DISS. ETH NO. 24311

Trading with Small Nonlinear Price Impact

Optimal Execution and Rebalancing of Active Investments

A dissertation submitted to attain the degree of

Doctor of Sciences of ETH Zürich
(Dr. sc. ETH Zürich)

presented by
Thomas Marie Léopold André Cayé

MSc in Applied Mathematics, ETH Zürich,
MSc in Quantitative Finance, UZH/ETH Zürich,
MSc in Engineering, École Polytechnique, Paris

born on October 22, 1987
citizen of France

Accepted on the recommendation of

Prof. Dr. Johannes Muhle–Karbe	examiner
Prof. Dr. H. Mete Soner	examiner
Prof. Dr. Paolo Guasoni	co-examiner
Dr. Martin Herdegen	co-examiner

2017

ISBN 978-3-906327-91-4

Acknowledgements

This thesis is the result of four years of effort that would never have been possible without the strong support of the people around me.

I want to start by expressing my deep and sincere gratitude to Johannes Muhle-Karbe for accepting to supervise my PhD studies. I very much profited from his experience, his advices and his deep knowledge of mathematics and economics. The topics he suggested for this dissertation fitted perfectly what I wished to work on, and I am happy that it lead to a fortunate conclusion. His commitment, availability and support had no limit and extended far beyond what should be expected of a supervisor. I am greatly indebted to him, and am aware of the luck that befell me in being able to work under his supervision. I am greatly thankful, as well, for the time and guidance of Martin Herdegen. The beginning of a learning process is often laborious and research is no exception. Our many discussions showed me what to improve and contributed strongly to make me a better researcher. All of this would not have happened without the support of Mete Soner, who accepted to co-supervise this thesis. I want to express my most genuine thanks to him for having made it possible for me to work in the best conditions, and permitting me to stay in the ETH Mathematics Department when Johannes Muhle-Karbe moved to Ann Arbor. I want to thank Paolo Guasoni as well, for having accepted to be a co-examiner for this thesis, for giving me the opportunity to give a talk at the seminar of Financial Mathematics in Dublin and for his numerous suggestions. Finally, I am grateful to Marko Weber for numerous discussions and for sharing insights on his work.

My years in Groupe 3 have had ups and downs, but always I could count on colleagues to make the rough times easier and the good times truly memorable. Time leads to take it for granted, but I truly consider having been able to work in such an environment and with such great people a luxury. I will remember fondly the many debates and discussions I had with my office mate Ibrahim Ekren, who over the years has always been ready to help, answer a question or share the delights of burger day. I am grateful to Ren Liu for having paved the way before me, and having shown me what my life would be a year ahead. Finally I want to thank my office mates over the years, (in chronological order) Ludovic Moreau, Albert Altarovici, Pierre-François Rodriguez, Ron Rosenthal, Xinyi Li, Mayra Bermúdes Contreras, Thomas Krabichler and Aditi Dandapani for having put up with my shenanigans, as well as all my colleagues of Groupe 3. Among them, Max Reppen for having introduced me to the delicate science of coffee brewing, Matti Kiiski for the music and literature discussions, Danijel Zivoi for his extensive knowledge and appetite for mathematical problems and brain teasers, Daniel Balint for always greeting me with a word or two of French, Mario Sikic for sharing his philosophical views of the world, Calypso for disagreeing with me in about all of our political discussions, Sara for her flawless organisation of the group and all the others.

I would be forgetting the most important if I omitted to thank my parents. Their unconditional support and love, through thick and thin, allowed me to start on a strong basis and to keep going forward when spirits were down. I dedicate this work to them.

Contents

Abstract

This work deals with nonlinear price impact and is composed of two parts. The first part extends the well known Almgren-Chriss price impact model. The aim is to find optimal liquidation strategies: an investor wants to sell a large amount of shares or build up a large position and needs to optimize the trading schedule in a market where trading moves the prices in the adverse direction. Here we do not consider linear price impacts as Almgren and Chriss, but price impacts whose intensity increases with the amount of shares traded: this is a reduced-form model to take into account predatory trading, the limited depth of the market and its finite resilience. We find that, even for a risk-neutral trader in such a setting, the best course is to liquidate more upfront to avoid front-runners as much as possible. However, for a risk-averse investor, the interplay between risk aversion and self-excitement is non-trivial, and the addition of self-excitement does not necessarily lead to faster initial trading. Self-exciting price impacts have a mitigating effect in that they increase effective risk aversion if the latter is relatively small and on the contrary decrease it if it is relatively large. We also provide a micro-foundation for our model, by arguing that it arises from an equilibrium between market makers, fundamental sellers and end users.

The second and main part of this work, deals with portfolio choice in general markets with nonlinear price impact. An investor with constant absolute risk aversion trades to maximize utility of terminal wealth. The gains of trading are diminished by transaction costs proportional to a power $p \in (1, 2)$ of the trading rate, and the price follows a general, not necessarily Markovian Itô diffusion. With careful estimates and the use of ergodic theorems for stochastic differential equations, we establish asymptotic optimality of an explicit family of strategies in the limit for small price impacts. These strategies track at a finite rate the optimizer of the frictionless version of the problem. This finite speed can be expressed in terms of the model primitives: the volatilities of the asset price and the frictionless optimizer as well as price impact intensity. The influence of the current position on the trading speed appears in its difference from the frictionless optimal position, and is modulated by the solution to

a universal ordinary differential equation depending only on p. The analysis requires relatively mild assumptions. This allows us to use the results for a wide range of applications: optimal portfolio choice and rebalancing of active investments as well as derivative pricing and hedging. Finally, similarly to the previous model, we propose a way to endogenize the price impacts through an equilibrium between market makers and end users, where transaction costs stem from the risk aversion or inventory costs of the market maker.

Résumé

Cette thèse traite des impacts de marché non-linéaires et se compose de deux parties. La première consiste en une extension du modèle d'Almgren et Chriss. Ce modèle vise à planifier de manière optimale l'achat et la vente d'actifs, pour un investisseur souhaitant liquider une importante position ou acheter un grand nombre de ces actifs sur un marché où l'achat et la vente provoquent un mouvement des prix dans la direction opposée à la transaction. Pour incorporer au modèle le trading aggressif, le "front-running", l'illiquidité et la résilience finie du marché, nous ne considérons pas d'impacts de marché linéaires comme dans le modèle d'Almgren et Chriss, mais des impacts dont l'intensité croit au cours du temps avec le nombre d'actifs achetés ou vendus. Dans ces conditions, la stratégie optimale est d'accélérer la liquidation au début de la période, et cela même en l'absence d'aversion au risque. Dans le cas d'un investisseur possédant une forte aversion au risque, l'interdépendence entre l'auto-stimulation des impacts de marché et l'aversion au risque n'engage pas nécéssairement à accélérer les opérations d'achat et de vente. L'auto-stimulation des impacts de marché atténue au contraire les effets de l'aversion au risque : une forte aversion est effectivement diminuée par l'introduction de l'auto-stimulation des impacts de marché alors qu'une aversion relativement faible semble augmentée. Enfin, nous fournissons une justification de ce modèle par un équilibre entre trois types d'agents opérant sur le marché financier : un teneur de marché servant d'intermédiaire entre un investisseur qui souhaite liquider sa position et les acheteurs "fondamentaux".

La seconde et principale partie de cette thèse traite d'un problème d'optimisation de portefeuille dans un marché sujet à des impacts de marché non-linéaires. L'investisseur, dont l'aversion au risque absolue est constante, souhaite maximiser l'utilité de sa richesse finale, obtenue par la gestion active du portefeuille et pénalisée par les coûts de transactions encourus lors des opérations financières. La dynamique des processus de prix est donnée par un processus d'Itô dont les caractéristiques ne sont pas nécéssairement Markoviennes. Au moyen d'estimations précises, nous prouvons l'optimalité asymptotique d'une famille de stratégies, à la limite de petits impacts de marchés. Ces stratégies suivent à vitesse finie la stratégie optimale sur le

même marché dépourvu de frictions. Cette vitesse peut être exprimée en fonction des données du problème : volatilité du prix et de la stratégie optimale obtenus s'il n'y avait pas d'impacts de marché et leur intensité. La position actuelle du portefeuille intervient à travers sa distance de la position optimale dans le marché sans frictions, cette distance étant modulée par l'application d'une fonction, solution d'une equation différentielle ordinaire universelle (ne dépendant que de l'élasticité des impacts de marché). L'analyse du problème nécessite des hypothèses relativement restreintes, ce qui permet l'application des résultats obtenus à de nombreuses questions de finance, telles que la gestion optimale de portefeuille, la réplication ou couverture de produits dérivés ainsi que la détermination de leur prix (par indifférence d'utilité). Enfin, comme pour le précédent modèle traité en première partie, nous fournissons un moyen de rendre ces impacts de marché endogènes au modèle, par un équilibre entre un teneur de marché ayant une aversion au risque ou un coût d'inventaire relativement faible, et le gérant de portefeuille.

Introduction

Starting from a description of market dynamics and agents' preferences, the purpose of Mathematical Finance is to determine the optimal way to trade and manage risk. Without simplifying assumptions, most such optimization problems are intractable. For this reason, classical Financial Mathematics focused first on idealized *frictionless* markets: no amount constraints on buying and selling, no bound on debts, trading happens continuously in time and share proportion, on markets where assets price processes are continuous and information is immediately and fully available. With this in mind, one typically considers a rational investor[1] that seeks to optimize a given criterion. Classical choices are mean-variance functionals and the expectation of increasing and concave, von Neumann-Morgenstern utility functions of consumption, terminal wealth or of a mixture of the two. In this setting, the aim is to obtain an optimal trading strategy and an optimal consumption policy.

Frictionless Models

Optimal Portfolio Choice and Utility Maximization. After the mean-variance portfolio theory of Markowitz [135] which focuses on one period markets, the first breakthrough in continuous-time Mathematical Finance is due to Merton and his pioneering works ([140], [141]), building on a first discrete approach by Samuelson [164]. Using optimal stochastic control methods, Merton found the optimal trading and consumption policies for an investor maximizing utility of instantaneous consumption. He first treated the cases of isoelastic utility and of constant absolute risk aversion (CARA), but later extended the method to more general utility functions. In a market where the price process of the only risky asset is a geometric Brownian

[1] A new strand of literature stemmed from the critics by Kahneman and Tversky of the theory of expected utility [100]: prospect theory tries to account for the observed lack of rationality demonstrated by human agents. For recent advances on portfolio optimization in such a setting see [154], [155] and references therein.

motion, he concludes that the trader should keep a constant proportion of wealth in the risky asset if his relative risk aversion is constant (power- or log-utility functions) or a constant value invested in the risky asset if his absolute risk aversion is constant (exponential utility function). These optimal solutions are obtained by solving a nonlinear partial differential equation, whose derivation crucially depends on the Markov property of the price process. To remove this constraint, a martingale duality methodology was developed by Pliska, first on a discrete probability space [147] and extended to adapted processes in [148]. On incomplete markets, he provides a characterization of the maximal utility, but needs completeness to exhibit a strategy achieving this value. Building on his results, Karatzas, Lehoczky, Shreve and Xu [111, 112] solve the problem with consumption in complete and incomplete markets where prices are Itô diffusions driven by Brownian motions. These advances permitted the resolution of the optimal portfolio problem in much more general frameworks, for exponential utility [56, 166] (by Delbaen, Grandits, Rheinländer, Samperi, Schweizer, Stricker and Schachermayer), for power utility [105] (by Kallsen and Muhle-Karbe), and culminated in the work of Kramkov and Schachermayer on incomplete markets with general utility functions [116] (see also [117, 165, 167], for additional properties of the solutions).

Replication and Optimal Hedging. A special case of these optimal trading problems is the replication and optimal hedging of derivative securities. Banks and financial institutions sell contracts whose value depends on underlying assets. The simplest examples of such products are European call and put options, giving their holder the right (but not the obligation) to respectively buy or sell the asset for a pre-agreed price, decided at initiation of the contract. In a complete market, any such claim is replicable, and particularly in the case of a single risky asset with geometric Brownian prices as shown in the seminal work of Black and Scholes [25] (later extended by Harrison and Pliska [84] to general complete markets). In these markets, the possibility to trade in the option does not offer any diversification gain, and so the portfolio optimization problem with initial holding of the derivative does not depart from the above setting: the investor perfectly hedges the claim, and applies the optimal portfolio strategy without claim.

However, real world markets are incomplete and a part of the claim cannot be hedged. The problem in turn becomes more challenging. One way to deal with this additional complexity is to transform the functional to maximize to include the payoff of the derivative security to the final wealth of the investor. The investor's optimal strategy then gives an optimal hedge given her initial position, and one can obtain the price of the derivative by equating the utility derived from owning the claim only to the utility derived from an initial capital without the claim. This method is called indifference pricing and is illustrated by [56] in the case of a general market for an investor with exponential utility function. For CARA investors a beautiful simplification indeed allows to solve the problem by a simple change of probability measure on the underlying probability space. To extend this to general utility functions, Cvitanić, Schachermayer and Wang [50], adapted the martingale duality method to utility maximization with random endowment in incomplete markets by enlarging the dual space of [116] to the dual of L^∞. The dual variable can indeed fail to be a martingale measure, and the loss of mass is accounted for by extending the admissibility space for the dual problem[2].

Portfolio and hedging problems are made considerably more arduous by frictions and numerous works have been devoted to solve the Merton problem (optimal investment in stochastic financial markets) and its variants in presence of frictions as we will shortly see.

Frictions in Financial Markets

Frictions refer to any real-world phenomenon that hinders the investor when trading. There is a minimum duration between two buying or selling orders, due to the physical transmission process, the strategy computation time and information gathering. Furthermore, it is not possible to divide indefinitely the traded assets. Hence, trading is inherently discrete.[3]

There is a limit to the amount of debts an investor can contract, and her wealth cannot reach high depth before she is forced to declare bankruptcy. In other words, her wealth process is bounded from below[4].

Of any asset (stocks, derivative products, currencies, commodities) there is finite supply, and the number of shares or units of assets available is bounded. Moreover, trading is further limited by availability of counterparties. The investor's position in assets is therefore constrained[5].

[2] Another pricing method is considered by Hugonnier, Kramkov and Schachermayer [94], the marginal utility based price. In an incomplete general semimartingale price model, they find a condition for the existence and uniqueness of this price.

[3] For a study of optimal discretisation of hedging strategies in continuous markets see [161, 62, 63, 33], for its extension to models with jumps cf. [162, 31].

[4] This also conveniently removes the doubling strategies that can generate arbitrages, even in well-behaved models; see [84].

[5] A significant strand of literature dedicated itself to solve the Merton problem and optimal hedging problems in presence of trading constraints, see for example [46, 47, 110, 30, 49, 39] and

Finally, trading induces costs that add to the quoted asset price. Fixed costs are charged by the broker or the bank acting as intermediary for the transaction, or represents the cost of collecting information necessary to decide on the trade. Proportional costs can represent the bid-ask spread - the difference between the best buy price and sell price offered for the asset - or fees the intermediary charges per transaction. Finite market depth is the source of a third type of costs: price impact. This is what we study in this thesis. Whether trading happens through a limit order book or over the counter, it is not possible to fill a large trading order at the quoted price, and on average the execution price is worse than what is observable. The extent of the price modification is typically found to be proportional to a power of the order size and execution speed; most practitioners agree that it is close to 0.5 - a rule of thumb dubbed "the square root law", cf. [123, 10].

Direct Transaction Costs

In continuous-time models of frictionless financial markets, the optimal trading strategies for utility maximization or hedging problems are typically semimartingales of infinite variation. A prototypical example is the world famous Black Scholes formula, giving the replication strategy as a time-dependent function of the price process (a geometric Brownian motion). However, actual trading incurs costs and using a strategy with infinite variation leads immediately to bankruptcy – for all types of transaction costs. The fundamental properties of the optimal control therefore change drastically. One obtains "local time" type controls for proportional costs, impulse controls for fixed costs, and finite trading rates for superlinear costs.

Proportional Transcation Costs and No-Trade Region. These are arguably the most studied of the transaction costs. Their investigation was spearheaded by Constantinides and Magill [43], who studied a finite horizon optimal consumption problem with proportional costs. They find that, in a Black-Scholes market, the aim of an investor should be to keep the proportion of wealth in the risky asset close to the one prescribed by Merton in [140, 141] (the Merton proportion). Effectively, the investor should monitor her investment in the stock and keep the proportion between two bounds around the Merton proportion, i.e. in a so-called no-trade region. The length of this interval is constant, independent of time and wealth. For the proportion to stay inside the region, the investor should just trade enough to bring it back when it breaches the boundary. In a later article [42], Constantinides finds that the welfare loss due to proportional transaction costs is actually small, as the utility derived from

the references therein.

trading only weakly depends on the displacement from the Merton proportion and agents can adapt the no-trade region to reduce effects of a large bid-ask spread. Once the milestone set, many extensions and generalizations emerged, pushed both by the needs of the industry [138, 136, 137], and academic curiosity. Financial economists such as Lynch, Balduzzi and Tan studied the question in a more applied fashion [14, 132, 133], while financial mathematicians pursued rigorous characterizations of the no-trade region and the optimal strategy in general markets. The first step toward a better comprehension of the no-trade frontiers in a Black-Scholes market was made by Davis and Norman [53], where the no-trade region is characterised by a nonlinear free-boundary problem[6]. This allowed to numerically obtain a solution. The second strand of literature set out to study the asymptotics of the problem in the limit of small transaction costs. As for many stochastic control problems, there are two main approaches. The first is to derive the Hamilton-Jacobi-Bellmann equation and wield the powerful machinery of viscosity solutions to handle the obtained partial differential equation. That is the one chosen by Shreve, Soner, Janeček, Possamaï, and Touzi [174, 97, 175, 151], where in increasing order of generality (first in a complete Black-Scholes market, then for general Markovian one-dimensional price processes, finally for multidimensional markets) they obtain the first order correction in the cost parameter of the value function as well as the no-trade region. Classical stochastic optimal control requires the Markov property for the processes involved in the analysis to hold. This condition was subsequently formally lifted by Kallsen, and Muhle-Karbe [104, 109] through the use of martingale duality. This was then rigorously proved by Ahrens & Kallsen [2] and Herdegen & Muhle-Karbe [87]. With the presence of proportional transaction costs, the dual variable becomes a couple: a so-called shadow price and a corresponding local martingale measure. The shadow price (this concept is originally due to Cvitanić, Karatzas and Loewenstein [48, 129]) is chosen in such a way that the optimal strategy in a frictionless market with this price process only buys (resp. sell) shares of the risky asset when it coincides with the ask (resp. bid) price (for results on the existence of shadow prices, see [106, 51, 20]). It is a sort of worse case frictionless price process that can be used to obtain the value function and trading strategy in the market with frictions. For its use in the derivation of solution to portfolio choice problems, see additionally [70, 1].

The introduction of frictions renders as well the hedging of claims considerably more difficult, even in complete markets, where every claim is replicable. Trading

[6] A simplification of this model where utility is derived only from wealth at the end of the trading period (which is subsequently postponed infinitely far into the future) is formally solved in [58] by Dumas and Luciano. A rigorous proof is given by Gerhold, Guasoni, Muhle-Karbe and Schachermayer [70].

costs make the optimal hedging strategies infeasible and a derivative product cannot be perfectly replicated anymore. Exponential utility indifference pricing with small transaction costs was first used by Hodges and Neuberger in [88] and by Davis, Panas and Zariphopoulou in [54] where the price and optimal hedging strategy are given by a three-dimensional free-boundary problem.[7] Asymptotic analysis was then heuristically applied by Whalley and Wilmott for the limit of small transaction costs [176]. This permits much faster computations for real world use. Precise and rigorous conditions for this derivation were recently given by Bichuch [22]. The assumption of a simple geometric Brownian price process was finally lifted by Kallsen and Muhle-Karbe in [107], where they formally derive asymptotically optimal hedging strategies as well as the welfare loss due to transaction costs for general Itô diffusions[8]. This was later proved by Ahrens and Kallsen in [2].

Among the extensions of the Merton problem with linear transaction costs, we can cite the addition of portfolio constraints by Liu and Muhle-Karbe [126], the introduction of dividends by Guasoni, Liu and Muhle-Karbe [75] and the extension to two assets by Bichuch, Guasoni and Shreve [24, 23].

In all the previously cited works, the price processes are exogenous. Yet they should emerge from the interactions of the different actors on the market. Then, understanding the influence of frictions on these equilibrium prices is of interest for market makers to decide on their fees, or for a policy maker to assess the true effects on welfare of new trading taxes[9]. Herdegen and Muhle-Karbe [86] find that frictionless equilibrium prices are robust to the introduction of small proportional transaction costs.

The study of proportional trading costs is of importance for actors of every size, and with the digitalisation of financial markets, volumes rose, intensifying the interest for those questions. This opened the way to a better understanding of frictions in financial market. Many new methods and techniques sprang into light and allowed in turn to solve problems arising from considering different frictions that were not tractable before.

[7] A characterisation of the optimal hedging strategy is given by Bouchard [26] in a general Markovian and multi-dimensional market.

[8] An alternative to utility indifference to price derivative products, is superreplication. Compare, e.g. [18], [27] and the references therein.

[9] In [86], trading taxes are levied in some way similar to a fee charged by a broker. This is not to be confused with the capital gain tax whose effects are investigated by Ben Tahar, Soner and Touzi in [19], as this tax is levied on each trade but depends on the average price obtained by the investor so far, and not only on the current trade.

Fixed Costs Models. Fixed transaction costs are such an example. Negligible for large investors, fixed costs are primordial for smaller agents. If each trade costs a fixed amount of money, none of the previously discussed strategies is applicable. Trading infinitely often is prohibited as the trader faces ruin at the beginning of the first trading episode. When fixed costs are of importance, investing in a market requires impulse control, where a strategy is composed of a finite family of trading dates and corresponding buy or sell orders. The first mention of impulse control in the literature is by Eastham and Hastings in 1988 [59]. The relative intractability of the problem lead first to resort to numerical solutions (see [173, 125]), or to consider formal asymptotic expansions for CARA investors (see [115, 128]). More recently, Oksendal and Sulem [144] used the method of viscosity solutions to treat the case of exponential utility in a Black-Scholes market with fixed and proportional costs, later extended by Altarovici, Reppen and Soner to the multi-asset case [12]. Exponential utility allows to discard the effects of the investor wealth on the portfolio problem. To allow for more general applications, the case of power utility in a Black-Scholes market is treated by Altarovici, Muhle-Karbe and Soner in [11]. Finally, Feodoria [60] solved the case of general price processes (with non necessarily Markovian dynamics) for a CARA investor by a delicate use of non-Markovian dynamic programming.

Price Impacts and Illiquidity

As amply shown, fixed and proportional costs have been extensively studied. We now turn to a more recent problem, the topic of this thesis, whose importance has become fundamental with the emergence of giant investment, pension and hedge funds. Such large investors are not only facing bid-ask spread and transaction fees, but effectively "move the market" every time they rebalance their portfolios. Indeed, market depth – a quantity relating the size of an order and its average execution price – is finite and a large enough order modifies temporarily the supply-demand balance. To find sufficiently many counterparties for a large trade, the investor needs to incrementally increase the price she offers. Consider a simple asset. Its shares are owned by the agents on the market, some trading actively (banks, institutional investors, etc.), other following simple buy and hold strategies. They have acquired the asset for a reason, investment or speculation, and may not be interested in selling: at every point in time only a fraction of the total amount of shares is available for trading, and not all at the same price. This takes the form, on electronically supported markets, of a *limit order book*. It is a database of *limit orders*, orders posted at a price (higher than the best ask for sell order, lower than the best bid for buy orders) that will be executed if a trader posts a *market order*, at the best price existing in the limit order book. This pairing is done at an extremly high frenquency (with a period under the milli-second) by an algorithm. Essentially, a limit order book for an asset can be visualised as stacks of orders lined up on the price line, waiting on both sides of the *midprice* for an agent to "cross the spread" and fill them. The price for a market order needing more than one stack to be filled is then an average of the prices of these piles weighted by their respective size. Obviously, this averaging is always detrimental to the investor crossing the bid-ask gap. The phenomenon just described is designed in the literature by price impact. The function linking the size of the order and the difference between the effective execution price and the best quoted bid or ask price is the object of importance. To mitigate the adverse consequences of price impacts on the agent portfolio, it falls onto her to estimate this function in a first step and then to take it into account in the resolution of her optimization problem. A significant group of researcher has tasked itself with modelling the microstructure of financial market: Bouchaud [28, 29], Cont [45, 44] Laruelle & Lehalle [122], Rosenbaum [90, 89] and many others. Their goal is to understand the inner working of the exchange platform so as to back out a usable model of the market depth and the ensuing price impacts. Others have tried to simply estimate the price impact function. As Lillo, Farmer & Mantegna [123] and Almgren, Thum, Hauptmann & Li [10] find, the price impact function must be close to a power function, and this power should be close to 0.5. This is in line with what the practitioners call the "square root law", that an optimized[10] trade raises an additional cost proportional to the square root of its size.

A second strand of research takes this impact function as given and seeks to derive the optimal portfolio or hedging strategy, as well as to price derivative in an illiquid market. This thesis endeavours to contribute to this study.

Liquidation and Optimal Execution. The first article featuring such price impact is the ground-breaking work of Kyle [119]. Taking the continuous limit of auctions to model the price formation process, he obtains that the price modifications observed by the noise traders are due to the market makers protecting themselves from adverse selection induced by the presence of insider traders. The price impact function is linear in the amount traded (by noise traders and informed traders alike), with proportionality constant usually denoted by λ, "Kyle's λ". This parameter increases with the trade volume variance, and decreases with price volatility, features that are still crucial in recent research (see [120] and [149]). Following Kyle's introduction of price impacts, financial economists concerned themselves with the optimal

[10] Traders routinely use ad hoc methods to reduce the impact of large trades on their P&L, developed through experience or in-house research.

execution problem[11]. Rather than finding how to design a flow of orders on a long time scale, a simpler first step is to consider a unique large order: executing a buy in or liquidating an obsolete position. The first such study was realised by Bertsimas and Lo [21], who in a market with linear price impacts (as in Kyle's model), find that a risk neutral investor just liquidates the position at a constant rate, so as to minimize the impact of her trade on her wealth. However it is observed on markets[12] that traders trade faster at the beginning of the liquidation period, be it for fear of a downward move of the market before the end of the liquidation or any other consideration. Progresses to include this fact into a model were made by Almgren and Chriss with their now famous price impact model in discrete time [7, 8]. In previous works (Kyle [119] and Bertsimas & Lo [21]), markets had infinite resilience: the price movement affects only the current trade, and reverts immediately back to its "fundamental" level. Almgren and Chriss make a distinction between temporary and permanent price impact: a part of the price modification is incorporated into the price process, a way to model information aggregation. They take both to be linear in the trade size, and show that (i) a risk averse investor trading off expected liquidation value against variance of the final wealth trades more upfront as is suggested by real-world observations, and (ii) permanent impacts, if they are linear do not influence the optimal strategy. The discrete-time model is later extended by Huberman and Stanzl to time varying volatility and impact intensity, and by Almgren to the continuous setting in [5, 6] (first for constant volatility and price impact, then for stochastically varying quantities).

Two remarks on these models need to be made. First, mean-variance criterions readily provide solvable problems: quadratic functions are the easiest kind to optimize. However this comes at a cost. Mean-variance functionals are not time-consistent and require the commitment of the agent to a rule at inception of the trading period. Indeed, the dynamic programming principle does not apply, and a different strategy can yield a strict improvement of the final objective function, somewhere along the trading period. This question is swept under the rug in [7, 8, 5] as the optimal strategies are anyway deterministic under the posited assumptions (see also [170, 171]). It is adressed in [131] where the authors show that pre-committing to a price dependent trading rule (that allows the trader to use new information), permits to reduce the ex-ante variance of the objective.[13] Second, linear permanent

[11] In the meantime, others studies paralleled Kyle's first model to explain the bid-ask spread formation through asymmetric information, see for example [71, 64].

[12] See for example [93] for a description of this stylized fact.

[13] This seems to depart from the rational investor hypothesis, but the agressive-in-the-money strategies obtained by Almgren and Lorenz are consistent with what seems to be practitioner's

and temporary impacts are chosen as a first approximation to provide tractable models. Still it appears, that linear permanent impacts are the only ones that rule out arbitrage in the sense that the expected gains of round-trips cannot be positive (see Huberman and Stanzl [91]), even if these impacts decay exponentially with time (see Gatheral [67]), or in the quasi-arbitrage sense in [92] where infinite profits with infinite Sharpe ratios are prohibited (then, large profits can only be obtained by accepting correspondingly large risks)[14].

The liquidation problem was then solved in a larger class of markets by Schied, Gatheral and Guéant [68, 69, 82], and its solution is proved to be robust to mispecifications of the price process [169].

An alternative, between the complete modelling of market microstructures and the assumption of a price impact function, is to model the limit order book shape. The price impact function is then the generalized inverse function of the antiderivative of this limit order book function. The seminal article of Obizheava and Wang [143] sets the ground for these kind of studies, followed by Predoiu, Shaiket, Shreve, Alfonsi, Fruth and Schied [152, 4]. The link between the two strands is made by Roch, Soner, Kallsen and Muhle-Karbe, who show in [159, 108] that in the limit of large resilience, the Almgren & Chriss model obtains.

The first part of this thesis extends one of the previous models (Almgren and Chriss's[8, 5]) to investigate the behaviour of an investor unloading a large position. The constant and unidirectional pressure applied to the market by the successive orders are bound to have an increasing effect on liquidity. Furthermore, high-frequency traders and aggressive agents might pick up on the buying (or selling) pattern and use it to their profit in order to front-run the large investor. That could be the case when trades need to be announced in advance, for insider trading for instance. An other example of such a situation, is the liquidation by Société Générale of Kerviel's positions. This is taken into account by replacing the proportionality constant in the linear price impact function of Almgren and Chriss by a function increasing in the volume traded. Price impacts then gradually increase over the liquidation.

Portfolio Choice and Optimal Investments. The study of linear price impacts was recently extended by financial economists to more complex portfolio choice problem in markets with more involved specifications. In factor models where returns are predictable and price impact is linear, Gârelanu and Pedersen [65, 66] find an explicit solution to the mean-variance optimization problem: the investor should trade

behaviour and is consistent with prospect theory.

[14] But, compare [81] for an argument against this.

at a finite rate, and aim in front of the target (in the sense of a weighted average of the frictionless optimal portfolio future expected positions). In a slightly more involved setting (volatility can be stochastic) Collin-Dufresne and co-authors [41] offer an efficient way to find an optimizer in a large sub-class of stragegies, that shows promises for practical use. The resolution of portfolio problems in market with quadratic transaction costs recently culminated in the work of Moreau, Muhle-Karbe and Soner [142] and Guasoni and Weber [79, 78]. In the first, using the powerful machinery of viscosity solutions the authors were able to exhibit a family of asymptotically optimal strategy for a CARA investor (and more generally, for a class of "well-behaved" utility functions) in a general market where the multi-dimensional price process is a Markovian Itô diffusion. In the second [79], Guasoni and Weber characterize the optimal policy of an investor with power utility trading in a single asset Black-Scholes market, and gave its asymptotic expansion for small price impact parameter. They later extended their asymptotical results to the case of several risky assets in the presence of cross-price impacts [78].

The aforementioned models all deal with linear impacts, leading to quadratic transaction costs. The main reason for this is mathematical tractability. Yet, real-world price impact are nonlinear, their elasticity between 0 and 1, which should lead to superlinear [15] (but not quadratic) costs. Very few articles treat optimal portfolio choice problems with nonlinear price impacts.

Structural results about duality for optimization problems with general superlinear frictions have only been recently uncovered by Guasoni and Rásonyi [77]. They define the duality variable as the execution price obtained for the optimal strategy coupled with an equivalent local martingale measure for this "shadow price". They use it to prove a characterization of the primal and dual optimizers in utility maximization problems with superlinear penalties. This result is a crucial tool to obtain an upper bound for the optimization problem of chapter 2. An other interesting fact they prove is that no additional care is necessary to restrict the admissible strategies for no-arbitrage considerations: superlinear frictions prevent the scaling of arbitrage strategies. There is a market bound on the profits an investor can make.

A first natural approximation for portfolio choices with nonlinear friction leads to investigate markets with both proportional and quadratic costs (see [127, 156]). The optimal strategy for an investor maximizing power utility over a long time horizon

[15]Price impacts typically induce nonlinear transaction costs: the price displacement, usually increasing in the amount of share purchased per time unit, is multiplied by the number of shares purchased or sold. This produce a cost increasing faster than linearly in the trading speed and volume. The square root law, for example, seems to indicate transaction costs proportional to the power $3/2$ of the trade size and its execution rate.

in a Black-Scholes market is then given by a no-trade region, whose width is smaller than in the setting without price impact. Indeed, the addition of this new friction prohibits the use of local time-type strategies to keep the wealth proportion inside its boundaries. Hence, trading must start earlier to ensure that the investor's position does not stray too far away from its optimal value.

The first article to solve the problem is the work of Guasoni and Weber [80]. They consider in a Black-Scholes market where trading induces price impact proportional to a power $p-1 \in (0,1)$ of the volume-renormalized trading rate, an investor who maximizes power utility from terminal wealth over a long horizon. The optimal strategy is characterized by a nonlinear partial differential equation, and in the limit of small impacts, asymptotically optimal strategies obtain. They consist in trading toward the Merton proportion at a finite rate given by the solution g_p to a universal (in that it depends only on p) nonlinear ordinary differential equation applied to a renormalization of the distance to the optimal proportion. As will be shown in Chapter 2 of this thesis, these ingredients also play a crucial role in the general case. Indeed, the function g_p drives the asymptotically optimal trading rates in a market with general non-Markovian dynamics as well, and the tracked target is also the optimizer of the frictionless problem, rescaled by the current market and cost parameters. It appears then, that the goal of a rational investor is to follow a known target while controlling a penalty on the "effort" needed. Hence, one can abstract the portfolio choice problem into a pure stochastic tracking problem setting. That is what Cai, Rosenbaum and Tankov study in their couple of articles [34, 35], where they reduce the tracking problem to a linear programming problem over a set of occupation measures. To achieve this reduction, careful ergodic estimates are made, in much the same way as the one we use in Chapter 2.

Option Pricing and Hedging. Few works deal with the hedging of derivative products in the presence of price impacts. Of note are [15, 16] by Bank, Soner and Voss, where surprisingly, an exact solution is found for the hedging of claims in the Bachelier model (where volatility can be stochastic) when trades are penalized by quadratic transaction costs. An important feature of the model is that it does not require the target strategy to be continuous, which allows to deal with contracts that depend on discretely sampled values of the underlying process, as is the case for Asian options. Parallelly, Almgren and Li derive in [9] the HJB equation satisfied by the solution of a hedging problem, where the expected terminal wealth of an investor is penalized by its variance. In this setting the investor is "long" the option and try to mitigate the loss realised at maturity while her trading is hindered by linear permanent and temporary price impacts. They also warn again naive discretisation of the strategy and offer an efficient scheme.

In general Markovian price models, and for general utility functions, Moreau, Muhle-Karbe and Soner [142] provide asymptotically optimal hedging strategies for contingent claims, when trading induces linear price impacts.

For general superlinear costs, Guéant and Pu obtain the optimal hedging strategy through dynamic programming and the use of viscosity solutions [83].

Finally, the problem of superhedging a claim in a Markovian diffusion market where trading entails liquidity costs (the larger the trade the further the price moves away from the quoted value: price impact is modelled by a supply curve) is studied by Çetin, Soner and Touzi in [38]. They find the optimal hedging strategy by dynamic programming methods.

Finding methods to price and hedge in the presence of nonlinear price impact and in markets with general specifications is an important issue, to which the present thesis contributes.

Overview of the Thesis

The rest of this thesis stems from two research articles, that make up the next two chapters. The first one, written with my supervisor Johannes Muhle-Karbe has already been published as "Liquidation with Self-Exciting Price Impact", in Mathematics and Financial Economics (Vol. 10 (2016), No.1, pp. 15-28.), [37]. The second, with Martin Herdegen and Johannes Muhle-Karbe, has not been submitted yet at the time of writing. They both deal with nonlinear price impact and optimal trading decisions.

The first chapter extends on the well-known Almgren-Chriss model to self-exciting price impacts. A large investor wishes to unload a sizeable position over a short time interval. The execution of the large order depletes the pool of fundamental buyers and the inventory of the market maker, may start stop-loss orders by driving prices down and attract predatory traders. It is therefore not unreasonable to consider price impacts whose intensity increases over the liquidation periode. We choose for tractability an intensity that increases linearly in the traded volume. For a risk-neutral investor a closed-form solution obtains. For risk averse investors, numerical methods are necessary. The conclusion we draw from the study is that self-excitement has a mitigating effect on the trading speed: if it is too low at the beginning trading is sped up, and on the contrary, if it is originally too fast, it is slowed down.

The second chapter of the thesis contains the asymptotic resolution of a portfolio choice problem set in a general single asset market where the price process is a general, non necessarily Markovian, Itô diffusion. Using probabilistic techniques, we show that the asymptotic solution for small impact can be described explicitly up to the solution of a nonlinear ODE, which identifies the optimal trading speed and the performance loss due to the trading friction. Previous asymptotic results for proportional and quadratic trading costs are obtained as limiting cases. As an illustration, we study how hedging strategies and active portfolio management are affected by the nonlinear trading cost.

Chapter 1

Optimal Liquidation with Self-Exciting Price Impact

Large trades executed quickly adversely affect execution prices. Whence, it is a key concern for portfolio managers to schedule the corresponding order flow in an efficient manner.

In the academic literature, the study of this "optimal execution problem" starts with the article of Bertsimas and Lo [21]. In a discrete-time model with linear price impact, they showed that – for risk-neutral traders – it is optimal to execute orders at a constant rate. If time is parametrized in volume rather than calendar time, this corresponds to a "VWAP" strategy, where the average execution price equals the Volume Weighted Average Price.

Yet, "institutions typically trade more up-front" [93], i.e., the initial trading rate is higher and then gradually decreases as the execution of the order is completed. Almgren and Chriss [7, 8] as well as Huberman and Stanzl [93] explain this using traders' risk aversion. Slower execution leads to larger risks due to uncertain future price moves. Therefore, risk-averse traders speed up liquidation initially to reduce this inventory risk. These models have been extended in various directions, see, e.g., [73, 69] and the references therein for an overview.

In the present chapter, we discuss a different kind of opportunity cost – "self-exciting" price impact. If a large sell order creates persistent selling pressure, it is reasonable to expect that the pool of available counterparties is diminished. As a consequence, each trade not only incurs price impact costs but, by depleting the liquidity supply, also increases the price impact of future trades in the same direction. This effect is exacerbated by predatory traders, who become aware of the intentions of the fundamental sellers, and front-run them by also selling.[1]

We model this effect in reduced form by an extension of the framework of Almgren and Chriss.[2] In our model, the price impact parameter is no longer constant but instead depends linearly on the number of shares already sold.[3] We motivate this specification by extending the micro-foundation for the Almgren-Chriss model proposed by Gârleanu and Pedersen [66]. There, price impact arises due to the inventory risk of market makers, who act as intermediaries between the fundamental seller and a group of end users. Gârleanu and Pedersen [66] show that this leads to the Almgren-Chriss model, if market makers need a constant amount of time to locate a suitable counterparty. Our model obtains if the search time increases with the number of shares sold already.

With self-exciting price impact, we find that a constant liquidation rate is no longer optimal even for risk-neutral traders. Instead, the trading rate is increased initially and then gradually slows down as the execution of the order is completed. This is qualitatively similar to the effect of risk aversion [7, 8, 93]. However, the interplay of self-excitement and risk aversion depends on their relative magnitude. If risk aversion is low, self-excitement reduces optimal positions, thereby increasing "effective" risk aversion. For moderate risk aversion, there is little change. With large risk aversion, self-excitement in fact increases positions, lowering effective risk aversion. In summary, self-excitement therefore moderates the optimal trading speed. The general convex shape of the optimal execution trajectory is a robust feature, however, in that it even obtains for risk-neutral agents here.

This chapter is organized as follows. In Section 1, we recall the Almgren-Chriss model, and extend it to include self-exciting price impact. Section 2 outlines the micro-foundation for the Almgren-Chriss model proposed by Gârleanu and Pedersen in [66], and adapts the argument to motivate our model. In Section 3, we show that the risk-neutral optimal execution problem in our setup can be solved in closed form. Then, we turn to risk averse investors, which we study using asymptotic expansions

[1] Front running becomes optimal in the presence of sufficiently large "preys" in the models of Brunnermeier and Pedersen [32] as well as Carlin, Lobo, and Viswanathan [36]. In the model of Schied and Schöneborn [170], potential predators can be either detrimental or beneficial depending on the model parameters.

[2] Very similar models were proposed and studied concurrently by Bertsimas and Lo [21], Madhavan [134], as well as Huberman and Stanzl [93]. To keep in line with most of the literature, we nevertheless stick to the nomenclature "Almgren-Chriss model".

[3] Recently, a different kind of "self-excitement" has also started to receive increasing attention, see, e.g., [3] and the references therein. In these models, the orders of other market participants are modeled by a Hawkes process, a counting process whose jump intensities are self-exciting in that they are influenced by the past jumps. Whence, self-excitement is produced by the trades of the other market participants in these models, whereas it is instigated by the large trader in ours.

and a numerical example.

1.1 Price-Impact Models

In this section, we first recall the widely used price-impact model of Almgren and Chriss [7, 8]. Then, we propose an extension that takes into account self-exciting price impact in reduced form. Throughout this chapter, fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$ supporting a standard Brownian motion $(W_t)_{t\in\mathbb{R}_+}$, and suppose the *unaffected price process* of a risky asset follows arithmetic Brownian motion:[4]

$$dS_t = \sigma dW_t, \tag{1.1.1}$$

for a constant volatility $\sigma > 0$.

1.1.1 The Almgren-Chriss Model

Consider a *large* trader, who needs to liquidate a substantial position $X > 0$ in the risky asset S. Her trades deplete the liquidity currently available in the market, and thereby adversely affect prices. To wit, a trade dx_t executed over a time interval dt moves the current market price by $\lambda \frac{dx_t}{dt}$, $\lambda > 0$, i.e., price impact is linear both in the size of the trade and the speed at which it is executed.[5] Therefore, the additional execution cost $\lambda(\frac{dx_t}{dt})^2 dt$ due to price impact is quadratic in the trading rate $\dot{x}_t = \frac{dx_t}{dt}$.

If the initial position X is liquidated using some absolutely continuous execution strategy $dx_t = \dot{x}_t dt$ satisfying $x_0 = X$ and $x_T = 0$, then the total proceeds from liquidation adjusted for trading costs are given by $-\int_0^T S_t dx_t - \int_0^T \lambda \dot{x}_t^2 dt = -\int_0^T (S_t \dot{x}_t + \lambda \dot{x}_t^2)\, dt$. Whence the *implementation cost* of an absolutely continuous liquidation strategy $dx_t = \dot{x}_t dt$ compared to immediate frictionless execution is

$$C_X(x) := XS_0 + \int_0^T \left(S_t \dot{x}_t + \lambda \dot{x}_t^2\right) dt. \tag{1.1.2}$$

[4]Since time horizons for liquidation programs are typically short, drifts are usually neglected (but cf. [130]) and it is reasonable to work with more tractable arithmetic Brownian motions rather than their geometric counterparts (but cf. [68]). In the case of a risk-neutral investor, the price process need not be an arithmetic Brownian motion, but can be a general martingale.

[5]This price impact is purely *temporary*, in that it only affects the current trade but not subsequent ones. Linear *permanent* price impact can also be accounted for by shifting the unaffected price quote, see [8]. However, like proportional transaction costs, linear permanent impact does not alter optimal execution strategies. Hence, we disregard these two frictions throughout. Nonlinear permanent price impact depending on the cumulated number of shares sold by the large trader is studied by [81].

1.1.2 A Reduced-Form Model for Self-Exciting Price Impact

In the Almgren-Chriss model discussed in the previous section, price impact is constant. However, if the position to be liquidated is very large, each order may increase the price impact for subsequent trades for at least three reasons. First, continued selling pressure makes it more and more difficult to find fundamental buyers who are willing to act as counterparties. Second, large sales and price drops may trigger stop-loss strategies leading to further selling. Third, the persistent order flow generated by the execution strategy may become visible to predatory traders, who can then front-run the large trader. In each case, the increased selling pressure depletes the liquidity available for selling, thereby increasing price impact for subsequent trades.

We try to capture these effects by a simple, tractable model. To wit, we suppose the temporary price impact parameter λ is no longer constant, but instead increases linearly with the number of shares sold:[6]

$$\lambda_t := \ell_0 + \ell_1(X - x_t),$$

for constants $\ell_0 > 0$ and $\ell_1 > 0$. With this specification, the implementation cost of an absolutely continuous, deterministic, decreasing execution path $dx_t = \dot{x}_t dt$ is given by

$$C_X^{\ell_1}(x) := XS_0 + \int_0^T \left(S_t\dot{x}_t + (\ell_0 + \ell_1(X - x_t))\dot{x}_t^2\right) dt. \tag{1.1.3}$$

Throughout, we focus on sell-only strategies, because otherwise the model admits quasi-arbitrages in the sense of Huberman and Stanzl [92].[7] Incorporating buy and sell trades in an economically plausible manner would necessitate a more involved specification that treats trading costs for buying and selling separately, ruling out analytical solutions.[8]

---

[6]The linear dependence is assumed for tractability. It allows for closed-form solutions and also can be seen as a first-order approximation for more general small self-excitement mechanisms.

