_\pi + L\delta. \end{aligned}$$

Taking the infimum over $\pi \in \Pi(\mu, \nu)$ concludes the proof. □

Proof of Theorem 5.2.11. For all $n \in \mathbb{N}$ we take $\pi^n \in \mathcal{M}(\mu^n, \nu^n)$ such that

$$\inf_{\pi \in \mathcal{M}(\mu^n,\nu^n)} \left(\int c(x_1, x_2)\, \pi(dx) \right) \geq \int c(x_1, x_2)\, \pi^n(dx) - 1/n$$

and note that (possibly after taking a subsequence) there exists $\tilde{\pi} \in \mathcal{M}(\mu, \nu)$ such that $\lim_{n\to\infty} \mathcal{W}^p(\pi^n, \tilde{\pi}) = 0$. Then

$$\begin{aligned} \liminf_{n\to\infty} \inf_{\pi \in \mathcal{M}(\mu^n,\nu^n)} \int c(x_1, x_2)\, \pi(dx_1, dx_2) &\geq \liminf_{n\to\infty} \left(\int c(x_1, x_2)\, \pi^n(dx_1, dx_2) - 1/n \right) \\ &= \int c(x_1, x_2)\, \tilde{\pi}(dx_1, dx_2) \\ &\geq \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c(x_1, x_2)\, \pi(dx_1, dx_2). \end{aligned}$$

For the converse inequality we note that for all $n \in \mathbb{N}$ there exists $\pi^n \in \mathcal{M}(\mu, \nu)$ such that

$$\inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c(x_1, x_2)\, \pi(dx_1, dx_2) \geq \int c(x_1, x_2)\, \pi^n(dx_1, dx_2) - 1/n.$$

We now apply Lemma 5.3.5 to conclude that for every $n \in \mathbb{N}$ there exists a coupling $\tilde{\pi}^n \in \Pi(\mu^n, \nu^n)$ such that $\mathcal{W}^p(\pi^n, \tilde{\pi}^n) \leq \mathcal{W}^p(\mu, \mu^n) + \mathcal{W}^p(\nu, \nu^n)$ and

$$\int \left| \int (x_2 - x_1)\, \tilde{\pi}^n_{x_1}(dx_2) \right| \mu^n(dx_1) \leq \mathcal{W}^p(\mu, \mu^n) + \mathcal{W}^p(\nu, \nu^n).$$

Note that as $\lim_{n\to\infty} \mathcal{W}^p(\nu_n, \nu) = 0$ the sequence $\{\nu_n\}_{n\in\mathbb{N}}$ is in particular uniformly integrable. By Theorem 5.2.8 there exists a sequence $\pi^n_m \in \mathcal{M}(\mu^n, \nu^n)$ such that $\lim_{n\to\infty} \mathcal{W}^1_{nd}(\tilde{\pi}^n, \pi^n_m) = 0$. Recall that the function $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies $|c(x_1, x_2)| \leq C(1 + |x_1|^p) + |x_2|^p)$. Then using Lemma 5.3.7

$$\begin{aligned} \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c(x_1, x_2)\, \pi(dx_1, dx_2) &\geq \limsup_{n\to\infty} \int c(x_1, x_2)\, \pi^n(dx_1, dx_2) \\ &\geq \limsup_{n\to\infty} \int c(x_1, x_2)\, \tilde{\pi}^n(dx_1, dx_2) \\ &\geq \limsup_{n\to\infty} \int c(x_1, x_2)\, \pi^n_m(dx_1, dx_2) \\ &\geq \limsup_{n\to\infty} \inf_{\pi \in \mathcal{M}(\mu^n,\nu^n)} \int c(x_1, x_2)\, \pi(dx_1, dx_2). \end{aligned}$$

This concludes the proof. □

Proof of Theorem 5.2.13. We only show sufficiency here. For a proof of necessity see e.g. Beiglböck and Griessler [2019]. Let us assume that $\pi \in \mathcal{M}(\mu, \nu)$ is not an optimiser of (5.1.1). We denote

$$\delta := \int c(x_1, x_2)\, \pi(dx_1, dx_2) - \inf_{\hat{\pi} \in \mathcal{M}(\mu,\nu)} \int c(x_1, x_2)\, \hat{\pi}(dx_1, dx_2) > 0.$$

Let $\Gamma \subseteq \mathbb{R}^2$ be a Borel set such that $\pi(\Gamma) = 1$. By Lemma 5.3.1 there exists a sequence of measures $(\pi^n)_{n \in \mathbb{N}}$, such that for each $n \in \mathbb{N}$ π^n is finitely supported on Γ, $\pi^n \in \mathcal{M}(\mu^n, \nu^n)$ for some sequences of measures $(\mu^n)_{n \in \mathbb{N}}$ and $(\nu^n)_{n \in \mathbb{N}}$ and $\lim_{n \to \infty} \mathcal{W}^p_{nd}(\pi^n, \pi) = 0$. Clearly $\mu^n \preceq_c \nu^n$ and

$$\lim_{n \to \infty} \mathcal{W}^p(\mu^n, \mu) = \lim_{n \to \infty} \mathcal{W}^p(\nu^n, \nu) = 0.$$

By Theorem 5.2.11 we also have

$$\lim_{n \to \infty} \inf_{\pi \in \mathcal{M}(\mu^n, \nu^n)} \int c(x, y)\, \pi(dx, dy) = \inf_{\pi \in \mathcal{M}(\mu,\nu)} \int c(x, y)\, \pi(dx, dy),$$

in particular there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\inf_{\pi \in \mathcal{M}(\mu^n, \nu^n)} \int c(x_1, x_2)\, \pi(dx_1, dx_2) \leq \int c(x_1, x_2)\, \pi^n(dx_1, dx_2) - 2\delta/3.$$

There exists a measure $\pi' \in \mathcal{M}(\mu^n, \nu^n)$ such that

$$\int c(x_1, x_2)\, \pi'(dx_1, dx_2) - \inf_{\pi \in \mathcal{M}(\mu^n, \nu^n)} \int c(x_1, x_2)\, \pi(dx_1, dx_2) \leq \delta/3.$$

In particular π' is a competitor of π^n and

$$\int c(x_1, x_2)\, \pi'(dx_1, dx_2) \leq \int c(x_1, x_2)\, \pi^n(dx_1, dx_2) - \delta/3,$$

showing that Γ is not finitely optimal. □

5.7 Proofs of approximation results

Let us first recall the following result:

Lemma 5.7.1 (General Tchakaloff's theorem, cf. [Bayer and Teichmann, 2006, Corollary 2]). *Let $\mu \in \mathcal{P}(\mathbb{R})$, $m \in \mathbb{N}$ and $f : \mathbb{R} \to \mathbb{R}^m$ such that $\int |f(x)|\, \mu(dx) < \infty$. Then there exists a probability measure $\tilde{\mu} \in \mathcal{P}(\mathbb{R})$ with finite support such that* $\operatorname{supp}(\tilde{\mu}) \subseteq \operatorname{supp}(\mu)$ *and* $\int f(x)\, \tilde{\mu}(dx) = \int f(x)\, \mu(dx)$.

Proof of Lemma 5.3.1. Throughout the proof we make the convention that $x/0 := 0$ for any $x \in \mathbb{R}$. We prove *(i)* via two discretisations: first we approximate the marginal μ and consecutively we approximate the disintegration $(\pi_{x_1})_{x_1 \in \mathbb{R}}$.