[7]Indeed, assume we start with X shares. Buying shares at a constant rate on the interval $[0, \frac{T}{2}]$ to reach $X + \frac{\ell_0}{\ell_1} + k$, for some integer k, and selling back these shares at the same speed on $[\frac{T}{2}, T]$ yields an expected profit of order $O(k^3)$ as k grows to infinity, while the standard deviation of the final profit is of order $O(k)$.

[8]To allow for buy orders in our model and rule out price manipulations, we could define the price impact parameter as $\lambda_t := \ell_0 + \ell_1(X - x_t)^+$ when selling ($\dot{x}_t < 0$) and $\lambda_t := \ell_0$ when buying ($\dot{x}_t > 0$). This means that, during a substantial liquidation, sales increase the price impact of further sales whereas the price impact of purchases remains constant. In such a model round-trips have a strictly positive expected cost, and optimal strategies are necessarily decreasing.

1.2 Micro-Foundation

We now provide a micro-foundation for the above models, by obtaining them as an equilibria between (i) the large trader liquidating her substantial position, (ii) market makers acting as intermediaries, and (iii) end users who eventually absorb the liquidated shares. We start by recalling the model of Gârleanu and Pedersen [66], which leads to the Almgren and Chriss model. Then, we adapt their argument to obtain our model with self-exciting price impact from Section 1.1.2.

In each case, we set up the model on a discrete time grid, and then pass to the continuous-time limit. In the discretely sampled version, trading takes place at equidistant time points $0, \Delta t, 2\Delta t, \ldots, N\Delta t$, where $\Delta t = T/N$ for some $N \in \mathbb{N}$. The unaffected price process follows an arithmetic random walk:

$$\Delta S_n = \sigma\sqrt{\Delta_t} Z_n, \quad n = 1, \ldots, N, \tag{1.2.1}$$

for independent, standard normally distributed random variables $Z_1, \ldots, Z_N$. Price impact is determined in equilibrium between three representative market participants. The first is a large trader, who starts with a substantial position to be liquidated. These risky shares are first transferred to mitigating "market makers", who then eventually sell them to a group of "end users". Price impact arises from the market makers' risk aversion: there is some search friction, so that market makers cannot locate suitable end users immediately. As a consequence, they have to hold non-trivial inventories. The compensation they demand for this liquidity provision is the price impact suffered by the large trader.

1.2.1 The Almgren-Chriss Model

To make this precise, we follow [66] and consider a continuum of market makers with mass one, indexed by the set $[0, h]$. This index represents the first time the respective market maker arrives at the market and trades. Thereafter, market makers need h time units to locate a suitable counterparty to lay off their inventory. Then, they only rejoin the market $h/\Delta t$ periods later.[9] Hence, the mass of market makers active in each given period is $\Delta t/h$.

Market makers sell their inventories to the end users at the competitive exogenous price S, the "fundamental" value of the risky asset. Assuming mean-variance

[9]For convenience, we assume that $h/\Delta t$ is an integer.

preferences for tractability, they therefore trade with the large trader to maximize[10]

$$\max_q \left(\mathbb{E}[q(S_{t+h} - \hat{S}_t)|\mathcal{F}_t] - \frac{\gamma^M}{2} \mathrm{Var}[q(S_{t+h} - \hat{S}_t)|\mathcal{F}_t] \right) \tag{1.2.2}$$

at time t. Here, S_{t+h} is the exogenous fundamental price at which the market maker sells her inventory to the end user once the latter is located at time $t+h$, whereas the price $\hat{S}_t$ charged to the large trader at time t is to be determined in equilibrium. γ^M is the market makers' risk aversion, which weighs the expected gains from trading against the variance of the corresponding position. Inserting the dynamics (1.2.1) of the fundamental price, (1.2.2) can be rewritten as

$$\max_q \left(q(S_t - \hat{S}_t) - \frac{\gamma^M}{2} q^2 \sigma^2 h \right). \tag{1.2.3}$$

As a result, the optimal position is

$$q^* = \frac{S_t - \hat{S}_t}{\gamma^M \sigma^2 h}.$$

Since a mass $\Delta t/h$ of market makers is active at each time interval, the total amount of liquidity provided therefore is $q^* \Delta t/h$. If the large trader wants to trade Δx_t shares at time t, market clearing (i.e., $\Delta x_t + q^* \Delta t/h = 0$) therefore determines the corresponding equilibrium price as

$$\hat{S}_t = S_t + \gamma^M \sigma^2 h^2 \frac{\Delta x_t}{\Delta t}.$$

Compared to the frictionless execution price $S_t \Delta x_t$, this linear price impact leads to trading costs $\gamma^M \sigma^2 h^2 (\frac{\Delta x_t}{\Delta t})^2 \Delta t$ quadratic in the trading rate $\Delta x_t/\Delta t$. If the search time h remains constant as the time-grid becomes finer and finer for $\Delta t \downarrow 0$, we have convergence to the Almgren-Chriss model from Section 1.1.1, compare [66]. The corresponding quadratic trading cost $\lambda = \gamma^M \sigma^2 h^2$ is determined by i) the market maker's risk aversion γ^M, ii) the variance σ^2 of the fundamental value, and iii) the (squared) search time h^2 needed to locate an end user.

1.2.2 Self-Exciting Price Impact

Now, we discuss how to adapt the model from Section 1.2.1 to obtain the self-exciting price impact proposed in Section 1.1.2. The basic idea is that very large

[10] Here, the risk-free rate is set equal to zero, because the time horizons we consider are short.

orders are more difficult to execute because with persistent large sales i) it becomes increasingly difficult to find end users willing to eventually take the other side of the trades and/or ii) the total selling flow is increased due to execution of stop-loss orders or front-running by predatory traders.[11]

In the above model, these effects can be accounted for in reduced form as follows. Suppose that either i) the (squared) search time increases in the number of shares sold already up to time t or ii) other (e.g., predatory) traders augment the order flow Δx_t of the large trader by a factor depending on $X - x_t$, so that we have $h_t^2 \Delta x_t = f(X - x_t)\Delta x_t$ for some increasing function f. Then, repeating the above arguments, one finds that in each case the equilibrium trading costs of a trade Δx_t at time t are given by

$$\gamma^M \sigma^2 f(X - x_t) \left(\frac{\Delta x_t}{\Delta t} \right)^2 \Delta t.$$

In the continuous-time limit ($\Delta t \downarrow 0$), we therefore obtain the path-dependent trading costs $\lambda_t = \gamma^M \sigma^2 f(X - x_t)$.

Choosing the function $f(X - x_t) = h^2 \Big(1 + C(X - x_t)\Big)$ we obtain the path-dependent trading costs from Section 1.1.2, with baseline trading cost $\ell_0 = \gamma^M \sigma^2 h^2$ as before, and an additional self-exciting parameter $\ell_1 = C\ell_0$. Taking the function f constant, we recover the original model of Almgren and Chriss.

Other dependences between search time and volume sold are of course possible. The liquidation problem, with linear self-excitement in the number of shares sold can be solved explicitly for risk-neutral investors. In contrast, closed-form formulas are not available for general relations between h_t and $(X - x_t)$.

1.3 Optimal Execution

We now determine optimal execution paths in the above models. Suppose the large trader has to unwind a substantial initial position $X > 0$ on some time interval $[0, T]$ using an absolutely continuous strategy $dx_t = \dot{x}_t dt$ satisfying $x_0 = X$ and $x_T = 0$. (That is, x_t denotes the number of risky shares still held at time t.)

[11] Note, however, that there is recent empirical evidence suggesting that the marginal impact of child orders decreases as metaorder execution proceeds [13, 74]. Whence, the argument presented here may only apply for very large execution programs.

1.3.1 The Almgren-Chriss Model

First, we recall the solution in the Almgren-Chriss model [8]. In this setting, the total proceeds from liquidation adjusted for trading costs are given by 1.1.2. For deterministic strategies x, a simple calculation using integration by parts yields the mean and variance of this random variable [8]:

$$\mathbb{E}[C_X(x)] = \int_0^T \lambda \dot{x}_t^2 dt,$$
$$\mathrm{Var}[C_X(x)] = \int_0^T \sigma^2 x_t^2 dt.$$

Hence, minimizing a mean-variance functional $\mathbb{E}[C_X(x)] + \frac{\gamma}{2}\mathrm{Var}[C_X(x)]$ of the implementation cost over all deterministic[12] decreasing execution strategies leads to a calculus of variation problem. Its solution x^* is characterized by the corresponding *Euler-Lagrange equation*:

$$\ddot{x}_t - \frac{\gamma\sigma^2}{2\lambda} x_t = 0, \quad x_0 = X, \quad x_T = 0. \tag{1.3.1}$$

Without risk aversion, this leads to a linear optimal execution path,

$$x_t^* = X\left(1 - \frac{t}{T}\right), \quad \text{for } \gamma = 0. \tag{1.3.2}$$

With risk aversion $\gamma > 0$ one obtains:

$$x_t^* = X\frac{\sinh(\kappa(T-t))}{\sinh(\kappa T)}, \quad \text{for } \gamma > 0, \text{ where } \kappa = \sqrt{\frac{\gamma\sigma^2}{2\lambda}}. \tag{1.3.3}$$

Whence, without risk aversion, the initial position is simply executed at a constant trading speed. Risk aversion discourages the investor from holding on to her risky position, so that liquidation is sped up initially and then gradually slowed down. This can be seen clearly through a Taylor expansion for small risk aversion ($\kappa = \sqrt{\gamma\sigma^2/2\lambda} \sim 0$):

$$x_t^* = X\left(1 - \frac{t}{T}\right) - \frac{1}{6}\kappa^2 XT\left(1 - \frac{t}{T}\right)\left(2t - \frac{t^2}{T}\right) + O(\kappa^4). \tag{1.3.4}$$

[12] Adaptive strategies are studied by Almgren and Lorenz [131] for a mean-variance criterion and by Schied and Schöneborn [172] for von Neumann-Morgenstern utilities. Here, we focus on deterministic strategies to obtain tractable solutions also with our more complicated price impact structure in Section 1.3.2. In the risk-neutral case, this entails no loss of generality, cf. Remark 1.3.2.

1.3.2 Self-Exciting Price Impact

Risk-Neutral Execution With self-exciting price impact, integration by parts and the martingale property of S show that the expected execution cost of a liquidation strategy x is given by

$$\mathbb{E}[C_X^{\ell_1}(x)] = \int_0^T (\ell_0 + \ell_1(X - x_t))\dot{x}_t^2 dt. \tag{1.3.5}$$

Hence, minimizing expected execution costs over deterministic, decreasing strategies again leads to a calculus of variations problem. This suggests that the solution should be determined by the corresponding Euler-Lagrange equation:

$$2\ell_1\ddot{x}_t x_t - 2(\ell_0 + \ell_1 X)\ddot{x}_t + \ell_1\dot{x}_t^2 = 0, \quad x_0 = X, \quad x_T = 0. \tag{1.3.6}$$

The general solution of this ordinary differential equation is given by

$$x_t = X + \frac{\ell_0}{\ell_1} - \left(\frac{9}{8\ell_1}\right)^{1/3} (A(B+t))^{2/3}.$$

The constants are determined by the initial and terminal conditions. We have

$$AB = \frac{(2\ell_0)^{3/2}}{3\ell_1} \quad \text{and} \quad A^2T^2 + 2A^2BT + A^2B^2 - \frac{8\ell_1}{9}\left(X + \frac{\ell_0}{\ell_1}\right)^3 = 0.$$

Hence, taking into account that x should be decreasing:

$$A = \frac{(2\ell_0)^{3/2}}{3\ell_1 T}\left[\left(\frac{X\ell_1}{\ell_0} + 1\right)^{3/2} - 1\right] \quad \text{and} \quad B = \frac{T}{(\frac{X\ell_1}{\ell_0} + 1)^{3/2} - 1}.$$

In summary, we obtain the following candidate optimal execution path:

$$x_t^* := X + \frac{\ell_0}{\ell_1} - \frac{\ell_0}{\ell_1}\left[1 + \left(\left(\frac{X\ell_1}{\ell_0} + 1\right)^{3/2} - 1\right)\frac{t}{T}\right]^{2/3}. \tag{1.3.7}$$

We now have to verify that this candidate indeed minimizes the execution costs over a suitable set of strategies. This is complicated by the fact that the goal functional (1.3.5) is no longer convex in this case.[13] Nevertheless, we can still establish optimality by a direct verification argument:

[13]The integrand to be minimized pointwise for each $t \in [0, T]$ is given by $F(t, x, v) = (\ell_0 + \ell_1 X)v^2 - \ell_1 x v^2$; its Hessian is $\begin{pmatrix} 0 & -2\ell_1 v \\ -2\ell_1 v & 2(\ell_0 + \ell_1(X - x)) \end{pmatrix}$. The sum of this matrix' eigenvalues is positive for $x \in [0, X]$ and their product negative. Whence the goal functional is not convex.

Theorem 1.3.1. *The strategy* x^* *from* (1.3.7) *minimizes the expected execution cost* (1.3.5) *over all deterministic, decreasing, absolutely continuous strategies* x *with square-integrable derivative, which satisfy* $x_0 = X$, $x_T = 0$.

Proof. For any decreasing, deterministic strategy x, one readily verifies that the corresponding expected execution cost can be written as

$$\mathbb{E}[C_X^{\ell_1}(x)] = (\ell_0 + \ell_1 X) \int_0^T \dot{x}_t^2 dt - \ell_1 \int_0^T x_t \dot{x}_t^2 dt := J(x).$$

Setting $y = x - x^*$, we therefore have

$$\begin{aligned}
J(x) &= J(x^* + y) \\
&= (\ell_0 + \ell_1 X) \int_0^T \left((\dot{x}_t^*)^2 + 2\dot{x}_t^* \dot{y}_t + \dot{y}_t^2\right) dt \\
&\quad - \ell_1 \int_0^T (x_t^* + y_t) \left((\dot{x}_t^*)^2 + 2\dot{x}_t^* \dot{y}_t + \dot{y}_t^2\right) dt \\
&= (\ell_0 + \ell_1 X) \int_0^T \left((\dot{x}_t^*)^2 + 2\dot{x}_t^* \dot{y}_t + \dot{y}_t^2\right) dt \\
&\quad - \ell_1 \int_0^T (x_t^*(\dot{x}_t^*)^2 + y_t(\dot{x}_t^*)^2 + 2x_t^*\dot{x}_t^*\dot{y}_t + 2y_t\dot{x}_t^*\dot{y}_t + x_t^*\dot{y}_t^2 + y_t\dot{y}_t^2) dt \\
&= \int_0^T (\ell_0 + \ell_1 X - \ell_1 x_t^*) (\dot{x}_t^*)^2 dt + \int_0^T (\ell_0 + \ell_1 X - \ell_1 x_t^* - \ell_1 y_t) \dot{y}_t^2 dt \\
&\quad + \int_0^T y_t \left(-2(\ell_0 + \ell_1 X)\ddot{x}_t^* + 2\ell_1 \ddot{x}_t^* x_t^* + 2\ell_1 (\dot{x}_t^*)^2 - \ell_1 (\dot{x}_t^*)^2\right) dt \\
&\quad - \ell_1 \int_0^T 2 y_t \dot{y}_t \dot{x}_t^* dt \\
&= J(x^*) + \int_0^T (\ell_0 + \ell_1 X - \ell_1 x_t^* - \ell_1 y_t) \dot{y}_t^2 dt + \ell_1 \int_0^T y_t^2 \ddot{x}_t^* dt \\
&\geqslant J(x^*),
\end{aligned}$$

proving the assertion. In this estimate, we have used integration by parts twice and also taken into account $y_0 = y_T = 0$ in the third step. In the fourth step, we have inserted the Euler-Lagrange equation (1.3.6) satisfied by x^*, and integrated by parts one more time. For the final inequality, we have used that $X - x_t^* - y_t = X - x_t \geqslant 0$ because $x_0 = X$, $x_T = 0$, and x is decreasing. Moreover, we have taken into account that $\ddot{x}^* \geqslant 0$, as is readily verified by differentiation of the explicit formula (1.3.7). □

Remark 1.3.2. An inspection of the proof shows that the strategy x^* is in fact optimal among all adapted, decreasing, absolutely continuous processes x with square-integrable derivative, which satisfy the boundary conditions $x_0 = X,\ x_T = 0$. Indeed, for these finite variation processes, all transformations can be performed pathwise and the result in turn follows from monotonicity of the expectation operator.

To shed some more light on the comparative statics of the optimal execution path (1.3.7), let us perform an asymptotic expansion where the self-excitement parameter ℓ_1 is sent to zero. This means that we consider our model as a perturbation of the Almgren-Chriss setup from Section 1.1.1, and study how the optimal execution strategy is corrected at the leading order in this case. Taylor expansion yields

$$x^*(t) = X\left(1 - \frac{t}{T}\right) - \frac{1}{4}X^2\frac{\ell_1}{\ell_0}\left(\frac{t}{T} - \frac{t^2}{T^2}\right) + O(\ell_1^2), \tag{1.3.8}$$

for $t \in [0, T]$, as $\ell_1 \downarrow 0$. The corresponding trading speed is given by

$$\dot{x}_t^* = -\frac{X}{T} - \frac{X^2}{4}\frac{\ell_1}{\ell_0}\frac{T - 2t}{T^2} + O(\ell_1^2).$$

For $\ell_1 = 0$, we recover the constant trading speed in the Almgren-Chriss model. With small self-exciting price impact $\ell_1 > 0$, trading is sped up in the first half of the trading interval ($t \in [0, T/2]$), and slowed down in the second half. The correction is large if i) self-excitement ℓ_1 is substantial relative to the baseline price impact ℓ_0, ii) the investor's initial position X is large, or iii) the execution horizon T is short.

These effects are qualitatively similar to the corresponding results for risk aversion (cf. Section 1.3.1). However, for very large self-excitement, the optimal execution is not carried out faster and faster as for high risk aversion. Instead, the optimal execution path (1.3.7) converges to a nontrivial finite limit as $\ell_1 \to \infty$. This shows that risk-aversion and self-exciting price impact are not simply substitutes for one another for optimal liquidation problems. Figure 1.1 illustrates these results. Let us briefly discuss the price impact parameters. Assume that the average traded volume for a stock is 100 000 shares per day, and that an investor wants to liquidate that volume over ten days. The baseline price impact ℓ_0 is chosen so that 1% price impact obtains if the trading rate equals 10% of the market rate. The three values of the self-excitement parameter ℓ_1 in turn lead to a 0%, 25%, resp. 75% increase of the price impact after half of the position has been liquidated.

Mean-Variance Optimization Let us now add an inventory penalty to the above optimization problem. As in Almgren and Chriss [8], we consider a mean-variance

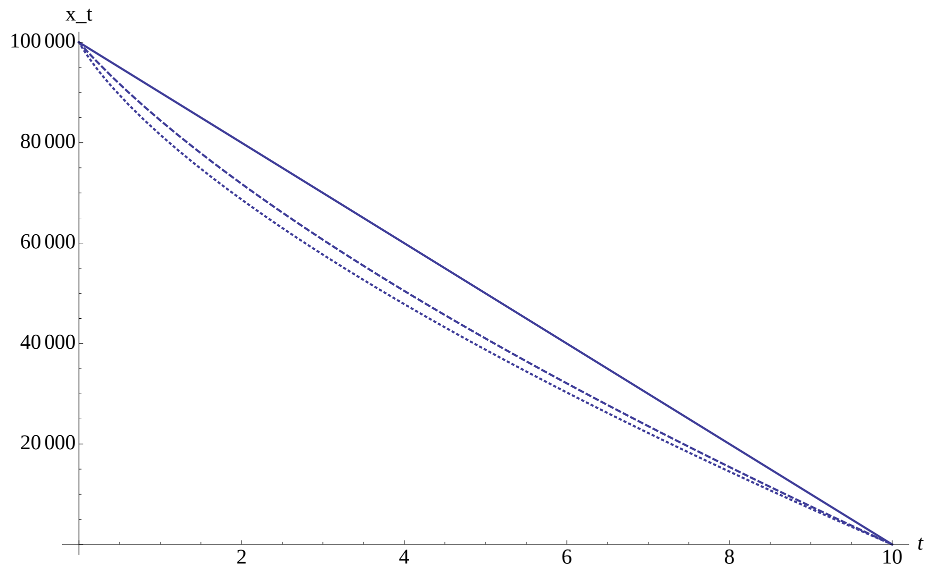

Figure 1.1: Liquidation strategies of a risk-neutral investor ($\gamma = 0$) for $X = 100\ 000$ risky shares, (daily) volatility $\sigma = 0.02$, baseline price impact $\ell_0 = 10^{-6}$, and self-excitement parameter $\ell_1 = 0$ (solid), $5 \cdot 10^{-11}$ (dashed), $1.5 \cdot 10^{-10}$ (dotted).

criterion:

$$\mathbb{E}\left[C_X^{\ell_1}(x)\right] + \frac{\gamma}{2}\mathrm{Var}\left[C_X^{\ell_1}(x)\right] \to \min! \tag{1.3.9}$$

Here, the execution paths again run through the set of deterministic, decreasing functions x satisfying $x_0 = X$ as well as $x_T = 0$. As in the Almgren-Chriss model, we have

$$\mathrm{Var}\left[C_X^{\ell_1}(x)\right] = \int_0^T \sigma^2 x_t^2 dt.$$

Therefore, we once again obtain a calculus of variations problem. The corresponding Euler-Lagrange equation reads as follows:[14]

$$2\ell_1 \ddot{x}_t x_t - 2(\ell_0 + \ell_1 X)\ddot{x}_t + \ell_1 \dot{x}_t^2 + \gamma\sigma^2 x_t = 0, \quad x_0 = X, \quad x_T = 0. \tag{1.3.10}$$

With both risk aversion and self-exciting price impact, an analytic solution is no longer available. However, it is still possible to perform an asymptotic expansion as in (1.3.4) and (1.3.8), respectively. Motivated by these separate expansions, whose first-order terms are linear in risk aversion γ and the self-excitement parameter ℓ_1, respectively, we look for a bivariate expansion of the form

$$x_t^{\gamma,\ell_1,*} = f^0 + \gamma f^\gamma + \ell_1 f^{\ell_1}, \tag{1.3.11}$$

which solves the Euler-Lagrange equation (1.3.10) "at the leading order", i.e., up to terms of order $o(\gamma + \ell_1)$. Evidently, f^0 should be the linear strategy (1.3.2) in the absence of risk aversion and self-exciting price impact. Plugging the ansatz (1.3.11) into (1.3.10) and neglecting terms of order $o(\gamma + \ell_1)$, we obtain

$$-2\ell_0\gamma\ddot{f}_t^\gamma - 2\ell_0\ell_1\ddot{f}_t^{\ell_1} + \ell_1(\dot{f}_t^0)^2 + \gamma\sigma^2 f^0 = 0.$$

Now, compare coefficients for γ and ℓ_1 and enforce the boundary conditions $f_0^\gamma = f_T^\gamma = f_0^{\ell_1} = f_T^{\ell_1} = 0$ derived from the boundary conditions for x and f_0. This leads to

$$f_t^\gamma = -\frac{X\sigma^2 T}{12\ell_0}\left(1 - \frac{t}{T}\right)\left(2t - \frac{t^2}{T}\right), \tag{1.3.12}$$

$$f_t^{\ell_1} = -\frac{X^2}{4\ell_0}\frac{t}{T}\left(1 - \frac{t}{T}\right). \tag{1.3.13}$$

To wit, the first-order corrections should therefore be given by the sum of the separate correction terms. By adapting the verification argument in Theorem 1.3.1, we can show that this candidate strategy is indeed optimal at the leading order $O(\gamma + \ell_1)$:

[14] Note that for $\ell_1 = 0$, we recover the Euler-Lagrange equation (1.3.1) in the classical Almgren-Chriss model. Conversely, for $\ell_1 > 0$ but $\gamma = 0$ we are back in the risk-neutral setting of Section 1.3.2.

Theorem 1.3.3. *The strategy $x^{\gamma,\ell_1,*}$ from* (1.3.11) *with f^γ, f^{ℓ_1} as in (1.3.12-1.3.13) minimizes the mean-variance criterion* (1.3.9) *over all deterministic, decreasing, absolutely continuous strategies x with square integrable derivative, which satisfy $x_0 = X$, $x_T = 0$, up to terms of order $o(\gamma + \ell_1)$.*

That is, for all competing strategies x:

$$\begin{aligned}\mathbb{E}\left[C_X^{\ell_1}(x^{\gamma,\ell_1,*})\right] &+ \frac{\gamma}{2}\,Var\left[C_X^{\ell_1}(x^{\gamma,\ell_1,*})\right] \\ &\leqslant \mathbb{E}\left[C_X^{\ell_1}(x)\right] + \frac{\gamma}{2}\,Var\left[C_X^{\ell_1}(x)\right] + o(\gamma+\ell_1), \quad \textit{as } \gamma, \ell_1 \downarrow 0.\end{aligned}$$

Proof. Abbreviate $x^{\gamma,\ell_1,*} =: x^*$ to ease notation. For any decreasing, deterministic strategy x, the corresponding expected execution cost can be written as

$$\begin{aligned}\mathbb{E}[C_X^{\ell_1}(x)] + \frac{\gamma}{2}\mathrm{Var}[C_X^{\ell_1}(x)] &= \int_0^T \left((\ell_0+\ell_1 X)\dot{x}_t^2 - \ell_1 x_t \dot{x}_t^2 + \frac{\gamma\sigma^2}{2}x_t^2\right)dt \\ &:= H(x).\end{aligned}$$

Setting $y = x - x^*$, we therefore have

$$\begin{aligned}H(x) &= H(x^*+y) \\ &= (\ell_0+\ell_1 X)\int_0^T \left((\dot{x}_t^*)^2 + 2\dot{x}_t^*\dot{y}_t + \dot{y}_t^2\right)dt \\ &\quad - \ell_1 \int_0^T (x_t^*+y_t)\left((\dot{x}_t^*)^2 + 2\dot{x}_t^*\dot{y}_t + \dot{y}_t^2\right)dt + \frac{\gamma\sigma^2}{2}\int_0^T (x_t^*+y_t)^2 dt \\ &= (\ell_0+\ell_1 X)\int_0^T \left((\dot{x}_t^*)^2 + 2\dot{x}_t^*\dot{y}_t + \dot{y}_t^2\right)dt + \frac{\gamma\sigma^2}{2}\int_0^T \left((x_t^*)^2 + 2x_t^* y_t + y_t^2\right)dt \\ &\quad - \ell_1 \int_0^T \left(x_t^*(\dot{x}_t^*)^2 + y_t(\dot{x}_t^*)^2 + 2x_t^*\dot{x}_t^*\dot{y}_t + 2y_t\dot{x}_t^*\dot{y}_t + x_t^*\dot{y}_t^2 + y_t\dot{y}_t^2\right)dt \\ &= \int_0^T (\ell_0+\ell_1 X - \ell_1 x_t^*)(\dot{x}_t^*)^2 dt + \frac{\gamma\sigma^2}{2}\int_0^T (x_t^*)^2\,dt \\ &\quad + \int_0^T (\ell_0+\ell_1 X - \ell_1 x_t^* - \ell_1 y_t)\,\dot{y}_t^2 dt + \frac{\gamma\sigma^2}{2}\int_0^T y_t^2 dt \\ &\quad + \int_0^T y_t\left(-2(\ell_0+\ell_1 X)\ddot{x}_t^* + 2\ell_1\ddot{x}_t^* x_t^* + 2\ell_1(\dot{x}_t^*)^2 - \ell_1(\dot{x}_t^*)^2 + \gamma\sigma^2 x_t^*\right)dt \\ &\quad - \ell_1\int_0^\infty 2y_t\dot{y}_t\dot{x}_t^* dt,\end{aligned}$$

which yields

$$
\begin{aligned}
H(x) = H(x^*) + \int_0^T (\ell_0 + \ell_1 X - \ell_1 x_t^* - \ell_1 y_t)\, \dot{y}_t^2 dt + \ell_1 \int_0^T y_t^2 \ddot{x}_t^* dt \\
+ \frac{\gamma\sigma^2}{2} \int_0^T y_t^2 dt + o(\gamma + \ell_1) \\
\geqslant H(x^*) + o(\gamma + \ell_1),
\end{aligned}
$$

proving the assertion. In this estimate, we have used integration by parts twice and also taken into account $y_0 = y_T = 0$ in the third step. In the fourth, we have inserted the Euler-Lagrange equation (1.3.6) satisfied by x^* up to terms of order $o(\gamma + \ell_1)$,[15] and integrated by parts one more time. For the final inequality, we have used that $X - x_t^* - y_t = X - x_t \geqslant 0$ because $x_0 = X$, $x_T = 0$ and x is decreasing, and also taken into account that $\ddot{x}^* \geqslant 0$ on $[0, T]$, as is readily verified by differentiation of the explicit formulae (1.3.2), (1.3.12), and (1.3.13). □

Theorem 1.3.3 shows that the effects of *small* risk aversion and self-excitement parameters overlap. In particular, with self-exciting price impact, the optimal liquidation trajectory again turns out to be steeper in the first half but flatter in the second half of the execution interval. Moreover, small self-excitement always compounds small risk aversion, in that the number of shares remaining in the portfolio is reduced at all times.

However, the comparative statics are more complex beyond this limiting regime. Then, numerical methods are needed to solve the boundary-value problem (1.3.10). Some sample results obtained using the Mathematica routine NDSolve are reported in Figures 1.2 and 1.3 for moderate and high risk aversion, respectively. Both figures show that self-excitement indeed speeds up liquidation initially and slows it down eventually. The size and overall effect of this change on the optimal position, however, crucially depend on the relative magnitudes of self-excitement and risk aversion. Recall from Figure 1.1 that self-excitement lowers the optimal position at all times for risk-neutral investors, thereby introducing some *effective risk aversion*. In contrast, with intermediate risk aversion, self-excitement has little effect, cf. Figure 1.2. In this case, the selling speed necessary to avoid large inventory risks is already sufficient, and does not need to be increased further. For large risk aversion, self-excitement – surprisingly – in fact increases the optimal position over virtually the whole execution interval, see Figure 1.3. Indeed, the initial period where liquidation is sped up is very

[15] Indeed, an inspection of the explicit formula shows that this holds uniformly on $[0, T]$, so that the claim for the remainder follows from the dominated convergence theorem.

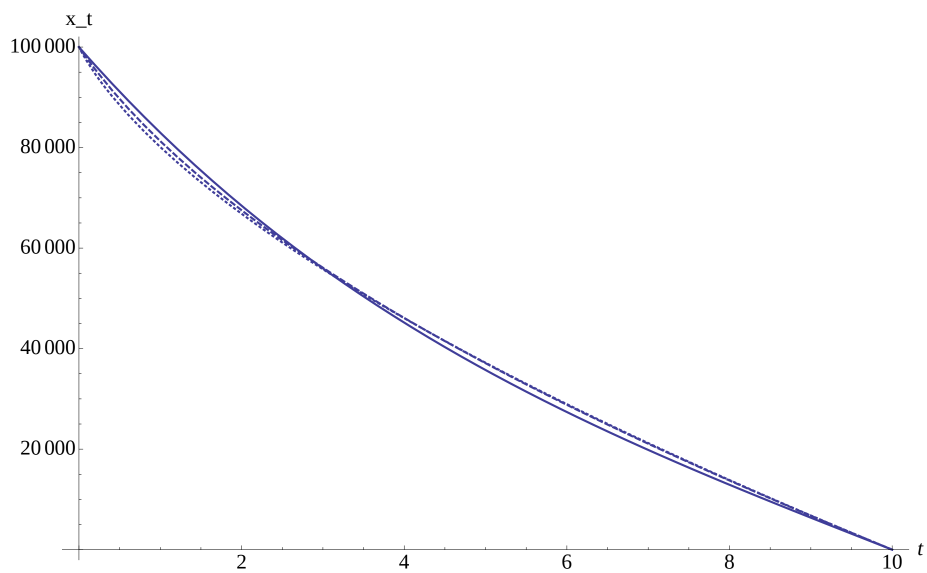

Figure 1.2: Liquidation strategies for an investor with risk aversion $\gamma = 1.5 \cdot 10^{-4}$ and $X = 100\ 000$ risky shares, (daily) volatility $\sigma = 0.02$, baseline price impact $\ell_0 = 10^{-6}$, and self-excitement parameter $\ell_1 = 0$ (solid), $5 \cdot 10^{-11}$ (dashed), $1.5 \cdot 10^{-10}$ (dotted).

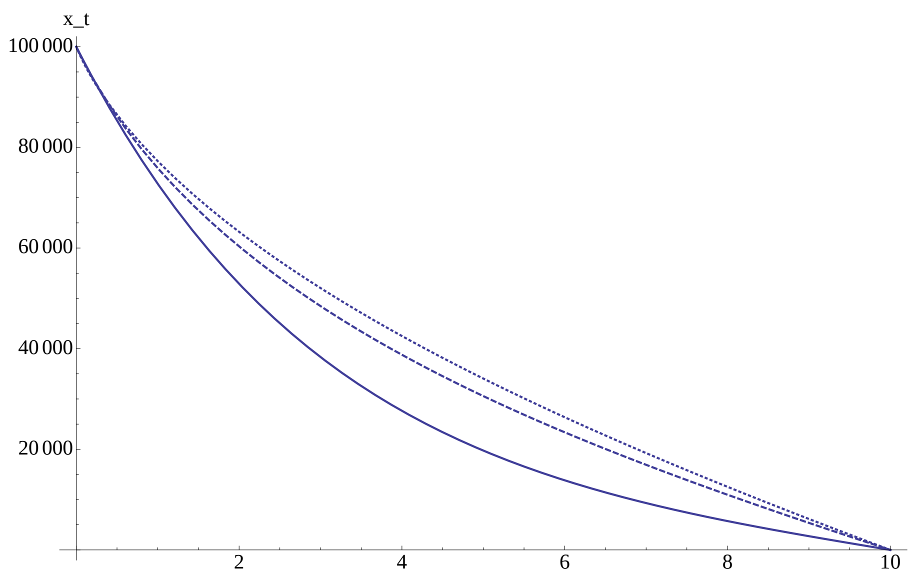

Figure 1.3: Liquidation strategies for an investor with risk aversion $\gamma = 5 \cdot 10^{-4}$ and $X = 100\ 000$ risky shares, (daily) volatility $\sigma = 0.02$, baseline price impact $\ell_0 = 10^{-6}$, and self-excitement parameter $\ell_1 = 0$ (solid), $5 \cdot 10^{-11}$ (dashed), $1.5 \cdot 10^{-10}$ (dotted).

short in this case, so that self-excitement actually *lowers* effective risk aversion over most of the execution interval.

This shift results from the interplay of two different effects. On the one hand, self-excitement provides an incentive to front load. On the other hand, it leads to higher levels of price impact that call for slower liquidation. Whence, in all cases, liquidation is initially sped up and eventually slowed down, but the quantitative properties of this effect depend on the relative magnitude of risk aversion and self-excitement. In particular, with high risk aversion, the large initial trading rate increases the price impact quickly, so that the execution path lags behind the one without self-excitement almost immediately.

Overall, our model for self-exciting price impact therefore has a “moderating” effect, in that very slow liquidation is sped up whereas very fast liquidation is slowed down. Nevertheless, the general convex shape of the optimal trajectory remains a robust feature of the solution.

Chapter 2

Utility Maximization in Markets with Small Nonlinear Price Impact

Classical financial theory is built on the paradigm of *frictionless* markets. By assuming that arbitrary quantities can be traded immediately at the quoted market price, many elegant and far-reaching results can be derived. Real financial markets, however, only supply limited liquidity. Accordingly, execution prices are adversely affected by large trades executed quickly. Optimally scheduling the order flow – to trade off displacement from the optimal frictionless risk-return profile against trading costs – is therefore a crucial concern for large investors such as trend-following hedge funds.

In this chapter, we study this problem in a general setting. We consider agents with constant absolute risk aversion,[1] who trade a risky asset with general, not necessarily Markovian, Itô dynamics to maximize their expected utility.

As in the model of Almgren [5], trades incur costs proportional to a power $p \in (1, 2)$ of the order flow, corresponding to a price impact proportional to the $(p-1)$-th power of both trade size and execution speed. A price impact elasticity of $p \approx 3/2$ is in line with the "square-root law" advocated by most practitioners (cf., e.g, [123, 10]). The limiting cases $p \to 1$ and $p \to 2$ lead to proportional and quadratic transaction costs – the two frictions that have been the focus of most of the academic research.[2]

To obtain tractable results in this general setting, we focus on *small* price impact,

[1] Our results formally extend to more general preferences, compare [109]. These arguments could be made rigorous similarly as for proportional transaction costs [1], but we do not pursue this here in order not to drown the new features of the model with nonlinear price impact in (even more) technical estimates arising from random and time-varying risk tolerances.

[2] See, e.g, [175, 107, 109, 136, 34, 35] and [79, 9, 65, 66, 142, 15, 34, 35] as well as the references therein for surveys of the large literatures on proportional and quadratic trading costs.

and perform a sensitivity analysis around the benchmark problem without trading costs. In frictionless diffusion models, optimal trading strategies $\hat{\varphi}_t$ are typically diffusion processes as well, and thereby generate infinite price impact costs. We show that, at the leading order, frictionless target strategies $\hat{\varphi}_t$ of this type are optimally tracked by smoothed strategies φ_t^λ satisfying (pathwise) the following ordinary differential equation (ODE):

$$\dot{\varphi}_t^\lambda = p^{-\frac{1}{p-1}} \left(\frac{\gamma c_t^S (c_t^{\hat{\varphi}})^2}{8\lambda_t} \right)^{\frac{1}{p+2}} g_p^{\frac{1}{p-1}} \left(\left(\frac{2^{p-1}\gamma c_t^S}{\lambda_t (c_t^{\hat{\varphi}})^p} \right)^{\frac{1}{p+2}} (\hat{\varphi}_t - \varphi_t^\lambda) \right). \tag{2.0.1}$$

Here, γ is the agents' risk aversion; $c_t^S = \frac{d\langle S\rangle_t}{dt}$ and $c_t^{\hat{\varphi}} = \frac{d\langle\hat{\varphi}\rangle_t}{dt}$ are the (squared) diffusion coefficients of the risky asset and the frictionless target strategy $\hat{\varphi}_t$; p is the elasticity of price impact, and λ_t is the corresponding constant of proportionality describing its magnitude at time t. Finally, g_p is the solution of a nonlinear ordinary differential equation, cf. Section 2.2.1.

For constant market and preference parameters, (2.0.1) formally recovers the trading speed that is asymptotically optimal in the Black-Scholes model of Guasoni and Weber [80]: a deterministic function of the current deviation from the frictionless optimizer.[3] In our general setting, the optimal trading speed remains "myopic", in that it is fully determined by current market and preference parameters, as well as the current displacement from the frictionless target. In particular, the function g_p is universal: it only depends on p, the elasticity of price impact, but not on the other primitives of the model. As price impact becomes linear for $p \to 2$, this function converges to $g_2(x) = 2x$. Accordingly, (2.0.1) reduces to the following much simpler formula from [142]:

$$\dot{\varphi}_t^\lambda = \sqrt{\frac{\gamma c_t^S}{2\lambda_t}} (\hat{\varphi}_t - \varphi_t^\lambda). \tag{2.0.2}$$

In this limiting case, the trading speed is independent of the volatility $c_t^{\hat{\varphi}}$ of the frictionless target. As a result, the (relative) trading speed for small linear price impact is universal, in that it does not depend on the optimal strategy for the concrete application at hand. In particular, the turnover rate (2.0.2) is the same as the initial optimal execution rate [8, 170], so that trading with small *linear* price impact can be interpreted as (locally) liquidating towards the frictionless target. Comparing our formula (2.0.1) to the results of [5], we find that this analogy breaks down with nonlinear price impact. This happens because the optimal trading rate now depends

[3]Since we use an absolute parametrization (for risk aversion, returns, etc.), it is difficult to make this connection to the relative quantities of [80] precise; compare [76] for more details.

on the target volatility $c_t^{\hat{\varphi}}$, which has no counterpart in optimal execution where the target is constant.

As $p \to 1$, the function g_p converges to the cubic polynomial $g_1(x) = x^3/3 - (3/2)^{2/3}x$. Whence, the associated trading speed (2.0.1) explodes once the deviation from the frictionless target exceeds

$$\hat{\varphi}_t - \varphi_t^\lambda = \pm \left(\frac{3}{2\gamma} \frac{c_t^{\hat{\varphi}}}{c_t^S} \lambda_t \right)^{1/3}.$$

Between these boundaries, the optimal trading speed converges to zero. This "bang-bang control" corresponds to the instantaneous reflection off these trading boundaries that is asymptotically optimal for small *proportional* transaction costs [175, 107, 109, 136].

The same interpolation between these limiting regimes obtains for the welfare effects of price impact. At the leading order, the certainty equivalent loss due to trading costs is

$$c_p \, \mathbb{E}_{\hat{\mathbb{Q}}} \left[\int_0^T \lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c_t^S (c_t^{\hat{\varphi}})^2}{8} \right)^{\frac{p}{p+2}} dt \right]. \tag{2.0.3}$$

Here, the constant c_p is obtained from the nonlinear ODE for g_p, cf. Section 2.2.1. The other terms show that price impact has a substantial effect if the market is illiquid (large λ_t), volatile (large c_t^S), or if the frictionless target strategy is difficult to track because it is very active (large $c_t^{\hat{\varphi}}$). If all of these quantities are constant, we formally recover an expression similar to the constant performance loss of Guasoni and Weber [80]. When these quantities are time dependent and random, they need to be averaged both across time and states. As for other frictions [107, 109, 142], averaging across states is performed under the frictionless agents' marginal pricing measure $\hat{\mathbb{Q}}$ – the small friction is priced like a marginal path-dependent option. The comparative statics of the certainty equivalent loss (2.0.3) also are the same as for proportional [107, 103], quadratic [142], or fixed transaction costs [11]; the elasticity of price impact p only governs the asymptotic order of the certainty equivalent loss, the power to which the inputs are raised, and contributes to the universal constant c_p.

To illustrate the wide scope of these results, we discuss a number of applications and implications. To wit, we first show how the general setting studied here allows to treat as special cases models for portfolio management with stochastic investment opportunities and models for the hedging of derivative securities. In both cases, active trading is crucial so that trading costs are of prime importance – but can still be handled in a tractable manner through our results. We also show how the trading cost itself can be endogenized by studying an equilibrium between risk-averse clients and dealers. For clients trading according to the asymptotically optimal rule (2.0.1), we determine the value of the trading cost λ_t that allows a dealer with small risk aversion to break even. Finally, we also discuss the performance of suboptimal policies. This allows to assess how much utility is lost by restricting to simpler trading rates such as (2.0.2).

To prove our results, we use the convex duality approach first used in a Mathematical Finance context by Henderson [85] for the indifference pricing of small unhedgeable claims.[4] More specifically, we obtain a lower bound for the value expansion (2.0.3) by analyzing a specific family of trading strategies. An upper bound can in turn be determined using convex duality. For proportional transaction costs, starting with the seminal work of Cvitanić and Karatzas [48], the corresponding duality has been studied intensely, and is used by Kallsen and Li [103] to obtain a tight upper bound for (2.0.3). In the present context with superlinear trading costs, duality results have only very recently been developed by Guasoni and Rásonyi [77]. These are the starting point for the construction of an asymptotically tight upper bound for the value expansion (2.0.3) in the present study. More specifically, the first-order condition of [77] allows us to derive a candidate dual minimizer from our candidate primal maximizer. By suitably localizing this "naive" dual candidate, we obtain a dual element that finally allows us to complete the asymptotic verification.

The computation of the primal and dual bounds proceeds in several steps. We first perform a second-order expansion of the primal and dual goal functionals, thereby reducing them to linear-quadratic functionals.[5] After renormalising time and space appropriately, we then show that "locally", i.e., on each small time interval, this simplified criterion converges to an ergodic mean-variance functional of a controlled diffusion process. In the present context, this controlled process is an Ornstein-Uhlenbeck-type process, with constant volatility but nonlinear mean reversion speed governed by the function g_p. With quadratic trading costs, the limiting process is a standard Ornstein-Uhlenbeck process; with proportional costs, it is a doubly reflected Brownian motion. Whereas these limiting processes are well understood, a number of delicate probabilistic estimates need to be developed from scratch here to establish

[4]Extensions of these results have been developed by [118]; similar arguments have also been used for the perturbation analysis of small variations of market prices of risk [121] or cumulative random endowments [86].

[5]Such simpler performance criteria are directly used in a number of articles, e.g., [65, 66, 9, 15, 162, 34, 35]. The same simplification for more general utilities also obtains for proportional costs [97, 160, 175, 109].

convergence to the limiting problem.

This connection between asymptotics of utility maximization problems with small transaction costs and ergodic control problems with "frozen coefficients" has first been established using PDE techniques by Soner and Touzi [175].[6] For proportional costs, non-Markovian extensions of these results have been obtained formally by [107, 109] and proved rigorously by [103]. A closely related strand of research studies "pathwise" criteria, where the goal is to trade off the error of tracking an exogenous target strategy against the trading costs incurred by the hedge. Building on work of [62, 63, 72], Cai, Rosenbaum and Tankov [34, 35] study such problems for quite general specifications of tracking errors and trading costs, and derive tight bounds for a number of examples, including proportional and quadratic costs. In their setting, the tight bound matching the performance of their asymptotically optimal family of trading strategies is obtained using weak-convergence techniques rather than convex duality. An advantage of the pathwise approach is that it avoids many of the integrability conditions that need to be imposed in the present expected utility-based setting. On the other hand, these models with exogenous target portfolios are not directly applicable for equilibrium analyses as in [87].

This chapter is organized as follows. In Section 1, we introduce the model. Section 2 collects our main results, and discusses their applications and implications. Section 3 gives the assumptions necessary to carry out the rigorous asymptotic analysis. The proofs can be found in Section 4, 5, 6 and 7: Section 4 contains results on the candidate strategy used to establish the primal lower bound in Section 5, and the dual upper bound in Section 6. For better readability, auxiliary estimates are delegated to the last section: Section 7.1 collects results about the function g_p; Section 7.2 contains probabilistic estimates for the (rescaled) deviations from the frictionless targets, their small-cost limits and bounds on the probability of stopping. Finally, Section 7.3 collects some auxiliary technical estimates used in the ergodic convergence proof.

Notations For an Itô process X, we denote by μ_t^X and $c_t^X = \frac{d\langle X\rangle_t}{dt}$ its drift rate (under the marginal pricing measure $\hat{\mathbb{Q}}$ from Section 2.1.2) and squared diffusion coefficient, respectively. The infinitesimal covariation process between two Itô processes X and Y is denoted by $c_t^{X,Y} = \frac{d\langle X,Y\rangle_t}{dt}$. We write $L(X)$ for the set of X-integrable processes, and denote by $L^2_{\mathbb{P}}(X)$ the subset of integrands H satisfying $\mathbb{E}_{\mathbb{P}}\left[\int_0^T H_t^2 d\langle X\rangle_t\right] < \infty$. For a continuous process X, the running maximum is

[6]Similar "homogenization results" have a long history in other contexts, cf., e.g., [145] and the references therein.

denoted by $X_t^* = \max\{|X_s| : s \in [0,t]\}$; more generally, for a positive real number q, we set $X_t^{q,*} = \max\{|X_s|^q : s \in [0,t]\}$ for $t \geqslant 0$. For an Itô process X, $\mathcal{E}(X) = \exp(X - \frac{1}{2}\langle X\rangle)$ denotes the corresponding stochastic exponential. To ease notation, we write C for a generic positive constant, independent of the asymptotic parameter λ, that may change from line to line. We also use the Landau notation; every time the symbols $O(\cdot)$ and $o(\cdot)$ appear, they refer to limits where the asymptotic parameter λ tends to zero.

2.1 Model

2.1.1 Financial Market

Fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ supporting a standard Brownian motion $W^{\mathbb{P}}$. We consider a financial market with two assets. The first one is safe, with price normalized to one. The second is risky, with general – not necessarily Markovian – Itô dynamics:

$$dS_t = \mu_t^S dt + \sqrt{c_t^S} dW_t^{\mathbb{P}}. \tag{2.1.1}$$

Here, we assume that the expected return μ^S is adapted, the (squared) diffusion coefficient c^S is a positive and continuous Itô process, and μ^S and c^S are such that the stochastic differential equation (2.1.1) has a unique strong solution on $[0,T]$.

2.1.2 Frictionless Portfolio Choice

In the above market, we study the portfolio choice problem of an agent with constant absolute risk aversion $\gamma > 0$. To wit, starting from an initial endowment $x \in \mathbb{R}$, the agent's goal is to choose a predictable trading strategy φ so as to maximize her expected exponential utility from terminal wealth:[7]

$$\mathbb{E}_{\mathbb{P}}\left[-\mathrm{e}^{-\gamma\left(x+\int_0^T \varphi_t dS_t\right)}\right] \to \max! \tag{2.1.2}$$

To ensure well-posedness of the maximization problem (2.1.2), we impose the following no-arbitrage condition [165, 56, 99, 167]:

[7] As is well known [56], random endowments can be readily absorbed into a change of probability.

Assumption 1. *There exists an equivalent local martingale measure $\mathbb{Q}$ for S, which has finite relative entropy with respect to $\mathbb{P}$, i.e.,*

$$H(\mathbb{Q}|\mathbb{P}) := \mathbb{E}_{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\log\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\right] < \infty.$$

In view of [61, Theorems 2.1, 2.2, and Remark 2.1], Assumption 1 implies that there exists a unique local martingale measure $\hat{\mathbb{Q}}$ equivalent to $\mathbb{P}$ that solves the dual problem of minimizing the relative entropy with respect to $\mathbb{P}$ among the absolutely continuous local martingale measures.

Assumption 1 also ensures the existence of an optimizer $\hat{\varphi}$ for (2.1.2) among all S-integrable processes φ whose gain processes $\int_0^\cdot \varphi_t dS_t$ are $\hat{\mathbb{Q}}$-martingales, compare [56, Theorem 1]. Henceforth, we therefore focus on such *admissible* trading strategies.[8] The primal maximizer $\hat{\varphi}$ is linked to the "minimal-entropy martingale measure" $\hat{\mathbb{Q}}$ by the first-order condition of convex duality [167, Equation (12)]:

$$U'\left(x + \int_0^T \hat{\varphi}_t dS_t\right) = \hat{y}\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}, \quad \text{for a constant } \hat{y} > 0. \tag{2.1.3}$$

2.1.3 Portfolio Choice with Superlinear Transaction Costs

As in [8, 10, 80, 77], we now assume that trades incur superlinear costs levied on the trading rate $\dot{\varphi}_t = \frac{d}{dt}\varphi_t$, i.e., trading costs increase with both trade size and speed.

More specifically, execution prices are shifted proportionally to a power $p-1 \in (0,1)$ of $\dot{\varphi}_t$, so that the corresponding trading costs accumulate at rate $\lambda_t|\dot{\varphi}_t|^p$. Here, the proportionality factor is of the form $\lambda_t = \lambda\Lambda_t$, for some small parameter $\lambda > 0$ that measures the magnitude of the trading costs, and a positive, continuous Itô process Λ that describes their dynamics. The constant p is the "elasticity of price impact"; proportional transaction costs correspond to the limiting case $p \to 1$, linear price impact (quadratic costs) to $p \to 2$. Empirical studies typically estimate values $p \approx 3/2$, compare [10, 123].

With trading costs, we need to specify how the agent's initial endowment x is allocated between her safe and risky accounts. For simplicity, we assume that the initial risky allocation equals the frictionless optimum $\hat{\varphi}_0$, so that $x_0 = x - \hat{\varphi}_0 S_0$ is the corresponding initial safe position.

Likewise, different terminal conditions are possible, cf. [15]. Here, as in [77], we impose that the risky position is eventually liquidated for consumption ($\varphi_T = 0$).

[8] Notions of admissibly for utility functions defined on the whole real line are delicate and usually involve dual objects like equivalent martingale measures. For more information and a discussion on some flexibility with respect to the precise choice of admissibility, we refer to [167].

The set of all absolutely continuous trading strategies $d\varphi_t = \dot{\varphi}_t dt$ that satisfy these requirements and belong to $L^2_{\hat{\mathbb{Q}}}(S)$ is denoted by Φ. The frictional wealth process corresponding to $\varphi \in \Phi$ is $X^\varphi = \left(x + \int_0^t \varphi_s dS_s - \int_0^t \lambda_s |\dot{\varphi}_s|^p ds\right)_{t\in[0,T]}$. Accordingly, in analogy to the frictionless case (2.1.2), the agent chooses $\varphi \in \Phi$ to maximize

$$\mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \varphi_t dS_t - \int_0^T \lambda_t |\dot{\varphi}_t|^p dt\right)\right] \to \max! \tag{2.1.4}$$

2.2 Main Results

The frictional portfolio choice problem (2.1.4) is intractable even in the simplest concrete models. We therefore study the asymptotic regime where the magnitude λ of the trading costs tends to zero. Results of this kind have recently been obtained by Guasoni and Weber [80] for a long-term portfolio choice problem in a Black-Scholes model with scale-invariant price impact. Here, we perform this sensitivity analysis in a general setting. This reveals the general underlying structure of the problem and identifies the relevant statistics that measure the susceptability of trading strategies with respect to small trading costs. Moreover, as discussed in Section 2.2.3 below, this allows to treat as special cases the trading problems where transaction costs are most relevant in practice: active portfolio management and hedging of derivatives.

2.2.1 Candidate Strategy

We now introduce the trading strategies that will turn out to be asymptotically optimal for (2.1.4) in the limit for small transaction costs λ, see Theorem 2.2.3.

A Nonlinear ODE A main ingredient is the solution to a nonlinear ODE, that already plays a central role in the work of Guasoni and Weber [80]:

Lemma 2.2.1. *There exists a unique positive constant c_p such that the ordinary differential equation*

$$g_p'(z) = (p-1)p^{-\frac{p}{p-1}} |g_p(z)|^{\frac{p}{p-1}} - z^2 + c_p \tag{2.2.1}$$

has a solution on $\mathbb{R}$ which satisfies the following growth conditions:

$$\lim_{z\to-\infty} \frac{g_p(z)}{|z|^{\frac{2(p-1)}{p}}} = -p(p-1)^{-\frac{p-1}{p}}, \quad \textit{and} \quad \lim_{z\to+\infty} \frac{g_p(z)}{|z|^{\frac{2(p-1)}{p}}} = p(p-1)^{-\frac{p-1}{p}}. \tag{2.2.2}$$

This solution is unique and is an odd, increasing function.

Proof. See [80, Lemmas 20 and 22] and Section 2.7.1. □

The growth conditions at $\pm\infty$ are somewhat ad hoc. In order to better motivate why the function g_p is a natural object in the present context, we provide the following alternative characterization.

Lemma 2.2.2. *The constant c_p from Lemma 2.2.1 is the smallest value for which the ODE* (2.2.1) *has a solution on the whole real line that is positive on $\mathbb{R}_+^*$ and negative on $\mathbb{R}_-^*$.*

Proof. See Section 2.7.1. □

The function g_p will turn out to parametrize the agent's trading speed as a function of her position's displacement from its frictionless target in Theorem 2.2.3. Accordingly, positivity on $\mathbb{R}_+$ and negativity on $\mathbb{R}_-$ translate to the natural property that one always trades towards the frictionless optimum. The corresponding constant c_p will turn out to describe the size of the corresponding utility loss. Since this needs to be minimized at the optimum, it is therefore natural that the smallest possible choice is the correct one. To simplify notation, we pass to the following "rescaled" version of g_p:

$$\widetilde{g}_p(x) = \operatorname{sgn}(x)\,|g_p(x)|^{\frac{1}{p-1}}\,, \quad x \in \mathbb{R}. \tag{2.2.3}$$

Asymptotically Optimal Strategies Using the function $\widetilde{g}_p$ from (2.2.3), we now define a family of strategies that will be shown to be asymptotically optimal in Theorem 2.2.3 below. Similarly as for models with linear price impact [66, 9, 142, 79], these strategies track the frictionless target portfolio $\hat{\varphi}$. Their fine structure in turn depends on the degree of activity exhibited by $\hat{\varphi}$.[9] Here, we focus on the generic case where the frictionless target $\hat{\varphi}$ is an Itô process[10] that has a strictly positive and continuous Itô process as diffusion coefficient ($c^{\hat{\varphi}} > 0$).

With small *linear* price impact ($p = 2$), it is asymptotically optimal to track the frictionless target at a trading speed that is given by $\sqrt{\gamma c_t^S / 2\lambda_t}$ times the deviation of the frictional portfolio from the target [142]. In the present nonlinear context ($1 < p < 2$), the impact of the deviation becomes nonlinear as well and governed by

[9]For example, if the target strategy is smoother than Brownian motion, then it can be tracked much more closely and with substantially smaller trading costs, compare [162].

[10]This obtains, for example, in Markovian models with sufficient regularity because the optimizer then is obtained by evaluating the derivatives of the value function along the Itô state variables.

the function $\widetilde{g}_p$. More specifically, the optimal trading rates are *essentially* characterized by the following pathwise ODE (existence and uniqueness are established in Section 2.4.1):

$$\dot{\varphi}_t^\lambda = \frac{1}{2} p^{-\frac{1}{p-1}} c_t^{\hat{\varphi}} m_t^\lambda \widetilde{g}_p \left(m_t^\lambda (\hat{\varphi}_t - \varphi_t^\lambda) \right), \tag{2.2.4}$$

where

$$m_t^\lambda = \left(\frac{2^{p-1} \gamma \, c_t^S}{\lambda_t (c_t^{\hat{\varphi}})^p} \right)^{\frac{1}{p+2}} = \lambda^{-\frac{1}{p+2}} m_t.$$

Comparative Statics The trading speed (2.2.4) is determined by the diffusion coefficients c_t^S, $c_t^{\hat{\varphi}}$ of the risky asset and the frictionless optimizer, the current trading cost λ_t, the risk aversion γ, and the elasticity of the price impact p. In addition, the function $\widetilde{g}_p$ determines how the deviation

$$\Delta \varphi_t^\lambda = \hat{\varphi}_t - \varphi_t^\lambda$$

of the actual position from the frictionless optimizer is incorporated. For $p \uparrow 2$, the function $\widetilde{g}_p$ becomes linear, and the optimal trading speed simplifies to the formula obtained by [142] – which no longer depends on the variation $c_t^{\hat{\varphi}}$ of the target strategy. For $p < 2$, the agent trades more slowly for small deviations (because the nonlinear trading cost is larger than quadratic in this case); conversely, more rapid trading is possible far from the frictionless optimum. For $p \downarrow 1$, the function g_p converges to a cubic polynomial, and $\widetilde{g}_p$ converges pointwise to 0 on an open interval around 0 and to infinity outside the closure of this interval. This limit therefore formally recovers the "bang-bang controls" for proportional costs, where the optimal trading rate is zero inside a "no-trade region" and infinite once its boundaries are breached [175, 107, 109, 136].

To understand the comparative statics of the trading rate, recall that the function $\widetilde{g}_p$ is increasing. Whence, the trading rate (2.2.4) remains increasing in the ratio of price volatility times risk aversion divided by the current price impact, as for quadratic trading costs [142]. The dependence on the volatility of the frictionless target is more complex. To wit, if the displacement of the frictional position is close to zero, then the ODE (2.2.1) shows that the function $\widetilde{g}_p$ is proportional to $x \mapsto x^{\frac{1}{p-1}}$. Hence, the trading rate (2.2.4) becomes close to proportional to $c_t^{\hat{\varphi}}$ raised to the power $\frac{2}{p+2} - \frac{p}{(p-1)(p+2)} < 0$. Thus, a large target volatility discourages the agent from trading when she has almost the optimal number of risky shares. The reason is that a price impact elasticity of $p \in (1, 2)$ leads to higher than quadratic trading costs for small trades. Thus, high tracking speeds are reduced near the frictionless optimum. On the other hand, when the displacement is large, the function $\widetilde{g}_p$ scales like $x \mapsto x^{\frac{2}{p}}$ (cf. (2.2.2)), so that the trading rate (2.2.4) no longer depends on $c_t^{\hat{\varphi}}$ like for quadratic costs. The intuition is that if the displacement is very large, the volatility of the target becomes insignificant relative to the displacement from the frictionless optimum.

Technical Modifications To obtain rigorous asymptotic results, the strategy (2.2.4) needs to be slightly modified by appropriate stopping. On the one hand, like for models with proportional and fixed transaction costs [103, 60], excessive deviations from the frictionless target need to be avoided. Therefore, liquidation is initiated immediately if the deviation from the frictionless target becomes too large or efficient tracking becomes impossible because the process m becomes too small.[11] The probability of these events becomes negligible for small λ (see Lemma 2.7.4 and Proposition 2.7.10), but the stopping is crucial to control the remainder terms.

On the other hand, we need to ensure here that the risky position is indeed liquidated until maturity T. (For proportional or fixed costs, this can be done by a single bulk trade at T, without affecting the asymptotic results at the leading order.) To make this precise, define

$$T^\lambda = T - \lambda^\eta, \quad \text{where } \eta \in \frac{1}{p+2}\left(2, \frac{p}{p-1}\right). \tag{2.2.5}$$

This will be the time at which the liquidation of the risky position is initiated at the latest. To also control large deviations from the frictionless target and small trading speeds, choose

$$\begin{aligned} &\kappa_1 \in \frac{1}{p+2}\left(\frac{2}{3}, 1\right), && \kappa_2 \in \left(0, \min\left\{\frac{1}{5(p+2)} - \frac{\kappa_1}{5}, \frac{\kappa_1}{5(2p+3)}\right\}\right), \\ &\kappa_3 \in \left(\frac{2}{3}\frac{1}{p+2}, \frac{2-2p}{p+2} + 2p\kappa_1\right), && \kappa_4 \in \left(0, \frac{1}{p+2} - \frac{p-1}{p}\eta\right), \end{aligned} \tag{2.2.6}$$

and define the stopping time

$$\begin{aligned} \tau^{\Delta\varphi} = T^\lambda \wedge \inf\Big\{ &t \in [0, T^\lambda] : \left|\Delta\varphi_t^\lambda\right| > \lambda^{\kappa_1} \text{ or } m_t < \lambda^{\kappa_2} \text{ or } m_t > \lambda^{-\kappa_2} \\ &\text{or } \int_0^t \lambda\Lambda_t \left|\dot{\varphi}_t\right|^p dt > \lambda^{\kappa_3} \text{ or } \left|\hat{\varphi}_t\right| > \lambda^{-\kappa_4} \Big\}. \end{aligned} \tag{2.2.7}$$

[11] This happens if the trading cost Λ becomes too large, or the target too volatile relative to the risky asset.

Our candidate asymptotic optimal policy is then defined as follows:

(i) On $[\![0, \tau^{\Delta\varphi}]\!]$, the trading rate $\dot{\varphi}_t$ is determined by the ODE (2.2.4).

(ii) On $[\![\tau^{\Delta\varphi}, \tau^{\Delta\varphi} + \lambda^\eta]\!]$ the risky position is liquidated at the constant rate $\dot{\varphi}^\lambda_t = -\lambda^\eta \varphi^\lambda_{\tau^{\Delta\varphi}}$.

(iii) On $[\![\tau^{\Delta\varphi} + \lambda^\eta, T]\!]$, no more trades are implemented ($\dot{\varphi}^\lambda_t = 0$) and the agent's position is $\varphi^\lambda_t = 0$.

2.2.2 Main Result

With the above preparations, we can now formulate our main result. For better readability, its long and technical proof is deferred to Sections 2.4, 2.5, and 2.6.

Theorem 2.2.3. *Suppose the no-arbitrage Assumption 1 holds, and the primitives of the model satisfy the continuity and integrability Assumptions 2 and 3 from Section 2.3. Then, the strategy φ^λ from Section 2.2.1 is asymptotically optimal for the expected utility maximisation problem* (2.1.4)*, in that*

$$\begin{aligned}
&\sup_{\varphi\in\Phi} \mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \varphi_t dS_t - \int_0^T \lambda_t \left|\dot{\varphi}_t\right|^p dt\right)\right] \\
&\quad = \mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \varphi^\lambda_t dS_t - \int_0^T \lambda_t \left|\dot{\varphi}^\lambda_t\right|^p dt\right)\right] + o\left(\lambda^{\frac{2}{p+2}}\right) \\
&\quad = \mathbb{E}_{\mathbb{P}}\left[U\left(x + \int_0^T \hat{\varphi}_t dS_t\right)\right] - \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c^S_t (c^{\hat{\varphi}}_t)^2}{8}\right)^{\frac{p}{p+2}} dt\right] c_p \\
&\qquad + o\left(\lambda^{\frac{2}{p+2}}\right).
\end{aligned}$$

Let us briefly discuss the interpretation of this result. The first term on the right-hand side of the last equality is the performance of the frictionless optimizer. Accordingly, the second corresponds to the minimal leading-order loss that can be achieved by applying the policy from Section 2.2.1. This minimal performance loss is of order $O\big(\lambda^{\frac{2}{p+2}}\big)$ for small tranding costs λ. In the limiting cases $p \downarrow 1$ (proportional costs) and $p \uparrow 2$ (quadratic costs), the orders $O(\lambda^{1/3})$ (cf. [97, 160]) and $O(\lambda^{1/2})$ (cf. [79, 142]) known from the literature obtain.

The factor multiplying this power of the trading cost has three components: the frictionless Lagrange multiplier $\hat{y}$, the constant c_p from Lemma 2.2.1, and an average of the model parameters with respect to time and space. In view of, e.g., [167, Theorem 1], $\hat{y}$ is the derivative of the frictionless performance with respect to the initial endowment. Whence, by Taylor's theorem, the other terms in the leading-order loss can be interpreted as a "certainty-equivalent loss" as in [109, 103, 142]. This means that they correspond to the amount of initial endowment the agent would give up in order to trade the risky asset without adjustment costs.

The first ingredient for this "cash equivalent of the small friction" is the constant c_p, which is universal in that it only depends on the elasticity of price impact p but none of the other model parameters. Its limiting values for $p \downarrow 1$ and $p \uparrow 2$ are $c_1 = (3/2)^{2/3} \approx 1.31$ and $c_2 = 2$, respectively, so that the value expansion in Theorem 2.2.3 reduces to the corresponding results for proportional costs [103] and quadratic costs [142] in these cases. For $p \in (1, 2)$ it needs to be computed numerically, cf. [80, Figure 2]. It turns out that $p \mapsto c_p$ is increasing; for the empirically most relevant case of $p \approx 3/2$, we have $c_p \approx 1.76$.

The final ingredient for the value expansion is the average of the other model parameters. In the Black-Scholes model of Guasoni and Weber [80] this term is constant. In the general model considered here, all these quantities are stochastic processes, and therefore need to be averaged appropriately both with respect to time and states. Like for proportional and quadratic costs [109, 103, 142], the averaging with respect to states is performed with respect to the frictionless agent's minimal entropy martingale measure $\hat{\mathbb{Q}}$. In view of [52], this means that the effect of the small friction is priced like a "marginal" path-dependent option. Like for other trading costs [109, 11, 142], this price is determined by i) the trading cost, ii) the volatility of the risky asset, iii) the volatility of the frictionless target strategy, and iv) the agent's risk aversion. The powers through which these quantities enter interpolate between the cases of proportional and quadratic costs. The comparative statics are the same in each case: the transaction costs cause a big welfare effect if i) trading costs are large, ii) the risky asset is volatile necessitating close tracking of the optimal risk-return allocation, iii) the frictionless target is volatile so that its tracking leads to substantial trading costs, and iv) risk aversion is high so that displacements from the optimal risk-return tradeoff have a big effect on welfare.

2.2.3 Examples and Applications

Let us now discuss some examples and applications for our main result, Theorem 2.2.3. First, we sketch how it can be used to study the effects of nonlinear trading costs in two of the settings where they are of crucial importance: active portfolio management and hedging of derivatives. Afterwards, we discuss how the trading cost can be endogenized in an equilibrium between risk-neutral clients and dealers. Finally, we also compare the asymptotic optimizers to the performance of the linear policies for quadratic costs.

Active Portfolio Management Let us first study a portfolio-choice model where randomly changing investment opportunities lead to active portfolio management. To wit, we consider the Kim-Omberg model [114] with mean-reverting returns:

$$dS_t = \mu_t dt + \sigma dW_t^{\mathbb{P}}, \tag{2.2.8}$$

where μ_t is an Ornstein-Uhlenbeck process,

$$d\mu_t = \kappa(\bar{\mu} - \mu_t)dt + \sigma_\mu dZ_t^{\mathbb{P}}.$$

Here, σ, $\bar{\mu}$, κ, σ_μ are positive constants and $W^{\mathbb{P}}$, $Z^{\mathbb{P}}$ are standard $\mathbb{P}$-Brownian motions with constant correlation $\rho \leqslant 0$.[12] In this setting, Assumption 1 is satisfied, and the frictionless optimal portfolio is [114]:

$$\hat{\varphi}_t = \frac{\mu_t}{\gamma\sigma^2} + \frac{\rho\sigma_\mu}{\gamma\sigma}(C(t)\mu_t + B(t)),$$

for nonpositive, smooth functions $C(t)$, $B(t)$ solving some Riccati equations.[13] Thus, $c_t^S = \sigma^2$ and $c_t^{\hat{\varphi}} = \frac{\sigma_\mu^2}{\gamma^2\sigma^4}(1 + \rho\sigma\sigma_\mu C(t))^2$ are deterministic, bounded and bounded away from zero here.

Next note that $\hat{\varphi}_t$ is the sum of an Ornstein-Uhlenbeck process (with bounded, time-dependent mean-reversion level and speed) and a bounded function under the minimal entropy measure $\hat{\mathbb{Q}}$. Whence, its supremum has finite $\hat{\mathbb{Q}}$-moments of all orders. Therefore, the moment conditions in Assumption 3 are satisfied for constant price impact, for example. The exponential moment conditions also hold if the time horizon is sufficiently short.[14] The asymptotically optimal trading rate (2.2.4) in turn is a deterministic function of the deviation from the frictionless target, similarly as in the Black-Scholes model of [80].

In the uncorrelated case ($\rho = 0$), the relative trading rate is constant, in line with the constant relative trading rates of [65, 66] and the no-trade regions of constant

[12] Empirical studies such as [17] typically find substantially negative values. For the uncorrelated case ($\rho = 0$), the optimal portfolio is the same as for the local mean-variance criteria of [102, 136, 65, 66].

[13] Assumption 1 follows from [124, Example 3 in Section 6.2]; admissibility can be established using [158, Lemma 2.12] and [124, Example 3 in Section 6.2].

[14] This restriction could be avoided by either directly working with quadratic rather than exponential preferences as in [66], or by truncating large values of the state variable as in [142, Section 8.1].

width in [55, 136]. The corresponding certainty equivalent loss from Theorem 2.2.3 then also accumulates at a constant rate here. It is given by

$$\lambda^{\frac{2}{2+p}} \left(\frac{\sigma_\mu^2}{8\gamma\sigma^2} \right)^{\frac{p}{p+2}} c_p,$$

and therefore is increasing in i) the trading cost, ii) the volatility of the signal μ_t, and iii) the inverse of the risk-adjusted asset volatility. The intuition for the last scaling is that the frictionless target is also inversely proportional to this term, and the resulting reduction of the frictionless target volatility overrides the increase of the tracking speed. In contrast, the mean-reversion level $\bar{\mu}$ and mean-reversion speed κ of the expected returns do not influence the leading-order term.

Hedging of Derivatives Let us now illustrate how to use Theorem 2.2.3 to implement hedging strategies in the presence of small nonlinear price impact. For concreteness, we consider a Bachelier model with dynamics

$$dS_t = \sigma dW_t^{\mathbb{P}}.$$

Here, $W^{\mathbb{P}}$ is a standard Brownian motion, and σ is a positive constant. Let us study the optimization problem of an agent that has sold a European option with payoff function $H = h(S_T)$, where $s \mapsto h(s)$ is four times differentiable, h', h''' are of linear growth, h, h'''' are of polynomial growth, and h'' is bounded and bounded away from zero. Then H is replicated by the delta hedge $\frac{\partial f(t,S_t)}{\partial s}$, where the option price $f(t, S_t)$ at time t is

$$f(t, s) = \int_{-\infty}^{\infty} h\left(s + x\sigma\sqrt{T-t}\right) \phi(x) dx.$$

(Here, ϕ denotes the density of the standard normal law.) Dominated convergence shows that f is four times differentiable. Moreover, f and $\frac{\partial f}{\partial s}$ are of polynomial and linear growth in s, respectively, and $\frac{\partial^2 f}{\partial s^2}$ is bounded and bounded away from zero.

Now, note that Jensen's inequality and the $\mathbb{P}$-martingale property of admissible strategies show that the replicating strategy $\frac{\partial f(t,S_t)}{\partial s}$ is optimal for the utility maximization problem (2.1.4) augmented by the short position in H. This problem is equivalent to the optimization problem without the claim H under the measure $\mathbb{P}^H$ with density $d\mathbb{P}^H/d\mathbb{P} = e^{-\gamma H}/\mathbb{E}_{\mathbb{P}}[e^{-\gamma H}]$. Whence, we can apply Theorem 2.2.3 for constant positive Λ, for example: Assumption 1 then holds with $\hat{\mathbb{Q}} = \mathbb{P}$; Assumption 3 is in turn satisfied because both the (constant) diffusion coefficient of the risky asset and the diffusion coefficient $\frac{\partial^2 f}{\partial s^2}\sigma$ of the frictionless target strategy are bounded and bounded away from zero, the drift rates of the frictionless target strategy is normal and the drift of its diffusion coefficient is bounded, and the supremum of the frictionless target strategy has all moments by the Burkholder-Davis-Gundy inequality because S is a Brownian motion and $\frac{\partial f}{\partial s}$ is of polynomial growth. Additionally, we assume that the trading horizon T or the risk aversion γ are small enough to ensure the existence of exponential moments for $(\mu^{\hat{\varphi}})^2$ and $(\hat{\varphi}^*)^2$.

As a concrete example, let us consider a "smoothed convexified put option" with strike K.[15] Here, smoothing refers to replacing the actual put payoff $(K - S_T)^+$ with its Bachelier price $(K - S_T)\Phi(\frac{K-S_T}{\sigma\sqrt{\vartheta}}) + \sigma\sqrt{\vartheta}\phi(\frac{K-S_T}{\sigma\sqrt{\vartheta}})$ with a very short maturity ϑ, say one day. This ensures that the payoff is not only of linear growth, but also has bounded smooth derivatives of all orders. Since the second derivative of this payoff and in turn the diffusion coefficient of the replicating strategy is not bounded away from zero, we slightly modify the payoff further by smoothly adding suitable parabolas for sufficiently large and small values of the terminal asset price. The resulting payoff function then satisfies all assumptions made above; it is depicted in the left panel of Figure 2.1. There, we also plot the corresponding Bachelier price and the illiquidity corrections derived from Theorem 2.2.3 by numerical integration for a long and a short position of one option, respectively. As (yearly) parameters, we use $\sigma = 0.2\times100$, which roughly corresponds to a Black-Scholes volatility of 20% at initial price 100 (compare [168]), $\gamma = 10$, $p = 3/2$, and the estimate $\lambda = 0.14 \times 1.57/250$ from [10].

To illustrate the nonlinear scaling induced by the nonlinear price impact, the right panel in Figure 2.1 plots the corresponding liquidity-adjusted price per share for various numbers of an at-the-money, smooth concavified put. The resulting nonlinear prices are compared to their counterparts in a model with linear price impact $p = 2$, keeping all other parameters the same.[16] Clearly, the prices with elasticity $p = 3/2$ are higher only for extremely small trade sizes; for the larger hedging trades necessitated by larger positions, the adjustments with linear costs quickly become substantially bigger.

Endogenous Trading Costs As another example for the scope of Theorem 2.2.3, let us now sketch how it can be used to endogenize the trading cost $(\lambda_t)_{t\in[0.T]}$ in

[15] Such regularity conditions are typical for most models with trading costs, compare [22, 150, 142]. With substantial additional technical effort, the case of a put option is worked out in a model with fixed costs in [60].

[16] In particular, we use the same value of λ as in [80]. Ideally, this scaling parameter of course should be estimated for each elasticity of price impact p from the same dataset, but such estimates do not seem to be available in the literature.

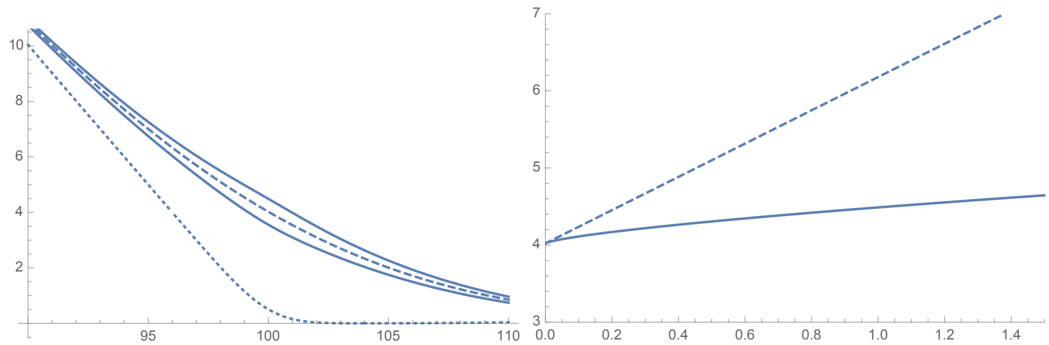

Figure 2.1: Left panel: payoff of the smooth concavified put (dotted), its frictionless Bachelier price (dashed) and its indifference prices for buying and selling one claim when hedging is subject to nonlinear price impact (solid) plotted against the initial price of the risky asset. Right panel: indifference price per claim plotted against the number of smooth, concavified puts traded, with nonlinear price impact (solid) and linear price impact (dashed).

an equilibrium between risk-averse dealers and clients trading according to Theorem 2.2.3. The idea is to choose λ_t so as to allow the (competitive) dealer that earns this cost to break even. If the dealer's risk aversion is small, then this leads to a small equilibrium cost, so that Theorem 2.2.3 is indeed applicable.

Let us now make this precise. To simplify the exposition, suppose the dealer has quadratic holding costs $\gamma_M \sigma^2/2$ per time unit for the position purchased by the client.[17] Then, the dealer's P&L on $[0, t]$ is

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^t \left(-\frac{\gamma_M c_s^S}{2}(\varphi_s^\lambda)^2 + \lambda_s |\dot{\varphi}_s^\lambda|^p\right) ds\right]. \tag{2.2.9}$$

Suppose that γ_M is small and define

$$\lambda_t = \frac{2^{p-1}}{b_p^{\frac{p+2}{2}}} \frac{\gamma_M^{\frac{p+2}{2}}}{\gamma^{\frac{p}{2}}} \frac{c_t^S}{(c_t^{\hat{\varphi}})^p} \hat{\varphi}_t^{p+2}, \tag{2.2.10}$$

[17]If the risky asset has martingale dynamics like in the hedging model discussed above, this is equivalent to assuming that the dealer has local mean-variance preferences as in [102, 136, 65, 66], with constant absolute risk aversion γ_M. In the portfolio choice model studied above, this equivalence remains true if the dealer disagrees with the client about the expected returns of the risky asset, and believes in unpredictable martingale dynamics rather than partial predictability.

where the constant b_p is universal in that it only depends on the elasticity of price impact p but not on the other model and preference parameters:

$$b_p = c_p \frac{\int_{\mathbb{R}} p^{-\frac{p}{p-1}} |\widetilde{g}_p(x)|^p \exp\left(-p^{-\frac{1}{p-1}} \widetilde{g}_p(x)\right) dx}{\int_{\mathbb{R}} \left(x^2 + p^{-\frac{p}{p-1}} |\widetilde{g}_p(x)|^p\right) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{g}_p(x)\right) dx}. \tag{2.2.11}$$

(For linear price impact ($p = 2$), we have $b_p = 1$.) We claim that the dealer's P&L vanishes at the leading order $O(\gamma_M)$ for this choice on all time intervals $[0, t]$. Indeed, for this small trading cost, Theorem 2.2.3 shows that the client's optimal trading strategy is of the form (2.2.4). As a consequence, the first part of the integral in (2.2.9) converges to

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^t -\frac{\gamma_M c_s^S}{2} \hat{\varphi}_s^2 ds\right] + o(\gamma_M), \quad \text{as } \gamma_M \to 0.$$

Similarly, the second part of the integral in (2.2.9) converges to

$$\mathbb{E}_{\mathbb{P}}\left[\int_0^t b_p \lambda_s^{\frac{2}{p+2}} \left(\frac{\gamma c_s^S (c_s^{\hat{\varphi}})^2}{8}\right)^{\frac{p}{p+2}} ds\right] + o(\gamma_M), \quad \text{as } \gamma_M \to 0.$$

Whence, the dealer's P&L (2.2.9) indeed vanishes at the leading order $O(\gamma_M)$ for the choice (2.2.10) of the price impact parameter.

The (asymptotic) equilibrium described above requires the dealer to break even on each time interval $[0, t]$, $t \in [0, T]$. This is akin, to the classical model of Kyle [119] and its descendants, where the break-even condition is also imposed in "each trading round". To obtain a simpler model with a constant trading cost, one can instead only require as in [163] that the dealer's expected risk-adjusted profits are zero over the entire trading interval $[0, T]$. The same arguments as above show that this obtains for

$$\lambda = \frac{2^{p-1}}{b_p^{\frac{p+2}{2}}} \frac{\gamma_M^{\frac{p+2}{2}}}{\gamma^{\frac{p}{2}}} \left(\frac{\mathbb{E}_{\mathbb{P}}\left[\int_0^T c_t^S \hat{\varphi}_t^2 dt\right]}{\mathbb{E}_{\mathbb{P}}\left[\int_0^T (c_t^S (c_t^{\hat{\varphi}})^2)^{\frac{p}{p+2}} dt\right]}\right)^{\frac{p+2}{2}}.$$

If c_t^S and $c_t^{\hat{\varphi}}$ are constant as in the model (2.2.8) with mean-reverting returns uncorrelated to the price shocks ($\rho = 0$), this simplifies to

$$\lambda = \frac{2^{p-1}}{b_p^{\frac{p+2}{2}}} \frac{\gamma_M^{\frac{p+2}{2}}}{\gamma^{\frac{p}{2}}} \frac{c^S}{(c^{\hat{\varphi}})^p} \left(\frac{1}{T} \mathbb{E}_{\mathbb{P}}\left[\int_0^T \hat{\varphi}_t^2 dt\right]\right)^{\frac{p+2}{2}}. \tag{2.2.12}$$

This formula displays the same comparative statics as its “local” version (2.2.10), except for the averaging of the frictionless target position with respect to time and states. As the frictionless optimal strategy $\hat{\varphi}_t = \mu_t/\gamma\sigma^2$ in this model is an Ornstein-Uhlenbeck process with mean $\bar{\mu}/\gamma\sigma^2$ and long-run variance $\sigma_\mu^2/2\kappa\gamma^2\sigma^4$, (2.2.12) becomes

$$\lambda = \frac{2^{p-1}}{b_p^{\frac{p+2}{2}}} \frac{\gamma_M^{\frac{p+2}{2}}}{\gamma^{\frac{p}{2}}} \frac{\sigma^2}{(\sigma_\mu)^{2p}} (\gamma\sigma^2)^{p-2} \times \left(\frac{\sigma_\mu^2}{2\kappa} + \bar{\mu}^2 - \frac{\bar{\mu}^2}{\kappa T}\left(1 - \mathrm{e}^{-\kappa T}\right) + \left(\frac{\bar{\mu}^2}{2\kappa T} - \frac{\sigma_\mu^2}{4\kappa^2 T}\right)\left(1 - \mathrm{e}^{-2\kappa T}\right) \right)^{\frac{p+2}{2}},$$

if $\mu_0 = 0$. For $\bar{\mu} = 0$[18] and trading costs that become quadratic ($p \uparrow 2$), this gives

$$\lambda = 2\frac{\gamma_M^2}{\gamma}\frac{\sigma^2}{4\kappa^2}\left(1 - \frac{1}{2\kappa T}\left(1 - \mathrm{e}^{-2\kappa T}\right)\right)^2.$$

The scaling with the variance of the risky asset is in line with the equilibrium model of Garleanu and Pedersen [66], where risk-averse dealers act as intermediaries between the clients and a group of “exogenous end-users”. (Note, however, that the scaling is typically much more complex for other settings here, e.g., in the hedging model considered above.) Here, there is an additional dependence on the client’s risk aversion and the persistence of her trading signal μ_t. Highly risk-averse client’s with short-lived signals do not take large positions, so that a small trading cost is sufficient to compensate the dealer. In contrast, the other model parameters such as the volatility of the signal do not enter into the above formula.

Performance of Linear Policies The methods developed for the proof of Theorem 2.2.3 can also be used to assess the asymptotic performance of suboptimal strategies of a similar form as (2.2.4). The most obvious candidate is the linear policy

$$\dot{\theta}_t^\lambda = \sqrt{\frac{\gamma c_t^S}{2\lambda_t}}(\hat{\varphi}_t - \theta_t^\lambda), \tag{2.2.13}$$

which is optimal for quadratic trading costs [142]. Arguing along the lines of Section 2.5 one can show that the corresponding deviation $\Delta\theta_t^\lambda = \hat{\varphi}_t - \theta_t^\lambda$ then approximately follows an Ornstein-Uhlenbeck process with mean zero and variance $\sqrt{\frac{\lambda_t}{2\gamma c_t^S}}c_t^{\hat{\varphi}}$

[18] This is the most reasonable case for the dealer to believe in local martingale dynamics.

"locally around t", so that the corresponding certainty-equivalent loss due to displacement from the frictionless target is

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{\gamma c_t^S}{2}(\Delta\theta_t^\lambda)^2\right] = \frac{1}{\sqrt{8}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\sqrt{\gamma c_t^S \lambda_t} c_t^{\hat{\varphi}}\right] + o(\lambda^{1/2}) = O(\lambda^{1/2}).$$

(Note that this term does not depend on the elasticity p of the price impact function.) Likewise, the associated leading-order transaction cost loss can be computed as

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\lambda_t|\dot{\theta}_t^\lambda|^p\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\lambda_t\left|\frac{\gamma c_t^S}{2\lambda_t}\right|^{\frac{p}{2}}|\Delta\theta_t^\lambda|^p\right] = O\left(\lambda_t^{\frac{4-p}{4}}\right).$$

Whence, if the actual price impact has elasticity $p \in (1,2)$, the performance loss for the linear policy (2.2.13) is dominated by the displacement in the small-cost limit, which amounts to half of the total asymptotic loss in [142]. As a consequence, this simpler formula for quadratic costs can serve as a conservative upper bound for other elasticities of the price impact function.

2.3 Assumptions

In order to establish our main result, Theorem 2.2.3, all primitives of the model and their drift and diffusion coefficients need to be continuous:

Assumption 2. *The processes S, $\hat{\varphi}$, Λ, $\mu^{\hat{\varphi}}$, $\mu^{c^{\hat{\varphi}}}$, μ^Λ, μ^m, c^S, $c^{\hat{\varphi}}$, $c^{c^{\hat{\varphi}}}$, c^Λ, c^m, $c^{\hat{\varphi},c^{\hat{\varphi}}}$, $c^{\hat{\varphi},\Lambda}$, $c^{\hat{\varphi},m}$, $c^{\hat{\varphi},S}$, $c^{c^{\hat{\varphi}},\Lambda}$, $c^{c^{\hat{\varphi}},m}$, $c^{c^{\hat{\varphi}},S}$, $c^{\Lambda,m}$, $c^{\Lambda,S}$, $c^{m,S}$ are continuous processes.*

In addition, the primitives of the model need to be integrable enough. For example, it is sufficient that the frictionless target strategy, its drift and diffusion coefficients, the diffusion coefficient of the risky asset, and the trading cost are all uniformly bounded and bounded away from zero. To succinctly formulate the integrability conditions, we introduce the following set:

$$\begin{aligned}\mathcal{X}^\varepsilon =& \Bigg\{ \left(\left((\hat{\varphi}_T^*)^2+1\right)c_T^{S*}\right)^3, (\varphi_T^*)^{\frac{2(1+2\varepsilon)}{(p+2)\kappa_4}}, \left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}, \exp\left(\varepsilon\Lambda_T^*\right), \\ & \exp\left(32\gamma^2\int_0^T(\hat{\varphi}_t^*)^2 c_t^S dt\right), \left(\frac{\left(c^S\right)^{1+\varepsilon}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}, \left(\frac{\left(\Lambda_{\lambda^\zeta u_\xi}\left(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi}\right)^p m^{3p}_{\lambda^\zeta u_\xi}\right)^{1+\varepsilon}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi}}\right)_T^{\frac{3}{2},*}\Bigg\},\end{aligned}$$

for a positive constant $0 < \varepsilon < \frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)} - 3$. Also define the process

$$m_t = \lambda^{\frac{1}{p+2}} m_t^\lambda = \left(\frac{2^{p-1}\gamma\, c_t^S}{\Lambda_t (c_t^{\hat\varphi})^p} \right)^{\frac{1}{p+2}}.$$

Assumption 3. *(i) For some* $0 < \varepsilon < \frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)} - 3$, *we have:* $\mathcal{X}^\varepsilon \subset L^1(\hat{\mathbb{Q}})$.

(ii)

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(8\int_0^T \frac{(\mu_t^{\hat\varphi})^2}{c_t^{\hat\varphi}} dt \right) \right] < \infty.$$

(iii) For all $n \leqslant 8\left\lceil \frac{27}{2\varepsilon} + \frac{9+\varepsilon}{2} \right\rceil$, *it holds*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[m_T^{\frac{4(1+2\varepsilon)}{\kappa_2(p+2)},*} \right] + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{1}{m}\right)_T^{\frac{4(1+2\varepsilon)}{\kappa_2(p+2)},*} \right] < \infty \quad \textit{and} \quad \mathbb{E}_{\hat{\mathbb{Q}}}\left[(m_T^*)^{-n\frac{2+p}{2-p}} \right] < \infty.$$

(iv)

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \Lambda_t^{\frac{2}{p+2}} \left(c_t^S (c_t^{\hat\varphi})^2 \right)^{\frac{p}{p+2}} dt \right] < \infty.$$

Remark 2.3.1. Henceforth, we always assume that Assumptions 1 and 2 are satisfied, but state precisely which parts of Assumption 3 is necessary for each result.

2.4 Properties of the Candidate Strategy

In Section 2.2.1, we defined a family of candidate strategies by means of the stochastic ODE (2.2.4). The function $\widetilde{g}_p$ appearing on its right-hand side has superlinear growth at infinity, so existence of this candidate is not straightforward. Furthermore, we need to make sure that the candidate is indeed admissible, i.e., in $L^2_{\hat{\mathbb{Q}}}(S)$. This is established in the next two subsections.

2.4.1 Existence of the Candidate Strategy

We first prove existence of the candidate strategies defined in Section 2.2.1. We start from the mean-reverting process that will turn out to be the deviation between the frictional trading strategy and the frictionless target.