For the first approximation, note that by [Stroock and Varadhan, 2007, Theorem 1.1.8] the property $\pi(\Gamma) = 1$ implies

$$\mu\left(\{x_1 \in \mathbb{R} \,:\, \pi_{x_1}(\Gamma_{x_1}\} = 1\}\right) = 1,$$

where Γ_{x_1} denotes the x_1-section of Γ. Without loss of generality we thus assume that $\pi_{x_1}(\Gamma_{x_1}) = 1$ for all $x_1 \in \Gamma^1$, where we recall that Γ^1 denotes the projection of Γ to the first coordinate. We can further assume without loss of generality that Γ^1 is not finite and fix some $0 \neq a_0 \in \Gamma^1$ satisfying $\int |x_2|^p \, \pi_{a_0}(dx_2) < \infty$ (see [Stroock and Varadhan, 2007, Cor. 1.1.7]). Choose $c_p \geq 1$ such that $(x+y)^p \leq c_p(x^p + y^p)$ for all $x, y \in \mathbb{R}$. As $\int |x_1|^p \, \mu(dx_1) < \infty$, $\int |x_2|^p \, \nu(dx_2) < \infty$ and $x_1 \mapsto \pi_{x_1}$ is Borel, an application of Lusin's theorem (see [Bogachev, 2007, Theorem 7.1.12]) to the measure ζ defined via

$$\zeta(A) := \frac{1}{3}\left(\frac{\int_A |x_1|^p \, \mu(dx_1)}{\int |x_1|^p \, \mu(dx_1)} + \frac{\int_A \int |x_2|^p \, \pi_{x_1}(dx_2) \, \mu(dx_1)}{\int |x_2|^p \, \nu(dx_2)} + \mu(A) \right)$$

for every Borel set $A \subseteq \mathbb{R}$, there exists a compact set K_1 such that

$$\int_{K_1^c} |x_1|^p \mu(dx_1) \leq \kappa^p/(6c_p), \quad \int_{K_1^c} \int |x_2|^p \, \pi_{x_1}(dx_2) \mu(dx_1) \leq \kappa^p/(6c_p),$$
$$\mu(K_1^c) \leq \frac{\kappa^p/(6c_p)}{|a_0|^p \vee \int |x_2|^p \, \pi_{a_0}(dx_2)}$$

and $x_1 \mapsto \pi_{x_1}$ is continuous in $\mathcal{W}^p$ on K_1. As K_1 is compact, $x_1 \mapsto \pi_{x_1}$ is uniformly continuous on K_1. Thus there exists $\delta > 0$ such that $\mathcal{W}^p(\pi_{x_1}, \pi_{y_1}) \leq \kappa/6$ for all $x_1, y_1 \in K_1$ with $|x_1 - y_1| \leq \delta$ and a finite partition $K_{1,\kappa} = \{a_1 \leq \cdots \leq a_N\}$ of $K_1 \cap \Gamma^1$ such that

$$\inf_{a \in K_{1,\kappa}} |a - y_1| \leq \kappa/6 \wedge \delta$$

for all $y_1 \in K_1$ and denote by a_0 and a_{N+1} the left and right end-points of $K_1 \cap \Gamma^1$. We now define

$$f^\kappa(x_1) := a_0 \mathbb{1}_{\{K_1^c \cap \Gamma^1\}}(x_1) + a_1 \mathbb{1}_{\{[a_0, a_2] \cap K_1 \cap \Gamma^1\}}(x_1) + \sum_{i=2}^{N-1} a_i \mathbb{1}_{\{[a_i, a_{i+1}) \cap K_1 \cap \Gamma^1\}}(x_1)$$
$$+ a_{N-1} \mathbb{1}_{\{[a_{N-1}, a_{N+1}] \cap K_1 \cap \Gamma^1\}}(x_1)$$

and set $\tilde{\pi} = (f^\kappa_* \pi^1) \otimes \pi_{x_1}$, where we recall that $(f_* \pi^1)$ denotes the push-forward of π^1 by the function f^κ. To estimate $\mathcal{W}^1_{nd}(\pi, \tilde{\pi})$ we set $\gamma^1 = (x, f^\kappa(x))_* \pi^1$ and note that $\gamma^1 \in \Pi(\pi^1, \tilde{\pi}^1)$. We define $\gamma^2 \in \Pi(\pi_{x_1}, \tilde{\pi}_{y_1})$ as a coupling which attains $\mathcal{W}^p(\pi_{x_1}, \tilde{\pi}_{y_1})$. Then by the triangle inequality and noting that $\pi^1(\Gamma^1) = 1$

$$\begin{aligned}
\mathcal{W}_{nd}^p(\pi, \tilde{\pi}) &\leq \left(\int |x_1 - y_1|^p + \int |x_2 - y_2|^p \, \gamma^2(dx_2, dy_2) \gamma^1(dx_1, dy_1) \right)^{1/p} \\
&\leq \left(c_p \int_{K_1^c} \left(|x_1|^p + |f^\kappa(x_1)|^p + \int |x_2|^p + |y_2|^p \, \gamma^2(dx_2, dy_2) \right) \pi^1(dx_1) \right. \\
&\quad \left. + \int_{K_1} \left(|x_1 - f^\kappa(x_1)|^p + \int |x_2 - y_2|^p \, \gamma^2(dx_2, dy_2) \right) \pi^1(dx_1) \right)^{1/p} \\
&\leq \left(c_p \int_{K_1^c} |x_1|^p \, \mu(dx_1) + c_p |a_0|^p \, \mu(K_1^c) + c_p \int_{K_1^c} \int |x_2|^p \, \pi_{x_1}(dx_2) \mu(dx_1) \right. \\
&\quad \left. + c_p \mu(K_1^c) \int |x_2|^p \, \pi_{a_0}(dx_2) + (\kappa/6)^p + (\kappa/6)^p \right)^{1/p} \leq \kappa. \qquad (5.7.1)
\end{aligned}$$

This concludes the first approximation step.

For the second approximation step we first fix $x_1 \in K_{1,\kappa} \cup \{a_0\}$. We now approximate the probability measure π_{x_1}: as $\int |x_2|^p \, \nu(dx_2) < \infty$ there exists a finite partition $K_{2,\kappa}(x_1) = \{b_1(x_1) \leq b_2(x_1) \leq \cdots \leq b_{N(x_1)}(x_1)\} \subseteq \mathbb{R}$ such that

$$\inf_{x_2 \in K_{2,\kappa}(x_1)} |x_2 - y_2| \leq \kappa/3$$

for all $y_2 \in [b_{1(x_1)}, b_{N(x_1)}]$ and $\int_{[b_1(x_1), b_N(x_1)]^c} |x_2|^p \, \pi_{x_1}(dx_2) \leq \kappa^p/(3c_p)$. Let us set $b_0(x_1) := -\infty$, $b_{N(x_1)+1} := \infty$ with the convention that $[-\infty, b_1(x_1)) := (-\infty, b_1(x_1))$. By Tchakaloff's theorem as stated in Lemma 5.7.1 there exist finitely supported measures $\{\hat{\pi}^{i,x_1} : i = 1, \ldots, N(x_1) + 1\}$ such that

$$\begin{aligned}
\operatorname{supp}\left(\hat{\pi}^{i,x_1}\right) &\subseteq \Gamma_{x_1} \cap [b_{i-1}(x_1), b_i(x_1)), \\
\hat{\pi}^{i,x_1}([b_{i-1}(x_1), b_i(x_1)) &= \pi_{x_1}([b_{i-1}(x_1), b_i(x_1)), \\
\int x_2 \, \hat{\pi}^{i,x_1}(dx_2) &= \int_{[b_{i-1}(x_1), b_i(x_1))} x_2 \, \pi_{x_1}(dx_2), \\
\int |x_2|^p \, \hat{\pi}^{i,x_1}(dx_2) &= \int_{[b_{i-1}(x_1), b_i(x_1))} |x_2|^p \, \pi_{x_1}(dx_2) \quad \text{for all } i = 1, \ldots, N(x_1) + 1.
\end{aligned}$$