To establish existence and uniqueness for the SDE that this process should solve, we first shift the mean-reversion level to zero by a change of measure, and then normalize the diffusion coefficient to one by a stochastic time change. Existence and uniqueness can in turn be established by a localization argument, cf. Section 2.7.2.1.

After reverting back to the original time scale and probability measure, this leads to the desired unique solution of the SDE for the displacement process. The corresponding unique solution of the stochastic ODE (2.2.4) is in turn obtained by adding this displacement process to the frictionless optimizer.

To carry out this program, recall that we are working on the probability space $\left(\Omega, \mathcal{F}, \hat{\mathbb{Q}}\right)$, let

$$\zeta = \frac{2}{p+2}.$$

and define $\xi^\lambda_T = \lambda^{-\zeta}\int_0^T c^{\hat\varphi}_s ds$ (the end of the trading interval in the new time scale) and $\xi^{\lambda,m} = \lambda^{-\zeta}\int_0^{\tau^{\lambda,m}} c^{\hat\varphi}_s ds$, where

$$\tau^{\lambda,m} := \inf\left\{t \in [0,T] : m_t < \lambda^{\kappa_2} \text{ or } m_t > \lambda^{-\kappa_2}\ \right\} \wedge T^\lambda. \tag{2.4.1}$$

With this notation, we first make the following standard observation:

Proposition 2.4.1. *The family of stopping times (indexed by λ)*

$$u^\lambda_\xi(\omega) := \begin{cases} \inf\left\{s \in \mathbb{R}_+ : \int_0^s c^{\hat\varphi}_{\lambda^\zeta r}(\omega)dr > \xi\right\} & \xi \leqslant \xi^\lambda_T, \\ u^\lambda_{\xi^\lambda_T} & \xi > \xi^\lambda_T, \end{cases}$$

is strictly increasing in ξ on $[\![0, \xi^\lambda_T]\!]$, forms a stochastic time change, and satisfies $u^\lambda_{\xi^\lambda_T} = \lambda^{-\zeta}T$. Moreover, $\xi \mapsto u^\lambda_\xi(\omega)$ is differentiable with derivative $1/c^{\hat\varphi}_{\lambda^\zeta u^\lambda_\xi}$ on $[\![0, \xi^\lambda_T]\!]$.

Following [95, Lemma 10.18], define

$$\widehat{W}^\lambda_\xi = \int_0^{\lambda^\zeta u^\lambda_\xi} \lambda^{-\frac{\zeta}{2}}\sqrt{c^{\hat\varphi}_s}dW_s \quad \text{for} \quad \xi \geqslant 0. \tag{2.4.2}$$

This is a Brownian motion on $[\![0, \xi^\lambda_T]\!]$ relative to the filtration $\mathbb{G}^\lambda = \left(\mathcal{G}^\lambda_\xi\right)_{\xi\in\mathbb{R}_+}$ with $\mathcal{G}^\lambda_\xi = \mathcal{F}_{\lambda^\zeta u^\lambda_\xi}$ for $\xi \in \mathbb{R}_+$. Since $\mathbb{E}_{\hat{\mathbb{Q}}}[\exp(\frac{1}{2}\int_0^T \frac{(\mu^{\hat\varphi}_t)^2}{c^{\hat\varphi}_t}dt)] < \infty$ by Assumption 3 (ii), Novikov's condition implies that the density

$$\begin{aligned}
\frac{d\widetilde{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} &= \exp\left(-\int_0^{\xi^{\lambda,m}} \lambda^{\frac{\zeta}{2}}\frac{\mu^{\hat\varphi}_{\lambda^\zeta u^\lambda_\xi}}{c^{\hat\varphi}_{\lambda^\zeta u^\lambda_\xi}}d\widehat{W}^\lambda_\xi - \frac{1}{2}\int_0^{\xi^{\lambda,m}} \lambda^\zeta\left(\frac{\mu^{\hat\varphi}_{\lambda^\zeta u^\lambda_\xi}}{c^{\hat\varphi}_{\lambda^\zeta u^\lambda_\xi}}\right)^2 d\xi\right) \\
&= \exp\left(-\int_0^{\tau^{\lambda,m}} \frac{\mu^{\hat\varphi}_t}{\sqrt{c^{\hat\varphi}_t}}dW_t - \frac{1}{2}\int_0^{\tau^{\lambda,m}} \frac{\left(\mu^{\hat\varphi}_t\right)^2}{c^{\hat\varphi}_t}dt\right)
\end{aligned} \tag{2.4.3}$$

defines a probability measure $\widetilde{\mathbb{Q}}^\lambda \approx \hat{\mathbb{Q}}$. Write $\widetilde{W}^\lambda$ for the $\widetilde{\mathbb{Q}}^\lambda$-Brownian motion on $[\![0, \xi_T^\lambda]\!]$, $\widetilde{W}_\xi^\lambda = \widehat{W}_\xi^\lambda + \int_0^{\xi \wedge \xi^{\lambda,m}} \lambda^{\frac{\zeta}{2}} \frac{\mu^{\hat{\varphi}}_{\lambda^\zeta u_y^\lambda}}{c^{\hat{\varphi}}_{\lambda^\zeta u_y^\lambda}} dy$, $\xi \geqslant 0$. With this notation, consider the following SDE under $\widetilde{\mathbb{Q}}^\lambda$,

$$d\widetilde{\Delta\varphi}_\xi^\lambda = -\frac{1}{2} p^{-\frac{1}{p-1}} m_{\lambda^\zeta u_\xi^\lambda} \widetilde{g}_p \left(m_{\lambda^\zeta u_\xi^\lambda} \widetilde{\Delta\varphi}_\xi^\lambda \right) \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\xi + \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\widetilde{W}_\xi^\lambda. \tag{2.4.4}$$

By Proposition 2.7.2, this SDE with initial condition $\widetilde{\Delta\varphi}_0^\lambda = 0$ has a unique strong solution on $\mathbb{R}_+$. The $\hat{\mathbb{Q}}$-dynamics of this process are in turn determined by Girsanov's theorem as

$$\begin{aligned} d\widetilde{\Delta\varphi}_\xi^\lambda = & \left(\lambda^{\frac{\zeta}{2}} \frac{\mu^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}} - \frac{1}{2} p^{-\frac{1}{p-1}} m_{\lambda^\zeta u_\xi^\lambda} \widetilde{g}_p \left(m_{\lambda^\zeta u_\xi^\lambda} \widetilde{\Delta\varphi}_\xi^\lambda \right) \right) \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\xi \\ & + \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\widehat{W}_\xi^\lambda. \end{aligned} \tag{2.4.5}$$

In particular, this SDE has a unique strong solution $\widetilde{\Delta\varphi}^\lambda$ under $\hat{\mathbb{Q}}$. We now change time back to the original time scale: set $t = \lambda^\zeta u_\xi^\lambda$ for $t \in [0, T]$ and define

$$\widehat{\Delta\varphi}_t^\lambda = \lambda^{\frac{\zeta}{2}} \widetilde{\Delta\varphi}^\lambda_{\lambda^{-\zeta} \int_0^{t \wedge \tau^{\lambda,m}} c_s^{\hat{\varphi}} ds}. \tag{2.4.6}$$

On $[0, T]$, this process satisfies

$$d\widehat{\Delta\varphi}_t^\lambda = \left(\mu_t^{\hat{\varphi}} - \frac{1}{2} p^{-\frac{1}{p-1}} \lambda^{-\frac{1}{p+2}} c_t^{\hat{\varphi}} m_t \widetilde{g}_p \left(\lambda^{-\frac{1}{p+2}} m_t \widehat{\Delta\varphi}_t^\lambda \right) \right) \mathbb{1}_{\left\{t \leqslant \tau^{\lambda,m}\right\}} dt \tag{2.4.7}$$

$$+ \sqrt{c_t^{\hat{\varphi}}} \mathbb{1}_{\left\{t \leqslant \tau^{\lambda,m}\right\}} dW_t, \tag{2.4.8}$$

compare [95, Theorem 10.18]. The time change $t = \lambda^\zeta u_\xi^\lambda$ is bijective from $[0, T]$ to $[\![0, \xi_T^\lambda]\!]$; whence, any solution of (2.4.7) is the unique solution of (2.4.5) after this time change. As a consequence, the SDE (2.4.7) has a unique strong solution.

We can now define the candidate strategy by

$$\varphi_t^\lambda = \left(\hat{\varphi}_t - \widehat{\Delta\varphi}_t^\lambda \right) \mathbb{1}_{\{t \leqslant \tau^{\Delta\varphi}\}} - \left(1 - \lambda^{-\eta} \left(t - \tau^{\Delta\varphi}\right)\right) \varphi^\lambda_{\tau^{\Delta\varphi}} \mathbb{1}_{\{\tau^{\Delta\varphi} < t \leqslant \tau^{\Delta\varphi} + \lambda^\eta\}},$$

for $t \in [0, T]$, and the actual displacement of this process as $\Delta\varphi^\lambda = \hat{\varphi} - \varphi^\lambda$. Note that for the stopping time $\tau^{\Delta\varphi}$ from (2.2.7) it holds that $\tau^{\Delta\varphi} \leqslant \tau^{\lambda,m}$, $\hat{\mathbb{Q}}$-a.s., $\Delta\varphi$ and $\widehat{\Delta\varphi}$ coincide on $[\![0, \tau^{\lambda,m}]\!]$ and $\tau^{\Delta\varphi}$ is indeed well defined.

One then easily verifies that this strategy satisfies the stochastic ODE (2.2.4) on $[\![0, \tau^{\Delta\varphi}]\!]$ and also the other specifications described at the end of Section 2.2.1.

2.4.2 Martingale Property of the Candidate Strategy

We now prove that the wealth process generated by the candidate strategy φ^λ is a true martingale under the frictionless dual martingale measure $\hat{\mathbb{Q}}$ and that φ^λ is admissible.

Proposition 2.4.2. *Suppose $((\hat{\varphi}^*)^2 c^S)^*_T \in L^1(\hat{\mathbb{Q}})$ (this is ensured by Assumption 3* (i)*). Then φ^λ belongs to $L^2_{\hat{\mathbb{Q}}}(S)$ and $\int_0^\cdot \varphi^\lambda_t dS_t$ is a true $\hat{\mathbb{Q}}$-martingale.*

Proof. Lemma 2.7.3 and the stated integrability assumptions give

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\varphi^\lambda_t\right)^2 d\langle S\rangle_t\right] \leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\hat{\varphi}^*_t\right)^2 c^S_t dt\right] < \infty. \qquad \square$$

2.5 Primal Considerations

2.5.1 Mean-Variance Tradeoff for the Candidate Strategy

In the proof of the asymptotic expansion from Theorem 2.2.3 (cf. Sections 2.5.2 and 2.6), we will see that a Taylor-expansion of the exponential utility generated by the candidate strategy from Section 2.2.1 is asymptotically equivalent to the following mean-variance tradeoff between squared displacement from the frictionless target and accumulated trading costs:[19]

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\frac{\gamma}{2}\left(\hat{\varphi}_t - \varphi^\lambda_t\right)^2 c^S_t + \lambda_t \left|\dot{\varphi}^\lambda_t\right|^p\right) dt\right] \qquad (2.5.1)\\ &\quad = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\frac{\gamma}{2}(\Delta\varphi^\lambda_t)^2 c^S_t + 2^{-p}p^{-\frac{p}{p-1}}\lambda^\zeta \Lambda_t (c^{\hat{\varphi}}_t)^p m^p_t \left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}} m_t \Delta\varphi^\lambda_t\right)\right|^p\right) dt\right].\end{aligned}$$

We now compute this mean-variance tradeoff; this will also provide crucial bounds for the remainder estimates in the proof of the expansion from Theorem 2.2.3. To this end, we proceed as follows. We first fix $t \in [0, T)$. Then, for λ small enough, $t + \lambda^{\frac{3\zeta}{4}} \leqslant T^\lambda$. In the following we assume that this relation holds. We then estimate (2.5.1) on the small interval $[t, t + \lambda^{\frac{3\zeta}{4}}]$ to the leading order $O(\lambda^\zeta)$. Here, the power $3\zeta/4$ is chosen so that the interval length converges to 0 slower than the time rescaling factor λ^ζ. We then integrate the result over $[0, \tau^{\Delta\varphi} - \lambda^{\frac{3\zeta}{4}}]$ to obtain the value of (2.5.1) on this interval. This approximation is done in probability under $\hat{\mathbb{Q}}$. We then prove a

[19] Similar goal functionals are directly used in a number of studies, cf., e.g., [136, 9]. Related pathwise criteria are studied in [72, 162, 34].

uniform integrability result for the integrated processes to obtain the limit in $L^1(\hat{\mathbb{Q}})$. Finally, we show that the remainder term that accumulates after the liquidation time $\tau^{\Delta\varphi}$ is negligible at the leading order $O(\lambda^\zeta)$.

Local estimation For the estimation on the small interval $[t, t+\lambda^{\frac{3\zeta}{4}}]$, we proceed in three steps. First, we use some rescaling and stopping arguments to bound each part of (2.5.1) by an expression that is easier to estimate because it separates the quickly oscillating displacement from the other "slow" processes. These expressions are in turn estimated using the solutions of some one-dimensional SDEs. In the third step, ergodic properties of these diffusions allow to complete the "local" computation of (2.5.1) on $[t, t+\lambda^{\frac{3\zeta}{4}}]$.

Step 1: Rescaling and stopping. Recall from Section 2.4.1 (Equation (2.4.6)), the definition of the rescaled displacement process $\widetilde{\Delta\varphi}^\lambda_\xi = \lambda^{-\frac{\zeta}{2}}\Delta\varphi^\lambda_{\lambda^\zeta u^\lambda_\xi}$ on $[\![0, \widetilde{\tau}^{\Delta\varphi}]\!]$, where

$$\widetilde{\tau}^{\Delta\varphi} = \lambda^{-\zeta}\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}} dt. \tag{2.5.2}$$

Define additionally the stopping time

$$\begin{aligned}\tau_t^{\lambda,\varepsilon} = \inf\Bigg\{ s \in \left[t, t+\lambda^{\frac{3\zeta}{4}}\right] : \frac{c_s^S}{c_s^{\hat{\varphi}}} \notin \left[\frac{c_t^S(1-\varepsilon)}{c_t^{\hat{\varphi}}}, \frac{c_t^S(1+\varepsilon)}{c_t^{\hat{\varphi}}}\right], \\ \text{or } \frac{\Lambda_s\big(c_s^{\hat{\varphi}}\big)^p m_s^p}{c_s^{\hat{\varphi}}} \notin \left[(1-\varepsilon)\frac{\Lambda_t\big(c_t^{\hat{\varphi}}\big)^p m_t^p}{c_t^{\hat{\varphi}}}, (1+\varepsilon)\frac{\Lambda_t\big(c_t^{\hat{\varphi}}\big)^p m_t^p}{c_t^{\hat{\varphi}}}\right], \\ \text{or } m_s \notin \left[\max\left\{m_t - \varepsilon, \frac{m_t}{2}\right\}, m_t + \varepsilon\right], \\ \text{or } \left|\mu_s^{\hat{\varphi}}\right| > \lambda^{-\frac{\zeta}{8}}, \text{or } c_s^{\hat{\varphi}} < \lambda^{\frac{\zeta}{8}} \Bigg\} \wedge \left(t+\lambda^{\frac{3\zeta}{4}}\right).\end{aligned} \tag{2.5.3}$$

This stopping time ensures that we are working on an interval where the relevant processes stay in small intervals around their value at t, and that the new probability introduced later on is close to $\hat{\mathbb{Q}}$ in term of the total variation distance (see [96, Chapter V.4]).

Let $\varepsilon_1 \in (0, 1/2)$. This constant is arbitrary and will be sent to 0 at the end of this section. By uniform continuity of $c^S_\cdot$, $c^{\hat{\varphi}}_\cdot$, $m_\cdot$ and $\Lambda_\cdot$ on $[0, T]$ there exists a random variable $\lambda^{\varepsilon_1} > 0$ such that for $0 < \lambda \leqslant \lambda^{\varepsilon_1}$ we have $\tau_t^{\lambda,\varepsilon_1} = t + \lambda^{\frac{3\zeta}{4}}$. If we denote by δ_1, δ_2 and δ_3 the respective moduli of uniform continuity of $c^S/c^{\hat\varphi}$, $\Lambda\left(c^{\hat\varphi}\right)^p m^p/c^{\hat\varphi}$ and m, we can take

$$\lambda^{\varepsilon_1} = \min\Bigg\{\delta_1\left(\varepsilon_1 \min_{t\in[0,T]}\frac{c^S_t}{c^{\hat\varphi}_t}\right),\ \delta_2\left(\varepsilon_1 \min_{t\in[0,T]}\frac{\Lambda_t\left(c^{\hat\varphi}_t\right)^p m^p_t}{c^{\hat\varphi}_t}\right),$$
$$\delta_3\left(\min\{\varepsilon_1, \min_{t\in[0,T]}\frac{m_t}{2}\}\right), \left(\max_{t\in[0,T]}\mu^{\hat\varphi}_t\right)^{-8/\zeta}, \left(\min_{t\in[0,T]}c^{\hat\varphi}_t\right)^{8/\zeta}\Bigg\}.$$

Crucially, λ^{ε_1} does not depend on t. For $0<\lambda\leqslant\lambda^{\varepsilon_1}$, it then follows that

$$\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi^\lambda_r\right)^2 c^S_r dr = \lambda^{-\frac{7\zeta}{4}}\int_t^{\tau^{\lambda,\varepsilon_1}_t}\left(\Delta\varphi^\lambda_r\right)^2 c^S_r dr, \tag{2.5.4}$$

and

$$\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\Lambda_r\left(c^{\hat\varphi}_r\right)^p m^p_r\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi^\lambda_r\right)\right|^p dr$$
$$= \lambda^{-\frac{3\zeta}{4}}\int_t^{\tau^{\lambda,\varepsilon_1}_t}\Lambda_r\left(c^{\hat\varphi}_r\right)^p m^p_r\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi^\lambda_r\right)\right|^p dr. \tag{2.5.5}$$

By passing to the rescaled displacement process and introducing the above stopping times, we can separate the quickly-oscillating displacement from the other, more slowly-varying processes in the estimation. To this end, we make the following definitions: the inverse of $\xi\mapsto u^\lambda_\xi$ is denoted by $u^{-1}: s\mapsto\int_0^s c^{\hat\varphi}_{\lambda^\zeta r}dr$, and we denote by $\xi^{\lambda,\varepsilon_1}_t$, $v^{\lambda,t}_\xi$ and $\widetilde{\Delta\varphi}^{\lambda,t}$ the following quantities

$$\begin{aligned}\xi^{\lambda,\varepsilon_1}_t &= u^{-1}(\lambda^{-\zeta}\tau^{\lambda,\varepsilon_1}_t) - u^{-1}(\lambda^{-\zeta}t) = \lambda^{-\zeta}\int_t^{\tau^{\lambda,\varepsilon_1}_t} c^{\hat\varphi}_s ds,\\ v^{\lambda,t}_\xi &= u^\lambda_{\xi+\lambda^{-\zeta}\int_0^t c^{\hat\varphi}_s ds},\\ \widetilde{\Delta\varphi}^{\lambda,t}_\xi &= \widetilde{\Delta\varphi}^\lambda_{\xi+\lambda^{-\zeta}\int_0^t c^{\hat\varphi}_s ds}.\end{aligned}$$

Lemma 2.5.1. *For $0 < \lambda \leqslant \lambda^{\varepsilon_1}$, the following inequalities hold:*

$$\begin{aligned}
&\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1-\varepsilon_1)\,c_t^S\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2d\xi\right)\\
&\qquad\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi_r^\lambda\right)^2c_r^Sdr \qquad (2.5.6)\\
&\qquad\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1+\varepsilon_1)\,c_t^S\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2d\xi\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1-\varepsilon_1)\,\Lambda_t\big(c_t^{\hat{\varphi}}\big)^pm_t^p\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^pd\xi\right)\\
&\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\Lambda_r\big(c_r^{\hat{\varphi}}\big)^pm_r^p\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi_r^\lambda\right)\right|^pdr \qquad (2.5.7)\\
&\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1+\varepsilon_1)\,\Lambda_t\big(c_t^{\hat{\varphi}}\big)^pm_t^p\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^pd\xi\right).
\end{aligned}$$

Proof. For $\lambda \in (0, \lambda^{\varepsilon_1}]$ and $t \leqslant \tau^{\Delta\varphi} - \lambda^{\frac{3\zeta}{4}}$, the change of variables $t = \lambda^\zeta u_\xi^\lambda$ gives

$$\lambda^{-\frac{7\zeta}{4}}\int_t^{\tau_t^{\lambda,\varepsilon_1}}\left(\Delta\varphi_r^\lambda\right)^2c_r^Sdr = \lambda^{\frac{\zeta}{4}}\int_{u^{-1}(\lambda^{-\zeta}t)}^{u^{-1}(\lambda^{-\zeta}\tau_t^{\lambda,\varepsilon_1})}\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2\frac{c^S_{\lambda^\zeta u_\xi^\lambda}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}}d\xi, \qquad (2.5.8)$$

and

$$\begin{aligned}
&\lambda^{-\frac{3\zeta}{4}}\int_t^{\tau_t^{\lambda,\varepsilon_1}}\Lambda_r\left(c_r^{\hat{\varphi}}\right)^pm_r^p\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi_r^\lambda\right)\right|^pdr \qquad (2.5.9)\\
&\qquad= \lambda^{\frac{\zeta}{4}}\int_{u^{-1}(\lambda^{-\zeta}t)}^{u^{-1}(\lambda^{-\zeta}\tau_t^{\lambda,\varepsilon_1})}\frac{\Lambda_{\lambda^\zeta u_\xi^\lambda}\big(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}\big)^pm^p_{\lambda^\zeta u_\xi^\lambda}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}}\left|\widetilde{g}_p\left(m_{\lambda^\zeta u_\xi^\lambda}\widetilde{\Delta\varphi}_\xi^\lambda\right)\right|^pd\xi.
\end{aligned}$$

On the interval $[t, \tau_t^{\lambda,\varepsilon_1}]$ (i.e., $[u^{-1}(\lambda^{-\zeta}t), u^{-1}(\lambda^{-\zeta}\tau_t^{\lambda,\varepsilon_1})]$ in the new time scale),

$$\begin{aligned}
&\frac{c_s^S}{c_s^{\hat{\varphi}}} \in \left[(1-\varepsilon_1)\frac{c_t^S}{c_t^{\hat{\varphi}}}, (1+\varepsilon_1)\frac{c_t^S}{c_t^{\hat{\varphi}}}\right]\\
&\qquad\text{and}\quad \frac{\Lambda_s\left(c_s^{\hat{\varphi}}\right)^pm_s^p}{c_s^{\hat{\varphi}}} \in \left[(1-\varepsilon_1)\frac{\Lambda_t\big(c_t^{\hat{\varphi}}\big)^pm_t^p}{c_t^{\hat{\varphi}}}, (1+\varepsilon_1)\frac{\Lambda_t\big(c_t^{\hat{\varphi}}\big)^pm_t^p}{c_t^{\hat{\varphi}}}\right].
\end{aligned}$$

Next, note that

$$\int_{u^{-1}(\lambda^{-\zeta}t)}^{u^{-1}(\lambda^{-\zeta}\tau_t^{\lambda,\varepsilon_1})} \left(\widetilde{\Delta\varphi}_\xi^{\lambda}\right)^2 d\xi = \int_0^{\xi_t^{\lambda,\varepsilon_1}} \left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi, \tag{2.5.10}$$

and

$$\int_{u^{-1}(\lambda^{-\zeta}t)}^{u^{-1}(\lambda^{-\zeta}\tau_t^{\lambda,\varepsilon_1})} \left|\widetilde{g}_p\left(m_{\lambda^\zeta u_\xi^\lambda}\widetilde{\Delta\varphi}_\xi^{\lambda}\right)\right|^p d\xi = \int_0^{\xi_t^{\lambda,\varepsilon_1}} \left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^p d\xi. \tag{2.5.11}$$

The assertion now follows by putting together the last four equations ((2.5.8), (2.5.9), (2.5.10), and (2.5.11)), the bounds for $c^S/c^{\hat{\varphi}}$, $\Lambda(c^{\hat{\varphi}})^p m^p/c^{\hat{\varphi}}$ on $[t,\tau_t^{\lambda,\varepsilon_1}]$, as well as equations (2.5.4-2.5.5). $\square$

Step 2: Majoring and minoring diffusions. After passing to the rescaled displacement process, we now turn to the estimation of the terms

$$\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi$$

$$\text{and}\qquad \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^p d\xi,$$

that appear as bounds for the mean-variance functional (2.5.1) in Lemma 2.5.1. The process $\widetilde{\Delta\varphi}^{\lambda,t}$ is the rescaled and time-changed displacement started at time $u^{-1}(\lambda^{-\zeta}t) = \int_0^{\lambda^{-\zeta}t} c^{\hat{\varphi}}_{\lambda^\zeta r}dr$ (corresponding to t in the new time scale). Recall from (2.4.5) that under $\hat{\mathbb{Q}}$, the process $\widetilde{\Delta\varphi}^{\lambda,t}$ satisfies the following SDE on $[\![0,\xi_t^{\lambda,\varepsilon_1}]\!]\cap\{t\leqslant \tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}$:

$$d\widetilde{\Delta\varphi}_\xi^{\lambda,t} = \left(\lambda^{\frac{\zeta}{2}}\frac{\mu^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}}{c^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}} - \frac{p^{-\frac{1}{p-1}}}{2} m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right)d\xi + d\widehat{W}_\xi^{\lambda,t},$$

where $\widehat{W}^{\lambda,t}$ is the $\hat{\mathbb{Q}}$-Brownian motion from (2.4.2) restarted at $u^{-1}(\lambda^{-\zeta}t)$. As in (2.4.3), we use a change of measure to shift the mean-reversion level of $\widetilde{\Delta\varphi}^{\lambda,t}$ to zero. Under the measure with density

$$\frac{d\widetilde{\mathbb{Q}}^{\lambda,t}}{d\hat{\mathbb{Q}}} = \exp\left(\left(-\int_0^{\xi_t^{\lambda,\varepsilon_1}} \lambda^{\frac{\zeta}{2}}\frac{\mu^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}}{c^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}} d\widehat{W}_\xi^{\lambda,t} - \frac{1}{2}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \lambda^\zeta \left(\frac{\mu^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}}{c^{\hat{\varphi}}_{\lambda^\zeta v_\xi^{\lambda,t}}}\right)^2 d\xi\right)\mathbb{1}_{\left\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\right\}}\right),$$

the SDE satisfied by the squared displacement $\left(\widetilde{\Delta\varphi}^{\lambda,t}\right)^2$ on $[\![0,\xi_t^{\lambda,\varepsilon_1}]\!]\cap\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}$ is

$$d\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2=\left(1-p^{-\frac{1}{p-1}}m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right)d\xi+2\widetilde{\Delta\varphi}_\xi^{\lambda,t}d\widetilde{W}_\xi^{\lambda,t},\quad(2.5.12)$$

where

$$\widetilde{W}_\xi^{\lambda,t}=\widehat{W}_\xi^{\lambda,t}+\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\int_0^\xi\lambda^{\frac{\zeta}{2}}\frac{\mu^{\hat{\varphi}}_{\lambda^\zeta v_y^t}}{c^{\hat{\varphi}}_{\lambda^\zeta v_y^t}}\mathbb{1}_{\{y\leqslant\xi_t^{\lambda,\varepsilon_1}\}}dy,\quad \xi\geqslant 0.$$

Define the process $B^{\lambda,t}$ by

$$B_\xi^{\lambda,t}=\int_0^\xi\left(\mathbb{1}_{\{y\leqslant\xi_t^{\lambda,\varepsilon_1}\}}\operatorname{sgn}(\widetilde{\Delta\varphi}_y^{\lambda,t})+\mathbb{1}_{\{y>\xi_t^{\lambda,\varepsilon_1}\}}\right)d\widetilde{W}_y^{\lambda,t}\text{ for }\xi\geqslant 0.$$

By Lévy's characterisation [113, Theorem 3.3.16] it is a Brownian motion on $[\![0,\xi_t^{\lambda,\varepsilon_1}]\!]$. As the function $\widetilde{g}_p$ is odd, we can then rewrite (2.5.12) as

$$\begin{aligned}d\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2=&\left(1-p^{-\frac{1}{p-1}}m_{\lambda^\zeta v_\xi^{\lambda,t}}\sqrt{\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2}\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\sqrt{\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2}\right)\right)d\xi\\&+2\sqrt{\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2}dB_\xi^{\lambda,t}.\end{aligned}$$

To establish ergodic theorems for $\left(\widetilde{\Delta\varphi}^{\lambda,t}\right)^2$, we want to sandwich it between the solutions of simpler SDEs that we define now. For an $\mathcal{F}_t$-measurable, positive random variable a with ess inf $a>0$, let $Y^{a,\lambda,\varepsilon_1,+}$, and $Y^{a,\lambda,\varepsilon_1,-}$ be the unique strong solutions of the following two SDE's:

$$\begin{aligned}dY_\xi^{a,\lambda,\varepsilon_1,+}=&\left(1+\varepsilon_1-p^{-\frac{1}{p-1}}b^{a,\varepsilon_1}\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+}}\widetilde{g}_p\left(b^{a,\varepsilon_1}\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+}}\right)\right)d\xi\\&+2\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+}}dB_\xi^{\lambda,t},&&(2.5.13)\\dY_\xi^{a,\lambda,\varepsilon_1,-}=&\left(1-\varepsilon_1-p^{-\frac{1}{p-1}}\left(a+\varepsilon_1\right)\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-}}\widetilde{g}_p\left(\left(a+\varepsilon_1\right)\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-}}\right)\right)d\xi\\&+2\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-}}dB_\xi^{\lambda,t},&&(2.5.14)\\Y_0^{a,\lambda,\varepsilon_1,+}=&Y_0^{a,\lambda,\varepsilon_1,-}=\left(\widetilde{\Delta\varphi}^{\lambda}_{u^{-1}(\lambda^{-\zeta}t)}\right)^2=\left(\widetilde{\Delta\varphi}_0^{\lambda,t}\right)^2=\lambda^{-\zeta}\left(\Delta\varphi_t^\lambda\right)^2,\end{aligned}$$

where $b^{a,\varepsilon_1} = \max\left\{a - \varepsilon_1, \frac{a}{2}\right\}$[20]. In these equations the only time-varying elements are the processes $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,+}$. If a is chosen non-random, we recover a simple SDE with local drifts and volatilities. The local volatility functions for these two processes are the same as the one of $(\widetilde{\Delta\varphi}^{\lambda,t})^2$. In contrast, the drift functionals are chosen such that on $[\![0, \xi_t^{\lambda,\varepsilon_1}]\!] \cap \{t \leqslant \tau^{\Delta\varphi} - \lambda^{\frac{3\zeta}{4}}\}$ the one of $Y^{a,\lambda,\varepsilon_1,+}$ (resp. of $Y^{a,\lambda,\varepsilon_1,-}$) dominates strictly (resp. is strictly dominated by) the one of $(\widetilde{\Delta\varphi}^{\lambda,t})^2$. This allows to use the comparison result of [139] and sandwich $(\widetilde{\Delta\varphi}^{\lambda,t})^2$ between $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$ on $[\![0, \xi_t^{\lambda,\varepsilon_1}]\!] \cap \{t \leqslant \tau^{\Delta\varphi} - \lambda^{\frac{3\zeta}{4}}\}$.

We now want to estimate the integrals from equations (2.5.10) and (2.5.11). We do so by replacing $\widetilde{\Delta\varphi}^{\lambda,t}$ in the integrals by processes of the same form as $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$, where the random variable a is replaced by an elementary random variable, as follows. For $n \in \mathbb{N}\backslash\{0\}$ and an $\mathcal{F}_t$-measurable, positive random variable a with ess inf $a > 0$, let $Y^{a,\lambda,\varepsilon_1,+,n}$ and $Y^{a,\lambda,\varepsilon_1,-,n}$ be the solutions of the following two SDEs with the same initial condition as $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$:

$$
\begin{aligned}
dY_\xi^{a,\lambda,\varepsilon_1,+,n} &= \left(1 + \frac{\varepsilon_1}{n} - p^{-\frac{1}{p-1}} b^{a^{n,-},\varepsilon_1} \sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+,n}}\, \widetilde{g}_p\left(b^{a^{n,-},\varepsilon_1} \sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+,n}}\right)\right) d\xi \\
&\quad + 2\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,+,n}}\, dB_\xi^{\lambda,t},
\end{aligned}
$$

$$
\begin{aligned}
dY_\xi^{a,\lambda,\varepsilon_1,-,n} &= \Bigg(1 - \frac{\varepsilon_1}{n} \\
&\quad - p^{-\frac{1}{p-1}} \left(a^{n,+} + \varepsilon_1\right) \sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-,n}}\, \widetilde{g}_p\Big(\left(a^{n,+} + \varepsilon_1\right) \sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-,n}}\Big)\Bigg) d\xi \\
&\quad + 2\sqrt{Y_\xi^{a,\lambda,\varepsilon_1,-,n}}\, dB_\xi^{\lambda,t},
\end{aligned}
$$

where

$$
\begin{aligned}
a^{n,+} &= \sum_{k=0}^{\infty} (k+1) 2^{-n} \mathbb{1}_{\{k2^{-n} < a \leqslant (k+1)2^{-n}\}}, \quad \text{and} \\
a^{n,-} &= \sum_{k=1}^{\infty} k 2^{-n} \mathbb{1}_{\{k2^{-n} \leqslant a < (k+1)2^{-n}\}} + \text{ess inf } a \mathbb{1}_{\{\text{ess inf } a \leqslant a < 2^{-n}\}}.
\end{aligned}
$$

Note that on each of the sets $\{k2^{-n} < a \leqslant (k+1)2^{-n}\}$ and on $\{\text{ess inf } a \leqslant a < 2^{-n}\}$, the SDEs satisfied by $Y^{a,\lambda,\varepsilon_1,+,n}$ and $Y^{a,\lambda,\varepsilon_1,-,n}$ are (2.5.13)

[20] Existence and uniqueness follow in the same way as in Section 2.7.2.1. Here, as in the proof of Proposition 2.7.2, we extend the probability space if necessary for $B^{\lambda,t}$ to be a Brownian motion on $\mathbb{R}_+$ and for the Y processes to be defined on $\mathbb{R}_+$.

and (2.5.14) with ε_1 replaced by $\frac{\varepsilon_1}{n}$ and a chosen constant. We will therefore prove the needed ergodic result first for these two SDEs for constant a and use a monotonicity argument to transfer the result to $Y^{a,\lambda,\varepsilon_1,+,n}$ and $Y^{a,\lambda,\varepsilon_1,-,n}$.

Indeed, by a comparison result for SDEs [139, Corollary of Theorem 2], we have $\widetilde{\mathbb{Q}}^{\lambda,t}$-a.s. on $[\![0,\xi_t^{\lambda,\varepsilon_1}]\!] \cap \{t \leqslant \tau^{\Delta\varphi} - \lambda^{\frac{3\zeta}{4}}\}$ for all $n \in \mathbb{N}\backslash\{0\}$,

$$Y_\xi^{m_t,\lambda,\varepsilon_1,-,n} \leqslant Y_\xi^{m_t,\lambda,\varepsilon_1,-,n+1} \leqslant \left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 \leqslant Y_\xi^{m_t,\lambda,\varepsilon_1,+,n+1} \leqslant Y_\xi^{m_t,\lambda,\varepsilon_1,+,n}.$$

For $n \in \mathbb{N}\backslash\{0\}$, $\varepsilon_1 \in (0,1)$ and $\lambda \in (0,\lambda^{\varepsilon_1})$, it therefore follows that

$$\begin{aligned}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} Y_\xi^{m_t,\lambda,\varepsilon_1,-,n}d\xi &\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi \\ &\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} Y_\xi^{m_t,\lambda,\varepsilon_1,+,n}d\xi\end{aligned} \tag{2.5.15}$$

and

$$\begin{aligned}&\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left|\widetilde{g}_p\left(b^{m_t^{n,-},\varepsilon_1}\sqrt{Y_\xi^{m_t,\lambda,\varepsilon_1,-,n}}\right)\right|^p d\xi \\ &\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^p d\xi \\ &\leqslant \mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}} \left|\widetilde{g}_p\left(\left(m_t^{n,+}+\varepsilon_1\right)\sqrt{Y_\xi^{m_t,\lambda,\varepsilon_1,+,n}}\right)\right|^p d\xi.\end{aligned} \tag{2.5.16}$$

Now, on any of the subsets of Ω appearing in the definition of $a^{n,+}$ and $a^{n,-}$ with positive measure under $\hat{\mathbb{Q}}$, the SDE satisfied by $Y^{a,\lambda,\varepsilon_1,+,n}$, and $Y^{a,\lambda,\varepsilon_1,-,n}$ have non-random drift and volatility, and we can use ergodic theorems to estimate the integrals.

Step 3: Computation by ergodic theorems. We now compute the integrals in the sandwiches (2.5.15-2.5.16) by means of ergodic theorems and the invariant distributions of one-dimensional diffusions. To this end, we first make the following observation:

Lemma 2.5.2. *Let $a > 0$ be a real number. Then, $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$ are regular, recurrent diffusions. Their speed measures are finite and have the following densities:*

$$\nu^{a,\lambda,\varepsilon_1,+}(x) = \frac{1}{2} y^{-\frac{1-\varepsilon_1}{2}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{y}\right) - p^{-\frac{1}{p-1}} \widetilde{G}_p\left(b^{a,\varepsilon_1}\right)\right) \text{ and}$$
$$\nu^{a,\lambda,\varepsilon_1,-}(x) = \frac{1}{2} y^{-\frac{1+\varepsilon_1}{2}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p\left((a+\varepsilon_1)\sqrt{y}\right) - p^{-\frac{1}{p-1}} \widetilde{G}_p\left(a+\varepsilon_1\right)\right).$$

Proof. See Section 2.7.3. □

In view of Lemma 2.5.2, the ergodic theorem as in [101, Theorem 23.14] is applicable.

Lemma 2.5.3. *Let $a > 0$ be a real number, $X \in \left\{Y^{a,\lambda,\varepsilon_1,+}, Y^{a,\lambda,\varepsilon_1,-}\right\}$, $\nu(x)$ the density of the associated speed measure, and f a measurable function such that*

$$\int_{\mathbb{R}} f(x)\nu(x)dx < \infty.$$

Then,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s)ds = \frac{\int_{\mathbb{R}} f(x)\nu(x)dx}{\int_{\mathbb{R}} \nu(x)dx}, \quad \widetilde{\mathbb{Q}}^{\lambda,t}\text{-a.s.}$$

Lemma 2.5.4. *Let $b > 0$ be a real number. The functions $x \mapsto x$ and $x \mapsto |\widetilde{g}_p(b\sqrt{x})|^p$ are integrable with respect to the two speed measures from Lemma 2.5.2. Furthermore, the ergodic limits for the processes $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$ as t goes to ∞ converge to the same value as ε_1 goes to 0,*

$$\begin{aligned}
\lim_{\varepsilon_1\to 0}\lim_{t\to\infty} \frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,+} ds &= \lim_{\varepsilon_1\to 0}\lim_{t\to\infty} \frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,-} ds \\
&= \frac{1}{a^2} \frac{\int_{\mathbb{R}} x^2 \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}{\int_{\mathbb{R}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx} \\
&:= v(a), \\
\lim_{\varepsilon_1\to 0}\lim_{t\to\infty} \frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,-}}\right)\right|^p ds &= \lim_{\varepsilon_1\to 0}\lim_{t\to\infty} \frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,+}}\right)\right|^p ds \\
&= \frac{\int_{\mathbb{R}} \left|\widetilde{g}_p(\frac{b}{a}x)\right|^p \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}{\int_{\mathbb{R}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}.
\end{aligned}$$

Proof. See Section 2.7.3. □

We now extend the limits from Lemma 2.5.4 from constants a to random variables that take countably many values and are measurable with respect to the initial σ-field. Recall that we are working on the filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{G}_\xi^{\lambda,t})_{\xi\geqslant 0}, \widetilde{\mathbb{Q}}^{\lambda,t}\right)$ equipped with the $\widetilde{\mathbb{Q}}^{\lambda,t}$-Brownian motion $B^{\lambda,t}$, where $\mathcal{G}_\xi^{\lambda,t} = \mathcal{G}^\lambda_{\xi+\lambda^{-\zeta}\int_0^t c_s^{\hat{\varphi}}ds}$. Let $(A_n)_{n\in\mathbb{N}}$ be a partition of Ω into sets of $\mathcal{G}_0^{\lambda,t}$ with positive measure under $\widetilde{\mathbb{Q}}^{\lambda,t}$. For positive constants a_k, b_k, $k\in\mathbb{N}$, define the following $\mathcal{G}_0^{\lambda,t}$-measurable random variables a and b,

$$a = \sum_{k=0}^\infty a_k \mathbb{1}_{A_k}, \quad \text{and} \quad b = \sum_{k=0}^\infty b_k \mathbb{1}_{A_k}.$$

Lemma 2.5.5. *The $\widetilde{\mathbb{Q}}^{\lambda,t}$-a.s. limits from Lemma 2.5.4 also hold for a and b as above:*

$$\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,+}ds = \lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,-}ds = v(a),$$

$$\begin{aligned}\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,-}}\right)\right|^p ds &= \lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,+}}\right)\right|^p ds\\ &= \frac{\int_{\mathbb{R}}\left|\widetilde{g}_p(\frac{b}{a}x)\right|^p \exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}.\end{aligned}$$

Proof. The Brownian motion $B^{\lambda,t}$ is independent from $\mathcal{G}_0^{\lambda,t}$, so it is a Brownian motion as well on any of the spaces $\left(A_k, \mathcal{F}^{(k)}, \left(\mathcal{G}_\xi^{\lambda,t,k}\right)_{\xi\geqslant 0}, \widetilde{\mathbb{Q}}^{\lambda,t}\right)$, where $\mathcal{F}^{(k)} = \{B\cap A_k \mid B\in\mathcal{F}\}$ and $\mathcal{G}_\xi^{\lambda,t,k} = \left\{B\cap A_k \mid B\in\mathcal{G}_\xi^{\lambda,t}\right\}$. On each of these spaces, we apply Lemma 2.5.4 and obtain the result. □

To transfer the above ergodic limits from the auxiliary measure $\widetilde{\mathbb{Q}}^{\lambda,t}$ to the frictionless dual martingale measure $\hat{\mathbb{Q}}$ appearing in the mean-variance tradeoff (2.5.1), we establish the following result:

Lemma 2.5.6. *For every $t\in[0,T)$, we have convergence in total variation of $\left(\widetilde{\mathbb{Q}}^{\lambda,t}\right)_{\lambda\in(0,1]}$ to $\hat{\mathbb{Q}}$:*

$$\lim_{\lambda\to 0}\sup_{A\in\mathcal{F}}\left|\widetilde{\mathbb{Q}}^{\lambda,t}(A) - \hat{\mathbb{Q}}(A)\right| = 0. \tag{2.5.17}$$

Proof. By [96, Theorem IV.1.33], the measure $\widetilde{\mathbb{Q}}^{\lambda,t}$ on $(\Omega,\mathcal{F})$ has the following Hellinger process with respect to $\hat{\mathbb{Q}}$:

$$h\left(\frac{1}{2},\mathbb{Q}^{\lambda,t},\hat{\mathbb{Q}}\right)_s=\frac{1}{8}\int_t^{s\wedge\tau_t^{\lambda,\varepsilon_1}}\frac{\left(\mu_u^{\hat{\varphi}}\right)^2}{c_u^{\hat{\varphi}}}du\mathbb{1}_{\{t\leqslant s\leqslant T\}}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}.$$

By definition of the stopping time $\tau_t^{\lambda,\varepsilon_1}$ in (2.5.3), the trajectory $\frac{(\mu_\cdot^{\hat{\varphi}})^2}{c_\cdot^{\hat{\varphi}}}\mathbb{1}_{[\![0,\tau_t^{\lambda,\varepsilon_1}]\!]}$ is uniformly bounded by $\lambda^{-\frac{3\zeta}{8}}$on [0,T]. Likewise, by definition of $\tau_t^{\lambda,\varepsilon_1}$,

$$\lim_{\lambda\to 0}h\left(\frac{1}{2},\mathbb{Q}^{\lambda,t},\hat{\mathbb{Q}}\right)_{\tau_t^{\lambda,\varepsilon_1}}=0\quad\hat{\mathbb{Q}}\text{-a.s.}$$

The assertion therefore follows from [96, Theorem V.4.31 (i) and V.4.3]. □

By combining Lemmas 2.5.5 and 2.5.6 with the bounds (2.5.15) and (2.5.16), we can now compute both components of the cost-displacement tradeoff (2.5.1):

Proposition 2.5.7. *The following limits hold in probability under* $\widetilde{\mathbb{Q}}$*:*

$$\begin{aligned}
&\lim_{\lambda\to 0}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2d\xi\\
&\qquad=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{m_t^2}\frac{\int_{\mathbb{R}}x^2\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}\\
&\qquad=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}v(m_t),\\
&\lim_{\lambda\to 0}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^pd\xi\\
&\qquad=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{\int_{\mathbb{R}}|\widetilde{g}_p(x)|^p\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p(x)\right)dx}\\
&\qquad=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}w.
\end{aligned}$$

Recall that by Lemma 2.5.1, these integrals provide estimates for the functional (2.5.1) on the small interval $[t,(t+\lambda^{\frac{3\zeta}{4}})]$ for $t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}$.