We set

$$\hat{\pi}(dx_1, dx_2) = \tilde{\pi}^1(dx_1) \left(\sum_{i=1}^{N(x_1)+1} \hat{\pi}^{i,x_1}(dx_2) \right),$$

which yields in particular

$$\begin{aligned}
\int x_2 \, \hat{\pi}_{x_1}(dx_2) &= \sum_{i=1}^{N(x_1)+1} \int x_2 \, \hat{\pi}^{i,x_1}(dx_2) \\
&= \sum_{i=1}^{N(x_1)+1} \int_{[b_{i-1}(x_1), b_i(x_1))} x_2 \, \pi_{x_1}(dx_2) = \int x_2 \, \pi_{x_1}(dx_2) \qquad (5.7.2)
\end{aligned}$$

and

$$
\begin{aligned}
\mathcal{W}^p_{nd}(\tilde{\pi},\hat{\pi}) &\le \left(\int \left(\inf_{\gamma^2\in\Pi(\tilde{\pi}_{x_1},\hat{\pi}_{x_1})}\int |x_2-y_2|^p\,\gamma^2(dx_2,dy_2)\right)\tilde{\pi}^1(dx_1)\right)^{1/p}\\
&\le \Bigg(\int \Bigg(c_p\int_{(-\infty,b_1(x_1))}|x_2|^p\;\pi_{x_1}(dx_2)+c_p\int_{(-\infty,b_1(x_1))}|x_2|^p\;\hat{\pi}^{i,x_1}(dx_2)\\
&\quad+\sum_{i=2}^{N(x_1)}\int_{[b_{i-1}(x_1),b_i(x_1))}|y_2-x_2|^p\;(\hat{\pi}^{i,x_1}\times\pi_{x_1})(dx_2,dy_2)\\
&\quad+c_p\int_{[b_N(x_1),\infty)}|x_2|^p\;\pi_{x_1}(dx_2)+c_p\int_{[b_N(x_1),\infty)}|x_2|^p\;\hat{\pi}_{\tilde{b}_N(x_1)}(dx_2)\Bigg)\tilde{\pi}^1(dx_1)\Bigg)^{1/p}\\
&\le \kappa. \qquad (5.7.3)
\end{aligned}
$$

This concludes the second approximation step. In particular the estimates above imply that $\mathcal{W}^p_{nd}(\pi,\hat{\pi})\le\mathcal{W}^p_{nd}(\pi,\tilde{\pi})+\mathcal{W}^p_{nd}(\tilde{\pi},\hat{\pi})\le 2\kappa$ and $\hat{\pi}$ is a probability measure finitely supported on Γ.

We now show (5.3.1) for $\hat{\pi}$. Indeed, using the same estimates as in (5.7.3) we obtain for $x\in\operatorname{supp}((\hat{\pi})^1)$

$$
\begin{aligned}
&\int_{\{x_1\ge x\}}(x_2-x_1)\,\hat{\pi}(dx_1,dx_2)-\int_{\{x_1\ge x\}}(x_2-x_1)\,\tilde{\pi}(dx_1,dx_2)\\
&\ge -\int \mathcal{W}^p(\tilde{\pi}_{x_1},\hat{\pi}_{x_1})\,\tilde{\pi}^1(dx_1)\ge-\kappa.
\end{aligned}
$$

Similarly using (5.7.1) we obtain for $x\in\{a_0,a_2,\dots,a_N\}$

$$
\begin{aligned}
&\int_{\{x_1\ge x\}}(x_2-x_1)\,\tilde{\pi}(dx_1,dx_2)-\int_{\{x_1\ge x\}}(x_2-x_1)\,\pi(dx_1,dx_2)\\
&\ge -\int_{K_1^c}\left(|x_1|+|f(x_1)|+\int |x_2|+|y_2|\,\gamma^2(dx_2,dy_2)\right)\,\pi^1(dx_1)\\
&\quad-\int_{\{x_1\ge x\}\cap K_1}\left(|x_1-f(x_1)|+\int |x_2-y_2|\,\gamma^2(dx_2,dy_2)\right)\,\pi^1(dx_1)\ge-\kappa.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\int_{\{x_1\ge x\}}(x_2-x_1)\,\hat{\pi}(dx_1,dx_2)-\int_{\{x_1\ge x\}}(x_2-x_1)\,\pi(dx_1,dx_2)\\
&\ge \int_{\{x_1\ge x\}}(x_2-x_1)\,\hat{\pi}(dx_1,dx_2)-\int_{\{x_1\ge x\}}(x_2-x_1)\,\tilde{\pi}(dx_1,dx_2)\\
&\quad+\int_{\{x_1\ge x\}}(x_2-x_1)\,\tilde{\pi}(dx_1,dx_2)-\int_{\{x_1\ge x\}}(x_2-x_1)\,\pi(dx_1,dx_2)\ge-2\kappa.
\end{aligned}
$$

This concludes the proof of *(i)*.

For *(ii)* we note that

$$\mu\left(\left\{x_1 \in \mathbb{R} \ : \ \int (x_2 - x_1)\,\pi_{x_1}(dx_2) = 0\right\}\right) = 1$$

for $\pi \in \mathcal{M}(\mu, \nu)$. We can thus proceed as in *(i)* on

$$\Gamma^1 \cap \left\{x_1 \in \mathbb{R} \ : \ \int (x_2 - x_1)\,\pi_{x_1}(dx_2) = 0\right\},$$

noting that (5.7.2) holds. This concludes the proof. □

Proof of Lemma 5.3.2. Let us adopt the same notation and conventions as in the proof of Lemma 5.3.1. Let us note that $\mu \preceq_c \nu$ implies $\mathcal{M}(\mu, \nu) \neq \emptyset$ and let us fix a martingale measure $\dot{\pi} \in \mathcal{M}(\mu, \nu)$. Similarly to the proof of Lemma 5.3.1 we also fix $0 \neq a_0 \in \Gamma^1$ satisfying $\int |x_2|^p\,\pi_{a_0}(dx_2) \vee \int |x_2|^p\,\dot{\pi}_{a_0}(dx_2) < \infty$. Then, applying Lusin's theorem to the measure ζ defined via

$$\zeta(A) := \frac{1}{4}\Bigg(\mu(A) + \frac{\int_A |x_1|^p\,\mu(dx_1)}{\int |x_1|^p\,\mu(dx_1)} + \frac{\int_A \int |x_2|^p\,\pi_{x_1}(dx_2)\,\mu(dx_1)}{\int |x_2|^p\,\nu(dx_2)} + \frac{\int_A \int |x_2|^p\,\dot{\pi}_{x_1}(dx_2)\,\mu(dx_1)}{\int |x_2|^p\,\nu(dx_2)}\Bigg)$$

for every Borel set $A \subseteq \mathbb{R}$, we can find a compact set K_1 such that

$$\int_{K_1^c} |x_1|^p \mu(dx_1) \le \kappa/(6c_p), \quad \int_{K_1^c}\int |x_2|^p\,\pi_{x_1}(dx_2)\mu(dx_1) \le \kappa/(6c_p),$$
$$\int_{K_1^c}\int |x_2|^p\,\dot{\pi}_{x_1}(dx_2)\mu(dx_1) \le \kappa/(6c_p), \quad \mu(K_1^c) \le \frac{\kappa/(6c_p)}{a_0 \vee \int |x_2|^p\,\pi_{a_0}(dx_2) \vee \int |x_2|^p\,\dot{\pi}_{a_0}(dx_2)}$$

and both $x_1 \mapsto \pi_{x_1}$ and $x_1 \mapsto \dot{\pi}_{x_1}$ are continuous in $\mathcal{W}^p$ on K_1. Now we proceed exactly as in the proof of Lemma 5.3.1 for $\Gamma = \mathbb{R}^2$: we conclude from the above that there exists $\delta > 0$ such that

$$\mathcal{W}^p(\pi_{x_1}, \pi_{y_1}) \vee \mathcal{W}^p(\dot{\pi}_{x_1}, \dot{\pi}_{y_1}) \le \kappa/6$$

for all $x_1, y_1 \in K_1$ with $|x_1 - y_1| \le \delta$ and a finite partition $K_{1,\kappa} = \{a_1 \le \cdots \le a_N\}$ of K_1 such that

$$\inf_{a \in K_{1,\kappa}} |a - y_1| \le \kappa/6 \wedge \delta$$

for all $y_1 \in K_1$, where a_1 and a_N are the left and right end-points of K_1. We now define

$$f^\kappa(x_1) = a_0 \mathbb{1}_{\{K_1^c\}}(x_1) + \sum_{i=1}^{N-2} a_i \mathbb{1}_{\{[a_i, a_{i+1}) \cap K_1\}}(x_1) + a_{N-1}\mathbb{1}_{\{[a_{N-1}, a_N] \cap K_1\}}(x_1).$$

Applying the second approximation step in the proof of Lemma 5.3.1 for both π and $\dot{\pi}$ individually and using the same estimates as in the proof of Lemma 5.3.1 we can thus find finitely supported measures, which we call $\hat{\pi} \in \Pi(\bar{\mu}, \hat{\nu})$ and $\pi' \in \mathcal{M}(\bar{\mu}, \bar{\nu})$, with the property $\mathcal{W}^p_{nd}(\pi, \hat{\pi}) \leq 2\kappa$ and $\mathcal{W}^p_{nd}(\dot{\pi}, \pi') \leq 2\kappa$. In particular $\bar{\mu} \preceq_c \bar{\nu}$ and $\hat{\pi}, \pi'$ have the same first marginals. We note that

$$\mathcal{W}^p(\hat{\nu}, \bar{\nu}) \leq \mathcal{W}^p(\hat{\nu}, \nu) + \mathcal{W}^p(\nu, \bar{\nu}) \leq 4\kappa.$$

Let ζ be an optimal coupling for $\mathcal{W}^p(\hat{\nu}, \bar{\nu})$. We define

$$\bar{\pi}(dx_1, dx_2) = \int \hat{\pi}(dx_1, dy_2)\zeta_{y_2}(dx_2) \in \Pi(\bar{\mu}, \bar{\nu})$$

and conclude

$$\mathcal{W}^p_{nd}(\pi, \bar{\pi}) \leq \mathcal{W}^p_{nd}(\pi, \hat{\pi}) + \mathcal{W}^p_{nd}(\hat{\pi}, \bar{\pi}) \leq 2\kappa + \left(\int |x_2 - y_2|^p \, \zeta(dx_2, dy_2) \right)^{1/p} \leq 6\kappa.$$

This proves the claim. $\square$

Proof of Lemma 5.3.3. We note that a disintegration of $\frac{1}{n}\sum_{i=1}^n \pi^i$ is given by

$$\left(\frac{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)\pi^i_{x_1}(dx_2)}{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)} \right)_{x_1 \in \cup_{i=1}^n \mathrm{supp}(\mu^i)}$$

and similarly

$$\left(\frac{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)\tilde{\pi}^i_{x_1}(dx_2)}{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)} \right)_{x_1 \in \cup_{i=1}^n \mathrm{supp}(\mu^i)}$$

is a disintegration of $\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i$. Given $\gamma^{2,i}_{x_1} \in \Pi(\pi^i_{x_1}, \tilde{\pi}^i_{x_1})$ for $i = 1, \dots, n$ and all $x_1 \in \cup_{i=1}^n \mathrm{supp}(\mu^i)$ we thus conclude

$$\frac{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)\gamma^{2,i}_{x_1}(dx_2, dy_2)}{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)} \in \Pi\left(\left(\frac{1}{n}\sum_{i=1}^n \pi^i \right)_{x_1}, \left(\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i \right)_{x_1} \right).$$

In particular choosing $\gamma^{2,i}_{x_1} \in \Pi(\pi^i_{x_1}, \tilde{\pi}^i_{x_1})$ optimal for $\mathcal{W}^1(\pi^i_{x_1}, \tilde{\pi}^i_{x_1})$ it follows

$$\begin{aligned} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \pi^i, \frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i \right) &\leq \int \int |x_2 - y_2| \frac{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)\gamma^{2,i}_{x_1}(dx_2, dy_2)}{\frac{1}{n}\sum_{i=1}^n \mu^i(x_1)} \left(\frac{1}{n}\sum_{i=1}^n \mu^i(dx_1) \right) \\ &= \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\pi^i_{x_1}, \tilde{\pi}^i_{x_1}) \, \mu^i(dx_1), \end{aligned}$$

which proves the claim. $\square$

Proof of Lemma 5.3.4. Throughout the proof we will adapt the notation of the proof of Lemma 5.3.1. The existence of the sequence $(\hat{\pi}^n)_{n\in\mathbb{N}}$ satisfying (5.3.1) follows directly from Lemma 5.3.1.*(i)* applied for $(\kappa_n)_{n\in\mathbb{N}}$. Thus we only need to show that

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i, \pi \right) \leq \frac{2}{n}\sum_{i=1}^n \kappa_i$$

for all $n \in \mathbb{N}$. Indeed by the triangle inequality

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i, \pi\right) \leq \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{k=1}^n \hat{\pi}^i, \frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i\right) + \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i, \pi\right), \tag{5.7.4}$$

where we used the same notation as in the proof of Lemma 5.3.1. For the first term in (5.7.4) we note that $\sum_{i=1}^n \hat{\pi}^i$ and $\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i$ have the same first marginals and thus by Lemma 5.3.3

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i, \frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i\right) \leq \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\hat{\pi}^i_{x_1}, \tilde{\pi}^i_{x_1})\, \mu(dx_1) \leq \frac{1}{n}\sum_{i=1}^n \kappa_i$$

as in (5.7.3) in the second step of the proof of Lemma 5.3.1.*(i)*. Next, choosing $\gamma^1 = \frac{1}{n}\sum_{i=1}^n (x, f^{\kappa_i}(x))_* \pi^1$ (where we set the ranges of f^{κ_i} without loss of generality disjoint) we obtain

$$\begin{aligned}\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i, \pi\right) &\leq \frac{1}{n}\sum_{i=1}^n \int |x_1 - f^{\kappa_i}(x_1)| + \mathcal{W}^1\Big(\big(\frac{1}{n}\sum_{i=1}^n \tilde{\pi}^i\big)_{f^{\kappa_i}(x_1)}, \pi_{x_1}\Big)\, \mu(dx_1) \\ &\leq \frac{1}{n}\sum_{i=1}^n \kappa_i.\end{aligned}$$

as in (5.7.1). Combining the two inequalities above concludes the proof of the first part of the lemma.