Proof of Proposition 2.5.7. Lemma 2.5.5 applied to $Y^{m_t,\lambda,\varepsilon_1,-,n}$ and $Y^{m_t,\lambda,\varepsilon_1,+,n}$ (replacing ε_1 by $\frac{\varepsilon_1}{n}$ for the drift in the computations) yields the following $\widetilde{\mathbb{Q}}^{\lambda,t}$-a.s. (and therefore $\hat{\mathbb{Q}}$-a.s.) limits:

$$
\begin{aligned}
&\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{m_t,\lambda,\varepsilon_1,+,n}ds = v(m_t^{n,-}),\\
&\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{m_t,\lambda,\varepsilon_1,-,n}ds = v(m_t^{n,+}),\\
&\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b^{m_t^{n,-},\varepsilon_1}\sqrt{Y_s^{m_t,\lambda,\varepsilon_1,-,n}}\right)\right|^p ds\\
&\qquad\qquad\qquad\qquad = \frac{\int_{\mathbb{R}}\left|\widetilde{g}_p(\frac{m^{n,-}}{m^{n,+}}x)\right|^p \exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx},\\
&\lim_{\varepsilon_1\to 0}\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(\left(m_t^{n,+}+\varepsilon_1\right)\sqrt{Y_s^{m_t,\lambda,\varepsilon_1,+,n}}\right)\right|^p ds\\
&\qquad\qquad\qquad\qquad = \frac{\int_{\mathbb{R}}\left|\widetilde{g}_p(\frac{m^{n,+}}{m^{n,-}}x)\right|^p \exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}.
\end{aligned}
$$

Evidently, $\lim_{n\to\infty} v(m_t^{n,-}) = \lim_{n\to\infty} v(m_t^{n,+}) = v(m_t)$, $\widetilde{\mathbb{Q}}^{\lambda,t}$-a.s (and therefore $\hat{\mathbb{Q}}$-a.s.). By the dominated convergence theorem, the other two sequences (denote them by $w(m,n,-)$ and $w(m,n,+)$, respectively) converge $\widetilde{\mathbb{Q}}^{\lambda,t}$-a.s. to w (and therefore $\hat{\mathbb{Q}}$-a.s.). Observe that as λ goes to 0, by definition of $\tau_t^{\lambda,\varepsilon_1}$ (see Equation (2.5.3)) and for $0<\lambda\leqslant\lambda^{\varepsilon_1}$, $\varepsilon_1\in(0,1/2)$,

$$
\begin{aligned}
\frac{1}{2}\lambda^{-\frac{\zeta}{4}}\min_{s\in[0,T]}c_s^{\hat{\varphi}} \leqslant (1-\varepsilon_1)c_t^{\hat{\varphi}}\lambda^{-\frac{\zeta}{4}} \leqslant \xi_t^{\lambda,\varepsilon_1} = \lambda^{-\zeta}\int_t^{\tau_t^{\lambda,\varepsilon_1}} c_r^{\hat{\varphi}}dr \qquad (2.5.18)\\
\leqslant (1+\varepsilon_1)c_t^{\hat{\varphi}}\lambda^{-\frac{\zeta}{4}} \leqslant \frac{3}{2}\lambda^{-\frac{\zeta}{4}}\max_{s\in[0,T]}c_s^{\hat{\varphi}}.
\end{aligned}
$$

Then, $\lim_{\lambda\to 0}\xi_t^{\lambda,\varepsilon_1} = +\infty$, $\hat{\mathbb{Q}}$-a.s., uniformly in t and ε_1. Together with the inequalities (2.5.15) and (2.5.16), this yields the following limits in probability with respect to $\hat{\mathbb{Q}}$:

$$\lim_{\varepsilon_1\to 0}\lim_{\lambda\to 0}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}v(m_t),$$

$$\lim_{\varepsilon_1\to 0}\lim_{\lambda\to 0}\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left|\widetilde{g}_p\left(m_{\lambda^\zeta v_\xi^{\lambda,t}}\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)\right|^p d\xi=\mathbb{1}_{\{t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}w.$$

Concatenation of the local estimates Recall that by equations (2.5.6) and (2.5.7) we have

$$\begin{aligned}0\leqslant\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1-\varepsilon_1)c_t^S\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}&\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi\right)\\&\leqslant\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi_r^\lambda\right)^2 c_r^S dr.\end{aligned}$$

In view of Lemma 2.5.8, the right-hand side (as a family indexed by λ) is uniformly integrable with respect to $\hat{\mathbb{Q}}\times\mathrm{Leb}_{|[0,T]}$. Furthermore, the right-hand side of (2.5.6) can also be bounded from above by

$$0\leqslant\frac{1+\varepsilon_1}{1-\varepsilon_1}\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi_r^\lambda\right)^2 c_r^S dr,$$

which is also uniformly integrable with respect to $\hat{\mathbb{Q}}\times\mathrm{Leb}_{|[0,T]}$. Now, Proposition 2.5.7, estimate (2.5.18) and Fubini's theorem give the following limits as λ goes to 0,

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1-\varepsilon_1)\,c_t^S\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi\right)dt\right]&\\\longrightarrow(1-\varepsilon_1)\,\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}c_t^S v(m_t)dt\right]&,\\\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}(1+\varepsilon_1)\,c_t^S\frac{\xi_t^{\lambda,\varepsilon_1}}{\lambda^{-\zeta}c_t^{\hat{\varphi}}\lambda^{\frac{3\zeta}{4}}}\left(\frac{1}{\xi_t^{\lambda,\varepsilon_1}}\int_0^{\xi_t^{\lambda,\varepsilon_1}}\left(\widetilde{\Delta\varphi}_\xi^{\lambda,t}\right)^2 d\xi\right)dt\right]&\\\longrightarrow(1+\varepsilon_1)\,\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}c_t^S v(m_t)dt\right]&.\end{aligned}$$

This, (2.5.6), and sending ε_1 to 0 gives

$$\lim_{\lambda\to 0}\left|\lambda^{-\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\left(\Delta\varphi_t^\lambda\right)^2 c_t^S dt\right]-\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} c_t^S v(m_t)dt\right]\right|=0. \tag{2.5.19}$$

Repeating the exact same reasoning for (2.5.7) yields

$$\lim_{\lambda\to 0}\left|\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\Lambda_t\big(c_t^{\hat{\varphi}}\big)^p m_t^p\left|\widetilde{g}_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)\right|^p dt\right]\right. \tag{2.5.20}$$
$$\left.-\,\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\Lambda_t\big(c_t^{\hat{\varphi}}\big)^p m_t^p w dt\right]\right|=0.$$

Uniform integrability and convergence in $L^1(\hat{\mathbb{Q}})$. To upgrade convergence in probability to convergence in $L^1(\hat{\mathbb{Q}})$ for the computation of the cost-displacement tradeoff (2.5.1), it remains to establish uniform integrability:

Lemma 2.5.8. *Suppose that Assumption 3*(iii) *holds, $\mathcal{X}^\varepsilon \subset L^1(\hat{\mathbb{Q}})$, and $x\mapsto\hat{\mathbb{Q}}\left[x\leqslant\int_0^T c_t^{\hat{\varphi}}dt\right]$ decays faster than $x^{-3-\delta}$ at infinity for some $\delta>0$ (Assumption 3*(i)*). Then:*

$$\sup_{0<\lambda\leqslant 1}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}}\left(\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi_r^\lambda\right)^2 c_r^S dr\right)^{1+\delta}dt\right]<\infty,$$
$$\sup_{0<\lambda\leqslant 1}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}}\left(\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi_r^\lambda\right)\right|^p dr\right)^{1+\delta}dt\right]<\infty.$$

In particular, the families

$$\left(\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\left(\Delta\varphi_r^\lambda\right)^2 c_r^S dr\right)_{0<\lambda\leqslant 1}\quad\text{and}$$
$$\left(\mathbb{1}_{\{0\leqslant t\leqslant\tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi_r^\lambda\right)\right|^p dr\right)_{0<\lambda\leqslant 1}$$

are uniformly integrable with respect to $\hat{\mathbb{Q}}\times Leb_{|[0,T]}$.

Note that the second assumption is satisfied by Markov's inequality if

$$\left(\int_0^T c_t^{\hat{\varphi}} dt\right)^{3+\delta} \in L^1(\hat{\mathbb{Q}}) \text{ for some } \delta > 0.$$

This follows from $\mathcal{X}^\varepsilon \subset L^1(\hat{\mathbb{Q}})$ and the definition of κ_1 and κ_3 for δ small enough. We now fix such a δ, additionally assumed to satisfy $\delta < 2\varepsilon/9$.

Proof of Lemma 2.5.8. First, note that Jensen's inequality, Fubini's theorem and $\tau^{\Delta\varphi} \leqslant T$ give

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\mathbb{1}_{\{0\leqslant t\leqslant \tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{7\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}} \left(\Delta\varphi_r^\lambda\right)^2 c_r^S dr\right)^{1+\delta} dt\right] \\ \leqslant C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\lambda^{-\frac{\zeta}{2}}\Delta\varphi_r^\lambda\right)^{2(1+\delta)} \left(c_r^S\right)^{1+\delta} dr\right].\end{aligned} \tag{2.5.21}$$

We compute, first using the time change $r = \lambda^\zeta u_\xi^\lambda$ (see Proposition 2.4.1, and recall from Equation (2.5.2) that $\widetilde{\tau}^{\Delta\varphi} = \lambda^{-\zeta}\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}} dt$), then Hölder and Jensen's inequality, and finally Fubini's theorem,

$$\begin{aligned}&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\lambda^{-\frac{\zeta}{2}}\Delta\varphi_r^\lambda\right)^{2(1+\delta)} \left(c_r^S\right)^{1+\delta} dr\right] \\ &= \lambda^\zeta\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\widetilde{\tau}^{\Delta\varphi}} \frac{\left(c^S_{\lambda^\zeta u_\xi^\lambda}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}} \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{2(1+\delta)} d\xi\right] \\ &\leqslant \lambda^\zeta\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(c^S\right)^{1+\delta}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\widetilde{\tau}^{\Delta\varphi}} \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{2(1+\delta)} d\xi\right)^3\right]^{\frac{1}{3}} \\ &\leqslant \lambda^\zeta\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(c^S\right)^{1+\delta}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^2 \int_0^{\widetilde{\tau}^{\Delta\varphi}} \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{6(1+\delta)} d\xi\right]^{\frac{1}{3}} \\ &\leqslant \lambda^\zeta\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(c^S\right)^{1+\delta}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}\right]^{\frac{2}{3}} \left(\int_0^\infty \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^2 \mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}} \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{6(1+\delta)}\right] d\xi\right)^{\frac{1}{3}}.\end{aligned}$$

Now apply Hölder's inequality twice (first with 3, second with $1+\delta' > 1$), obtaining

$$\begin{aligned} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^2 \mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi_\xi^\lambda}\right)^{6(1+\delta)}\right] \leqslant & \hat{\mathbb{Q}}\left[\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\right]^{\frac{1}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^{3(1+\delta')}\right]^{\frac{2}{3(1+\delta')}} \\ & \times \mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi_\xi^\lambda}\right)^{9(1+\delta)\frac{1+\delta'}{\delta'}}\right]^{\frac{2\delta'}{3(1+\delta')}}. \end{aligned} \tag{2.5.22}$$

With $\widetilde{\tau}^{\Delta\varphi} = \lambda^{-\zeta}\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}}\,dt$, it in turn follows that

$$\begin{aligned} &\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\left(\lambda^{-\frac{\zeta}{2}}\Delta\varphi_r^\lambda\right)^{2(1+\delta)}\left(c_r^S\right)^{1+\delta}dr\right] \\ &\quad\leqslant \lambda^{\frac{\zeta}{3}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(c^S\right)^{1+\delta}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}\right]^{\frac{2}{3}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}}\,dt\right)^{3(1+\delta')}\right]^{\frac{2}{9(1+\delta')}} \\ &\quad\times \sup_{\xi\in\mathbb{R}_+}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi_\xi^\lambda}\right)^{9(1+\delta)\frac{1+\delta'}{\delta'}}\right]^{\frac{2\delta'}{9(1+\delta')}}\left(\int_0^\infty\hat{\mathbb{Q}}\left[\xi\leqslant\lambda^{-\zeta}\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}}\,dt\right]^{\frac{1}{3}}d\xi\right)^{\frac{1}{3}} \\ &\quad\leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(c^S\right)^{1+\delta}}{c^{\hat{\varphi}}}\right)_T^{\frac{3}{2},*}\right]^{\frac{2}{3}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{T} c_t^{\hat{\varphi}}\,dt\right)^{3(1+\delta')}\right]^{\frac{2}{9(1+\delta')}} \\ &\quad\times \sup_{\xi\in\mathbb{R}_+}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi_\xi^\lambda}\right)^{9(1+\delta)\frac{1+\delta'}{\delta'}}\right]^{\frac{2\delta'}{9(1+\delta')}}\left(\int_0^\infty\hat{\mathbb{Q}}\left[\xi\leqslant\int_0^{\tau^{\Delta\varphi}} c_t^{\hat{\varphi}}\,dt\right]^{\frac{1}{3}}d\xi\right)^{\frac{1}{3}}. \end{aligned}$$

The parameters $\delta' > 0$ was arbitrary and we can choose it such that $3(1+\delta') < 3+\varepsilon$ (recall as well that $3\delta/2 < \varepsilon$), and such that $9(1+\delta)\frac{1+\delta'}{\delta'} \leqslant 27/\varepsilon + 9 + \varepsilon$. Then, by the integrability assumptions, the first and second term of the product are finite. Lemma 2.7.11 shows that this also holds for the third term. Finally, the growth condition for $x \mapsto \hat{\mathbb{Q}}\left[x \leqslant \int_0^T c_t^{\hat{\varphi}}\,dt\right]$ ensures that the last term is finite as well.

The argument for the second family is similar. First, observe that Jensen's inequality and Fubini's theorem give

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\mathbb{1}_{\{0\leqslant t\leqslant \tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}} \Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p \left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}} m_r \Delta\varphi_r^\lambda\right)\right|^p dr\right)^{1+\delta} dt\right]$$
$$\leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p\right)^{1+\delta} \left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}} m_r \Delta\varphi_r^\lambda\right)\right|^{p(1+\delta)} dr\right].$$

Now, a change of variables, Corollary 2.7.1, and the elementary inequality $(a+b)^{1+\delta} \leqslant C(a^{1+\delta}+b^{1+\delta})$ (for $a, b \geqslant 0$ and some $C > 0$) yield

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\mathbb{1}_{\{0\leqslant t\leqslant \tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}} \Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p \left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}} m_r \Delta\varphi_r^\lambda\right)\right|^p dr\right)^{1+\delta} dt\right]$$
$$\leqslant \lambda^\zeta \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\widetilde{\tau}^{\Delta\varphi}} \frac{\left(\Lambda_{\lambda^\zeta u_\xi^\lambda}\left(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}\right)^p m^p_{\lambda^\zeta u_\xi^\lambda}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}} \left|\widetilde{g}_p\left(m_{\lambda^\zeta u_\xi^\lambda}\widetilde{\Delta\varphi}_\xi^\lambda\right)\right|^{p(1+\delta)} d\xi\right]$$
$$\leqslant \lambda^\zeta \mathbb{E}_{\hat{\mathbb{Q}}} C\left[\int_0^{\widetilde{\tau}^{\Delta\varphi}} \frac{\left(\Lambda_{\lambda^\zeta u_\xi^\lambda}\left(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}\right)^p m^{3p}_{\lambda^\zeta u_\xi^\lambda}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}} \left|\widetilde{\Delta\varphi}_\xi^\lambda\right|^{2p(1+\delta)} d\xi\right]$$
$$+ \lambda^\zeta C \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\widetilde{\tau}^{\Delta\varphi}} \frac{\left(\Lambda_{\lambda^\zeta u_\xi^\lambda}\left(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}\right)^p m^{p}_{\lambda^\zeta u_\xi^\lambda}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}} d\xi\right].$$

The second term of the last inequality is readily bounded using Hölder's inequality by

$$C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(\Lambda_{\lambda^\zeta u_\xi^\lambda}\left(c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}\right)^{p-1} m^{p}_{\lambda^\zeta u_\xi^\lambda}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^\zeta u_\xi^\lambda}}\right)_T^{*,\frac{3}{2}}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T c_t^{\hat{\varphi}} dt\right)^3\right]^{\frac{1}{3}},$$

which is finite by the integrability assumptions on $\mathcal{X}^\varepsilon$, assuming $3\delta < 2\varepsilon$. Then it remains to bound

$$
\begin{aligned}
&\lambda^{\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\widetilde{\tau}^{\Delta\varphi}} \frac{\left(\Lambda_{\lambda^{\zeta}u_{\xi}^{\lambda}}\left(c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^p m^{3p}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}}\left|\widetilde{\Delta\varphi}_{\xi}^{\lambda}\right|^{2p(1+\delta)} d\xi\right] \\
&\leqslant \lambda^{\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(\Lambda_{\lambda^{\zeta}u_{\xi}^{\lambda}}\left(c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^p m^{3p}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}}\right)_T^{*,\frac{3}{2}}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\widetilde{\tau}^{\Delta\varphi}}\left|\widetilde{\Delta\varphi}_{\xi}^{\lambda}\right|^{2p(1+\delta)} d\xi\right)^3\right]^{\frac{1}{3}} \\
&\leqslant \lambda^{\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(\Lambda_{\lambda^{\zeta}u_{\xi}^{\lambda}}\left(c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^p m^{3p}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}}\right)_T^{*,\frac{3}{2}}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^2\int_0^{\widetilde{\tau}^{\Delta\varphi}}\left|\widetilde{\Delta\varphi}_{\xi}^{\lambda}\right|^{6p(1+\delta)} d\xi\right]^{\frac{1}{3}} \\
&\leqslant \lambda^{\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(\Lambda_{\lambda^{\zeta}u_{\xi}^{\lambda}}\left(c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^p m^{3p}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}}\right)_T^{*,\frac{3}{2}}\right]^{\frac{2}{3}} \\
&\qquad\qquad \times\left(\int_0^{\infty}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\widetilde{\tau}^{\Delta\varphi}\right)^2\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi}_{\xi}^{\lambda}\right)^{6p(1+\delta)}\right] d\xi\right)^{\frac{1}{3}},
\end{aligned}
$$

where we have used successively Hölder's inequality, Jensen's inequality, and Fubini's theorem. Together with Estimate (2.5.22) and the computation that follows it, we obtain

$$
\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T\left(\mathbb{1}_{\{0\leqslant t\leqslant \tau^{\Delta\varphi}-\lambda^{\frac{3\zeta}{4}}\}}\lambda^{-\frac{3\zeta}{4}}\int_t^{t+\lambda^{\frac{3\zeta}{4}}}\Lambda_r\left(c_r^{\hat{\varphi}}\right)^p m_r^p\left|\widetilde{g}_p\left(\lambda^{-\frac{\zeta}{2}}m_r\Delta\varphi_r^{\lambda}\right)\right|^p dr\right)^{1+\delta} dt\right] \\
&\leqslant C\lambda^{\zeta}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{\left(\Lambda_{\lambda^{\zeta}u_{\xi}^{\lambda}}\left(c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^p m^{3p}_{\lambda^{\zeta}u_{\xi}^{\lambda}}\right)^{1+\delta}}{c^{\hat{\varphi}}_{\lambda^{\zeta}u_{\xi}^{\lambda}}}\right)_T^{*,\frac{3}{2}}\right]^{\frac{2}{3}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T c_t^{\hat{\varphi}}dt\right)^{3(1+\delta')}\right]^{\frac{2}{9(1+\delta')}} \\
&\quad\times\sup_{\xi\in\mathbb{R}_+}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi}_{\xi}^{\lambda}\right)^{9(1+\delta)\frac{1+\delta'}{\delta'}}\right]^{\frac{2\delta'}{9p(1+\delta')}}\left(\int_0^{\infty}\hat{\mathbb{Q}}\left[\xi\leqslant\int_0^{\tau^{\Delta\varphi}}c_t^{\hat{\varphi}}dt\right]^{\frac{1}{3}} d\xi\right)^{\frac{1}{3}}.
\end{aligned}
$$

The integrability of $\mathcal{X}^\varepsilon$ and Lemma 2.7.11 ensure that the above upper bound is finite, provided that $3\delta/2 < \varepsilon$, $3(1+\delta') < 3+\varepsilon$ and $9(1+\delta)\frac{1+\delta'}{\delta'} \leqslant 27/\varepsilon + 9 + \varepsilon$, which we assumed in the first part of this proof. This establishes the first part of the assertion. Uniform integrability in turn follows from the de la Vallée-Poussin's theorem. □

At the optimum, the limiting expressions (2.5.19) and (2.5.20) for the expected transaction costs and squared displacements simplify considerably:

Lemma 2.5.9. *Suppose Assumption 3*(iv) *is satisfied. With the constant c_p from Lemma 2.2.1, we have*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\frac{\gamma}{2} c_t^S v(m_t) + 2^{-p} p^{-\frac{p}{p-1}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^p m_t^p w\right) dt\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \Lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c_t^S \left(c_t^{\hat{\varphi}}\right)^2}{8}\right)^{\frac{p}{p+2}} dt\right] c_p.$$

Proof. First, insert the definitions of v, w and m_t, obtaining

$$\begin{aligned}
&\frac{\gamma}{2} c_t^S v(m_t) + 2^{-p} p^{-\frac{p}{p-1}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^p m_t^p w \\
&= \frac{\int_{\mathbb{R}} \left(\frac{\gamma c_t^S}{2m_t^2} x^2 + 2^{-p} p^{-\frac{p}{p-1}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^p m_t^p \left|\widetilde{g}_p(x)\right|^p\right) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}{\int_{\mathbb{R}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx} \\
&= \Lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c_t^S \left(c_t^{\hat{\varphi}}\right)^2}{8}\right)^{\frac{p}{p+2}} \frac{\int_{\mathbb{R}} \left(x^2 + p^{-\frac{p}{p-1}} \left|\widetilde{g}_p(x)\right|^p\right) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}{\int_{\mathbb{R}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx} \\
&= \Lambda_t^{\frac{2}{p+2}} \left(\frac{\gamma c_t^S \left(c_t^{\hat{\varphi}}\right)^2}{8}\right)^{\frac{p}{p+2}} \frac{\int_{\mathbb{R}} \left(x^2 - (p-1) p^{-\frac{p}{p-1}} \left|\widetilde{g}_p(x)\right|^p + p^{-\frac{1}{p-1}} \left|\widetilde{g}_p(x)\right|^p\right) e^{-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)} dx}{\int_{\mathbb{R}} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx}.
\end{aligned}$$

Now, notice that by Lemma 2.7.1, the function $\widetilde{g}_p$ and its antiderivative $\widetilde{G}_p$ are bounded from above and below by polynomials of non-zero degree in a neighbourhood of infinity. Whence, integration by part yields

$$\begin{aligned}
&\int_{\mathbb{R}} p^{-\frac{1}{p-1}} \left|\widetilde{g}_p(x)\right|^p \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx \\
&\qquad = -\int_{\mathbb{R}} \left|\widetilde{g}_p(x)\right|^{p-1} \operatorname{sgn}(x) \left(\exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right)\right)' dx \\
&\qquad = -\int_{\mathbb{R}} g_p(x) \left(\exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right)\right)' dx \\
&\qquad = \int_{\mathbb{R}} g_p'(x) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) dx.
\end{aligned}$$

Inserting this together with the ODE 2.2.1 for g_p, this yields

$$\frac{\gamma}{2}c_t^S v(m_t)+2^{-p}p^{-\frac{p}{p-1}}\Lambda_t\left(c_t^{\hat{\varphi}}\right)^p m_t^p w = \Lambda_t^{\frac{2}{p+2}}\left(\frac{\gamma c_t^S\left(c_t^{\hat{\varphi}}\right)^2}{8}\right)^{\frac{p}{p+2}} c_p\frac{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\int_{\mathbb{R}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx},$$

and hence the assertion. □

Bounding the liquidation cost. So far, we have computed the cost-displacement tradeoff (2.5.1) of the strategy φ^λ from Section 2.2.1 until the liquidation time $\tau^{\Delta\varphi}$ defined in (2.2.7). We now show that the trading costs and displacement accumulated on the remaining interval $\left[\tau^{\Delta\varphi},T\right]$ is negligible at the leading order. To wit, liquidating the position at a constant rate $-\lambda^{-\eta}\varphi^\lambda_{\tau^{\Delta\varphi}}$ on the interval $\left[\tau^{\Delta\varphi},\tau^{\Delta\varphi}+\lambda^\eta\right]$ and holding 0 shares between $\tau^{\Delta\varphi}+\lambda^\eta$ and T generates transaction and displacement costs of higher order $o(\lambda^\zeta)$.

Recall from (2.2.5) that $\eta\in\left(\frac{2}{p+2},\frac{p}{(p-1)(p+2)}\right)$ and $T^\lambda=T-\lambda^\eta$.

Lemma 2.5.10. *Suppose that Assumptions 3*(i), (ii), (iii) *are satisfied. Then:*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^T\frac{\gamma}{2}\left(\hat{\varphi}_t-\varphi_t^\lambda\right)^2c_t^S dt\right]=o\left(\lambda^{\frac{2}{p+2}}\right), \tag{2.5.23}$$

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^T\lambda\Lambda_t\left|\dot{\varphi}_t^\lambda\right|^p dt\right]=o\left(\lambda^{\frac{2}{p+2}}\right). \tag{2.5.24}$$

Proof. To establish (2.5.23), observe that Lemma 2.7.3 and Hölder's inequality, the fact that $\eta>\frac{2}{p+2}$ and $\hat{\mathbb{Q}}\left[\tau^{\Delta\varphi}<T^\lambda\right]=o\left(\lambda^{\frac{2(1+\varepsilon)}{p+2}}\right)$ (cf. Lemma 2.7.10), and the stated integrability conditions imply

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^T\frac{\gamma}{2}\left(\varphi_t^\lambda-\hat{\varphi}_t\right)^2c_t^S dt\right]&\leqslant 2\gamma\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^{T^\lambda}\left(\hat{\varphi}_t^*\right)^2c_t^S dt\right]+2\gamma\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{T^\lambda}^T\left(\hat{\varphi}_t^*\right)^2c_t^S dt\right]\\&\leqslant 2\gamma\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T\left(\hat{\varphi}_t^*\right)^2c_t^S dt\right)^{1+\frac{1}{\varepsilon}}\right]^{\frac{1}{1+\frac{1}{\varepsilon}}}\hat{\mathbb{Q}}\left[\tau^{\Delta\varphi}<T^\lambda\right]^{\frac{1}{1+\varepsilon}}\\&\quad+2\gamma\lambda^\eta\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\hat{\varphi}_T^*\right)^2c_T^{S*}\right]\\&=o\left(\lambda^{\frac{2}{p+2}}\right).\end{aligned}$$

To establish (2.5.24), note that $\left|\lambda^{-\eta}\varphi^{\lambda}_{\tau^{\Delta\varphi}}\right| \leqslant \lambda^{-\eta-\kappa_4}$, $1-(p-1)\eta-p\kappa_4 > \frac{2}{p+2}$ and the stated integrability conditions give

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^{T} \lambda\Lambda_t \left|\dot{\varphi}^{\lambda}_t\right|^p dt\right] \leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^{\tau^{\Delta\varphi}+\lambda^{\eta}} \lambda\Lambda_t \left|\lambda^{-\eta}\varphi^{\lambda}_{\tau^{\Delta\varphi}}\right|^p dt\right]$$
$$\leqslant \lambda^{1-(p-1)\eta-p\kappa_4}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\Lambda^*_T\right] = o\left(\lambda^{\frac{2}{p+2}}\right).$$

□

Combining (2.5.19-2.5.20), Lemma 2.5.9, Lemma 2.5.10 and dominated convergence, we finally obtain the following asymptotic expression for the cost-displacement tradeoff (2.5.1):

Proposition 2.5.11. *Suppose Assumption 3 holds. Then, the leading-order asymptotics of the cost-displacement tradeoff* (2.5.1) *for the candidate strategy* φ^{λ} *from Section 2.2.1 are*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\frac{\gamma}{2}\left(\hat{\varphi}_t-\varphi^{\lambda}_t\right)^2 c^S_t + \lambda_t \left|\dot{\varphi}^{\lambda}_t\right|^p\right)dt\right]$$
$$= \lambda^{\frac{2}{p+2}}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \Lambda_t^{\frac{2}{p+2}}\left(\frac{\gamma c^S_t\left(c^{\hat{\varphi}}_t\right)^2}{8}\right)^{\frac{p}{p+2}} dt\right] c_p + o\left(\lambda^{\frac{2}{p+2}}\right),$$

where c_p *is the constant defined in Lemma 2.2.1.*

2.5.2 Leading-Order Equivalence of Utility Loss and Cost-Displacement Tradeoff

We now show that the cost-displacement tradeoff (2.5.1) is indeed asymptotically equivalent to the expected utility lost compared to the frictionless optimum when applying the strategy φ^{λ} from Section 2.3.

Proposition 2.5.12. *Suppose Assumption 3 holds. Then:*

$$\mathbb{E}_{\mathbb{P}}\left[U\left(x+\int_0^T \varphi^{\lambda}_t dS_t - \int_0^T \lambda_t \left|\dot{\varphi}^{\lambda}_t\right|^p dt\right)\right]$$
$$= \mathbb{E}_{\mathbb{P}}\left[U\left(x+\int_0^T \hat{\varphi}_t dS_t\right)\right] - \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\frac{\gamma}{2}\left(\hat{\varphi}_t-\varphi_t\right)^2 c^S_t + \lambda\Lambda_t\left|\dot{\varphi}_t\right|^p\right)dt\right]$$
$$+ o\left(\lambda^{\frac{2}{p+2}}\right).$$

Proof. To ease notation, set $V_T = x + \int_0^T \hat{\varphi}_t dS_t$, $V_T^\lambda = x + \int_0^T \varphi_t^\lambda dS_t - \lambda \int_0^T \Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt$ as well as $\bar{V}_T^\lambda = x + \int_0^T \varphi_t^\lambda dS_t - \lambda \int_0^{\tau^{\Delta\varphi}} \Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt$.

By concavity of U, the fact that U is an exponential function, and the first-order condition $U'(V_T) = \hat{y} d\hat{\mathbb{Q}}/d\mathbb{P}$,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[U(V_T^\lambda)] &\geqslant \mathbb{E}_{\mathbb{P}}[U(\bar{V}_T^\lambda)] + \mathbb{E}_{\mathbb{P}}[U'(V_T^\lambda)(V_T^\lambda - \bar{V}_T^\lambda)] \\ &= \mathbb{E}_{\mathbb{P}}[U(\bar{V}_T^\lambda)] + \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}[\exp(-\gamma(V_T^\lambda - V_T))(V_T^\lambda - \bar{V}_T^\lambda)].\end{aligned}$$

We proceed to show that $\mathbb{E}_{\hat{\mathbb{Q}}}[\exp(-\gamma(V_T^\lambda - V_T))(V_T^\lambda - \bar{V}_T^\lambda)]$ is of order $o\left(\lambda^{\frac{2}{p+2}}\right)$ and therefore negligible at the leading order for small λ. Since $\lambda \int_0^{\tau^{\Delta\varphi}} \Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p \leqslant \lambda^{\kappa_3}$, a similar estimate as in the proof of Lemma 2.5.10, the elementary inequality $x \leqslant \frac{2(1+\epsilon)}{\epsilon^2} \exp\left(\frac{\epsilon^2}{2(1+\epsilon)} x\right)$ for $x \geqslant 0$ and Hölder's inequality with exponents $1+\epsilon$ and $1+\frac{1}{\epsilon}$, [153, Theorem III.43] and Lemma 2.7.3, we obtain for λ sufficiently small:

$$\begin{aligned}
&\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(-\gamma(V_T^\lambda - V_T)\right)|V_T^\lambda - \bar{V}_T^\lambda|\right] \\
&\quad = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\int_0^T -\gamma(\varphi_t^\lambda - \hat{\varphi}_t) dS_t + \int_0^T \gamma\lambda\Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right) \int_{\tau^{\Delta\varphi}}^T \lambda\Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right] \\
&\quad \leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\int_0^T -\gamma(\varphi_t^\lambda - \hat{\varphi}_t) dS_t\right) \exp(\gamma\lambda^{\kappa_3}) \exp\left(\gamma\lambda^{1-(p-1)\eta-\kappa_4 p}\Lambda_T^*\right) \lambda^{1-(p-1)\eta-\kappa_4 p}\Lambda_T^*\right] \\
&\quad \leqslant \exp(\gamma\lambda^{\kappa_3})\lambda^{1-(p-1)\eta-\kappa_4 p}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\int_0^T -\gamma(\varphi_t^\lambda - \hat{\varphi}_t) dS_t\right) \exp\left(\frac{\epsilon^2}{2(1+\epsilon)}\Lambda_T^*\right)\Lambda_T^*\right] \\
&\quad \leqslant \exp(\gamma\lambda^{\kappa_3})\lambda^{1-(p-1)\eta-\kappa_4 p}\frac{2(1+\epsilon)}{\epsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\int_0^T -\gamma(\varphi_t^\lambda - \hat{\varphi}_t) dS_t\right) \exp\left(\frac{\epsilon^2}{(1+\epsilon)}\Lambda_T^*\right)\right] \\
&\quad \leqslant \exp(\gamma\lambda^{\kappa_3})\lambda^{1-(p-1)\eta-\kappa_4 p}\frac{2(1+\epsilon)}{\epsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(-(1+\epsilon)\gamma\int_0^T (\varphi_t^\lambda - \hat{\varphi}_t) dS_t\right)\right]^{\frac{1}{1+\epsilon}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\epsilon\Lambda_T^*\right)\right]^{\frac{\epsilon}{1+\epsilon}} \\
&\quad \leqslant \exp(\gamma\lambda^{\kappa_3})\lambda^{1-(p-1)\eta-\kappa_4 p}\frac{2(1+\epsilon)}{\epsilon^2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(8(1+\epsilon)^2\gamma^2\int_0^T (\hat{\varphi}_t^*)^2 c_t^S dt\right)\right]^{\frac{1}{2(1+\epsilon)}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\epsilon\Lambda_T^*\right)\right]^{\frac{\epsilon}{1+\epsilon}} \\
&\quad = O\left(\lambda^{1-(p-1)\eta-\kappa_4 p}\right) = o\left(\lambda^{\frac{2}{p+2}}\right). \qquad (2.5.25)
\end{aligned}$$

Next, a second-order Taylor expansion with Cauchy remainder term of $U(\bar{V}_T^\lambda)$

around the frictionless optimizer V_T yields

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}[U(\bar{V}_T^{\lambda})] =& \mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right] + \mathbb{E}_{\mathbb{P}}\left[U'\left(V_T\right)\left(\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)dS_t - \lambda\int_0^{\tau^{\Delta\varphi}}\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right)\right] \\
&+ \frac{1}{2}\mathbb{E}_{\mathbb{P}}\left[U''\left(V_T\right)\left(\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)dS_t - \lambda\int_0^{\tau^{\Delta\varphi}}\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right)^2\right] \\
&+ \frac{1}{6}\mathbb{E}_{\mathbb{P}}\left[U'''\left(V_T + \chi_1\left(\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)dS_t - \lambda\int_0^{\tau^{\Delta\varphi}}\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right)\right)\right. \\
&\qquad \left.\times\left(\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)dS_t - \lambda\int_0^{\tau^{\Delta\varphi}}\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right)^3\right],
\end{aligned}
$$

for some random $\chi_1 \in [0,1]$. By the first-order condition $U'\left(V_T\right) = \hat{y}d\hat{\mathbb{Q}}/d\mathbb{P}$, $U''/U' \equiv -\gamma$, the fact that $\varphi^{\lambda}\cdot S$ and $\hat{\varphi}\cdot S$ are $\hat{\mathbb{Q}}$-martingales (cf. Section 2.4.2), the fact that $\lambda\int_0^{\tau^{\Delta\varphi}}\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p \leqslant \lambda^{\kappa_3}$, the elementary inequality $(a+b)^3 \leqslant 4(a^3+b^3)$ for $a,b \geqslant 0$ and the fact that $\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T\lambda\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right] = O\left(\lambda^{\frac{2}{p+2}}\right)$ by Proposition 2.5.11,

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[U\left(\bar{V}_T^{\lambda}\right)\right] \geqslant& \mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right] - \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\lambda\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right] - \frac{\gamma\hat{y}}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)^2 c_t^S dt\right] \\
&+ \hat{y}\gamma\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T(\varphi_t^{\lambda}-\hat{\varphi}_t)dS_t\right)\left(\int_0^{\tau^{\Delta\varphi}}\lambda\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right)\right] \\
&- \frac{\hat{y}\gamma}{2}\lambda^{\kappa_3}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}}\lambda\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right] \\
&- \frac{2\hat{y}\gamma^2}{3}\exp\left(\gamma\lambda^{\kappa_3}\right)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\gamma\left|\int_0^T(\varphi_t^{\lambda}-\hat{\varphi}_t)dS_t\right|\right)\left|\int_0^{\tau^{\Delta\varphi}}\lambda\Lambda_t\left|\dot{\varphi}_t^{\lambda}\right|^p dt\right|^3\right] \\
&- \frac{2\hat{y}\gamma^2}{3}\exp\left(\gamma\lambda^{\kappa_3}\right)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\gamma\left|\int_0^T(\varphi_t^{\lambda}-\hat{\varphi}_t)dS_t\right|\right)\left|\int_0^T\left(\varphi_t^{\lambda}-\hat{\varphi}_t\right)dS_t\right|^3\right] \\
:=& \mathbb{E}_{\mathbb{P}}\left[U\left(V_T\right)\right] - \hat{y}C(\varphi^{\lambda},T) + \hat{y}\gamma A_1 - o\left(\lambda^{\frac{2}{p+2}}\right) \\
&- \frac{2\hat{y}\gamma^2}{3}\exp\left(\gamma\lambda^{\kappa_3}\right)(A_2+A_3). \qquad (2.5.26)
\end{aligned}
$$

We now prove that A_1, A_2 and A_3 are each of order $o\left(\lambda^{\frac{2}{p+2}}\right)$ and therefore negligible at the leading order for small λ.

By the Cauchy-Schwarz inequality, Itô isometry, the fact that $\lambda \int_0^{\tau^{\Delta\varphi}} \Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p \leqslant \lambda^{\kappa_3}$ and Proposition 2.5.11, we obtain

$$\begin{aligned} A_1 &\leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \left(\Delta\varphi_t^\lambda\right)^2 c_t^S dt\right]^{\frac{1}{2}} \lambda^{\frac{\kappa_3}{2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda\Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right]^{\frac{1}{2}} \\ &= O\left(\lambda^{\frac{1}{p+2}+\frac{\kappa_3}{2}+\frac{1}{p+2}}\right) \\ &= o\left(\lambda^{\frac{2}{p+2}}\right). \end{aligned}$$

Next, $\lambda \int_0^{\tau^{\Delta\varphi}} \Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p \leqslant \lambda^{\kappa_3}$, $3\kappa_3 > \frac{2}{p+2}$, [153, Theorem III.43] (combined with the elementary inequality $\exp(|x|) \leqslant \exp(x) + \exp(-x)$ for $x \in \mathbb{R}$) and Lemma 2.7.3 give

$$\begin{aligned} A_2 &\leqslant \lambda^{3\kappa_3}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(\gamma\left|\int_0^T (\varphi_t^\lambda - \hat{\varphi}_t)dS_t\right|\right)\right] \\ &\leqslant 2\lambda^{3\kappa_3}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(2\gamma^2 \int_0^T (\varphi_t^\lambda - \hat{\varphi}_t)^2 c_t^S dt\right)\right] \\ &\leqslant 2\lambda^{3\kappa_3}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(8\gamma^2 \int_0^T (\hat{\varphi}_t^*)^2 c_t^S dt\right)\right] = O\left(\lambda^{3\kappa_3}\right) = o\left(\lambda^{\frac{2}{p+2}}\right). \end{aligned}$$

Finally, it follows from the Cauchy-Schwarz inequality, the Burkholder-Davis-Gundy inequality with constant C, [153, Theorem III.43] (combined with the elementary inequality $\exp(|x|) \leqslant \exp(x) + \exp(-x)$ for $x \in \mathbb{R}$), and Lemma 2.7.3 that

$$\begin{aligned} A_3 &\leqslant C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(2\gamma\left|\int_0^T (\varphi_t^\lambda - \hat{\varphi}_t)dS_t\right|\right)\right]^{\frac{1}{2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|\int_0^T \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right|^3\right]^{\frac{1}{2}} \\ &\leqslant \sqrt{2}C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\exp\left(32\gamma^2 \int_0^T (\hat{\varphi}_t^*)^2 c_t^S dt\right)\right]^{\frac{1}{2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right]^{\frac{1}{2}}. \end{aligned}$$

To conclude the proof, it suffices to establish that

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right] = o\left(\lambda^{\frac{4}{p+2}}\right).$$

To this end, using the elementary inequality $(a+b+c)^3 \leqslant 9(a^3+b^3+c^3)$ for $a, b, c \geqslant 0$, it suffices to show that the terms

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\Delta\varphi}} \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right], \quad \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{T^\lambda} \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right]$$
$$\text{and} \quad \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{T^\lambda}^{T} \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right]$$

are all of order $o\left(\lambda^{\frac{4}{p+2}}\right)$.

For the first term, this follows from the fact that $|\varphi_t^\lambda - \hat{\varphi}_t| \leqslant \lambda^{\kappa_1}$ on $[\![0, \tau^{\Delta\varphi}]\!]$ and $6\kappa_1 > \frac{4}{p+2}$. For the second term, Lemma 2.7.3, Proposition 2.7.10 and Hölder's inequality give

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{T^\lambda} \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right] &\leqslant 2^6 \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T (\hat{\varphi}_t^*)^2 c_t^S dt\right)^{3(1+\frac{1}{\varepsilon})}\right]^{\frac{1}{1+\frac{1}{\varepsilon}}} \hat{\mathbb{Q}}\left[\tau^{\Delta\varphi} < T^\lambda\right]^{\frac{1}{1+\varepsilon}} \\ &= o\left(\lambda^{\frac{4}{p+2}}\right).\end{aligned}$$

For the third term, Lemma 2.7.3, $T - T^\lambda = \lambda^\eta$ and $\eta > \frac{2}{p+2}$ yield

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{T^\lambda}^{T} \left(\varphi_t^\lambda - \hat{\varphi}_t\right)^2 c_t^S dt\right)^3\right] \leqslant 2^6 \lambda^{3\eta} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left((\hat{\varphi}_T^*)^2 c_T^{S*}\right)^3\right] = O\left(\lambda^{3\eta}\right) = o\left(\lambda^{\frac{4}{p+2}}\right).$$

This completes the proof. □

2.6 Dual Considerations

2.6.1 Asymptotic Duality Bound

To complete the proof of Theorem 2.2.3, we now complement Proposition 2.5.12 by the following asymptotic duality bound, valid for *any* admissible strategy:

Proposition 2.6.1. *Under Assumption 3, the following upper duality bound is valid for any admissible strategy φ:*

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[U\left(X_T^\varphi\right)\right] \leqslant x\hat{y} + \mathbb{E}_{\mathbb{P}}\left[\widetilde{U}\left(\hat{y}\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] - \hat{y}\frac{\gamma}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] \\ - \hat{y}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda\Lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right).\end{aligned}$$

Once this result is established, it suffices to note that (see [167, Equation (18)])

$$x\hat{y} + \mathbb{E}_{\mathbb{P}}\left[\widetilde{U}\left(\hat{y}\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] = \mathbb{E}_{\mathbb{P}}\left[U\left(X_T^{\hat{\varphi}}\right)\right].$$

Theorem 2.2.3 then follows from Proposition 2.6.1 combined with the primal lower bound established in Propositions 2.5.11 and 2.5.12 for the candidate strategy from Section 2.2.1.

2.6.2 Proof of Proposition 2.6.1

It remains to prove Proposition 2.6.1. To this end, we use the duality theory for superlinear frictions developed very recently by Guasoni and Rasonyi [77]. They argue that in this context, the dual measures do not turn the frictionless price into a martingale, but much rather the actual execution price with transaction costs.[21] This characterization is apparently difficult to apply, since it links the primal and dual optimizers, but both are unknown. However, it is extremely useful for the present asymptotic verification, because we already have a candidate *asymptotic* optimal strategy φ^λ at hand. We use the execution price corresponding to the latter as a substitute for the exact optimizer. We then stop the contribution of the friction to the dynamics of the shadow price appropriately (cf. Equation (2.6.8)), in order to keep the processes involved in the ensuing estimates bounded by carefully chosen powers of λ. By establishing an upper bound for the probability that this stopping times is strictly smaller than $\tau^{\Delta\varphi}$, we ensure that this modification of the dual variable does not affect the duality bound at the leading order.

In a second step, we then introduce the dual element for the duality bound in Proposition 2.6.1 – an equivalent martingale measure for the asymptotic execution price.

Finally, convex duality (both for the utility function and the trading cost functional), Taylor expansions, and careful remainder estimates yield the desired duality bound.