Denote the first marginals of π^n by $\bar{\mu}^n$ for all $n \in \mathbb{N}$. The proof of the second part of the lemma now follows as in the proof of Lemma 5.3.2, by constructing another sequence $(\bar{\pi}^n)_{n\in\mathbb{N}}$ such that $\bar{\pi}^n \in \Pi(\bar{\mu}^n, \bar{\nu}^n)$ with $\bar{\mu}^n \preceq_c \bar{\nu}^n$ and $\mathcal{W}^1_{nd}(\pi, \bar{\pi}^n) \leq \kappa_n$ for all $n \in \mathbb{N}$. As $\hat{\pi}^n$ and $\bar{\pi}^n$ have the same first marginals we now conclude from Lemma 5.3.3

$$\begin{aligned}\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i,\ \frac{1}{n}\sum_{i=1}^n \bar{\pi}^i\right) &\leq \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\hat{\pi}^i_{x_1}, \bar{\pi}^i_{x_1})\, \mu(dx_1) \\ &\leq \frac{1}{n}\sum_{i=1}^n \int \mathcal{W}^1(\hat{\pi}^i_{x_1}, \pi_{x_1}) + \mathcal{W}^1(\pi_{x_1}, \bar{\pi}^i_{x_1})\, \mu(dx_1) \leq \frac{2}{n}\sum_{i=1}^n \kappa^i.\end{aligned}$$

This gives

$$\mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i, \pi\right) \leq \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i,\ \frac{1}{n}\sum_{i=1}^n \hat{\pi}^i\right) + \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \hat{\pi}^i, \pi\right) \leq \frac{3}{n}\sum_{i=1}^n \kappa_i,$$

which concludes the proof. □

Proof of Lemma 5.3.5. Let us denote by $\zeta \in \Pi(\tilde{\mu}, \mu)$ an optimal coupling for $\mathcal{W}^p(\tilde{\mu}, \mu)$ and by $\eta \in \Pi(\nu, \tilde{\nu})$ an optimal coupling for $\mathcal{W}^p(\nu, \tilde{\nu})$. Now we define $\hat{\rho} \in \mathcal{P}(\mathbb{R}^4)$ via

$$\hat{\rho}(dx_1, dx_2, dy_1, dy_2) = \zeta_{x_1}(dy_1)\, \eta_{x_2}(dy_2)\, \pi(dx_1, dx_2).$$

Let

$$\tilde{\pi}(dy_1, dy_2) := \int_{\mathbb{R}\times\mathbb{R}} \hat{\rho}(dx_1, dx_2, dy_1, dy_2)$$

be its projection to the third and fourth component. We compute

$$\begin{aligned}
\mathcal{W}^p(\pi, \tilde{\pi}) &\leq \left(\int |x_1 - y_1|^p + |x_2 - y_2|^p \, \hat{\rho}(dx, dy) \right)^{1/p} \\
&= \left(\int \left(|x_1 - y_1|^p + |x_2 - y_2|^p\right) \, \zeta_{x_1}(dy_1)\eta_{x_2}(dy_2)\pi(dx_1, dx_2) \right)^{1/p} \\
&= ((\mathcal{W}^p(\mu, \tilde{\mu}))^p + (\mathcal{W}^p(\nu, \tilde{\nu}))^p)^{1/p} \\
&\leq \mathcal{W}^p(\mu, \tilde{\mu}) + \mathcal{W}^p(\nu, \tilde{\nu}).
\end{aligned}$$

The proof of (5.3.2) follows by use of the triangle inequality and Jensen's inequality as in [Guo and Obłój, 2019, proof of Prop. 4.2, p.20]: indeed,

$$\begin{aligned}
&\int \left| \int (y_2 - y_1) \, \hat{\rho}_{y_1}(dy_2) \right| \, \tilde{\mu}(dy_1) \\
&= \int \left| \int (y_2 - y_1) \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \right| \, \tilde{\mu}(dy_1) \\
&\leq \int \left| \int (y_2 - x_2) \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \right| \, \tilde{\mu}(dy_1) + \int \left| \int (x_2 - x_1) \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \right| \, \tilde{\mu}(dy_1) \\
&\quad + \int \left| \int (x_1 - y_1) \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \right| \, \tilde{\mu}(dy_1) \\
&\leq \int |y_2 - x_2| \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \, \tilde{\mu}(dy_1) + \int \left| \int (x_2 - x_1) \, \zeta_{y_1}(dx_1)\pi_{x_1}(dx_2) \right| \, \tilde{\mu}(dy_1) \\
&\quad + \int |x_1 - y_1| \, \hat{\rho}_{y_1}(dx_1, dx_2, dy_2) \, \tilde{\mu}(dy_1) \\
&\leq \int |y_2 - x_2| \, \eta(dx_2, dy_2) + 0 + \int |x_1 - y_1| \, \zeta(dx_1, dy_1) \\
&\leq \left(\int |y_2 - x_2|^p \, \eta(dx_2, dy_2) \right)^{1/p} + \left(\int |x_1 - y_1|^p \, \zeta(dx_1, dy_1) \right)^{1/p} \\
&= \mathcal{W}^p(\nu, \tilde{\nu}) + \mathcal{W}^p(\mu, \tilde{\mu}).
\end{aligned}$$

This concludes the proof. □

Proof of Lemma 5.3.6. As we have coupled μ and μ^n by an explicit construction in the proof of Lemma 5.3.1.*(i)* for $\Gamma = \mathbb{R}^2$ it is straightforward to extend the disintegration $x_1 \mapsto \check{\pi}^n_{x_1}$ such that it is defined μ-a.e: indeed using the notation of the proof of Lemma 5.3.1 we set

$$\begin{aligned}
\check{\pi}^n_{x_1}\big|_{[a_i, a_{i+1})\cap K_1} &:= \check{\pi}^n_{a_i}, \qquad i = 1, \ldots N-2 \\
\check{\pi}^n_{x_1}\big|_{[a_{N-1}, a_N]\cap K_1} &:= \check{\pi}^n_{a_{N-1}} \\
\check{\pi}^n_{x_1}\big|_{K_1^c} &:= \check{\pi}^n_{mr, a_0}.
\end{aligned}$$

Defining $\bar{\pi}^n := \mu \otimes \breve{\pi}^n_{x_1}$ we note that by the proof of Lemma 5.3.1 we have

$$\limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \breve{\pi}^i, \frac{1}{n}\sum_{i=1}^n \bar{\pi}^i\right) = 0.$$

By [Balder, 1995, Theorem 3.15, p.18] there exists a disintegration $x_1 \mapsto \breve{\pi}_{x_1}$ such that (after taking a subsequence) the measures

$$\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i_{x_1}\right)_{n\in\mathbb{N}}$$

converge weakly to $\breve{\pi}_{x_1}$ for μ-a.e. $x_1 \in \mathbb{R}$. Setting $\breve{\pi} := \mu \otimes \breve{\pi}_{x_1}$ this implies in particular

$$\begin{aligned}
&\limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i, \breve{\pi}\right) \\
&\quad \le \limsup_{n\to\infty} \int \left(\inf_{\gamma^2\in\Pi\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i_{x_1}, \breve{\pi}_{x_1}\right)} \int |x_2 - y_2|\,\gamma^2(dx_2, dy_2)\right) \mu(dx_1) = 0.
\end{aligned}$$

Thus also

$$\begin{aligned}
&\lim_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \breve{\pi}^i, \breve{\pi}\right) \\
&\quad \le \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \breve{\pi}^i, \frac{1}{n}\sum_{i=1}^n \bar{\pi}^i\right) + \limsup_{n\to\infty} \mathcal{W}^1_{nd}\left(\frac{1}{n}\sum_{i=1}^n \bar{\pi}^i, \breve{\pi}\right) = 0.
\end{aligned}$$

This shows *(2)* and concludes the proof. □

Proof of Lemma 5.3.7. As $\lim_{n\to\infty} \mathcal{W}^p(\mu^n, \mu) = 0 = \lim_{n\to\infty} \mathcal{W}^p(\nu^n, \nu)$ we have

$$\begin{aligned}
&\lim_{l\to\infty}\lim_{n\to\infty} \int \left((1 + |x_1|^p - l)^+ + (|x_2|^p - l)^+\right) \pi^n(dx_1, dx_2) \\
&= \lim_{l\to\infty}\lim_{n\to\infty} \int \left((1 + |x_1|^p - l)^+ + (|x_2|^p - l)^+\right) \tilde{\pi}^n(dx_1, dx_2) \\
&= \lim_{l\to\infty}\left(\lim_{n\to\infty} \int (1 + |x_1|^p - l)^+)\,\mu^n(dx_1) + \lim_{n\to\infty} \int (|x_2|^p - l)^+\,\nu^n(dx_2)\right) \\
&= \lim_{l\to\infty}\left(\int (1 + |x_1|^p - l)^+\,\mu(dx_1) + \int (|x_2|^p - l)^+\,\nu(dx_2)\right) = 0.
\end{aligned}$$

Furthermore the Portmanteau theorem and $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$ implies

$$\begin{aligned}
\lim_{l\to\infty}\limsup_{n\to\infty} l\mu_n(1 + |x_1|^p \ge l) &\le \lim_{l\to\infty} l\mu(1 + |x_1|^p \ge l) \\
&\le \lim_{l\to\infty} \int_{\{1+|x_1|^p\ge l\}} (1 + |x_1|^p)\,\mu(dx_1) = 0
\end{aligned}$$

and

$$\lim_{l\to\infty}\limsup_{n\to\infty} l\nu_n(|x_2|^p \ge l) \le \lim_{l\to\infty} l\nu(|x_2|^p \ge l) \le \lim_{l\to\infty} \int_{\{|x_2|^p\ge l\}} |x_2|^p\,\nu(dx_2) = 0.$$

Writing

$$\begin{aligned}\{|c(x_1,x_2)|>k\} = \{&1+|x_1|^p > k/(2C), |x_2|^p \le k/(2C)\}\\ &\cup \{1+|x_1|^p \le k/(2C), |x_2|^p > k/(2C)\}\\ &\cup \{1+|x_1|^p > k/(2C), |x_2|^p > k/(2C)\},\end{aligned}$$

for $k \in \mathbb{N}$, choosing $l = k/(2C)$ and using the estimates above, this yields

$$\begin{aligned}&\left|\lim_{n\to\infty} \int c(x_1,x_2)\,\pi^n_{mr}(dx_1,dx_2) - \lim_{n\to\infty}\int c(x_1,x_2)\,\tilde{\pi}^n(dx_1,dx_2)\right|\\ &\le \lim_{k\to\infty}\Big|\lim_{n\to\infty}\int (c(x_1,x_2)\vee(-k)\wedge k)\ \pi^n(dx_1,dx_2)\\ &\quad - \lim_{n\to\infty}\int (c(x_1,x_2)\vee(-k)\wedge k)\ \tilde{\pi}^n(dx_1,dx_2)\Big|\\ &= 0.\end{aligned}$$

as $\lim_{n\to\infty}\mathcal{W}^1(\pi^n,\tilde{\pi}^n)=0$ and $(c(x_1,x_2)\vee(-k)\wedge k)$ is bounded and continuous. □

Bibliography

B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer. A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Math. Finance*, 26(2):233–251, 2013.

A. Aksamit, S. Deng, J. Obłój, and X. Tan. The robust pricing–hedging duality for American options in discrete time financial markets. *Math. Finance*, pages 1–37, 2018. Published Online.

C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis : A Hitchhiker's Guide*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 3rd edition, 2006.

M. Avellaneda, A. Levy, and A. Paras. Pricing and hedging derivatives securities in markets with uncertain volatilities. *Appl. Math. Finance*, 2(2):73–88, 1996.

J. Backhoff, D. Bartl, M. Beiglböck, and J. Wiesel. Estimating processes in adapted wasserstein distance. *arXiv preprint arXiv:2002.07261*, 2020.

J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *arXiv preprint arXiv:1904.04171.*, 2019.

J. Backhoff-Veraguas, M. Beiglböck, M. Eder, and A. Pichler. Fundamental Properties of Process Distances. *arXiv preprint arXiv:1701.03955*, 2017a.

J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM J. Optim.*, 27(4):2528–2562, 2017b.

J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein Distances and Stability in Mathematical Finance. *arXiv preprint arXiv:1901.07450*, 2019.

E. J. Balder. Lectures on young measures. *Cahiers de Mathématiques de la Décision*, 9517, 1995.

D. Bartl. Exponential utility maximization under model uncertainty for unbounded endowments. *Ann. Apl. Prob.*, 29(1):577–612, 2019.

D. Bartl, P. Cheridito, M. Kupper, and L. Tangpi. Duality for increasing convex functionals with countably many marginal constraints. *Banach J. Math. Anal*, 11(1):72–89, 2017.

D. Bartl, M. Kupper, and A. Neufeld. Pathwise superhedging on prediction sets. *Financ. Stoch.*, 24(1):215–248, 2020.

Basel Committee on Banking Supervision. Fundamental review of the trading book: A revised market risk framework. Technical report, 2013.

Basel Committee on Banking Supervision. Minimum capital requirements for market risk. *Bank for International Settlements, Basel, Switzerland*, 2019.

B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Process. Appl*, 99(1):95–115, 2002.

C. Bayer and J. Teichmann. The proof of Tchakaloff's theorem. *Proc. Amer. Math. Soc*, 134(10):3035–3040, 2006.

E. Bayraktar and Y. Zhang. Fundamental theorem of asset pricing under transaction costs and model uncertainty. *Math. Oper. Res.*, 41(3):1039–1054, 2016.

E. Bayraktar and Z. Zhou. On arbitrage and duality under model uncertainty and portfolio constraints. *Math. Finance*, 27(4):988–1012, 2017.

E. Bayraktar, Y. Zhang, and Z. Zhou. A note on the fundamental theorem of asset pricing under model uncertainty. *Risks*, 2(4):425–433, 2014.

M. Beiglböck and C. Griessler. A land of monotone plenty. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 19(1):109–127, 2019.

M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016.

M. Beiglböck and M. Nutz. Martingale inequalities and deterministic counterparts. *Electron. J. Probab.*, 19(94):1–15, 2014.

M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance Stoch.*, 17(3):477–501, 2013.

M. Beiglböck, A. M. G. Cox, M. Huesmann, N. Perkowski, and D. Prömel. Pathwise superreplication via vovk's outer measure. *Finance Stoch.*, 21(4):1141–1166, 2017.

M. Beiglböck, M. Nutz, and N. Touzi. Complete duality for martingale optimal transport on the line. *Ann. Apl. Prob.*, 45(5):3038–3074, 2017.

D. Bertsekas and S. Shreve. *Stochastic optimal control: The discrete time case*, volume 23. Academic Press New York, 1978.

S. Biagini, B. Bouchard, C. Kardaras, and M. Nutz. Robust fundamental theorem for continuous processes. *Math. Finance*, 27(4):963–987, 2017.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Polit. Econ.*, 81(3):637–654, 1973a.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Political Econ.*, pages 637–654, 1973b.