Shadow price. Recall that under the frictionless dual martingale measure $\hat{\mathbb{Q}}$, the risky asset has dynamics $dS_t = \sqrt{c_t^S}dW_t^{\hat{\mathbb{Q}}}$. Inspired by the first-order condition

[21] For models with proportional transaction costs, this leads to a “shadow price” in the spirit of [98, 48] that coincides with the bid- or ask-price, respectively, whenever the optimal strategy sells or purchases. Between these trading times, the shadow price needs to be chosen so that it is indeed optimal not to alter the portfolio at hand, compare [104]. In the present context, it is optimal to trade at all times, so that the execution price is directly linked to the optimal strategy.

of [77], define the following execution price:

$$S_t^\lambda = S_t + \lambda_t p \operatorname{sgn}\left(\dot{\varphi}_t^\lambda\right) \left|\dot{\varphi}_t^\lambda\right|^{p-1} = S_t + 2^{-(p-1)} \lambda^{\frac{3}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^{p-1} m_t^{p-1} g_p \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right),$$

where $\dot{\varphi}^\lambda$ is the candidate trading rate from Section 2.2.1. Since the function g_p is twice continuously differentiable (cf. (2.2.1)), Itô's formula readily yields the corresponding $\hat{\mathbb{Q}}$ dynamics on $[\![0, \tau^{\Delta\varphi}]\!]$:

$$dS_t^\lambda = \mu_t^{S^\lambda} dt + \frac{c^{S^\lambda, S}}{\sqrt{c_t^S}} dW_t^{\hat{\mathbb{Q}}} + dM_t^\lambda,$$

where μ^{S^λ} the drift process of S^λ under $\hat{\mathbb{Q}}$ is given by

$$\begin{aligned}
\mu_t^{S^\lambda} &= 2^{1-p} \lambda^{\frac{3}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^{p-1} m_t^{p-1} g_p \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right) \\
&\quad \times \left[\frac{\mu_t^\Lambda}{\Lambda_t} + (p-1)\left(\frac{\mu_t^{c^{\hat{\varphi}}}}{c_t^{\hat{\varphi}}} + \frac{\mu_t^m}{m_t} + \frac{c_t^{\Lambda, c^{\hat{\varphi}}}}{\Lambda_t c_t^{\hat{\varphi}}} + \frac{c_t^{\Lambda, m}}{\Lambda_t m_t} + \frac{(p-1) c_t^{c^{\hat{\varphi}}, m}}{c_t^{\hat{\varphi}} m_t} + \frac{(p-2)}{2}\left(\frac{c_t^{c^{\hat{\varphi}}}}{(c_t^{\hat{\varphi}})^2} + \frac{c_t^m}{m_t^2}\right)\right)\right] \\
&\quad + 2^{1-p} \lambda^{\frac{2}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^{p-1} m_t^{p-1} g_p' \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right) \\
&\quad \times \left(\Delta\varphi_t^\lambda \mu_t^m + m_t \mu^{\hat{\varphi}} + \frac{\Delta\varphi_t^\lambda c_t^{\Lambda, m}}{\Lambda_t} + \frac{m_t c_t^{\Lambda, \hat{\varphi}}}{\Lambda_t} + (p-1)\left(\frac{\Delta\varphi_t^\lambda c_t^m}{m_t} + \frac{\Delta\varphi_t^\lambda c_t^{c^{\hat{\varphi}}, m}}{c_t^{\hat{\varphi}}} + \frac{m_t c_t^{c^{\hat{\varphi}}, \hat{\varphi}}}{c_t^{\hat{\varphi}}} + c_t^{m, \hat{\varphi}}\right)\right) \\
&\quad - 2^{-p} \lambda^{\frac{1}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^p m_t^{p+1} \Bigg(p^{-\frac{1}{p-1}} \left(\widetilde{g}_p g_p'\right) \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right) \\
&\quad - \left(\frac{\left(\Delta\varphi_t^\lambda\right)^2}{m_t^2 c_t^{\hat{\varphi}}} c_t^m + \frac{2\Delta\varphi_t^\lambda}{m_t c_t^{\hat{\varphi}}} c_t^{m, \hat{\varphi}} + 1\right) g_p'' \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right)\Bigg),
\end{aligned}$$

the quadratic covariation of S^λ with respect to S is

$$\begin{aligned}
c_t^{S^\lambda, S} &= c_t^S + 2^{1-p} \lambda^{\frac{3}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^{p-1} m_t^{p-1} g_p \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right) \left(\frac{c_t^{\Lambda, S}}{\Lambda_t} + (p-1)\left(\frac{c_t^{c^{\hat{\varphi}}, S}}{c_t^{\hat{\varphi}}} + \frac{c_t^{m, S}}{m_t}\right)\right) \\
&\quad + 2^{1-p} \lambda^{\frac{2}{p+2}} \Lambda_t \left(c_t^{\hat{\varphi}}\right)^{p-1} m_t^{p-1} g_p' \left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi_t^\lambda\right) \left(\Delta\varphi_t^\lambda c_t^{m, S} + m_t c_t^{\hat{\varphi}, S}\right),
\end{aligned}$$

and the orthogonal local martingale M^λ has the following instantaneous quadratic variation:

$$
\begin{aligned}
c_t^{M^\lambda} &= 2^{2(1-p)}\lambda^{\frac{6}{p+2}}\Lambda_t^2\big(c_t^{\hat\varphi}\big)^{2(p-1)}m_t^{2(p-1)}g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)^2\left[\frac{c_t^\Lambda}{\Lambda_t^2}\left(1-\frac{\big(c_t^{\Lambda,S}\big)^2}{c_t^\Lambda c_t^S}\right)\right. \qquad (2.6.1)\\
&\quad+(p-1)^2\frac{c_t^{c^{\hat\varphi}}}{\big(c_t^{\hat\varphi}\big)^2}\left(1-\frac{\big(c_t^{c^{\hat\varphi},S}\big)^2}{c_t^{c^{\hat\varphi}}c_t^S}\right)+(p-1)^2\frac{c_t^m}{m_t}\left.\left(1-\frac{\big(c_t^{m,S}\big)^2}{c_t^m c_t^S}\right)\right]\\
&\quad+2^{2(1-p)}\lambda^{\frac{4}{p+2}}\Lambda_t^2\big(c_t^{\hat\varphi}\big)^{2(p-1)}m_t^{2(p-1)}g_p'\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)^2\left[\big(\Delta\varphi_t^\lambda\big)^2c_t^m\left(1-\frac{\big(c_t^{m,S}\big)^2}{c_t^m c_t^S}\right)\right.\\
&\quad+m_t^2c_t^{\hat\varphi}\left.\left(1-\frac{\big(c^{\hat\varphi,S}\big)^2}{c_t^{\hat\varphi}c_t^S}\right)\right].
\end{aligned}
$$

The ODE (2.2.1) shows first that g_p is infinitely differentiable, because its right-hand side is differentiable. We can therefore express g_p' and g_p'' as functions of its argument and p, obtaining

$$
g_p''(z)=p^{-\frac{1}{p-1}}\widetilde g_p(z)\left[(p-1)p^{-\frac{p}{p-1}}\left|\widetilde g_p(z)\right|^p-z^2+c_p\right]-2z. \qquad (2.6.2)
$$

This allows us to simplify the above expressions for μ^{S^λ} and $c^{S^\lambda,S}$ to

$$
\begin{aligned}
\mu_t^{S^\lambda} &= -\gamma c_t^S\Delta\varphi_t^\lambda+2^{1-p}\lambda^{\frac{3}{p+2}}\Lambda_t\big(c_t^{\hat\varphi}\big)^{p-1}m_t^{p-1}g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right) \qquad (2.6.3)\\
&\times\left[\frac{\mu_t^\Lambda}{\Lambda_t}+(p-1)\left(\frac{\mu_t^{c^{\hat\varphi}}}{c_t^{\hat\varphi}}+\frac{\mu_t^m}{m_t}+\frac{c_t^{\Lambda,c^{\hat\varphi}}}{\Lambda_tc_t^{\hat\varphi}}+\frac{c_t^{\Lambda,m}}{\Lambda_tm_t}+\frac{(p-1)c_t^{c^{\hat\varphi},m}}{c_t^{\hat\varphi}m_t}+\frac{(p-2)}{2}\left(\frac{c_t^{c^{\hat\varphi}}}{\big(c_t^{\hat\varphi}\big)^2}+\frac{c_t^m}{m_t^2}\right)\right)\right]\\
&+2^{1-p}\lambda^{\frac{2}{p+2}}\Lambda_t\big(c_t^{\hat\varphi}\big)^{p-1}m_t^{p-1}\left((p-1)p^{-\frac{p}{p-1}}\left|\widetilde g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)\right|^p-\lambda^{-\frac{2}{p+2}}m_t^2\left(\Delta\varphi_t^\lambda\right)^2+c_p\right)\\
\times&\left(\Delta\varphi_t^\lambda\mu_t^m+m_t\mu_t^{\hat\varphi}+\frac{\Delta\varphi_t^\lambda c_t^{\Lambda,m}}{\Lambda_t}+\frac{m_tc_t^{\Lambda,\hat\varphi}}{\Lambda_t}+(p-1)\left(\frac{\Delta\varphi_t^\lambda c_t^m}{m_t}+\frac{\Delta\varphi_t^\lambda c_t^{c^{\hat\varphi},m}}{c_t^{\hat\varphi}}+\frac{m_tc_t^{c^{\hat\varphi},\hat\varphi}}{c_t^{\hat\varphi}}+c_t^{m,\hat\varphi}\right)\right)\\
&+2^{-p}\lambda^{\frac{1}{p+2}}\Lambda_t\big(c_t^{\hat\varphi}\big)^{p-1}m_t^p\left(\frac{\big(\Delta\varphi_t^\lambda\big)^2}{m_t}c_t^m+2\Delta\varphi_t^\lambda c_t^{m,\hat\varphi}\right)\Bigg(-2\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\\
&+p^{-\frac{1}{p-1}}\widetilde g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)\left[(p-1)p^{-\frac{p}{p-1}}\left|\widetilde g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)\right|^p-\lambda^{-\frac{2}{p+2}}m_t^2\left(\Delta\varphi_t^\lambda\right)^2+c_p\right]\Bigg)
\end{aligned}
$$

and

$$c_t^{S^\lambda,S} = c_t^S + 2^{1-p}\lambda^{\frac{3}{p+2}}\Lambda_t\big(c_t^{\hat{\varphi}}\big)^{p-1}m_t^{p-1}g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right) \tag{2.6.4}$$

$$\left(\frac{c_t^{\Lambda,S}}{\Lambda_t}+(p-1)\left(\frac{c_t^{c^{\hat{\varphi}},S}}{c_t^{\hat{\varphi}}}+\frac{c_t^{m,S}}{m_t}\right)\right)$$

$$+\,2^{1-p}\lambda^{\frac{2}{p+2}}\Lambda_t\big(c_t^{\hat{\varphi}}\big)^{p-1}m_t^{p-1}\left((p-1)p^{-\frac{p}{p-1}}\left|\widetilde{g}_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)\right|^p-\lambda^{-\frac{2}{p+2}}m_t^2\left(\Delta\varphi_t^\lambda\right)^2+c_p\right)$$

$$\left(\Delta\varphi_t^\lambda c_t^{m,S}+m_tc^{\hat{\varphi},S}\right)$$

$$:= c_t^S\left(1+z_t^\lambda\right). \tag{2.6.5}$$

Stopping and a Dual Martingale To obtain an upper duality bound in the spirit of [77], we now need to define an equivalent measure that turns the execution price S^λ into a martingale. In view of analogous constructions for proportional costs [103, 87], we would like to use the density process $\mathcal{E}\left(-\int_0^{\cdot\wedge\tau^{\Delta\varphi}}Z_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)=\exp(N^\lambda)$, where

$$Z_t^\lambda=\frac{\mu_t^{S^\lambda}\sqrt{c_t^S}}{c_t^{S^\lambda,S}},\qquad N_t^\lambda=-\int_0^{t\wedge\tau^{\Delta\varphi}}Z_t^\lambda dW_t^{\hat{\mathbb{Q}}}-\frac{1}{2}\int_0^{t\wedge\tau^{\Delta\varphi}}\left(Z_t^\lambda\right)^2dt. \tag{2.6.6}$$

This probability change evidently removes the $\hat{\mathbb{Q}}$-drift by Girsanov's theorem and the definition of $c^{S^\lambda,S}$ in (2.6.4). Since N^λ may generally not be integrable enough to define a valid change of measure (and to ensure enough integrability for the ensuing estimates), we introduce the following dual stopping time:

$$\tau^{\lambda,\mathrm{dual}}=\inf\Bigg\{t\in[0,\tau^{\Delta\varphi}]:\left|z_t^\lambda\right|>\lambda^{\kappa_5},\ \text{or}\ \left|Z_t^\lambda\right|>\lambda^{\kappa_6},\ \text{or}\ c_t^{M^\lambda}>c_t^S, \tag{2.6.7}$$

$$\text{or}\ \left|\frac{\mu_t^{S^\lambda}+\gamma c_t^S\Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right|>\lambda^{\kappa_7},\ \text{or}\ \left|N_t^\lambda\right|^2>\lambda^{\kappa_8}\Bigg\}\wedge\tau^{\Delta\varphi},$$

where the κ_i are chosen as follows (one can readily see from the definition of κ_2 in (2.2.7) that these intervals are not empty)

$$\kappa_5\in\left(\frac{2}{p+2}-2\kappa_1,\frac{2}{p+2}-10\kappa_2\right),\quad \kappa_6\in\left(\frac{1}{2(p+2)},\frac{1}{p+2}-5\kappa_2\right),$$
$$\kappa_7\in\left(\frac{2}{p+2}-\kappa_1,\frac{2}{p+2}-5(2p+3)\kappa_2\right),\quad \kappa_8\in\left(\frac{4}{3(p+2)},\frac{2}{p+2}-10\kappa_2\right).$$

Now define the following "stopped" change of measure:

$$\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} = \exp\left(N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right).$$

The corresponding "modified" execution price is in turn defined as

$$d\widetilde{S}^\lambda_t = dS^\lambda_t \mathbb{1}_{\{t\leqslant\tau^{\lambda,\mathrm{dual}}\}} + dS_t\mathbb{1}_{\{t>\tau^{\lambda,\mathrm{dual}}\}}. \tag{2.6.8}$$

Under $\hat{\mathbb{Q}}^\lambda$, the modified execution price then evidently is a martingale. Moreover, by definition of $\tau^{\lambda,\mathrm{dual}}$, for λ small enough

$$\frac{1}{2} \leqslant \frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} \leqslant \frac{3}{2}, \quad \frac{1}{2} \leqslant 1 + z^\lambda_t \leqslant \frac{3}{2}, \quad \text{and} \quad \frac{1}{2} \leqslant \exp\left(2N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right) \leqslant \frac{3}{2}. \tag{2.6.9}$$

The stopping times are crucial to control the remainders in the ensuing estimates. However, their precise choice does not affect the dual bound at the leading order, due to the convergence of the stopping time $\tau^{\lambda,\mathrm{dual}}$ to $\tau^{\Delta\varphi}$ in probability as λ goes to 0 (cf. Lemma 2.6.5).

Convex Duality Estimates. To derive the upper duality bound in Proposition 2.6.1, we follow [77] in using the convex conjugates of both the utility function U and the instantaneous trading cost $x \mapsto \Psi_t(x) = \lambda_t |x|^p$:[22]

$$\widetilde{\Psi}_t(y) = \sup_{x\in\mathbb{R}}\{xy - \Psi_t(x)\} = \frac{p-1}{p\,(\lambda_t p)^{\frac{1}{p-1}}}\,|y|^{\frac{p}{p-1}}, \qquad \widetilde{\Psi}'_t(y) = \frac{1}{(\lambda_t p)^{\frac{1}{p-1}}}\,|y|^{\frac{1}{p-1}}\,\mathrm{sgn}(y), \tag{2.6.10}$$

$$\widetilde{U}(y) = \sup_{x\in\mathbb{R}}\{U(x) - xy\} = \frac{y}{\gamma}\left(\log\left(\frac{y}{\gamma}\right) - 1\right), \quad \widetilde{U}'(y) = \frac{1}{\gamma}\log\left(\frac{y}{\gamma}\right), \quad \widetilde{U}''(y) = \frac{1}{\gamma y}. \tag{2.6.11}$$

Let φ be any admissible trading strategy. As the terminal risky position is zero and the initial wealth is x, two integrations by parts, the definition of $\widetilde{\Psi}$, and $S^\lambda_0 = S_0$

[22] The importance of the dual friction $\widetilde{\Psi}$ was first recognized in [57], where it is used to establish a superhedging theorem.

yield

$$
\begin{aligned}
X_T^\varphi &= x + \int_0^T \varphi_t dS_t - \lambda \int_0^T |\dot{\varphi}_t|^p \, dt \\
&= x - \varphi_0 S_0 - \int_0^T \dot{\varphi}_t S_t dt - \lambda \int_0^T |\dot{\varphi}_t|^p \, dt \\
&= x - \varphi_0 S_0 - \int_0^T \dot{\varphi}_t \widetilde{S}_t^\lambda dt + \int_0^T \dot{\varphi}_t \left(\widetilde{S}_t^\lambda - S_t \right) dt - \lambda \int_0^T |\dot{\varphi}_t|^p \, dt \\
&\leqslant x + \int_0^T \varphi_t d\widetilde{S}_t^\lambda + \int_0^T \widetilde{\Psi}_t (\widetilde{S}_t^\lambda - S_t) dt. \qquad (2.6.12)
\end{aligned}
$$

To proceed, we next show that the wealth process of any admissible strategy remains a martingale (under the corresponding dual measure $\hat{\mathbb{Q}}^\lambda$) when evaluated with the (modified) execution price $\widetilde{S}^\lambda$ of our candidate strategy:

Lemma 2.6.2. *Suppose that $(\hat{\varphi}_T^*)^2 (c^S)_T^* \in L^1(\hat{\mathbb{Q}})$ (Assumption 3*(i)*). Then, the stochastic integral $\int_0^\cdot \varphi_t d\widetilde{S}_t^\lambda$ is a $\hat{\mathbb{Q}}^\lambda$-martingale for any admissible strategy φ.*

Proof. We compute

$$
\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \varphi_t^2 d\langle \widetilde{S}^\lambda \rangle_t \right] =& \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \varphi_t^2 \frac{\left(c_t^{S^\lambda,S}\right)^2}{c_t^S} dt \right] + \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \varphi_t^2 d\langle M^\lambda \rangle_t \right] \\
&+ \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^T \varphi_t^2 c_t^S dt \right]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \varphi_t^2 d\langle \widetilde{S}^\lambda \rangle_t \right] &\leqslant \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \varphi_t^2 c_t^S \left(1 + z_t^\lambda\right)^2 dt \right] + \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \varphi_t^2 c_t^S dt \right] \\
&\quad + \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^T \varphi_t^2 c_t^S dt \right] \\
&\leqslant C \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \varphi_t^2 c_t^S dt \right] < \infty,
\end{aligned}
$$

where we have used the definition of $\widetilde{S}^\lambda$ and M^λ (which is orthogonal to $W^{\hat{\mathbb{Q}}}$ by definition), the uniform bound for $d\hat{\mathbb{Q}}^\lambda / d\hat{\mathbb{Q}}$, and the admissibility of φ. Therefore, $\varphi \in L^2_{\hat{\mathbb{Q}}^\lambda}\left(\widetilde{S}^\lambda\right)$ and the assertion follows. $\square$

In view of Lemma 2.6.2, taking $\hat{\mathbb{Q}}^\lambda$-expectation in (2.6.12) yields

$$\mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[x - X_T^\varphi + \int_0^T \widetilde{\Psi}_t(\widetilde{S}_t^\lambda - S_t)dt\right] \geqslant 0.$$

For $y > 0$, the definition of the convex conjugate (2.6.11), and a third-order Taylor expansion of $\widetilde{U}$ with Cauchy remainder give

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[U\left(X_T^\varphi\right)\right] &\leqslant \mathbb{E}_{\mathbb{P}}\left[U\left(X_T^\varphi\right) + y\frac{d\hat{\mathbb{Q}}^\lambda}{d\mathbb{P}}\left(x - X_T^\varphi + \int_0^T \widetilde{\Psi}_t(\widetilde{S}_t^\lambda - S_t)dt\right)\right] \\
&\leqslant \mathbb{E}_{\mathbb{P}}\left[\widetilde{U}\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\right) + y\frac{d\hat{\mathbb{Q}}^\lambda}{d\mathbb{P}}\int_0^T \widetilde{\Psi}_t(\widetilde{S}_t^\lambda - S_t)dt\right] + xy \\
&= xy + \mathbb{E}_{\mathbb{P}}\left[\widetilde{U}\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] + y\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\widetilde{U}'\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] \\
&\quad + \frac{y^2}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^2\widetilde{U}''\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] \\
&\quad + \frac{y^3}{6}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)^2\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^3\widetilde{U}'''\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}} + \chi_2 y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\right)\right] \\
&\quad + y(p-1)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \lambda_t\left|\dot{\varphi}_t^\lambda\right|^p dt\right],
\end{aligned}$$

for a random $\chi_2 \in [0,1]$. Replacing the functions $\widetilde{U}''$ and $\widetilde{U}'''$ by their expressions, we obtain

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[U\left(X_T^\varphi\right)\right] \leqslant{}& xy + \mathbb{E}_{\mathbb{P}}\left[\widetilde{U}\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] + y\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\widetilde{U}'\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] \\
&+ \frac{y}{2\gamma}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^2\right] + y(p-1)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda_t\left|\dot{\varphi}_t^\lambda\right|^p dt\right] \\
&- \frac{y}{6\gamma}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{1}{\left(1 + \chi_2\left(d\hat{\mathbb{Q}}^\lambda/d\hat{\mathbb{Q}} - 1\right)\right)^2}\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^3\right] \\
&+ y(p-1)\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\left(\int_0^T \lambda_t\left|\dot{\varphi}_t^\lambda\right|^p dt\right)\right].
\end{aligned} \tag{2.6.13}$$

The first two terms correspond to the value of the frictionless optimum for $y = \hat{y} = \mathbb{E}_{\mathbb{P}}\left[U'\left(X_T^{\hat{\varphi}}\right)\right]$ (see [167, Equation (12)]). Since $-\widetilde{U}' = (U')^{-1}$, the frictionless first-order condition (2.1.3) in turn yields $\widetilde{U}'\left(\hat{y}\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right) = -x - \int_0^T \hat{\varphi}_t dS_t$. As a consequence,

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\widetilde{U}'\left(y\frac{d\hat{\mathbb{Q}}}{d\mathbb{P}}\right)\right] &= -\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\left(x + \int_0^T \hat{\varphi}_t dS_t\right)\right] \\
&= -\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \hat{\varphi}_t dS_t\right] \\
&= -\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \varphi_t^\lambda dS_t\right] - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \Delta\varphi_t^\lambda dS_t\right] \\
&= -\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \varphi_t^\lambda dS_t\right] - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\int_0^T \Delta\varphi_t^\lambda dS_t\right].
\end{aligned}$$

Here, the second equality holds because $\int_0^\cdot \hat{\varphi}_t dS_t$ is a $\hat{\mathbb{Q}}$-martingale. For the last equality, we have used that $\int_0^\cdot \varphi_t^\lambda dS_t$ is a $\hat{\mathbb{Q}}$-martingale as well (cf. Proposition 2.4.2), so that $\int_0^\cdot \Delta\varphi_t^\lambda dS_t$ is a $\hat{\mathbb{Q}}$-martingale, too.

Furthermore, by Proposition 2.4.2, φ^λ is an admissible strategy, and therefore in $L^2_{\hat{\mathbb{Q}}^\lambda}\left(\widetilde{S}^\lambda\right)$ by Lemma 2.6.2. In particular, $\int_0^\cdot \varphi_t^\lambda d\widetilde{S}_t^\lambda$ is a $\hat{\mathbb{Q}}^\lambda$-martingale. Introducing $\widetilde{S}^\lambda$ in the previous equation and integrating by part yields (noting that $\varphi_T^\lambda = 0$ and $\Delta\varphi_0^\lambda = 0$, so that $\dot{\varphi}_0^\lambda = 0$), we obtain

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}\int_0^T \varphi_t^\lambda dS_t\right] &= \mathbb{E}_{\hat{\mathbb{Q}}^\lambda}\left[\int_0^T \varphi_t^\lambda d\widetilde{S}_t^\lambda\right] \\
&\quad - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} p \int_0^T \varphi_t^\lambda d\left(\lambda_t \operatorname{sgn}\left(\dot{\varphi}_t^\lambda\right)\left|\dot{\varphi}_t^\lambda\right|^{p-1}\right)\right] \\
&= p\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right] + p\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\int_0^T \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right].
\end{aligned}$$

It now remains to estimate the following terms:

$$A_4 = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\left(\int_0^T \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right)\right],\ A_6 = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)\int_0^T \Delta\varphi_t^\lambda dS_t\right],$$

$$A_5 = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^2\right], \quad A_7 = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\frac{1}{\left(1+\chi_2\left(d\hat{\mathbb{Q}}^\lambda/d\hat{\mathbb{Q}} - 1\right)\right)^2}\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right)^3\right].$$

The computation of these terms completes the proof of Proposition 2.6.1:

Lemma 2.6.3. *Suppose Assumption 3 is satisfied. Then:*

$$\begin{aligned} A_4 + A_7 &= o\left(\lambda^{\frac{2}{p+2}}\right), \\ A_5 &= \gamma^2 \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right), \\ A_6 &= \gamma \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right). \end{aligned}$$

Proof. We first estimate A_4. By the Cauchy-Schwarz inequality and the definition of A_5,

$$A_4 \leqslant (A_5)^{\frac{1}{2}}\, \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right)^2\right]^{\frac{1}{2}}.$$

Using the elementary inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$, that $\lambda_t|\dot{\varphi}_t^\lambda|^p \leqslant \lambda^{1-p(\eta+\kappa_4)}\Lambda_T^*$ on $[\![\tau^{\Delta\varphi}, \tau^{\Delta\varphi} + \lambda^\eta]\!]$, the second part of the right-hand side can be estimated as follows:

$$\begin{aligned} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right)^2\right] \leqslant& 2\left(\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\Delta\varphi}} \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right)^2\right]\right. \\ &\left. + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\Delta\varphi}}^{\tau^{\Delta\varphi}+\lambda^\eta} \lambda_t \left|\dot{\varphi}_t^\lambda\right|^p dt\right)^2\right]\right) \\ \leqslant& C\lambda^{\kappa_3 + \frac{2}{p+2}} + \lambda^{2(1-\eta(p-1)-p\kappa_4)}\mathbb{E}_{\hat{\mathbb{Q}}}\left[(\Lambda_T^*)^2\right] \\ =& o\left(\lambda^{\frac{2}{p+2}}\right), \end{aligned}$$

where we obtained the estimate for the first term of the sum from the definition of $\tau^{\Delta\varphi}$ in (2.2.7) and Proposition 2.5.11. The second part comes from the definition of

η (2.2.5) and the integrability Assumption 3(i). Now the claim follows from the fact that $A_5 = O\left(\lambda^{\frac{2}{p+2}}\right)$ by the argument below and Proposition 2.5.11.

We next estimate A_5. By a use of Itô formula on $\mathcal{E}(\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}})$ between 0 and $\tau^{\lambda,\text{dual}}$, we obtain,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}-1\right)^2\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\text{dual}}} \mathcal{E}\left(-\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}}\right)_t^2 \left(Z_t^\lambda\right)^2 dt\right]$$
$$- 2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\text{dual}}} \left(\mathcal{E}\left(-\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}}\right)_t - 1\right) Z_t^\lambda \mathcal{E}\left(-\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}}\right)_t dW_t^{\hat{\mathbb{Q}}}\right].$$

Now, note that by definition of $\tau^{\lambda,\text{dual}}$ in (2.6.7), on $[\![0,\tau^{\lambda,\text{dual}}]\!]$ we have

$$\mathcal{E}\left(-\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}}\right)_{t\wedge\tau^{\lambda,\text{dual}}} = \exp\left(N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}}\right) \leqslant \frac{3}{2} \quad \text{and} \quad Z_t^\lambda \leqslant C.$$

Moreover, by a Taylor expansion with Cauchy remainder,

$$\begin{aligned}\mathcal{E}\left(-\int_0^\cdot Z_s^\lambda dW_s^{\hat{\mathbb{Q}}}\right)_{t\wedge\tau^{\lambda,\text{dual}}} &= \exp\left(N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}}\right) \\ &= 1 + \exp\left(\chi_4 N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}}\right) N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}}.\end{aligned} \tag{2.6.14}$$

for some random $\chi_4 \in [0,1]$. Note that for λ small enough, $\exp\left(\chi_4 N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}}\right) N^\lambda_{t\wedge\tau^{\lambda,\text{dual}}} \in \left[\lambda^{\frac{\kappa_8}{2}}\frac{1}{2}, \lambda^{\frac{\kappa_8}{2}}\frac{3}{2}\right]$. Thus, by the definition of Z^λ in (2.6.6), it remains to estimate

$$A_8 := \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\text{dual}}} \frac{\left(\mu_t^{S^\lambda}\right)^2 c_t^S}{\left(c_t^{S^\lambda,S}\right)^2} dt\right].$$

Simple algebraic manipulations give

$$\begin{aligned}
A_8 = {} & \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \gamma^2 c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^T \gamma^2 c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] \\
& + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma^2 c_t^S \left(\Delta\varphi_t^\lambda\right)^2 \left(\frac{\left(c_t^S\right)^2}{\left(c_t^{S^\lambda,S}\right)^2} - 1\right) dt\right] \\
& + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} c_t^S \left(\frac{\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right)^2 dt\right] \\
& - 2\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma \left(c_t^S\right)^2 \Delta\varphi_t^\lambda \frac{\left(\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda\right)}{\left(c_t^{S^\lambda,S}\right)^2} dt\right].
\end{aligned}$$

We now consider the last three terms of the previous equation. By a Taylor expansion with Cauchy remainder of $x \mapsto \frac{1}{(1+x)^2}$ around 0 and the bound on z^λ ensured by the definition of $\tau^{\lambda,\mathrm{dual}}$, for λ small enough we have

$$\left|\frac{\left(c_t^S\right)^2}{\left(c_t^{S^\lambda,S}\right)^2} - 1\right| = \left|\frac{\left(c_t^S\right)^2}{\left(c_t^S(1+z_t^\lambda)\right)^2} - 1\right| = z_t^\lambda \frac{2}{(1+\chi_3 z_t^\lambda)^3} \leqslant 16\lambda^{\kappa_5}. \tag{2.6.15}$$

Recall that on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$, $\Delta\varphi^\lambda$ is bounded by λ^{κ_1} and

$$\begin{aligned}
\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma^2 c_t^S \left(\Delta\varphi_t^\lambda\right)^2 \left(\frac{\left(c_t^S\right)^2}{\left(c_t^{S^\lambda,S}\right)^2} - 1\right) dt\right] & \leqslant 16\gamma^2\lambda^{2\kappa_1+\kappa_5}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S dt\right] \\
& = o\left(\lambda^{\frac{2}{p+2}}\right),
\end{aligned}$$

by definition of κ_5 and Assumption 3(i). By definition of the stopping time, on $[\![0, \tau^{\lambda,\mathrm{dual}}]\!]$ we have

$$\left|\frac{\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right| \leqslant \lambda^{\kappa_7}$$

and we obtain with Assumption 3(i) and the definition of κ_7 that

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} c_t^S \left(\frac{\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right)^2 dt\right] \leqslant \lambda^{2\kappa_7}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S dt\right] = o\left(\lambda^{\frac{2}{p+2}}\right).$$

Similarly,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma\left(c_t^S\right)^2 \Delta\varphi_t^\lambda \frac{\left(\mu_t^{S^\lambda}+\gamma c_t^S\Delta_t^\lambda\right)}{\left(c_t^{S^\lambda,S}\right)^2}dt\right] \leqslant \gamma\lambda^{\kappa_1+\kappa_7}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T \frac{c_t^S}{1+z_t^\lambda}dt\right]$$
$$= o\left(\lambda^{\frac{2}{p+2}}\right).$$

It remains to estimate

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^T c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right] + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\Delta\varphi}}^T c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right].$$

First, by Lemma 2.5.10, the second part is of order $o(\lambda^{\frac{2}{p+2}})$. Second, for fixed $\lambda \in (0,1]$,

$$\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt \leqslant \int_0^{\tau^{\Delta\varphi}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt.$$

The family

$$\left(\int_0^{\tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right)_{\lambda\in(0,1]}$$

is uniformly integrable with respect to $\hat{\mathbb{Q}}$ by Jensen's inequality and the estimates made in the proof of Lemma 2.5.8 (cf. (2.5.21)). Indeed for some $\delta > 0$ as in the proof of Lemma 2.5.8,

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right)^{1+\delta}\right] \leqslant C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\Delta\varphi}} \left(\lambda^{-\frac{2}{p+2}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2\right)^{1+\delta} dt\right],$$

which was proved to be uniformly bounded in λ. Then, the family

$$\left(\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 dt\right)_{\lambda\in(0,1]}$$

is uniformly integrable too. By Lemma 2.6.5, $\tau^{\lambda,\mathrm{dual}}$ converges in probability under $\hat{\mathbb{Q}}$ to $\tau^{\Delta\varphi}$, so we have

$$\mathbb{1}_{\left\{\tau^{\lambda,\mathrm{dual}}\leqslant t\leqslant\tau^{\Delta\varphi}\right\}}\lambda^{-\frac{2}{p+2}} c_t^S\left(\Delta\varphi_t^\lambda\right)^2 \longrightarrow 0 \quad \text{in probability under } \hat{\mathbb{Q}} \text{ as } \lambda\to 0.$$

By the previous statement, indexed by λ this family is also uniformly integrable with respect to $\text{Lebesgue}_{|[0,T]} \times \hat{\mathbb{Q}}$ and by [101, Proposition 4.12] we obtain

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{\tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}} c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] \longrightarrow 0 \text{ as } \lambda \to 0.$$

and

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^{T} c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] = o\left(\lambda^{\frac{2}{p+2}}\right). \tag{2.6.16}$$

This yields

$$A_8 = \gamma^2 \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right)$$

as well as

$$\begin{aligned}
A_4 &= o\left(\lambda^{\frac{2}{p+2}}\right), \\
A_5 &= \gamma^2 \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right).
\end{aligned}$$

We now turn to A_6. A Taylor expansion with Cauchy remainder of the exponential function around 0, the definition of N^λ (cf. (2.6.6)) and $c_t^{S^\lambda,S} = c_t^S(1+z_t^\lambda)$ give

$$\begin{aligned}
A_6 &= \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(N^\lambda_{\tau^{\lambda,\mathrm{dual}}} + \frac{1}{2}\left(N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)^2 \exp\left(\chi_5 N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)\right)\left(\int_0^T \Delta\varphi_t^\lambda \sqrt{c_t^S} dW_t^{\hat{\mathbb{Q}}}\right)\right] \\
&= \gamma \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)^2\right] \\
&\quad - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_{\tau^{\lambda,\mathrm{dual}}}^T \gamma \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)\left(\int_0^T \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)\right] \\
&\quad - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma \sqrt{c_t^S} \Delta\varphi_t^\lambda \left(1 - \frac{1}{1+z_t^\lambda}\right) dW_t^{\hat{\mathbb{Q}}}\right)\left(\int_0^T \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)\right] \\
&\quad - \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\lambda,\mathrm{dual}}} \sqrt{c_t^S} \frac{\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}} dW_t^{\hat{\mathbb{Q}}}\right)\left(\int_0^T \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)\right] \\
&\quad - \frac{1}{2} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \frac{\left(\mu_t^{S^\lambda}\right)^2 c_t^S}{\left(c_t^{S^\lambda,S}\right)^2} dt \int_0^T \sqrt{c_t^S} \Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right] \\
&\quad + \frac{1}{2} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)^2 \exp\left(-\chi_5 N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)\left(\int_0^T \Delta\varphi_t^\lambda \sqrt{c_t^S} dW_t^{\hat{\mathbb{Q}}}\right)\right].
\end{aligned}$$

We want now to estimate the last five terms. Define

$$A_9 := A_6 - \gamma \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^T \sqrt{c_t^S}\Delta\varphi_t^\lambda dW_t^{\hat{\mathbb{Q}}}\right)^2\right].$$

The Cauchy-Schwarz inequality and the Itô isometry, applied to the remaining terms, give

$$\begin{aligned}
A_9 \leqslant & \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right]^{\frac{1}{2}} \Bigg(\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_{\tau^{\lambda,\mathrm{dual}}}^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right]^{\frac{1}{2}} \\
&+ \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} \gamma^2 c_t^S \left(\Delta\varphi_t^\lambda\right)^2 \left(1 - \frac{1}{1+z_t^\lambda}\right)^2 dt\right]^{\frac{1}{2}} \\
&+ \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{\tau^{\lambda,\mathrm{dual}}} c_t^S \left(\frac{\mu_t^{S^\lambda} + \gamma c_t^S \Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right)^2 dt\right]^{\frac{1}{2}} \\
&+ \frac{1}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{\tau^{\lambda,\mathrm{dual}}} \frac{\left(\mu_t^{S^\lambda}\right)^2 c_t^S}{\left(c_t^{S^\lambda,S}\right)^2} dt\right)^2\right]^{\frac{1}{2}} \\
&+ \frac{1}{2}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\left(N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)^2 \exp\left(-\chi_5 N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)\right)^2\right]^{\frac{1}{2}}\Bigg).
\end{aligned}$$

In view of Proposition 2.5.11, the first term is of order $O(\lambda^{\frac{1}{p+2}})$. By Equation (2.6.16), the second term is of order $o(\lambda^{\frac{1}{p+2}})$. By definition of $\tau^{\lambda,\mathrm{dual}}$, κ_5, κ_7, κ_8 (and a Taylor expansion similar to (2.6.15) for the third term) the last four terms are of order $o(\lambda^{\frac{1}{p+2}})$, too. In summary,

$$A_6 = \gamma \mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^T c_t^S \left(\Delta\varphi_t^\lambda\right)^2 dt\right] + o\left(\lambda^{\frac{2}{p+2}}\right).$$

Finally, (2.6.9) implies that, for some positive constant C and any $\chi_6 \in [0,1]$:

$$\left|\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}} - 1\right| \leqslant \frac{1}{2} \quad \text{and} \quad \exp\left(3\chi_6 N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right) \leqslant C.$$

Together with the bound on $N^\lambda_{\tau^{\lambda,\mathrm{dual}}}$ induced by stopping (see the definition of $\tau^{\lambda,\mathrm{dual}}$ in (2.6.7) and of κ_8), it follows that

$$\begin{aligned} A_7 &\leqslant 4\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|\frac{d\hat{\mathbb{Q}}^\lambda}{d\hat{\mathbb{Q}}}-1\right|^3\right] = 4\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right|^3 \exp\left(3\chi_6 N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right)\right] \\ &\leqslant C\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|N^\lambda_{\tau^{\lambda,\mathrm{dual}}}\right|^3\right] = o\left(\lambda^{\frac{2}{p+2}}\right), \end{aligned}$$

for some (random) $\chi_6 \in [0,1]$, where the order comes from the definition of κ_8. This completes the proof. □

We now need to prove the result used in the derivation of estimate (2.6.16), that $\tau^{\lambda,\mathrm{dual}}$ converges in probability under $\hat{\mathbb{Q}}$ to $\tau^{\Delta\varphi}$. To this end, we first proove the following lemma.

Lemma 2.6.4. *Assume Assumption 3* (i) *and* (ii) *holds. For arbitrarily small $\varepsilon' > 0$, the following limits hold in probability under $\hat{\mathbb{Q}}$ as $\lambda \to 0$:*

(i) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{1}{p+2}+5\kappa_2+\varepsilon'}\Delta\varphi^\lambda_t \longrightarrow 0,$

(ii) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{10(p-1)\kappa_2+\varepsilon'} g_p\left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi^\lambda_t\right) \longrightarrow 0,$

(iii) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{10\kappa_2+\varepsilon'} \widetilde{g}_p\left(\lambda^{-\frac{1}{p+2}} m_t \Delta\varphi^\lambda_t\right) \longrightarrow 0,$

(iv) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}+10\kappa_2+\varepsilon'}\left|z^\lambda_t\right| \longrightarrow 0,$

(v) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{1}{p+2}+5\kappa_2+\varepsilon'}\left|Z^\lambda_t\right| \longrightarrow 0,$

(vi) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{2}{p+2}+5(2p+3)\kappa_2+\varepsilon'}\left|\frac{\mu^{S^\lambda}_t+\gamma c^S_t\Delta\varphi^\lambda_t}{c^{S^\lambda,S}}\right| \longrightarrow 0,$

(vii) $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\varepsilon'}\left|\frac{c^{M^\lambda}_t}{c^S_t}\right| \longrightarrow 0,$

(viii) $\max_{0\leqslant t\leqslant \tau^{\lambda,dual}} \lambda^{-\frac{2}{p+2}+10\kappa_2+\varepsilon'}\left|N^\lambda_t\right|^2 \longrightarrow 0.$

Proof. First, note that by Lemma 2.7.7, we have for any $\varepsilon' > 0$

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{1}{p+2}}\Delta\varphi^\lambda_t\right] \leqslant \mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant \xi\leqslant \xi^{\lambda,m}} \widetilde{\Delta\varphi}^\lambda_\xi\right] = o\left(\lambda^{-5\kappa_2-\varepsilon'}\right).$$

Whence, $\max_{0\leqslant t\leqslant \tau^{\Delta\varphi}} \lambda^{-\frac{1}{p+2}+5\kappa_2+\varepsilon'}\Delta\varphi_t^\lambda$ converges to 0 in $L^1(\hat{\mathbb{Q}})$ and therefore in probability under $\hat{\mathbb{Q}}$. By Corollary 2.7.1,

$$g_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right) \leqslant C\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)^{2(p-1)} + C,$$
$$\widetilde{g}_p\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right) \leqslant C\left(\lambda^{-\frac{1}{p+2}}m_t\Delta\varphi_t^\lambda\right)^{2} + C.$$

Note as well that for a finite continuous process B on $[0,T]$,

$$\max_{0\leqslant t\leqslant T} \lambda^{\varepsilon'}B_t \longrightarrow 0 \quad \text{in probability under } \hat{\mathbb{Q}}, \text{ as } \lambda\to 0, \tag{2.6.17}$$

for any positive constant $\varepsilon' > 0$. The process m is continuous and finite for all $t\in[0,T]$, so the second and third limits follow from the bound on the function g_p and $\widetilde{g}_p$ and the first limit.

The limits for z^λ, Z^λ, $\left|\frac{\mu_t^{S^\lambda}+\gamma c_t^S\Delta\varphi_t^\lambda}{c^{S^\lambda,S}}\right|$ and $\left|\frac{c_t^{M^\lambda}}{c_t^S}\right|$ come from the definitions (2.6.5), (2.6.6), (2.6.3), (2.6.4), (2.6.1), the fact that the processes Λ, m, $\hat{\varphi}$, $c^{\hat{\varphi}}$, c^S, their drifts and instantaneous quadratic variations are continuous, (2.6.17), and the first three limits already established.

Finally, Itô's isometry shows that

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|N^\lambda_{t\wedge\tau^{\lambda,\mathrm{dual}}}\right|^2\right] &\leqslant C\left(\mathbb{E}_{\hat{\mathbb{Q}}}\left[\int_0^{t\wedge\tau^{\lambda,\mathrm{dual}}}\left(Z_t^\lambda\right)^2dt\right] + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\int_0^{t\wedge\tau^{\lambda,\mathrm{dual}}}\left(Z_t^\lambda\right)^2dt\right)^2\right]\right)\\ &\leqslant CT\lambda^{2\kappa_6} + CT^2\lambda^{4\kappa_6}.\end{aligned}$$

Therefore, $\max_{0\leqslant t\leqslant\tau^{\lambda,\mathrm{dual}}}\lambda^{-2\kappa_6+\varepsilon''}|N_t|$ converges to 0 in $L^1(\hat{\mathbb{Q}})$ and so in probability under $\hat{\mathbb{Q}}$, for some $\varepsilon''>0$ arbitrarily small. This completes the proof. □

With this, we can now establish the last missing piece for our tight upper duality bound:

Lemma 2.6.5. *Suppose Assumptions 3* (i) *and* (ii) *is satisfied. Then the stopping time $\tau^{\lambda,dual}$ converges in probability under $\hat{\mathbb{Q}}$ to $\tau^{\Delta\varphi}$,*

$$\hat{\mathbb{Q}}\left[\tau^{\lambda,dual} < \tau^{\Delta\varphi}\right] \longrightarrow 0 \ \text{as}\ \lambda\to 0.$$

Proof. This is a simple consequence of Lemma 2.6.4,

$$\begin{aligned}\left\{\tau^{\lambda,\mathrm{dual}}<\tau^{\Delta\varphi}\right\} \subseteq &\left\{\sup_{t\in[0,T]}\left|z_t^\lambda\right|>\lambda^{\kappa_5}\right\}\cup\left\{\sup_{t\in[0,T]}\left|N_t^\lambda\right|^2>\lambda^{\kappa_8}\right\}\cup\left\{\sup_{t\in[0,T]}\left|Z_t^\lambda\right|>\lambda^{\kappa_6}\right\}\\ &\cup\left\{\sup_{t\in[0,T]}\frac{c_t^{M^\lambda}}{c_t^S}>1\right\}\cup\left\{\sup_{t\in[0,T]}\left|\frac{\mu_t^{S^\lambda}+\gamma c_t^S\Delta\varphi_t^\lambda}{c_t^{S^\lambda,S}}\right|>\lambda^{\kappa_7}\right\},\end{aligned}$$

and the fact that

$$\kappa_5 < \frac{2}{p+2} - 10\kappa_2 - \varepsilon', \quad \kappa_6 < \frac{1}{p+2} - 5\kappa_2 - \varepsilon',$$
$$\kappa_7 < \frac{2}{p+2} - 5(2p+3)\kappa_2 - \varepsilon', \quad \kappa_8 < \frac{2}{p+2} - 10\kappa_2 - \varepsilon'$$

for ε' small enough. □

2.7 Technical Results

2.7.1 Results on the Function g_p.

For the convenience of the reader, we recall the main steps of the proof in [80], which will be used to prove the second characterisation of c_p below.

Remark on the proof of Lemma 2.2.1 [80, Lemma 20] shows that for every $c > 0$ there is a unique solution with the required growth at $+\infty$. By symmetry, there exists a unique solution for each $c > 0$ that has the required growth condition at $-\infty$. [80, Lemma 22] in turn establishes that there is a unique c_p for which the corresponding solution matches both growth conditions. As a by-product of the proof, the solution g_p is odd and satisfies $g_p(0) = 0$.

Proof of Lemma 2.2.2 To prove this second characterization, let $c < c_p$, and consider the ODE

$$g'(z) = (p-1)p^{-\frac{p}{p-1}} |g(z)|^{\frac{p}{p-1}} - z^2 + c \tag{2.7.1}$$

with initial condition $g(0) = 0$. Then a comparison argument yields that $g_p(z) > g(z)$ for all $z > 0$. Note that a solution of (2.7.1) starting at 0 is necessarily odd. Since (g_p, c_p) is the only solution of (2.2.1) that satisfies the growth condition (2.2.2) and such that $g_p(0) = 0$, there exists a positive constant δ such that

$$\liminf_{z\to\infty} \frac{g(z)}{|z|^{\frac{2(p-1)}{p}}} = M =: p(p-1)^{-\frac{p-1}{p}} - 2\delta.$$

For some sufficiently large $z_0 \in \mathbb{R}_+$, plugging this back in the ODE we then have $g(z_0) < (M+\delta)\,|z_0|^{\frac{2(p-1)}{p}}$ and $g'(z_0) < m < 0$. Since g satisfies (2.7.1), the function g is decreasing for all $z \geqslant z_0$ such that $g(z) > -p\left|\frac{1}{p-1}\left(z^2 - c\right)\right|^{\frac{p-1}{p}}$. It holds additionally that $g'(z) < m$ for all $z > z_0$ such that $g(z) \geqslant 0$. This yields that for some

$z > z_0$, $g(z) < 0$ and g is not positive on $\mathbb{R}_+$. $\square$

The next result establishes some further properties of g_p that are used in the proof of Theorem 2.2.3:

Corollary 2.7.1. *(i) There exists a constant $C > 0$ such that*

$$|\widetilde{g}_p(z)|^j = |g_p(z)|^{\frac{j}{p-1}} \leqslant C\left(|z|^{2j}+1\right), \ z \in \mathbb{R}, \ j \in \{1, p-1, p, p+1, 2(p-1), 2p\}. \tag{2.7.2}$$

(ii) There exists a constant $C > 0$ such that

$$|z\tilde{g}_p(z)| \geqslant C\left(|z|^{\frac{2+p}{p}} - 1\right), \quad z \in \mathbb{R} \tag{2.7.3}$$

$$|z\tilde{g}_p(z)| \geqslant C\,|z|^2\,\mathbb{1}_{\{|z|\geqslant 1\}}, \quad z \in \mathbb{R}. \tag{2.7.4}$$

Proof. Item (i) follows from continuity of g_p, its growth rate at infinity (2.2.2), and $2/p \in (1,2)$.

For Item (ii), recall from Lemma 2.2.1 that $g_p(0) = 0$ and $c_p > 0$. Hence, since g_p satisfies (2.2.1), it holds that $g_p(z) > \frac{c_p}{2}z$ on a sufficiently small interval around 0. On $(0, \sqrt{c_p}]$ we then have $g_p'(z) = -z^2 + c_p + (p-1)p^{-\frac{p}{p-1}}\,|g_p(z)|^{\frac{p}{p-1}} > 0$. On $[0, \sqrt{c_p}]$, the function $z \mapsto g_p(z)/z$ therefore is continuous and bounded from below by a constant C. The function g_p is odd, so that $|g_p(z)| \geqslant C\,|z|$ on $\left[-\sqrt{c_p}, \sqrt{c_p}\right]$.

The growth condition for g_p at infinity gives the existence of constants C' and K such that $|g_p(z)| \geqslant C'\,|z|^{\frac{2(p-1)}{p}}$ for $|z| \geqslant K$. Moreover, on $[\sqrt{c_p}, K]$ it holds that $g_p'(z) \geqslant 0$. Then $|g_p(z)|\,/\,|z|$ is bounded from below on $[0, \sqrt{c_p}]$ by C, is continuous and bounded from below by $g_p(\sqrt{c_p})/K$ on $[\sqrt{c_p}, K]$. This implies that the following holds for some $C'' \leqslant (C \wedge C' \wedge g_p(\sqrt{c_p})/K)^{\frac{1}{p-1}}$:

$$|\widetilde{g}_p(z)| = |g_p(z)|^{\frac{1}{p-1}} \geqslant C''\,|z|^{\frac{1}{p-1}}, \quad z \in [-K, K], \tag{2.7.5}$$

$$|\widetilde{g}_p(z)| = |g_p(z)|^{\frac{1}{p-1}} \geqslant C''\,|z|^{\frac{2}{p}}, \quad z \in \mathbb{R}\setminus[-K, K]. \tag{2.7.6}$$

Then, the function $|g_p(z)|\,/\,|z|^{\frac{2(p-1)}{p}}$ is strictly positive, continuous and therefore bounded from below on $\mathbb{R}\setminus[-1,1]$. Similarly, $|g_p(z)|\,/\,|z|$ is strictly positive, continuous and therefore bounded from below on $[-1,1]$. This gives the inequalities (2.7.3) and (2.7.4).

To show that $g_p'(z) \geqslant 0$ on $[\sqrt{c_p}, K]$, assume on the contrary that there is a z_1 in the interval such that $g_p'(z_1) < 0$. Then for all $z \geqslant z_1$ with $g_p(z) \geqslant -p|\frac{1}{p-1}(z^2-c_p)|^{\frac{p-1}{p}}$, it holds that $g_p'(z) < 0$, which contradicts the growth condition for g_p at $+\infty$. $\square$

2.7.2 Results on the Displacement Process and its Rescaled Version

This section gathers results on the displacement process, its rescaled version, and estimates on the different stopping times used to define the candidate strategies.

We first prove an existence and uniqueness result for the SDE driving the displacement process. This is necessary to define rigorously the candidate strategies. These strategies follow at a finite rate the frictionless optimizer, therefore the stock holdings are always bounded by the running maximum of the frictionless optimum, cf. Lemma 2.7.3.

Then we bound the probability that $\tau^{\Delta\varphi}$ occurs strictly before the end of the trading interval $[0, T^\lambda]$. This is done by applying a maximal inequality of Peskir [146] to the rescaled displacement.

Finally, we prove a uniform moment estimate for $\widetilde{\Delta\varphi}^\lambda$, independent of λ, that allows us to establish a uniform integrability result necessary for the ergodic convergence statements of Section 2.5.1.

2.7.2.1 Existence Result for an SDE

Proposition 2.7.2. *Under $\widetilde{\mathbb{Q}}^\lambda$, the SDE* (2.4.4),

$$d\widetilde{\Delta\varphi}_\xi^\lambda = -\frac{1}{2}p^{-\frac{1}{p-1}} m_{\lambda\varsigma u_\xi^\lambda} \widetilde{g}_p \left(m_{\lambda\varsigma u_\xi^\lambda} \widetilde{\Delta\varphi}_\xi^\lambda \right) \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\xi + \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\widetilde{W}_\xi^\lambda,$$

with initial condition $\widetilde{\Delta\varphi}_0^\lambda = 0$, has a unique strong solution on $\mathbb{R}_+$.

Proof. Define for $n \in \mathbb{N}$, the bounded function

$$\widetilde{g}_p^{(n)}(x) = \operatorname{sgn}(x) \left(|\widetilde{g}_p(x)| \wedge n \right),$$

and consider the same SDE as above with truncated drift,

$$d\widetilde{\Delta\varphi}_\xi^{\lambda,n} = -\frac{1}{2}p^{-\frac{1}{p-1}} m_{\lambda\varsigma u_\xi^\lambda} \widetilde{g}_p^{(n)} \left(m_{\lambda\varsigma u_\xi^\lambda} \widetilde{\Delta\varphi}_\xi^\lambda \right) \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\xi + \mathbb{1}_{\left\{\xi \leqslant \xi^{\lambda,m}\right\}} d\widetilde{W}_\xi^\lambda. \quad (2.7.7)$$

The function $\widetilde{g}_p^{(n)}$ is Lipschitz continuous with Lipschitz constant $K_{p,n}$ (because g_p is continuously differentiable and monotone on $\mathbb{R}$). The function $f : (\xi, \omega, x) \mapsto -m_{\lambda\varsigma u_\xi^\lambda} \widetilde{g}_p^{(n)}(m_{\lambda\varsigma u_\xi^\lambda} x)$ satisfies

$$|f(t,\omega,x) - f(t,\omega,y)| \leqslant K_{p,n} m_{\lambda\varsigma u_\xi^\lambda}(\omega)^2 |x-y|, \quad \text{for all } (x,y) \in \mathbb{R}^2.$$

The random variable $K = K_{p,n} \sup_{t\in[0,T]} m_t^2$ is $\hat{\mathbb{Q}}$-almost surely finite by almost sure continuity of m on $[0,T]$. By [153, Theorem V.7], there exists a unique strong solution of the SDE (2.7.7) on $(\Omega, \mathcal{G}, \mathbb{G}^\lambda, \widetilde{\mathbb{Q}}^\lambda, \widetilde{W}^\lambda)$. Define the stopping time

$$\tau^n = \inf\left\{\xi \in [0,\xi^{\lambda,m}] \;\middle|\; \left|m_{\lambda\varsigma u_\xi^\lambda}\Delta\varphi_\xi^{\lambda,n}\right| \geqslant \widetilde{g}_p^{-1}(n)\right\} \wedge \xi^{\lambda,m}. \tag{2.7.8}$$

On $[\![0,\tau^n]\!]$, the processes $\widetilde{\Delta\varphi}^\lambda$ and $\widetilde{\Delta\varphi}^{\lambda,n}$ satisfy the same SDE with the same initial condition and are therefore equal $\widetilde{\mathbb{Q}}^\lambda$-a.s. The process $((\widetilde{\Delta\varphi_t}^{\lambda,n})^2)_{\xi\in\mathbb{R}_+}$ satisfies the following SDE:

$$\begin{aligned} d\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2 &= \left(1 - p^{-\frac{1}{p-1}} m_{\lambda\varsigma u_\xi^\lambda}\widetilde{\Delta\varphi_\xi}^{\lambda,n}\widetilde{g}_p^{(n)}\left(m_{\lambda\varsigma u_\xi^\lambda}\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)\right)\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}}d\xi \\ &\quad + 2\sqrt{\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2}\operatorname{sgn}\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}}d\widetilde{W}_\xi^\lambda. \end{aligned}$$

The process $W^{(\lambda,n)}$ defined as

$$\widetilde{W}_\xi^{(\lambda,n)} = \int_0^\xi \left(\operatorname{sgn}\left(\widetilde{\Delta\varphi_y}^{\lambda,n}\right)\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}} + \mathbb{1}_{\{\xi>\xi^{\lambda,m}\}}\right)d\widetilde{W}_y^\lambda$$

is a $\left(\widetilde{\mathbb{Q}}^\lambda, \mathbb{G}^\lambda\right)$-Brownian motion on $[\![0,\xi_T^\lambda]\!]$ by Lévy's characterisation [113, Theorem 3.3.16], and $x \mapsto x\widetilde{g}_p^{(n)}(x)$ is an even function. Therefore, we can rewrite the previous equation as

$$\begin{aligned} d\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2 &= \left(1 - p^{-\frac{1}{p-1}} m_{\lambda\varsigma u_\xi^\lambda}\sqrt{\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2}\widetilde{g}_p^{(n)}\left(m_{\lambda\varsigma u_\xi^\lambda}\sqrt{\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2}\right)\right)\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}}d\xi \\ &\quad + 2\sqrt{\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2}\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}}d\widetilde{W}_\xi^{\lambda,n}. \end{aligned}$$

Define now the process $Y^{(\lambda,n)}$ as the unique strong solution of the following SDE (cf. [157, Chapter XI, p. 439]), with initial condition $Y_0^{(\lambda,n)} = 0$,

$$dY_\xi^{(\lambda,n)} = 2d\xi + 2\sqrt{Y_\xi^{(\lambda,n)}}\; d\widetilde{W}_\xi^{(\lambda,n)}.$$

This process is the square of a 2-dimensional Bessel process started at 0. It therefore has moments of all orders at all finite times [157, Chapter XI]. By the comparison

Theorem 2 in [139],[23] we obtain

$$\widetilde{\mathbb{Q}}^\lambda\left[Y_\xi^{(\lambda,n)} \geqslant \left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2, \forall \xi \in \left[0, \xi^{\lambda,m}\right]\right] = 1.$$

Hence, for $x > 0$, and n large enough,

$$\begin{aligned}
\widetilde{\mathbb{Q}}^\lambda\left[\tau^n < x \wedge \xi^{\lambda,m}\right] &\leqslant \widetilde{\mathbb{Q}}^\lambda\left[\sup\left\{\left|m_{\lambda\varsigma u^\lambda_\xi}\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right| \;\middle|\; \xi \in \left[0, x \wedge \xi^{\lambda,m}\right]\right\} \geqslant \widetilde{g}_p^{-1}(n)\right] \\
&\leqslant \widetilde{\mathbb{Q}}^\lambda\left[\sup\left\{\left(\widetilde{\Delta\varphi_\xi}^{\lambda,n}\right)^2 \;\middle|\; \xi \in \left[0, x \wedge \xi^{\lambda,m}\right]\right\} \geqslant \lambda^{2\kappa_2}\widetilde{g}_p^{-1}(n)^2\right] \\
&\leqslant \widetilde{\mathbb{Q}}^\lambda\left[\sup\left\{Y_\xi^{(\lambda,n)} \;\middle|\; \xi \in [0,x]\right\} \geqslant \lambda^{2\kappa_2}\widetilde{g}_p^{-1}(n)^2\right] \\
&\leqslant \frac{\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\left(Y_x^{(\lambda,n)}\right)^+\right]}{\lambda^{2\kappa_2}\widetilde{g}_p^{-1}(n)^2} \longrightarrow 0 \text{ as } n \to \infty,
\end{aligned}$$

where the third inequality follows from the above comparison argument, and the fourth from Doob's martingale inequality applied to the non-negative submartingale $Y^{(\lambda,n)}$. This shows that for arbitrary $x > 0$,

$$\widetilde{\mathbb{Q}}^\lambda\left[\lim_{n\to\infty} \tau^n \wedge x \wedge \xi^{\lambda,m} = x \wedge \xi^{\lambda,m}\right] = 1,$$

and the solution of (2.4.4) exists $\widetilde{\mathbb{Q}}^\lambda$-a.s. on $\mathbb{R}_+$ and in particular, on $[\![0, \xi^{\lambda,m}]\!]$ ($\xi^{\lambda,m}$ is almost surely finite by the continuity Assumption 2). □

2.7.2.2 Bounding the Candidate Strategy by the Frictionless Optimizer

Lemma 2.7.3. *Suppose that $\left|\varphi_0^\lambda\right| \leqslant |\hat{\varphi}_0|$. Then the candidate strategy φ_t^λ satisfies*

$$\left|\varphi_t^\lambda\right| \leqslant \hat{\varphi}_t^*, \ \forall t \in [0,T].$$

As a consequence,

$$\left|\hat{\varphi}_t - \varphi_t^\lambda\right| \leqslant 2\hat{\varphi}_t^*, \ \forall t \in [0,T]. \tag{2.7.9}$$

In other words, existence of moments for the supremum of the frictionless optimal strategy on $[0,T]$ implies the existence of the corresponding moments for the displacement $\Delta\varphi_t^\lambda$.

[23]Taking, following the notation of the Theorem in [139], $g(s,\omega,x) = \sqrt{x}$, $\rho(x) = \sqrt{x}$, $C(s,\omega) = 1$, $\beta(s,\omega) = f(s,\omega,x) = 1 - p^{-\frac{1}{p-1}} m_{\lambda\varsigma u_s^\lambda(\omega)}\sqrt{x}\widetilde{g}_p^{(n)}\left(m_{\lambda\varsigma u_s^\lambda(\omega)}\sqrt{x}\right)$, $\widetilde{\beta}(s,\omega) = \widetilde{f}(s,\omega,x) = 2$, the assumptions on the drift and volatility are satisfied.

Proof of Lemma 2.7.3. It suffices to show that $\left|\varphi^{\lambda}\right| \leqslant \hat{\varphi}^*$ on $[\![0,\tau^{\Delta\phi}]\!]$ since $|\varphi^{\lambda}| \leqslant |\varphi^{\lambda}_{\tau^{\Delta\varphi}}|$ on $[\![\tau^{\Delta\phi},T]\!]$ by definition of φ^{λ}. Fix $\omega\in\Omega$ and let $\tau_0=\inf\{t\in[0,\tau^{\Delta\varphi}] \mid |\varphi^{\lambda}_t| > \hat{\varphi}^*_t\}$. Assume that $\tau_0\in[0,\tau^{\Delta\varphi})$. By continuity of φ^{λ} and $\hat{\varphi}^*$, we have $|\varphi^{\lambda}_{\tau_0}|=\hat{\varphi}^*_{\tau_0}$. Furthermore, by definition of the infimum, there exist $\varepsilon>0$ and $\tau_1\in(\tau_0,\tau^{\Delta\varphi})$ such that $|\varphi^{\lambda}_{\tau_1}| > \hat{\varphi}^*_{\tau_1}+\varepsilon$. Let $\tau_2=\inf\{t\in[0,\tau_1] \bigm| |\varphi^{\lambda}_t| > \hat{\varphi}^*_t+\frac{\varepsilon}{2}\}$. By continuity of $\hat{\varphi}^*$ and φ^{λ} and the definition of τ_0 and τ_1, it holds that $\tau_0<\tau_2<\tau_1$. Without loss of generality we can assume that $\varphi^{\lambda}_{\tau_2} > \hat{\varphi}^*_{\tau_2} \geqslant \hat{\varphi}_{\tau_2}$ (the case where $-\varphi^{\lambda}_{\tau_2} > \hat{\varphi}^*_{\tau_2} \geqslant \hat{\varphi}_{\tau_2}$ is treated similarly), which implies by definition of $\dot{\varphi}^{\lambda}$ (see (2.2.4)) that $\dot{\varphi}^{\lambda}_{\tau_2}=\lim_{t\to\tau_2^+}\frac{\varphi^{\lambda}_t-\varphi^{\lambda}_{\tau_2}}{t-\tau_2}<0$. However, by definition of τ_2, for every $\delta>0$ there exists $\tau^{\delta}\in(\tau_2,\tau_2+\delta)$ such that $\varphi^{\lambda}_{\tau^{\delta}} > \varphi^{\lambda}_{\tau_2}$. This contradicts the existence of a negative limit. $\square$

2.7.2.3 Stopping Time Bounds

We have defined the stopping time $\tau^{\Delta\varphi}$ in Section 2.2.1 by

$$\tau^{\Delta\varphi}=\inf\Bigg\{t\in[0,T^{\lambda}]:\left|\Delta\varphi^{\lambda}_t\right|>\lambda^{\kappa_1} \text{ or } m_t<\lambda^{\kappa_2} \text{ or } m_t>\lambda^{-\kappa_2} \\ \text{or } \int_0^t \lambda\Lambda_t\left|\dot{\varphi}_t\right|^p dt>\lambda^{\kappa_3} \text{ or } |\hat{\varphi}_t|\geqslant\lambda^{-\kappa_4}\Bigg\}\wedge T^{\lambda}.$$

We can rewrite this stopping time as the minimum of four simpler stopping times:

$$\tau^{\Delta\varphi}=\tau^{\lambda,m}\wedge\tau^{\lambda,\Delta\varphi}\wedge\tau^{\lambda,\text{cost}}\wedge\tau^{\lambda,\hat{\varphi}},$$

where

$$\begin{aligned}
\tau^{\lambda,m}&=\inf\left\{t\in[0,T]:m_t<\lambda^{\kappa_2} \text{ or } m_t>\lambda^{-\kappa_2}\ \right\}\wedge T^{\lambda},\\
\tau^{\lambda,\Delta\varphi}&=\inf\left\{t\in[0,\tau^{\lambda,m}]\ \middle|\ \left|\Delta\varphi^{\lambda}\right|>\lambda^{\kappa_1}\right\}\wedge\tau^{\lambda,m},\\
\tau^{\lambda,\text{cost}}&=\inf\left\{t\in[0,\tau^{\lambda,\Delta\varphi}]:\int_0^t\lambda\Lambda_u\left|\dot{\varphi}^{\lambda}_u\right|^p du\geqslant\lambda^{\kappa_3}\right\}\wedge\tau^{\lambda,\Delta\varphi},\\
\tau^{\lambda,\hat{\varphi}}&=\inf\left\{t\in[0,T]:|\hat{\varphi}_t|\geqslant\lambda^{-\kappa_4}\right\}\wedge T^{\lambda}.
\end{aligned}$$

Lemma 2.7.4. *Suppose that $(m^{-1})^*_T$ has $\frac{4(1+2\varepsilon)}{\kappa_2(p+2)}$-th moment (Assumption 3 (iii)), then the following estimate holds:*

$$\hat{\mathbb{Q}}\left[\tau^{\lambda,m}<T^{\lambda}\right]=O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right). \tag{2.7.10}$$

Proof. The process m is $\hat{\mathbb{Q}}$-a.s. positive on $[0,T]$, and m_T^* and $(m^{-1})_T^*$ has $\frac{4(1+2\varepsilon)}{\kappa_2(p+2)}$-th moment by assumption. Therefore, Markov's inequality yields

$$\begin{aligned}
\hat{\mathbb{Q}}\left[\tau^{\lambda,m} < T^\lambda\right] &\leqslant \hat{\mathbb{Q}}\left[\inf_{t\in[0,T]} m_t < \lambda^{\kappa_2}\right] + \hat{\mathbb{Q}}\left[\sup_{t\in[0,T]} m_t > \lambda^{-\kappa_2}\right] \\
&= \hat{\mathbb{Q}}\left[\sup_{t\in[0,T]} \frac{1}{m_t} > \lambda^{-\kappa_2}\right] + \hat{\mathbb{Q}}\left[\sup_{t\in[0,T]} m_t > \lambda^{-\kappa_2}\right] \\
&\leqslant \lambda^{\frac{4(1+2\varepsilon)}{p+2}} \left(\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\sup_{t\in[0,T]} \frac{1}{m_t}\right)^{\frac{4(1+2\varepsilon)}{\kappa_2(p+2)}}\right] + \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\sup_{t\in[0,T]} m_t\right)^{\frac{4(1+2\varepsilon)}{\kappa_2(p+2)}}\right]\right) \\
&= O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right).
\end{aligned}$$

□

Lemma 2.7.5. *Suppose that* $\left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}$ *and* $\exp\left(\varepsilon\Lambda_T^*\right)$ *belong to* $L^1(\hat{\mathbb{Q}})$ *(Assumption 3(i)). Then the stopping time*

$$\tau^{\lambda,cost} = \inf\left\{t\in[0,\tau^{\lambda,\Delta\varphi}] : \int_0^t \lambda\Lambda_u \left|\dot{\varphi}_u^\lambda\right|^p du \geqslant \lambda^{\kappa_3}\right\} \wedge \tau^{\lambda,\Delta\varphi}$$

satisfies $\hat{\mathbb{Q}}[\tau^{\lambda,cost} < \tau^{\lambda,\Delta\varphi}] = O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right)$.

Proof. Taking into account the ODE for φ^λ and Corollary 2.7.1, it follows that – for $\lambda \leqslant 1$, a positive constant C which may change from line to line, $t \leqslant \tau^{\lambda,\Delta\varphi}$ and the fact that $\kappa_1 < 1/(p+2)$– we have

$$\begin{aligned}
\left|\dot{\varphi}_t^\lambda\right| &\leqslant C c_t^{\hat{\varphi}} \lambda^{-\frac{1}{p+2}} m_t \left(1 + m_t^2 \lambda^{-\frac{2}{p+2}} \left(\Delta\varphi_t^\lambda\right)^2\right) \leqslant C c_t^{\hat{\varphi}} \lambda^{-\frac{1}{p+2}} m_t \left(1 + m_t^2 \lambda^{-\frac{2}{p+2}+2\kappa_1}\right) \\
&\leqslant C\lambda^{-\frac{3}{p+2}+2\kappa_1} c_t^{\hat{\varphi}} (1+m_t^3) \leqslant C\lambda^{-\frac{3}{p+2}+2\kappa_1}\left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right).
\end{aligned}$$

Thus, by the inequalities of Markov and Hölder, we obtain under the stated integrability assumptions that

$$
\begin{aligned}
\hat{\mathbb{Q}}[\tau^{\lambda,\mathrm{cost}} < \tau^{\lambda,\Delta\varphi}] &\leqslant \hat{\mathbb{Q}}\left[\int_0^{\tau^{\lambda,\Delta\varphi}} \lambda\Lambda_u \left|\dot{\varphi}_t^\lambda\right|^p du \geqslant \lambda^{\kappa_3}\right] \\
&\leqslant \hat{\mathbb{Q}}\left[\Lambda_T^* C \lambda^{1-\frac{3p}{p+2}+2p\kappa_1}\left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^p \geqslant \lambda^{\kappa_3}\right] \\
&= C\lambda^{\frac{4(1+2\varepsilon)}{p+2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(\Lambda_T^*\left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^p\right)^{\frac{4(1+2\varepsilon)}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\right] \\
&= C\lambda^{\frac{4(1+2\varepsilon)}{p+2}} \mathbb{E}_{\hat{\mathbb{Q}}}\left[(\Lambda_T^*)^{\frac{4(1+2\varepsilon)(1+\frac{1}{\varepsilon})}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\right]^{\frac{1}{1+\frac{1}{\varepsilon}}} \\
&\quad \times \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(2p\kappa_1-\kappa_3)}}\right]^{\frac{1}{1+\varepsilon}} \\
&= O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right). \qquad \square
\end{aligned}
$$

The following lemma follows immediately from Markov's inequality:

Lemma 2.7.6. *Suppose that* $(\varphi_T^*)^{\frac{4(1+2\varepsilon)}{(p+2)\kappa_4}} \in L^1(\hat{\mathbb{Q}})$ *(Assumption 3* (i)*). Then the stopping time*

$$
\tau^{\lambda,\hat{\varphi}} = \inf\left\{t \in [0,T] : |\hat{\varphi}_t| \geqslant \lambda^{-\kappa_4}\right\} \wedge T^\lambda
$$

satisfies $\hat{\mathbb{Q}}[\tau^{\lambda,\hat{\varphi}} < T^\lambda] = O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right)$.

Finally, we prove a maximal inequality for the displacement process and its corollary, a bound for the probability that the stopping time $\tau^{\lambda,\Delta\varphi}$ is strictly smaller than T^λ.

Lemma 2.7.7 (Maximal inequality for $\Delta\varphi$)**.** *Supose that Assumptions 3* (i) *and* (ii) *are satisfied. Let $n \in \mathbb{N}$. For $0 < \lambda \leqslant 1$, and $\varepsilon' > 0$ small enough*

$$
\mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant\xi\leqslant\xi^{\lambda,m}}\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^n\right] = o\left(\lambda^{-5n\kappa_2+\varepsilon'}\right). \tag{2.7.11}
$$

Proof. The SDE satisfied by $\left(\widetilde{\Delta\varphi}^\lambda\right)^2$ under the probability $\widetilde{\mathbb{Q}}^\lambda$ is (cf. (2.4.4) and

Section 2.7.2.1):

$$d\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2 = \left(1 - p^{-\frac{1}{p-1}} m_{\lambda^\zeta u_\xi^\lambda}\sqrt{\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2}\widetilde{g}_p\left(m_{\lambda^\zeta u_\xi^\lambda}\sqrt{\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2}\right)\right)\mathbb{1}_{\left\{\xi\leqslant\xi^{\lambda,m}\right\}}d\xi$$
$$+ 2\sqrt{\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2}\mathbb{1}_{\left\{\xi\leqslant\xi^{\lambda,m}\right\}}dB_\xi^\lambda, \qquad \widetilde{\Delta\varphi}_0^\lambda = 0,$$

where B^λ is the $\widetilde{\mathbb{Q}}^\lambda$-Brownian motion on $[\![0,\xi_T^\lambda]\!]$ defined by

$$B_\xi^\lambda = \int_0^\xi \left(\operatorname{sgn}\left(\widetilde{\Delta\varphi}_y^\lambda\right)\mathbb{1}_{\{\xi\leqslant\xi^{\lambda,m}\}} + \mathbb{1}_{\{\xi>\xi^{\lambda,m}\}}\right)d\widetilde{W}_y^\lambda. \tag{2.7.12}$$

We have, by Corollary 2.7.1 (ii), that for $x \in \mathbb{R}$, $x\widetilde{g}_p(x) \geqslant C\left(|x|^2 - 1\right)$. Note that C can be chosen arbitrarily small. On $[\![0,\tau^{\lambda,m}]\!]$, we have $m_t \geqslant \lambda^{\kappa_2}$. Let $C' = p^{-\frac{1}{p-1}}C\lambda^{2\kappa_2}$, $C'' = p^{-\frac{1}{p-1}}C$, $0 < \varepsilon < 1$ and Z^ε, be the unique strong solution of the following SDE with initial condition $Z_0^\varepsilon = x_0 \geqslant 1$ (existence is obtained as in Section 2.7.2.1):

$$dZ_\xi^\varepsilon = \left(1 + \varepsilon + C'' - C'Z_\xi^\varepsilon\right)d\xi + 2\sqrt{Z^\varepsilon}dB_\xi^\lambda.$$

By a comparison argument [139, Corollary of Theorem 2], it holds that $\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^2 \leqslant Z_\xi^\varepsilon$ on $[\![0,\xi^{\lambda,m}]\!]$, $\widetilde{\mathbb{Q}}^\lambda$-a.s., where $\xi^{\lambda,m} = \lambda^{-\zeta}\int_0^{\tau^{\lambda,m}} c_s^{\hat{\varphi}}ds$. The SDE satisfied by $(Z^\varepsilon)^k$ for $k \in \mathbb{N}\backslash\{0\}$ is

$$d\left(Z_\xi^\varepsilon\right)^k = \left(k(2k-1+\varepsilon+C'')\left(\left(Z_\xi^\varepsilon\right)^k\right)^{\frac{k-1}{k}} - kC'\left(Z_\xi^\varepsilon\right)^k\right)d\xi + 2k\left(\left(Z_\xi^\varepsilon\right)^k\right)^{\frac{k-\frac{1}{2}}{k}}dB_\xi^\lambda.$$

Now we use the process Z^ε to prove that $\sup_{0\leqslant\xi\leqslant\widetilde{\tau}^{\Delta\varphi}}\left|\widetilde{\Delta\varphi}_\xi^\lambda\right|^n$, $n \in \mathbb{N}$, is in $L^1(\hat{\mathbb{Q}})$ using a result of Peskir [146, Theorem 2.5]. Following his proof, we define the scale function of the diffusion $(Z^\varepsilon)^k$,

$$q'(z) = \exp\left(-2\int_{x_0}^z \frac{k(2k-1+\varepsilon+C'')y^{\frac{k-1}{k}} - kC'y}{4k^2y^{\frac{2k-1}{k}}}dy\right) = C^{x_0}z^{-\frac{2k-1+\varepsilon+C''}{2k}}\exp\left(\frac{C'}{2}z^{\frac{1}{k}}\right),$$

where $C^{x_0} = x_0^{\frac{2k-1+\varepsilon+C''}{2k}}\exp\left(-\frac{C'}{2}x_0^{\frac{1}{k}}\right)$ and the speed measure

$$m(x_0,z] = \int_{x_0}^z \frac{2}{4k^2q'(y)}y^{-2+\frac{1}{k}}dy = \frac{1}{2k^2C^{x_0}}\int_{x_0}^z y^{\frac{1+\varepsilon+C''}{2k}-1}\exp\left(-\frac{C'}{2}y^{\frac{1}{k}}\right)dy$$
$$= \frac{1}{2kC^{x_0}}\int_{x_1}^{z^{\frac{1}{k}}} y^{\frac{1+\varepsilon+C''}{2}-1}\exp\left(-\frac{C'}{2}y\right)dy, \quad z \geqslant x_0.$$

where $x_1 = x_0^{\frac{1}{k}}$. We now define the function F (as in Peskir's article [146]) by,

$$\begin{aligned}
F(x) &= \int_{x_0}^{x} m(x_0, z]q'(z)dz \\
&= \frac{1}{2k}\int_{x_0}^{x} z^{-1-\frac{-1+\varepsilon+C''}{2k}} \exp\left(\frac{C'}{2}z^{\frac{1}{k}}\right)\int_{x_1}^{z^{\frac{1}{k}}} y^{\frac{1+\varepsilon+C''}{2}-1}\exp\left(-\frac{C'}{2}y\right)dy\ dz \\
&= \frac{1}{2}\int_{x_1}^{x^{\frac{1}{k}}} \mathrm{e}^{\frac{C'}{2}z} z^{-1+\frac{1-\varepsilon-C''}{2}}\int_{x_1}^{z} y^{\frac{1+\varepsilon+C''}{2}-1}\mathrm{e}^{-\frac{C'}{2}y}dy\ dz.
\end{aligned}$$

The next step is to prove that

$$\sup_{x>x_0}\frac{F(x)}{x}\int_x^\infty \frac{dz}{F(z)} < \infty, \tag{2.7.13}$$

in order to use [146, Theorem 2.5]. Choosing C'' such that $1+\varepsilon+C'' < 2$, we have the following estimates

$$\begin{aligned}
F(x) &\leqslant \frac{1}{C'}x_1^{\frac{1+\varepsilon+C''}{2}-1}\int_{x_1}^{x^{\frac{1}{k}}} \mathrm{e}^{\frac{C'}{2}z}z^{-1+\frac{1-\varepsilon-C''}{2}}\left(\mathrm{e}^{-\frac{C'}{2}x_1}-\mathrm{e}^{-\frac{C'}{2}z}\right)dz \\
&\leqslant \frac{2}{C'^2x_1}\mathrm{e}^{-\frac{C'}{2}x_1}\mathrm{e}^{\frac{C'}{2}x^{\frac{1}{k}}}
\end{aligned} \tag{2.7.14}$$

and for $x \geqslant \bar{x} := \left(\left(x_1 + 4\frac{\log(2)}{C'}\right)\vee\frac{32}{C'^2}\right)^k$

$$\begin{aligned}
F(x) &\geqslant \frac{1}{C'}\int_{x_1}^{x^{\frac{1}{k}}}\frac{1}{z}\mathrm{e}^{\frac{C'}{2}z}\left(\mathrm{e}^{-\frac{C'}{2}x_1}-\mathrm{e}^{-\frac{C'}{2}z}\right)dz \\
&\geqslant \frac{x^{-\frac{1}{k}}}{C'}\int_{x_1+2\frac{\log(2)}{C'}}^{x^{\frac{1}{k}}}\mathrm{e}^{\frac{C'}{2}z}\left(\mathrm{e}^{-\frac{C'}{2}x_1}-\mathrm{e}^{-\frac{C'}{2}z}\right)dz \\
&\geqslant \frac{x^{-\frac{1}{k}}\mathrm{e}^{-\frac{C'}{2}x_1}}{2C'}\int_{x_1+2\frac{\log(2)}{C'}}^{x^{\frac{1}{k}}}\mathrm{e}^{\frac{C'}{2}z}dz \\
&\geqslant \frac{x^{-\frac{1}{k}}\mathrm{e}^{-\frac{C'}{2}x_1}}{C'^2}\left(\mathrm{e}^{\frac{C'}{2}x^{\frac{1}{k}}}-\mathrm{e}^{\frac{C'}{2}\left(x_1+2\frac{\log(2)}{C'}\right)}\right) \\
&\geqslant \frac{x^{-\frac{1}{k}}\mathrm{e}^{-\frac{C'}{2}x_1}}{2C'^2}\mathrm{e}^{\frac{C'}{2}x^{\frac{1}{k}}},
\end{aligned} \tag{2.7.15}$$

using the fact that for $x \geqslant y + 2\log(2)/C'$ it holds $\exp(-C'y/2) - \exp(-C'x/2) \geqslant \frac{1}{2}\exp(-C'y/2)$ and $\exp(C'x/2) - \exp(C'y/2) \geqslant \frac{1}{2}\exp(C'x/2)$. Moreover, using that for $y \geqslant \frac{32}{C'^2}$ it holds that $\exp\left(\frac{C'}{4}y\right) \geqslant y$, we obtain

$$F(x) \geqslant \frac{\mathrm{e}^{-\frac{C'}{2}x_1}}{2C'^2}\mathrm{e}^{\frac{C'}{4}x^{\frac{1}{k}}}. \tag{2.7.16}$$

Integrating the previous estimate (2.7.15), we obtain for $x \geqslant \bar{x}$ that

$$\begin{aligned}\int_x^\infty \frac{dz}{F(z)} &\leqslant 2C'^2\mathrm{e}^{\frac{C'}{2}x_1}\int_x^\infty z^{\frac{1}{k}}\mathrm{e}^{-\frac{C'}{2}z^{\frac{1}{k}}}dz\\ &= 2kC'^2\mathrm{e}^{\frac{C'}{2}x_1}\int_{x^{\frac{1}{k}}}^\infty z^k\mathrm{e}^{-\frac{C'}{2}z}dz\\ &\leqslant CC'\mathrm{e}^{\frac{C'}{2}x_1}\mathrm{e}^{-\frac{C'}{2}x^{\frac{1}{k}}}x,\end{aligned} \tag{2.7.17}$$

where C is a positive constant that only depends on k. Here, the second inequality follows from successive integrations by part and $x \geqslant 1$. Putting together Equations (2.7.14) and (2.7.17) we get, for $x \geqslant \bar{x}$,

$$\frac{F(x)}{x}\int_x^\infty \frac{dz}{F(z)} \leqslant \frac{2C}{C'x_1} \tag{2.7.18}$$

Moreover, for $x \in [x_0, \bar{x}]$, using that $\frac{1}{x} \leqslant 1$, that F is increasing, and the estimate (2.7.18), we obtain

$$\begin{aligned}\frac{F(x)}{x}\int_x^\infty \frac{dz}{F(z)} &\leqslant \int_x^{\bar{x}} \frac{F(x)}{F(z)}dz + F(\bar{x})\int_{\bar{x}}^\infty \frac{dz}{F(z)}\\ &\leqslant \int_x^{\bar{x}} dz + \bar{x}\frac{F(\bar{x})}{\bar{x}}\int_{\bar{x}}^\infty \frac{dz}{F(z)} \leqslant \bar{x}\left(1 + \frac{2C}{C'x_1}\right).\end{aligned} \tag{2.7.19}$$

Combining (2.7.18) and (2.7.19) gives

$$\sup_{x \geqslant x_0} \frac{F(x)}{x}\int_x^\infty \frac{dz}{F(z)} + 2 \leqslant 2 + \bar{x}\left(1 + \frac{2C}{C'x_1}\right) \tag{2.7.20}$$

Therefore (2.7.13) is satisfied. Let B be the inverse of F on $[x_0, \infty)$. As proved above (2.7.16), for $x \geqslant \bar{x}$,

$$F(x) \geqslant \frac{\mathrm{e}^{-\frac{C'}{2}x_1}}{2C'^2}\mathrm{e}^{\frac{C'}{4}x^{\frac{1}{k}}}.$$

Therefore, $B(y) \leqslant \bar{x} + \left(\frac{4}{C'} \ln\left(2\,(C')^2\, \mathrm{e}^{\frac{C'}{2}x_1} y\right)\right)^k$. By [146, Theorem 2.5], we obtain that for λ small enough

$$\begin{aligned}
\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\max_{0\leqslant\xi\leqslant\xi^{\lambda,m}} \left(Z^\varepsilon_\xi\right)^k\right] &\leqslant \bar{x}\left(2 + \frac{2C}{C'x_1}\right)\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[B\left(\xi^{\lambda,m}\right)\right] \\
&\leqslant \bar{x}\left(2 + \frac{2C}{C'x_1}\right)\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\bar{x} + \left(\frac{4}{C'}\ln\left(2\,(C')^2\,\mathrm{e}^{\frac{C'}{2}x_1}\lambda^{-\zeta}\int_0^T c_t^{\hat{\varphi}}\,dt\right)\right)^k\right] \\
&= O\left(\lambda^{-2k\kappa_2\left(4+\frac{1}{k}\right)}\right).
\end{aligned}$$

Here, the last equality follows from Assumptions 3 (i) and (ii) and the Cauchy-Schwarz inequality. Together with the comparison result established above, this yields

$$\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\max_{0\leqslant\xi\leqslant\xi^{\lambda,m}}\left(\widetilde{\Delta\varphi}^\lambda_\xi\right)^{2k}\right] = o\left(\lambda^{-10k\kappa_2+\varepsilon'}\right),$$

for any $\varepsilon' > 0$ small enough. With Assumption 3 (ii) and Cauchy-Schwarz' inequality to obtain the result under $\hat{\mathbb{Q}}$, the assertion follows. □

The estimate (2.7.11) yields the following uniform bound in λ for the moments of the displacement $\Delta\varphi^\lambda$:

Corollary 2.7.8. *Suppose that Assumptions 3* (i) *and* (ii) *are satisfied. Then there exists a family of constants $C_n,\ n \in \mathbb{N}$ such that, for $\lambda \in (0,1]$,*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left|\Delta\varphi^\lambda_{t\wedge\tau^{\lambda,m}}\right|^n\right] \leqslant C_n,\ \forall\ t \in [0,T]. \tag{2.7.21}$$

Proof. For $n \in \mathbb{N}$, it follows from Lemma 2.7.7 that

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant t\leqslant\tau^{\lambda,m}}\left(\Delta\varphi^\lambda_t\right)^n\right] = \mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant\xi\leqslant\xi^{\lambda,m}}\left(\lambda^{\frac{\zeta}{2}}\widetilde{\Delta\varphi}^\lambda_\xi\right)^n\right] = o\left(\lambda^{n\left(\frac{\zeta}{2}-5\kappa_2\right)}\right). \tag{2.7.22}$$

□

As a consequence:

Lemma 2.7.9. *Suppose that Assumptions 3* (i) *and* (ii) *are satisfied. Then, for $n \in \mathbb{N}$:*

$$\hat{\mathbb{Q}}\left[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m}\right] = O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right). \tag{2.7.23}$$

Proof. Markov's inequality shows that, for $n \in \mathbb{N}$, we have

$$\begin{aligned}\hat{\mathbb{Q}}\left[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m}\right] &\leqslant \hat{\mathbb{Q}}\left[\sup_{0\leqslant t\leqslant \tau^{\lambda,m}} \left|\Delta\varphi_t^\lambda\right| > \lambda^{\kappa_1}\right] \\ &\leqslant \hat{\mathbb{Q}}\left[\sup_{0\leqslant \xi\leqslant \xi^{\lambda,m}} \left|\widetilde{\Delta\varphi}_\xi^\lambda\right| > \lambda^{\kappa_1-\frac{1}{p+2}}\right] \\ &\leqslant \lambda^{n\left(\frac{1}{p+2}-\kappa_1\right)}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\max_{0\leqslant \xi\leqslant \xi^{\lambda,m}} \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^n\right] \\ &= o\left(\lambda^{n\left(\frac{1}{p+2}-\kappa_1-5\kappa_2\right)}\right).\end{aligned}$$

The assertion now follows from Lemma 2.7.7 and the fact that $\frac{1}{p+2}-\kappa_1-5\kappa_2 > 0$. $\square$

Combining the Lemmas 2.7.4, 2.7.5, 2.7.6, and 2.7.9, we finally obtain the following important estimate which is used in the proof of Proposition 2.5.12:

Proposition 2.7.10. *Suppose that Assumptions 3* (ii)*, and* (iii) *are satisfied and that* $(\hat{\varphi}_T^*)^{\frac{4(1+2\varepsilon)}{(p+2)\kappa_4}}, \left(c_T^{\hat{\varphi}*}\left(1+(m_T^*)^3\right)\right)^{\frac{4(1+2\varepsilon)(1+\varepsilon)p}{2-2p+(p+2)(p\kappa_1-\kappa_3)}}, \exp\left(\varepsilon\Lambda_T^*\right) \in L^1(\hat{\mathbb{Q}})$ *(Assumption 3* (i)*). Then:*

$$\hat{\mathbb{Q}}\left[\tau^{\Delta\varphi} < T^\lambda\right] = O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right), \ \textit{and} \ \mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(T^\lambda-\tau^{\Delta\varphi}\right)^n\right] = O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right), \quad n \in \mathbb{N}. \tag{2.7.24}$$

Proof. By Lemmas 2.7.4, 2.7.5, 2.7.6, and 2.7.9, it holds

$$\begin{aligned}\hat{\mathbb{Q}}\left[\tau^{\Delta\varphi} < T^\lambda\right] &\leqslant \hat{\mathbb{Q}}\left[\tau^{\lambda,m} < T^\lambda\right] + \hat{\mathbb{Q}}\left[\tau^{\lambda,\Delta\varphi} < \tau^{\lambda,m}\right] \\ &\quad + \hat{\mathbb{Q}}\left[\tau^{\lambda,\mathrm{cost}} < \tau^{\lambda,\Delta\varphi}\right] + \hat{\mathbb{Q}}\left[\tau^{\lambda,\hat{\varphi}} < T^\lambda\right] \\ &= O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right).\end{aligned}$$

Then,

$$\begin{aligned}\mathbb{E}_{\hat{\mathbb{Q}}}\left[\left(T^\lambda-\tau^{\Delta\varphi}\right)^n\right] &\leqslant T^n\hat{\mathbb{Q}}\left[\tau^{\Delta\varphi} < T^\lambda\right] \\ &= O\left(\lambda^{\frac{4(1+2\varepsilon)}{p+2}}\right).\end{aligned}$$

$\square$

2.7.2.4 Uniform Moment Estimates for the Renormalised Displacement

Lemma 2.7.7 proves a maximal inequality for the renormalized displacement. However, the estimate obtained is of order negative in λ as λ goes to 0. The aim of the following lemma is to show that one can actually bound the moment of any power of the renormalized displacement on $[\![0, \widetilde{\tau}^{\Delta\varphi}]\!]$ by a constant, uniformly in λ. This is important for the proof of Lemma 2.5.8 and therefore for the ergodic estimates of Proposition 2.5.11.