R. Blanchard and L. Carassus. Multiple-priors investment in discrete time for unbounded utility function. *Ann. Apl. Prob.*, 2017.

R. Blanchard and L. Carassus. No-arbitrage with multiple-priors in discrete time. *arXiv preprint arXiv:1904.08780*, 2019.

V. Bogachev. *Measure theory*, volume 1. Springer-Verlag, Berlin, 2007.

W. Bonnice and J. Reay. Relative interiors of convex hulls. *Proc. Am. Math. Soc.*, 20(1): 246–250, 1969.

B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.*, 25:823–859, 2015.

B. Bouchard, S. Deng, and X. Tan. Super-replication with proportional transaction cost under model uncertainty. *Math. Finance*, 29(3):837–860, 2019.

F. Boussama. *Ergodicité, mélange et estimation dans les modeles GARCH.* PhD thesis, Paris 7, 1998.

M. Burzoni, M. Frittelli, and M. Maggis. Universal arbitrage aggregator in discrete-time markets under uncertainty. *Finance Stoch.*, 20(1-50), 2016.

M. Burzoni, M. Frittelli, and M. Maggis. Model-free superhedging duality. *Ann. Appl. Probab.*, 27(3):1452–1477, 2017a.

M. Burzoni, F. Riedel, and H. Soner. Viability and arbitrage under Knightian Uncertainty. 2017b.

M. Burzoni, M. Frittelli, Z. Hou, M. Maggis, and J. Obłój. Pointwise arbitrage pricing theory in discrete time. *Math. Oper. Res.*, 2019.

L. Carassus and T. Vargiolu. Super-replication price: it can be ok. *ESAIM Proc. Surv.*, 65: 241–281, 2018.

L. Carassus, E. Gobet, and E. Temam. A class of financial products and models a class of financial products and models where super-replication prices are explicit. In *International Symposium on Stochastic Processes and Mathematical Finance*. Ritsumeikan University, Tokyo, 2006.

L. Carassus, J. Obłój, and J. Wiesel. The robust superreplication problem: A dynamic approach. *SIAM J. Financial Math.*, 10(4):907–941, 2019.

M. Carrasco and X. Chen. Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39, 2002.

A. Casal. Set estimation under convexity type assumptions. In *Ann. Inst. H. Poincaré Probab. Statist.*, volume 43, pages 763–774. Elsevier, 2007.

P. Cheridito, M. Kupper, and L. Tangpi. Duality formulas for robust pricing and hedging in discrete time. *SIAM J. Financial Math.*, 8(1):738–765, 2017.

J. Chevalier. Estimation du support et du contenu de support d'une loi de probabilité. *Ann. Inst. Henri Poincaré Probab. Stat.*, 12:339–364, 1976.

D. Cohn. *Measure theory.* Birkhäuser, 1980.

R. Cont. Model uncertainty and its impact on pricing derivative instruments. *Math. Finance*, 16(3):519–547, 2006.

R. Cont, R. Deguest, and G. Scandolo. Robustness and sensitivity analysis of risk measurement procedures. *Quant. Finance*, 10(6):593–606, 2010.

A. Cox and J. Obłój. Robust pricing and hedging of double no-touch options. *Finance Stoch.*, 15(3):573–605, 2011.

A. Cox, Z. Hou, and J. Obłój. Robust pricing and hedging under trading restrictions and the emergence of local martingale models. *Financ. Stoch.*, 20(3):669–704, 2016.

J. Cox, S. Ross, and M. Rubistein. Option pricing: a simplified approach. *J. Financial Econ.*, 7(229-264), 1979.

A. Cuevas. On pattern analysis in the non-convex case. *Kybernetes*, 19(6):26–33, 1990.

A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Adv. Appl. Prob.*, 36(2): 340–354, 2004.

M. Davis and D. Hobson. The range of traded option prices. *Math. Finance*, 17(1):1–14, 2007.

F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(1):463–520, 1994.

F. Delbaen and W. Schachermayer. *The Mathematics of Arbitrage*. Springer-Verlag, Berlin, 2006.

L. Denis and M. Kervarec. Utility functions and optimal investment in non-dominated models. 2007.

L. Denis and C. Martini. A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *Ann. Appl. Probab.*, 16(2):827–852, 2006.

L. Devroye and G. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.*, 38(3):480–488, 1980.

Y. Dolinsky and H. Soner. Robust hedging with proportional transaction costs. *Finance Stoch.*, 18(2):327–347, 2014.

R. M. Dudley. *Real analysis and probability*. Chapman and Hall/CRC, 2004.

B. Dupire. Talk and discussion during the “Robust Techniques in Quantitative Finance” Conference, Oxford-Man Institute of Quantitative Finance. March 2010.

S. Eckstein, M. Kupper, and M. Pohl. Robust risk aggregation with neural networks. *arXiv preprint arXiv:1811.00304*, 2018.

P. Embrechts, B. Wang, and R. Wang. Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stoch.*, 19(4):763–790, 2015.

P. Embrechts, A. Schied, and R. Wang. Robustness in the optimization of risk measures. *arXiv preprint arXiv:1809.09268*, 2018.

L. G. Epstein and S. Ji. Ambigous volatility, possibility and utility in continuous time. *J. Math. Econ.*, 50:269–282, 2014.

P. Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming*, 171:115–166, 2018.

H. Föllmer and D. Kramkov. Optional decompositions under constraints. *Probab. Theory Related Fields*, 109(1025), 1997.

H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, 2004.

N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4):707–738, 2015.

C. Francq and J.-M. Zakoïan. Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica*, 80(2):821–861, 2012.

C. Francq and J.-M. Zakoian. *GARCH models: structure, statistical inference and financial applications*. Wiley, 2019.

A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Ann. Appl. Probab.*, 24(1):312–336, 2014.

J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quant. Finance*, 18(6): 933–949, 2018.

J. Geffroy. Sur un problème d'estimation géométrique. *Publications de l'Institut de Statistique des Universités de Paris*, 13:191–210, 1964.

N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017.

U. Grenander. Abstract inference. Technical report, 1981.

C. Griessler. An extended footnote on finitely minimal martingale measures. *arXiv preprint arXiv:1606.03106*, June 2016.

G. Guo and J. Obłój. Computational Methods for Martingale Optimal Transport problems. *Ann. Apl. Prob.*, 29(6):3311–3347, 2019.

P. Hall and Q. Yao. Inference in arch and garch models with heavy–tailed errors. *Econometrica*, 71(1):285–317, 2003.

F. Hampel. A general qualitative definition of robustness. *Ann. Math. Stat.*, 42(6):1887–1896, 1971.

W. Hardle, B. Park, and A. Tsybakov. Estimation of non-sharp support boundaries. *J. Multivariate Anal.*, 55(2):205–218, 1995.

C. He and T. Teräsvirta. Fourth moment structure of the GARCH (p, q) process. *Econom. Theory*, 15(6):824–846, 1999.

D. Hobson. Robust hedging of the lookback option. *Finance Stoch.*, 2:329–347, 1998.

D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015.

Z. Hou and J. Obłój. Robust pricing-hedging dualities in continuous time. *Finance Stoch.*, 22:511–567, 2018.

F. Hubalek and W. Schachermayer. When does convergence of asset price processes imply convergence of option prices? *Math. Finance*, 8(4):385–403, 1998.

P. Huber. *Robust statistical procedures.* SIAM, 1996.

P. Huber and E. Ronchetti. *Robust statistics.* Wiley, 2009.

E. Jouini, W. Schachermayer, and N. Touzi. Law invariant risk measures have the Fatou property. In *Adv. Math. Econ.*, pages 49–71. Springer, 2006.

B. Jourdain and W. Margheriti. A new family of one dimensional martingale couplings. *arXiv preprint arXiv:1808.01390.*, Aug. 2018.

N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016.

Y. Kabanov and D. Kramkov. Large financial markets: asymptotic arbitrage and contiguity. *Theory Probab. Appl.*, 39(1):182–187, 1995.

O. Kallenberg. *Foundations of Modern Probability.* Springer Berlin, 2002.

H. Kellerer. Duality theorems and probability metrics. In *Proceedings of the seventh conference on probability theory (Brasov, 1982)*, pages 211–220, 1982.

F. Knight. *Risk, Uncertainty, and Profit.* Boston, MA: Hart, Schaffner Marx; Houghton Mifflin Co, 1921.

A. Korostelëv, L. Simar, and A. Tsybakov. Efficient estimation of monotone boundaries. *Ann. Statist.*, pages 476–489, 1995.

M. Kosorok. *Introduction to empirical processes and semiparametric inference.* Springer, 2008.

S. Kou, X. Peng, and C. Heyde. External risk measures and basel accords. *Math. Oper. Res.*, 38(3):393–417, 2013.

V. Krätschmer, A. Schied, and H. Zähle. Qualitative and infinitesimal robustness of tail-dependent statistical functionals. *J. Multivariate Anal.*, 103(1):35–47, 2012.

V. Krätschmer, A. Schied, and H. Zähle. Comparative and qualitative robustness for law-invariant risk measures. *Finance Stoch.*, 18(2):271–295, 2014.

S. Kusuoka. On law invariant coherent risk measures. In *Adv. Math. Econ.*, pages 83–95. Springer, 2001.

K. Lange. Borels sets of probability measures. *Pacific J. Math.*, 48(1):141–161, 1973.

R. Lassalle. Causal transport plans and their Monge-Kantorovich problems. *Stoch. Anal. Appl.*, 36(3):452–484, 2018.

L. Le Cam and G. Yang. *Locally asymptotically normal families*, pages 52–98. Springer-Verlag, Berlin, 1990.

X. Li and Y. Lin. Generalized Wasserstein distance and weak convergence of sublinear expectations. *J. Theoret. Probab.*, 30(2):581–593, 2017.

A. M. Lindner. Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer, 2009.

F. Lyons. Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance*, 2:117–133, 1995.

E. Mammen and A. Tsybakov. Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.*, pages 502–524, 1995.

R. C. Merton. Theory of rational option pricing. *Bell J. Econ.*, 4(1):141–183, 1973.

A. Mokkadem. Propriétés de mélange des processus autorégressifs polynomiaux. In *Ann. Inst. Henri Poincaré Probab. Stat*, volume 26, pages 219–260, 1990.

G. Monge. *Mémoire sur la théorie des déblais et des remblais.* Histoire de l'Académie Royale des Sciences de Paris., 1781.

P. Mykland. Conservative delta hedging. *Ann. Appl. Probab.*, pages 664–683, 2000.

P. Mykland. Financial options and statistical prediction intervals. *Ann. Statist.*, 31(5): 1413–1438, 2003a.

P. Mykland. The interpolation of options. *Finance Stoch.*, 7(4):417–432, 2003b.

A. Neufeld and M. Sikic. Robust utility maximization in discrete time with friction. *SIAM J. Control Optim.*, 56(3):1912–1937, 2018.

M. Nutz. Utility maximisation under model uncertainty in discrete time. *Math. Finance*, 26(2):252–268, 2016.

J. Obłój and J. Wiesel. A unified framework for robust modelling of financial markets in discrete time. 2018. arXiv prerprint arXiv:1808.06430.

J. Obłój and J. Wiesel. Robust estimation of superhedging prices. *Ann. Stat. (to appear)*, 2020.

J. Obłój and J. Wiesel. Supplement to "Robust Estimation of Superhedging Prices". *Ann. Stat. (to appear)*, 2020.

S. Peng. Nonlinear expectations, nonlinear evaluations and risk measures. In *Stoch. Methods Finance*, pages 165–253. Springer, 2004.

G. C. Pflug. Version-independence and nested distributions in multistage stochastic optimization. *SIAM J. Optim.*, 20(3):1406–1420, 2009.

G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.

A. Pichler. Evaluations of risk measures for different probability measures. *SIAM J. Optim.*, 23(1):530–551, 2013.

W. Polonik. Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *Ann. Statist.*, 23(3):855–881, 1995.

M. Rásonyi. A note on martingale measures with bounded densities. *Tr. Mat. Inst. Steklova*, 237(0):212–216, 2002.

M. Rásonyi and L. Stettner. On the existence of optimal portfolios for the utility maximization problem in discrete time financial models. *In: Kabanov, Y.; Lipster, R.; Stoyanov,J. (Eds), From Stochastic Calculus to Mathematical Finance, Springer.*, pages 589–608, 2006.

F. Riedel. Financial economics without probabilistic prior assumptions. *Decis. Econ. Finance*, 38(1):75–91, 2015.

R. Rockafellar and R. Wets. *Variational Analysis*, volume 317. Springer-Verlag, Berlin, 2009.

R. T. Rockafellar. *Convex Analysis*. Princeton, 1970.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1998. ISBN 3-540-62772-3.

D. Rokhlin. A proof of the Dalang-Morton-Willinger theorem. *arXiv preprint arXiv:0804.3308*, 2008.

M.-F. Sainte-Beuve. On the extension of von Neumann-Aumann's theorem. *J. Functional Analysis*, 17(1):112–129, 1974.

P. Samuelson. Rational theory of warrant pricing. *Ind. Manag. Rev.(pre-1986)*, 6(2):13, 1965.

P. A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Rev. Econom. Statist.*, 51:239–246, 1969.

H. Sartipizadeh and T. Vincent. Computing the approximate convex hull in high dimensions. *arXiv preprint arXiv:1603.04422*, 2016.

A. Schied and C.-T. Wu. Duality theory for optimal investments under model uncertainty. *Stat. Decis.*, 23(3/2005):199–217, 2005.

G. Shorack and J. Wellner. *Empirical processes with applications to statistics*, volume 59. Siam, 2009.

V. Strassen. The existence of probability measures with given marginals. *Ann. Statist.*, pages 423–439, 1965.

D. W. Stroock and S. S. Varadhan. *Multidimensional diffusion processes*. Springer, 2007.

R. Takahashi. Normalizing constants of a distribution which belongs to the domain of attraction of the Gumbel distribution. *Stat. Probab. Lett.*, 5(3):197–200, 1987.

F. Terkelsen. Some minimax theorems. *Math. Scand.*, 31(2):405–413, 1973.

N. Trillos and D. Slepčev. On the rate of convergence of empirical measures in infinity-transportation distance. *Canad. J. Math.*, 67:1358–1383, 2015.

A. Tsybakov. On nonparametric estimation of density level sets. *Ann. Statist.*, 25(3): 948–969, 1997.

A. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge University Press, 1998.

R. Vanderbei. Uniform continuity is almost Lipschitz continuity. Technical report, SOR-91-11, Statistics and Operations Research Series, 1997.

C. Villani. *Optimal Transport: Old and New*, volume 338. Springer, 2008.

J. Wiesel. Continuity of the martingale optimal transport problem on the real line. *arXiv preprint arXiv:1905.04574*, 2019.