Lemma 2.7.11. *Let $n \leqslant 2\lceil \frac{27}{2\varepsilon} + \frac{9+\varepsilon}{2} \rceil$ and suppose that Assumption 3*(iii) *is satisfied. Then there exists a positive constant $\widetilde{C}_n$ such that, for any $\xi \geqslant 0$,*

$$\mathbb{E}_{\hat{\mathbb{Q}}}\left[\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left|\widetilde{\Delta\varphi}_\xi^\lambda\right|^n\right] \leqslant \widetilde{C}_n. \tag{2.7.25}$$

Proof. By Jensen's inequality it suffices to establish the claim for even n. We argue by induction. The induction basis $n = 0$ is trivial. So let $2 \leqslant n \leqslant 4\lceil \frac{27}{2\varepsilon} + \frac{9+\varepsilon}{2} \rceil$ be an even number and assume that the claim is true for $n - 2$.

Using that $\widetilde{\Delta\varphi}^\lambda$ is continuous, it follows that $\mathbb{1}_{\{\cdot\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi}_\cdot^\lambda\right)^n$ satisfies under $\widetilde{\mathbb{Q}}^\lambda$ the SDE

$$\begin{aligned} d\left(\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\widetilde{\Delta\varphi}_\xi^\lambda\right)^n = &-\left(\frac{n}{2}p^{-\frac{1}{p-1}}m_{\lambda^\zeta u_\xi^\lambda}\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-1}\widetilde{g}_p\left(m_{\lambda^\zeta u_\xi^\lambda}\widetilde{\Delta\varphi}_\xi^\lambda\right)\right)\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}d\xi \\ &+\frac{n(n-1)}{2}\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-2}\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}d\xi + n\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-1}\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}d\widetilde{W}_\xi^\lambda \\ &+\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^n d\mathbb{1}_{\{\xi\leqslant\widetilde{\tau}^{\Delta\varphi}\}}. \end{aligned} \tag{2.7.26}$$

On $[\![0, \widetilde{\tau}^{\Delta\varphi}]\!]$, the process $\widetilde{\Delta\varphi}^\lambda$ is bounded by $\lambda^{\kappa_1-\frac{1}{p+2}}$, and so the process $(\int_0^{\xi\wedge\widetilde{\tau}^{\Delta\varphi}}(\widetilde{\Delta\varphi}_y^\lambda)^n d\widetilde{W}_y^\lambda)_{\xi\geqslant0}$ is a true martingale on $[0,\infty)$.

Define the function $h(\xi) := \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}[\mathbb{1}_{\{\cdot\leqslant\widetilde{\tau}^{\Delta\varphi}\}}(\widetilde{\Delta\varphi}_\cdot^\lambda)^n]$, and note that $h(0) = 0$. Integrate the SDE (2.7.26) from 0 to ξ for $\xi \geqslant 0$, take expectations, and use the induction hypothesis and Fubini's theorem to obtain

$$\begin{aligned} h(\xi) \leqslant &-\frac{n}{2}p^{-\frac{1}{p-1}}\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi\left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-2}\left|m_{\lambda^\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda\widetilde{g}_p\left(m_{\lambda^\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda\right)\right|\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}dy\right] \\ &+\frac{n(n-1)}{2}\widetilde{C}_{n-2}\xi + \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi\left(\widetilde{\Delta\varphi}_y^\lambda\right)^n d\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\right]. \end{aligned} \tag{2.7.27}$$

As the process $\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}$ has only (one) downwards jump, the third term on the right hand side of (2.7.27) is negative. We proceed to estimate the first expectation on the right hand side of (2.7.27). Corollary 2.7.1 gives $|x\widetilde{g}_p(x)| \geqslant C\left(|x|^{\frac{2+p}{p}}-1\right)$ for $x \in \mathbb{R}$ and some constant $C > 0$. Together with the induction hypothesis and Fubini's theorem, this yields

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-2}\left|m_{\lambda\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda \widetilde{g}_p\left(m_{\lambda\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda\right)\right| \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}dy\right]\\
&\quad\geqslant C\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-2}\left(\left|m_{\lambda\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda\right|^{\frac{2+p}{p}}-1\right)\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}dy\right]\\
&\quad\geqslant C\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi \left(\widetilde{\Delta\varphi}_\xi^\lambda\right)^{n-2}\left|m_{\lambda\zeta u_y^\lambda}\widetilde{\Delta\varphi}_y^\lambda\right|^{\frac{2+p}{p}}\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}dy\right]-C\widetilde{C}_{n-2}\xi. \qquad (2.7.28)
\end{aligned}$$

We proceed to estimate the expectation in (2.7.28). To this end, note that

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^\xi \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda\zeta u_y^\lambda}^{\frac{2+p}{p}}\left|\widetilde{\Delta\varphi}_y^\lambda\right|^{n-2+\frac{2+p}{p}}dy\right]\\
&\qquad\qquad = \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^{\lceil\xi\rceil} \mathbb{1}_{\{y\leqslant\xi\}}\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda\zeta u_y^\lambda}^{\frac{2+p}{p}}\left|\widetilde{\Delta\varphi}_y^\lambda\right|^{n-2+\frac{2+p}{p}}dy\right]
\end{aligned}$$

Then by Hölder's inequality, for processes A and B, and constants $c, d > 1$ with $\frac{1}{c}+\frac{1}{d}=1$ we have for $l \in \mathbb{N}$

$$\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} A_y^c dy\right] \geqslant \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} A_y B_y dy\right]^c \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} B_y^d dy\right]^{-\frac{c}{d}}.$$

Applying this with $c = \frac{n-2+\frac{2+p}{p}}{n}$, $d = \frac{2+(n-1)p}{2-p}$, $A_y = \mathbb{1}_{\{y\leqslant\xi\}}\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda\zeta u_y^\lambda}^{\frac{n(2+p)}{2+(n-1)p}}\left(\widetilde{\Delta\varphi}_y^\lambda\right)^n$, $B_y = m_{\lambda\zeta u_y^\lambda}^{-\frac{n(2+p)}{2+(n-1)p}}$, and $l \in \{0, 1, \ldots\lfloor\xi\rfloor - 1\}$ we obtain

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} \mathbb{1}_{\{y\leqslant\xi\}}\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda\zeta u_y^\lambda}^{\frac{2+p}{p}}\left|\widetilde{\Delta\varphi}_y^\lambda\right|^{n-2+\frac{2+p}{p}}dy\right]\\
&\quad\geqslant \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}}\left(\widetilde{\Delta\varphi}_y^\lambda\right)^n dy\right]^{\frac{n-2+\frac{2+p}{p}}{n}} \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} m_{\lambda\zeta u_y^\lambda}^{-n\frac{2+p}{2-p}}dy\right]^{-\frac{2-p}{pn}}.
\end{aligned}$$

Next, for $n \leqslant 2\lceil\frac{27+\varepsilon}{2}+\frac{9}{\varepsilon}\rceil$,

$$\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_l^{l+1} m_{\lambda\zeta u_y^\lambda}^{-n\frac{2+p}{2-p}}dy\right]^{-\frac{2-p}{pn}} \geqslant \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[(m_T^*)^{-n\frac{2+p}{2-p}}\right]^{-\frac{2-p}{np}} =: C_{m,n},$$

which is finite by Cauchy-Schwarz' inequality as well as Assumptions 3(ii) and (iii). Similarly, for $\xi \in \mathbb{R}\backslash\mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_{\lfloor\xi\rfloor}^{\lceil\xi\rceil} \mathbb{1}_{\{y\leqslant\xi\}}\mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda^\zeta u_y^\lambda}^{\frac{2+p}{p}} \left|\widetilde{\Delta\varphi}_y^\lambda\right|^{n-2+\frac{2+p}{p}} dy\right] \\
&\qquad\geqslant \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_{\lfloor\xi\rfloor}^{\xi} \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} \left(\widetilde{\Delta\varphi}_y^\lambda\right)^n dy\right]^{\frac{n-2+\frac{2+p}{p}}{n}} \mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_{\lfloor\xi\rfloor}^{\lceil\xi\rceil} m_{\lambda^\zeta u_y^\lambda}^{-n\frac{2+p}{2-p}} dy\right]^{-\frac{2-p}{pn}} \\
&\qquad\geqslant C_{m,n}\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_{\lfloor\xi\rfloor}^{\xi} \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} \left(\widetilde{\Delta\varphi}_y^\lambda\right)^n dy\right]^{\frac{n-2+\frac{2+p}{p}}{n}}.
\end{aligned}$$

Using the inequality $x^{\frac{n-2+\frac{2+p}{p}}{n}} \geqslant x-1$, for $x \geqslant 0$, we obtain

$$\begin{aligned}
&\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^{\xi} \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} m_{\lambda^\zeta u_y^\lambda}^{\frac{2+p}{p}} \left|\widetilde{\Delta\varphi}_y^\lambda\right|^{n-2+\frac{2+p}{p}} dy\right] \\
&\qquad\geqslant C_{m,n}\mathbb{E}_{\widetilde{\mathbb{Q}}^\lambda}\left[\int_0^{\xi} \mathbb{1}_{\{y\leqslant\widetilde{\tau}^{\Delta\varphi}\}} \left(\widetilde{\Delta\varphi}_y^\lambda\right)^n dy\right] - C_{m,n}(\xi+1).
\end{aligned}$$

Putting everything together and applying Fubini's theorem, it follows that

$$\begin{aligned}
&h(\xi) + \frac{n}{2}p^{-\frac{1}{p-1}}CC_{m,n}\int_0^{\xi} h(y)dy \\
&\qquad\leqslant \frac{n(n-1)}{2}\tilde{C}_{n-2}\xi + \frac{n}{2}p^{-\frac{1}{p-1}}C(\widetilde{C}_{n-2}+C_{m,n})\xi + \frac{n}{2}p^{-\frac{1}{p-1}}CC_{m,n}.
\end{aligned}$$

We obtain that, for three positive constants C' C'' and C''' which depend neither on λ nor ξ,

$$h(\xi) + C'\int_0^{\xi} h(y)dy \leqslant C''\xi + C'''.$$

By Gronwall's inequality, this implies that

$$h(\xi) \leqslant \sup_{y\in\mathbb{R}_+}(C''y + C''')\exp(-C'y) =: \widetilde{C}_n, \quad \text{for all } \xi \in \mathbb{R}_+.$$

This with Assumption 3 (ii) and Cauchy-Schwarz' inequality to obtain the result under $\hat{\mathbb{Q}}$ completes the proof. □

2.7.3 Speed Measures and Convergence

This section contains the proof of Lemmas 2.5.2 and 2.5.4 from Section 2.5.1. The first states some properties of the diffusions $Y^{a,\lambda,\varepsilon_1,+}$ and $Y^{a,\lambda,\varepsilon_1,-}$, provides their speed measure, and establishes that these measures are finite. The second is a convergence result necessary for the proof of the ergodic estimates.

Proof of Lemma 2.5.2 To prove that the diffusions are regular and reccurent, we use the tools from [40, Chapter 5]. Denote the drifts of the two diffusions by $\mu^{a,\lambda,\varepsilon_1,+}$ and $\mu^{a,\lambda,\varepsilon_1,-}$ and their volatilities by $\sigma^{a,\lambda,\varepsilon_1,+}$ and $\sigma^{a,\lambda,\varepsilon_1,-}$. Then we have the following behaviours around 0 and ∞:

$$\frac{\mu^{a,\lambda,\varepsilon_1,+}(x)}{(1+\varepsilon_1)x^0} \xrightarrow{x\to 0} 1, \quad \frac{\mu^{a,\lambda,\varepsilon_1,-}(x)}{(1-\varepsilon_1)x^0} \xrightarrow{x\to 0} 1, \quad \frac{\sigma^{a,\lambda,\varepsilon_1,+}(x)}{2x^{\frac{1}{2}}} \xrightarrow{x\to 0} 1, \quad \frac{\sigma^{a,\lambda,\varepsilon_1,-}(x)}{2x^{\frac{1}{2}}} \xrightarrow{x\to 0} 1,$$
$$\frac{\mu^{a,\lambda,\varepsilon_1,+}(x)}{-(p-1)^{-\frac{1}{p}}\left(\left(b^{a,\varepsilon_1}\right)^2 x\right)^{\frac{2+p}{2p}}} \xrightarrow{x\to\infty} 1, \quad \frac{\mu^{a,\lambda,\varepsilon_1,-}(x)}{-(p-1)^{-\frac{1}{p}}\left(\left(a+\varepsilon_1\right)^2 x\right)^{\frac{2+p}{2p}}} \xrightarrow{x\to\infty} 1,$$
$$\frac{\sigma^{a,\lambda,\varepsilon_1,+}(x)}{2x^{\frac{1}{2}}} \xrightarrow{x\to\infty} 1, \quad \frac{\sigma^{a,\lambda,\varepsilon_1,-}(x)}{2x^{\frac{1}{2}}} \xrightarrow{x\to\infty} 1.$$

Following the notations of [40], by an adaptation of [40, Theorem 5.3], 0 is a singular point of the SDE of type 2, and by [40, Theorem 5.5], ∞ is of type A. This means according to the discussions of Theorem 2.12, Section 2.4, Theorem 4.1, Section 4.2 in [40], the diffusions are recurrent in the sense of [157, Definitions X.3.5 and X.3.8]: a process X is recurrent under $\hat{\mathbb{Q}}$ on $\mathbb{R}_+$ if

$$\text{for all } B \in \mathcal{B}\left(\mathbb{R}_+\right) \text{ with Leb}\left(B\right) > 0,\ \hat{\mathbb{Q}}\left[\limsup_{t\to\infty} \mathbb{1}_B(X_t) = 1\right] = 1.$$

Note that this definition of recurrence implies the one used by [101], (defined for Theorem 20.12, p.399) see [157, Proposition X.3.11] .

Now, define the scale functions $q^{a,\lambda,\varepsilon_1,+}$ and $q^{a,\lambda,\varepsilon_1,-}$ of the two processes:

$$
\begin{aligned}
q^{a,\lambda,\varepsilon_1,+}(x) &= \int_1^x \exp\left(-2\int_1^y \frac{1+\varepsilon_1 - p^{-\frac{1}{p-1}} b^{a,\varepsilon_1}\sqrt{z}\widetilde{g}_p\left(b^{a,\varepsilon_1}\sqrt{z}\right)}{4z} dz\right) dy \\
&= \int_1^x y^{-\frac{1+\varepsilon_1}{2}} \exp\left(p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{y}\right) - p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\right)\right) dy \\
q^{a,\lambda,\varepsilon_1,-}(x) &= \int_1^x \exp\left(-2\int_1^y \frac{1-\varepsilon_1 - p^{-\frac{1}{p-1}}\left(a+\varepsilon_1\right)\sqrt{z}\widetilde{g}_p\left(\left(a+\varepsilon_1\right)\sqrt{z}\right)}{4z} dz\right) dy \\
&= \int_1^x y^{-\frac{1-\varepsilon_1}{2}} \exp\left(p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{y}\right) - p^{-\frac{1}{p-1}}\widetilde{G}_p\left(a+\varepsilon_1\right)\right) dy,
\end{aligned}
$$

where $\widetilde{G}_p(x) = \displaystyle\int_0^x \widetilde{g}_p(y)dy$ and $b^{a,\varepsilon_1} = \max\left\{a-\varepsilon_1, \frac{a}{2}\right\}$. The corresponding speed measures of the two processes are (see [146, Equation (2.4)], for example)

$$
\begin{aligned}
\nu^{a,\lambda,\varepsilon_1,+}(x) &= \frac{1}{2}x^{-\frac{1-\varepsilon_1}{2}} \exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{x}\right) + p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\right)\right), \\
\nu^{a,\lambda,\varepsilon_1,-}(x) &= \frac{1}{2}x^{-\frac{1+\varepsilon_1}{2}} \exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{x}\right) + p^{-\frac{1}{p-1}}\widetilde{G}_p\left(a+\varepsilon_1\right)\right).
\end{aligned}
$$

Since $\widetilde{g}_p$ is odd and positive on $\mathbb{R}_+$, its antiderivative $\widetilde{G}_p$ is continuous, even, nonnegative and 0 at 0. With Corollary 2.7.1, we obtain $\widetilde{G}_p(x) \geqslant C\left(1+|x|^{\frac{2+p}{p}}\right)$ for $|x| \geqslant K_p$ and some positive constant C. This implies that the two speed measures are finite. □

Proof of Lemma 2.5.4 Corollary 2.7.1 shows that the function $\widetilde{g}_p$ is of polynomial growth. To prove the integrability statement, we use the same estimate as in the previous proof: for $|x| \geqslant K_p$, $\widetilde{G}_p(x) \geqslant C\left(1+|x|^{\frac{2+p}{p}}\right)$ for some positive constant C.

For the limits we use the dominated convergence theorem. First, notice that by Lemma 2.5.3 and the integrability of the various functions with respect to the speed measures,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,+}ds = \frac{\int_{\mathbb{R}_+} x^{\frac{\varepsilon_1}{2}}\sqrt{x}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{x}\right)\right)dx}{\int_{\mathbb{R}_+} x^{-\frac{1-\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{x}\right)\right)dx}$$
$$= \frac{\int_{\mathbb{R}_+} x^{2+\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\left(b^{a,\varepsilon_1}\right)^2\int_{\mathbb{R}_+} x^{\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx},$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t Y_s^{a,\lambda,\varepsilon_1,-}ds = \frac{\int_{\mathbb{R}_+} x^{\frac{1-\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{x}\right)\right)dx}{\int_{\mathbb{R}_+} x^{-\frac{1+\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{x}\right)\right)dx}$$
$$= \frac{\int_{\mathbb{R}_+} x^{2-\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\left(a+\varepsilon_1\right)^2\int_{\mathbb{R}_+} x^{-\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx},$$

and

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,+}}\right)\right|^p ds$$
$$= \frac{\int_{\mathbb{R}_+}\left|\widetilde{g}_p\left(b\sqrt{x}\right)\right|^p x^{-\frac{1-\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{x}\right)\right)dx}{\int_{\mathbb{R}_+} x^{-\frac{1-\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(b^{a,\varepsilon_1}\sqrt{x}\right)\right)dx}$$
$$= \frac{\int_{\mathbb{R}_+}\left|\widetilde{g}_p\left(\frac{b}{b^{a,\varepsilon_1}}x\right)\right|^p x^{\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\int_{\mathbb{R}_+} x^{\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx},$$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\widetilde{g}_p\left(b\sqrt{Y_s^{a,\lambda,\varepsilon_1,-}}\right)\right|^p ds$$
$$= \frac{\int_{\mathbb{R}_+}\left|\widetilde{g}_p\left(b\sqrt{x}\right)\right|^p x^{-\frac{1+\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{x}\right)\right)dx}{\int_{\mathbb{R}_+} x^{-\frac{1+\varepsilon_1}{2}}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(\left(a+\varepsilon_1\right)\sqrt{x}\right)\right)dx}$$
$$= \frac{\int_{\mathbb{R}_+}\left|\widetilde{g}_p\left(\frac{b}{a+\varepsilon_1}x\right)\right|^p x^{-\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}{\int_{\mathbb{R}_+} x^{-\varepsilon_1}\exp\left(-p^{-\frac{1}{p-1}}\widetilde{G}_p\left(x\right)\right)dx}.$$

To obtain the limit for $\varepsilon_1 \to 0$, we first bound the integrands by functions integrable on $\mathbb{R}_+$. For $x > 0$ and $0 < \varepsilon_1 < \frac{1}{2}$, we have

$$
\begin{aligned}
x^{2+\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant (1+x^3) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right), \\
x^{2-\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant (1+x^2) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right), \\
x^{\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant (1+x) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right), \\
x^{-\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant \left(1+x^{-\frac{1}{2}}\right) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right), \\
\left|\widetilde{g}_p\left(\frac{b}{b^{a,\varepsilon_1}} x\right)\right|^p x^{\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant \left|\widetilde{g}_p\left(\frac{2b}{a} x\right)\right|^p (1+x) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right), \\
\left|\widetilde{g}_p\left(\frac{b}{a+\varepsilon_1} x\right)\right|^p x^{-\varepsilon_1} \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right) &\leqslant \left|\widetilde{g}_p\left(\frac{b}{a} x\right)\right|^p \left(1+x^{-\frac{1}{2}}\right) \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right).
\end{aligned}
$$

Moreover, as $\varepsilon_1 \to 0$, the left-hand sides of these inequalities converge pointwise to $x^2 \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right)$ for the first two, to $\exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right)$ for the next two, and to $\left|\widetilde{g}_p(\frac{b}{a} x)\right|^p \exp\left(-p^{-\frac{1}{p-1}} \widetilde{G}_p(x)\right)$ for the last two. The assertion in turn follows from the dominated convergence theorem. $\square$

Bibliography

[1] L. Ahrens. *On using shadow prices for the asymptotic analysis of portfolio optimization under proportional transaction costs.* PhD thesis, Christan-Albrechts-Universität zu Kiel, 2015.

[2] L. Ahrens and J. Kallsen. Portfolio optimization under small transaction costs: a convex duality approach. Preprint, 2015.

[3] A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. Preprint, 2014.

[4] A. Alfonsi, A. Fruth, and A. Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[5] R. F. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1):1–18, 2003.

[6] R. F. Almgren. Optimal trading with stochastic liquidity and volatility. *SIAM Journal of Financial Mathematics*, 3(1):163–181, 2012.

[7] R. F. Almgren and N. Chriss. Value under liquidation. *Risk*, 12(12):61–63, 1999.

[8] R. F. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[9] R. F. Almgren and T. M. Li. Option hedging with smooth market impact. *Market microstructure and liquidity*, 2(1), 2016.

[10] R. F. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *RISK*, July, 2005.

[11] A. Altarovici, J. Muhle-Karbe, and H. M. Soner. Asymptotics for fixed transaction costs. *Finance and Stochastics*, 19(2):363–414, 2015.

[12] A. Altarovici, M. Reppen, and H. M. Soner. Merton problem with fixed and proportional costs. preprint, 2015.

[13] E. Bacry, A. Iuga, M. Lasnier, and C.-A. Lehalle. Market impacts and the life cycle of investors orders. Preprint, 2014.

[14] P. Balduzzi and A. Lynch. Transaction costs and predictability: some utility cost calculations. *Journal of Financial Economics*, 52(1):47–78, 1999.

[15] P. Bank, H. Soner, and M. Voss. Hedging with temporary price impact. *Mathematics and Financial Economics*, 11(2):215–239, 2016.

[16] P. Bank and M. Voß. Linear quadratic stochastic control problems with singular stochastic terminal constraint. preprint, 2016.

[17] N. Barberis. Investing for the long run when returns are predictable. *Journal of Finance*, 55(1):225–264, 2000.

[18] G. Barles and H. M. Soner. Option pricing with transaction costs and a nonlinear black-scholes equation. *Finance and Stochastics*, 2(4):269–397, 1998.

[19] I. Ben Tahar, H. M. Soner, and N. Touzi. Merton problem with taxes: characterization, computation, and approximation. *SIAM Journal of Financial Mathematics*, 1(1):366–395, 2010.

[20] G. Benedetti, L. Campi, J. Kallsen, and J. Muhle-Karbe. On the existence of shadow prices. *Finance and Stochastics*, 17(4):801–818, 2013.

[21] D. Bertsimas and A. W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[22] M. Bichuch. Pricing a contingent claim liability with transaction costs using asymptotic analysis for optimal investment. *Finance and Stochastics*, 18(3):651–694, 2014.

[23] M. Bichuch and P. Guasoni. Investing with liquid and illiquid assets. *Mathematical Finance*, to appear, 2016.

[24] M. Bichuch and S. E. Shreve. Utility maximization trading two futures with transaction costs. *SIAM Journal on Financial Mathematics*, 4(1):26–85, 2013.

[25] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[26] B. Bouchard. Option pricing via utility maximization in the presence of transaction costs: am asymptotic analysis. Preprint, 2000.

[27] B. Bouchard, L. Moreau, and H. M. Soner. Hedging under an expected loss constraint with small transaction costs. Preprint, 2013.

[28] J.-P. Bouchaud, M. Mézard, and M. Potters. Statistical properties of stock order

books: empirical results and models. *Quantitative finance*, 2(4):251–256, 2002.

[29] J.-P. Bouchaud and M. Potters. More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1):133–140, 2003.

[30] M. Broadie, J. Cvitanić, and H. M. Soner. Optimal replication of contingent claims under portfolio constraints. *Review of Financial Studies*, 11(1):59–79, 1998.

[31] M. Brodén and P. Tankov. Tracking errors from discrete hedging in exponential lévy models. *International Journal of Theoretical and Applied Finance*, 14(6):803–837, 2011.

[32] M. K. Brunnermeier and L. H. Pedersen. Predatory trading. *Journal of Finance*, 60(4):1825–1863, 2005.

[33] J. Cai, M. Fukasawa, M. Rosenbaum, and P. Tankov. Optimal discretization of hedging strategies with directional views. *SIAM Journal on Financial Mathematics*, 7(1):34–69, 2016.

[34] J. Cai, M. Rosenbaum, and P. Tankov. Asymptotic lower bounds for optimal tracking: a linear programming approach. *Annals of Applied Probability*, to appear, 2016.

[35] J. Cai, M. Rosenbaum, and P. Tankov. Asymptotic optimal tracking: feedback strategies. *Stochastics*, to appear, 2016.

[36] B. I. Carlin, M. S. Lobo, and S. Viswanathan. Episodic liquidity crises: cooperative and predatory trading. *Journal of Finance*, 62(5):2235–2274, 2007.

[37] T. Cayé and J. Muhle-Karbe. Liquidation with self-exciting price impact. *Mathematics and Financial Economics*, 10(1):15–28, 2016.

[38] U. Çetin, H. M. Soner, and N. Touzi. Option hedging for small investors under liquidity costs. *Finance and Stochastics*, 14(3):317–341, 2010.

[39] J. F. Chassagneux, R. Elie, and I. Kharroubi. When terminal facelift enforces delta constraints. *Finance and Stochastics*, 19(2):329–362, 2015.

[40] A.-S. Cherny and H.-J. Engelbert. *Singular Stochastic Differential Equations.*, volume 1858 of *Lecture notes in mathematics.* Springer, Berlin, 2005.

[41] P. Collin-Dufresne, K. Daniel, C. Moallemi, and M. Saglam. Strategic asset allocation with predictable returns and transaction costs. Preprint, 2012.

[42] G. M. Constantinides. Capital market equilibrium with transaction costs. *Journal of Political Economy*, 94(4):842–862, 1986.

[43] G. M. Constantinides and M. J. P. Magill. Portfolio selection with transaction costs.

Journal of Economic Theory, 13(2):245–263, 1976.

[44] R. Cont and A. de Larrard. Price dynamics in a markovian limit order market. *SIAM Journal of Financial Mathematics*, 4(1):1–25, 2013.

[45] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operation research*, 58(3):549–563, 2010.

[46] J. Cvitanić and I. Karatzas. Convex duality in constrained portfolio optimization. *Annals of Applied Probability*, 2(4):767–818, 1992.

[47] J. Cvitanić and I. Karatzas. Hedging contingent claims with constrained portfolio. *Annals of Applied Probability*, 3(3):652–681, 1993.

[48] J. Cvitanić and I. Karatzas. Hedging and portfolio optimization under transaction costs: a martingale approach. *Mathematical Finance*, 6(2):133–165, 1996.

[49] J. Cvitanić, I. Karatzas, and H. M. Soner. Backward stochastic differential equations with constraints on the gain-process. *Annals of Probability*, 26(4):1522–1551, 1998.

[50] J. Cvitanić, W. Schachermayer, and H. Wang. Utility maximization in incomplete markets with random endowment. *Finance and Stochastics*, 5:259–272, 2001.

[51] C. Czichowsky, J. Muhle-Karbe, and W. Schachermayer. Transaction costs, shadow prices, and duality in discrete time. *SIAM Journal of Financial Mathematics*, 5(1):258–277, 2014.

[52] M. H. A. Davis. Option pricing in incomplete markets. In *Mathematics of Derivative Securities*, pages 216–226. Cambridge University Press, Cambridge, 1997.

[53] M. H. A. Davis and Norman. Portfolio selection with transaction costs. *Mathematics of Operation Research*, 15(4):676–713, 1990.

[54] M. H. A. Davis, V. G. Panas, and T. Zariphopoulou. European option pricing with transaction costs. *SIAM Journal on Control and Optimization*, 31(2):470–493, 1993.

[55] J. De Lataillade, C. Deremble, M. Potters, and J.-P. Bouchaud. Optimal trading with linear costs. *Journal of Investment Strategies*, 1(3):91–115, 2012.

[56] F. Delbaen, P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer, and C. Stricker. Exponential hedging and entropic penalties. *Mathematical Finance*, 12(2):99–123, 2002.

[57] Y. Dolinsky and H. M. Soner. Duality and convergence for binomial markets with friction. *Finance and Stochastics*, 17(3):447–475, 2013.

[58] B. Dumas and E. Luciano. An exact solution to a dynamic portfolio choice problem

under transaction costs. *The Journal of Finance*, 46(2):577–595, 1991.

[59] J. Eastham and K. Hastings. Optimal impulse control of portfolios. *Mathematics of Operation Research*, 13(4):588–605, 1988.

[60] M. R. Feodoria. *Optimal investment and utility indifference pricing in the presence of small fixed transaction costs*. PhD thesis, Christian-Albrechts-Universität zu Kiel, 2016.

[61] M. Fritelli. The minimal entropy martingale measure and the valuation problem in incomplete markets. *Mathematical Finance*, 10(1):39–52, 2000.

[62] M. Fukasawa. Asymptotically efficient discrete hedging. In *Stochastic Analysis with Financial Applications*, pages 331–346. Springer, 2011.

[63] M. Fukasawa. Efficient discretization of stochastic integrals. *Finance and Stochastics*, 18(1):175–208, 2014.

[64] N. Garleanu and L. H. Pedersen. Adverse selection and the required return. *Review of Financial Studies*, 17(3):643–665, 2004.

[65] N. Garleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *Journal of Finance*, 68(6):2309–2340, 2013.

[66] N. Garleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016.

[67] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, 2010.

[68] J. Gatheral and A. Schied. Optimal trade execution under geometric brownian motion in the Almgren and Chriss framework. *International Journal of Theoretical and Applied Finance*, 14(03):353–368, 2011.

[69] J. Gatheral and A. Schied. Dynamical models of market impact and algorithms for order execution. In J. P. Fouque and J. Langsam, editors, *Handbook on Systemic Risk*, pages 579–602. Cambridge University Press, Cambridge, 2013.

[70] S. Gerhold, P. Guasoni, J. Muhle-Karbe, and W. Schachermayer. Transaction costs, trading volume, and the liquidity premium. *Finance and Stochastics*, 18(1):1–37, 2014.

[71] L. R. Glosten and P. R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

[72] E. Gobet and N. Landon. Almost sure optimal hedging strategy. *Annals of Applied Probability*, 24(4):1652–1690, 2014.

[73] S. Gökay, A. F. Roch, and H. M. Soner. Liquidity models in continuous and discrete time. In *Advanced Mathematical Methods for Finance*, pages 333–365. Springer, Heidelberg, 2011.

[74] C. Gomes and H. Waelbroeck. Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5):773–793, 2014.

[75] P. Guasoni, R. Liu, and J. Muhle-Karbe. Who should sell stocks ? Preprint, 2014.

[76] P. Guasoni and J. Muhle-Karbe. Long horizons, high risk aversion, and endogenous spreads. *Mathematical Finance*, 25(4):724–753, 2015.

[77] P. Guasoni and M. Rásonyi. Hedging, arbitrage and optimality with superlinear frictions. *Annals of Applied Probability*, 25(4):2066–2095, 2015.

[78] P. Guasoni and M. Weber. Optimal trading with multiple assets and cross-price impact. Preprint, 2014.

[79] P. Guasoni and M. Weber. Dynamic trading volume. *Mathematical Finance*, 27(2):313–349, 2015.

[80] P. Guasoni and M. Weber. Nonlinear price impact and portfolio choice. Preprint, 2015.

[81] O. Guéant. Permanent market impact can be nonlinear. Preprint, 2013.

[82] O. Guéant. Optimal execution and block trade pricing: a general framework. *Applied Mathematical Finance*, 22(4):336–365, 2015.

[83] O. Guéant and J. Pu. Option pricing and hedging with execution costs and market impact. *Mathematical Finance*, to appear, 2015.

[84] J. M. Harrison and S. R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:215–260, 1981.

[85] V. Henderson. Valuation of claims on nontraded assets using utility maximization. *Mathematical Finance*, 12(4):351–373, 2002.

[86] M. Herdegen and J. Muhle-Karbe. Sensitivity of optimal consumption streams. Preprint, 2015.

[87] M. Herdegen and J. Muhle-Karbe. Stability of Radner equilibria with respect to small frictions. Preprint, 2016.

[88] S. Hodges and A. Neuberger. Optimal replication of contingent claims under transaction costs. *Review of Futures Markets*, 8:222–239, 1989.

[89] W. Huang, C.-A. Lehalle, and M. Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.

[90] W. Huang and M. Rosenbaum. Ergodicity and diffusivity of markovian order book models: a general framework. preprint, 2015.

[91] G. Huberman and W. Stanzl. Arbitrage-free price-update and price-impact functions. 2000.

[92] G. Huberman and W. Stanzl. Price manipulation and quasi-arbitrage. *Econometrica*, 72(4):1247–1275, 2004.

[93] G. Huberman and W. Stanzl. Optimal liquidity trading. *Review of Finance*, 9(2):165–200, 2005.

[94] J. Hugonnier, D. Kramkov, and M. Sirbu. On utility based pricing of contingent claims in incomplete markets. *Mathematical Finance*, 15(2):203–212, 2005.

[95] J. Jacod. *Calcul stochastique et problèmes de martingales.*, volume 714 of *Lecture notes in mathematics.* Springer, Berlin, 1979.

[96] J. Jacod and N. Shiryaev. *Limit theorems for stochastic processes.* Springer, Berlin, second edition, 2003.

[97] K. Janeček and S. E. Shreve. Asymptotic analysis for optimal investment and consumption with transaction costs. *Finance and Stochastics*, 8(2):181–206, 2004.

[98] E. Jouini and H. Kallal. Martingale and arbitrage in securities markets with transaction costs. *Journal of Economic Theory*, 66(1):178–197, 1995.

[99] Y. Kabanov and C. Stricker. On the optimal portfolio for the exponential utility maximization: remarks to the six-authors paper. *Mathematical Finance*, 12(2):125–134, 2002.

[100] D. Kahneman and A. Tversky. Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2):263–292, March 1979.

[101] O. Kallenberg. *Foundations of modern probability.* Springer, Berlin, second edition, 2002.

[102] J. Kallsen. Derivative pricing based on local utility maximization. *Finance and Stochastics*, 6(1):115–140, 2002.

[103] J. Kallsen and S. Li. Portfolio optimization under small transaction costs: a convex duality approach. Preprint, 2013.

[104] J. Kallsen and J. Muhle-Karbe. On using shadow prices in portfolio optimization with transaction costs. *Annals of Applied Probabilities*, 20(4):1341–1358, 2010.

[105] J. Kallsen and J. Muhle-Karbe. Utility maximization in models with conditionally independent increments. *Annals of Applied Probabilities*, 20(6):2162–2177, 2010.

[106] J. Kallsen and J. Muhle-Karbe. Existence of shadow prices in finite probability spaces. *Mathematical Methods of Operations Research*, 73(2):251–262, 2011.

[107] J. Kallsen and J. Muhle-Karbe. Option pricing and hedging with small transaction costs. *Mathematical Finance*, 25(4):702–723, 2013.

[108] J. Kallsen and J. Muhle-Karbe. High-resilience limits of block-shaped order books. preprint, 2014.

[109] J. Kallsen and J. Muhle-Karbe. The general structure of optimal investment and consumption with small transaction costs. *Mathematical Finance*, to appear, 2015.

[110] I. Karatzas and S. G. Kou. On the pricing of contingent claims under constraints. *Annals of Applied Probability*, 6(2):321–369, 1996.

[111] I. Karatzas, J. P. Lehoczky, and S. E. Shreve. Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM journal on control and optimization*, 25(6):1557–1586, 1987.

[112] I. Karatzas, J. P. Lehoczky, S. E. Shreve, and G. Xu. Martingale and duality methods for utility maximization in an incomplete market. *SIAM Journal of Control and Optimization*, 29(3):702–730, 1991.

[113] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*. Springer, New York, second edition, 1991.

[114] T. Kim and E. Omberg. Dynamic nonmyopic portfolio behavior. *Review of Financial Studies*, 9(1):141–161, 1996.

[115] R. Korn. Portfolio optimisation with strictly positive transaction costs and impulse control. *Finance and Stochastics*, 2(2):85–114, 1998.

[116] D. Kramkov and W. Schachermayer. The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Annals of Applied Probability*, 9(3):904–950, 1999.

[117] D. Kramkov and W. Schachermayer. Necessary and sufficient conditions in the

problem of optimal investment in incomplete markets. *Annals of Applied Probability*, 13(4):1504–1516, 2003.

[118] D. Kramkov and M. Sirbu. Sensitivity analysis of utility-based prices and risk-tolerance wealth processes. *Annals of Applied Probability*, 16(4):2140–2194, 2006.

[119] A. S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[120] A. S. Kyle and A. A. Obizhaeva. Market microstructure invariants: empirical evidence from portfolio transitions. Preprint, 2011.

[121] K. Larsen, O. Mostovyi, and G. Žitković. An expansion in the model space in the context of utility maximization. Preprint, 2015.

[122] S. Laruelle and C.-A. Lehalle. *Market microstructures in practice*. World Scientific, 2013.

[123] F. Lillo, J. D. Farmer, and R. N. Mantegna. Master curve for price-impact function. *Nature*, 421:129–130, 2003.

[124] R. Liptser and A. N. Shiryaev. *Statistics of random processes: I. general theory*. Springer, Berlin, second edition, 2013.

[125] H. Liu. Optimal consumption and investment with transaction costs and multiple risky assets. *Journal of Finance*, 59(1):289–338, 2004.

[126] R. Liu and M.-K. J. Portfolio selection with small transaction costs and binding portfolio constraints. *SIAM Journal of Financial Mathematics*, 4(1):203–227, 2013.

[127] R. Liu, M.-K. J., and M. Weber. Rebalancing with linear and quadratic costs. 2014.

[128] A. Lo, H. Mamaysky, and J. Wang. Asset prices and trading volume under fixed transactions costs. *Journal of Political Economy*, 112(5):1054–1090, 2004.

[129] M. Loewenstein. On optimal portfolio trading strategies for an investor facing transactions costs in a continuous trading market. *Journal of Mathematical Economics*, 33(2):209–228, 2000.

[130] C. Lorenz and A. Schied. Drift dependence of optimal trade execution strategies under transient price impact. *Finance and Stochastics*, 17(4):743–770, 2013.

[131] J. Lorenz and R. F. Almgren. Mean–variance optimal adaptive execution. *Applied Mathematical Finance*, 18(5):395–422, 2011.

[132] A. Lynch and P. Balduzzi. Predictability and transaction costs: the impact on rebalancing rules and behavior. *Journal of Finance*, 55(5):2285–2310, 2000.

[133] A. Lynch and S. Tan. Explaining the magnitude of liquidity premia: the roles of return predictability, wealth shocks, and state dependent transaction costs. *Journal of Finance*, 66(4):1329–1368, 2011.

[134] A. Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258, 2000.

[135] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[136] R. Martin. Optimal trading under proportional transaction costs. *RISK*, August:54–59, 2014.

[137] R. Martin. Universal trading under proportional costs. preprint, 2016.

[138] R. Martin and T. Schöneborn. Mean reversion pays, but costs. *RISK*, February:96–101, 2011.

[139] A. V. Melnikov. On the theory of stochastic equations in components of semimartingales. *Mathematics USSR Sbornik*, 38(3), 1981.

[140] R. Merton. Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, 1969.

[141] R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, 1971.

[142] L. Moreau, J. Muhle-Karbe, and H. M. Soner. Trading with small price impact. *Mathematical Finance*, 27(2):350–400, 2015.

[143] A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

[144] B. Oksendal and A. Sulem. Optimal consumption and portfolio with both fixed and proportional transaction costs. *SIAM Journal on Control and Optimization*, 40(6):1765–1790, 2002.

[145] G. A. Pavliotis and A. Stuart. *Multiscale methods: averaging and homogenization.* Springer, New York, 2008.

[146] G. Peskir. Bounding the maximal height of a diffusion by the time elapsed. *Journal of Theoretical Probability*, 14(3), 2015.

[147] S. R. Pliska. A discrete stochastic decision model. *Advances in Filtering and Optimal Stochastic Control*, pages 290–304, 1982.

[148] S. R. Pliska. A stochastic calculus model of continuous trading: optimal portfolios. *Mathematics of Operation Research*, 2(11):371–382, 1986.

[149] M. Pohl, A. Ristig, W. Schachermayer, and L. Tangpi. The amazing power of dimensional analysis: quantifying market impact. preprint, 2017.

[150] D. Possamaï and G. Royer. General indifference pricing with small transaction costs. *Asymptotic Anal.*, to appear, 2014.

[151] D. Possamaï, H. M. Soner, and N. Touzi. Homogenization and asymptotics for small transaction costs: the multidimensional case. *Communications in Partial Differential Equations*, 40(11):2005–2046, 2015.

[152] S. Predoiu, G. Shaikhet, and S. E. Shreve. Optimal execution in a general one-sided limit-order book. *SIAM Journal on Financial Mathematics*, 2(1):183–212, 2011.

[153] P. E. Protter. *Stochastic integration and differential equations.* Springer, Berlin, second edition, 2005.

[154] M. Rásonyi. Optimal investment with nonconcave utilities in discrete-time markets. *SIAM Journal of Financial Mathematics*, 6(1):517–529, 2015.

[155] M. Rásonyi and R. A. M. Continuous-time portfolio optimisation for a behavioural investor with bounded utility on gains. *Electronic Communications in Probability*, 19(38):1–13, 2014.

[156] A. Rej, R. Benichou, J. de Lataillade, G. Zérah, and J.-P. Bouchaud. Optimal trading with linear and (small) non-linear costs. preprint, 2015.

[157] D. Revuz and M. Yor. *Continuous martingale and Brownian motion.* Springer, third edition, 1999.

[158] T. Rheinländer and G. Steiger. The minimal entropy martingale measure for general Barndorff-Nielsen/Shephard models. *Annals of Applied Probability*, 16(3):1319–1351, 2006.

[159] A. F. Roch and H. M. Soner. Resilient price impact of trading and the cost of illiquidity. *International Journal of Theoretical and Applied Finance*, 16(6):1350037 (27 pages), 2013.

[160] L. C. G. Rogers. Why is the effect of proportional transaction costs $O(\delta^{2/3})$? In *Mathematics of Finance*, pages 303–308. Amer. Math. Soc., Providence, RI, 2004.

[161] M. Rosenbaum. Integrated volatility and round-off error. *Bernoulli*, 15(3):687–720, 2009.

[162] M. Rosenbaum and P. Tankov. Asymptotically optimal discretization of hedging strategies with jumps. *Annals of Applied Probability*, 24(3):1002–1048, 2014.

[163] I. Roşu. Fast and slow informed trading. Preprint, 2016.

[164] P. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, August 1969.

[165] W. Schachermayer. Optimal investment in incomplete markets when wealth may become negative. *Annals of Applied Probability*, 11(3):694–734, 2001.

[166] W. Schachermayer. How potential investments may change the optimal portfolio for the exponential utility. Working Paper of the SFB, 2002.

[167] W. Schachermayer. A super-martingale property of the optimal portfolio process. *Finance and Stochastics*, 7(4):433–456, 2003.

[168] W. Schachermayer and J. Teichmann. How close are the option pricing formulas of Bachelier and Black–Merton–Scholes? *Mathematical Finance*, 18(1):155–170, 2008.

[169] A. Schied. Robust strategies for optimal order execution in the almgren?chriss framework. *Applied Mathematical Finance*, 20(3):264–286, 2013.

[170] A. Schied and T. Schöneborn. Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance and Stochastics*, 13(2):181–204, 2009.

[171] A. Schied, T. Schöneborn, and M. Tehranchi. Optimal basket liquidation for CARA investors is deterministic. *Applied Mathematical Finance*, 17(6):471–489, 2010.

[172] T. Schöneborn and A. Schied. Liquidation in the face of adversity: stealth vs. sunshine trading. Preprint, 2009.

[173] S. Schroder. Optimal portfolio selection with fixed transaction costs: numerical solutions. preprint, 1995.

[174] S. E. Shreve and H. M. Soner. Optimal investment and consumption with transaction costs. *Annals of Applied Probability*, 4(3):609–692, 1994.

[175] H. M. Soner and N. Touzi. Homogenization and asymptotics for small transaction costs. *SIAM Journal on Control and Optimization*, 51(4):2893–2921, 2013.

[176] A. E. Whalley and P. Wilmott. An asymptotic analysis of an optimal hedging model for option pricing with transaction costs. *Mathematical Finance*, 7(3):307–324, 1997.