EQUILIBRIUM ASSET PRICING WITH TRANSACTION COSTS

by

XIAOFEI SHI

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN MATHEMATICAL SCIENCES

ACADEMIC ADVISOR: JOHANNES MUHLE-KARBE

DEPARTMENT OF MATHEMATICAL SCIENCES
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PA, USA

Monday May 4th 2020

Acknowledgements

Above all, I would like to express my deepest gratitude to my thesis advisor, Johannes Muhle-Karbe for making all of this possible. Johannes has always helped me find important and interesting problems and has had very insightful comments about all of my work. This thesis would not have been possible without his unparalleled support and encouragement.

My thanks extend to my thesis committee members, Prof. Steven Shreve, Prof. Martin Larsson, and Prof. Hao Xing, for their commitment to serve on my thesis committee. I sincerely appreciate their helpful and detailed comments on this thesis, from which I benefited a lot.

Next, I would thank all the people who contributed to this thesis. I am grateful to Prof. Robert Pego for insightful discussions on the formulation of FBSDE system, and to Prof. Steven Shreve for his comments on the proof for the Riccati-type ODE. I also thank Prof. Yang Chen for help with the data analysis, Dr. Jiequn Han and my coauthor Dr. Lukas Gonon for helpful discussion on the numerical experiments, and David Itkin for pointing out the integral version of Minkowski's inequality. Further discussions with Prof. Martin Herdegen, Prof. Paolo Guasoni, Prof. Dmitry Kramkov, and Prof. Marko Weber have also been extremely helpful.

A special thank you goes to Prof. Robert Pego, Prof. William Hrusa, Prof. Steven Shreve and Prof. Dejan Slepcev for providing me with fantastic training throughout my PhD. Teaching a physicist to write a rigorous proof is a challenging task, and I am also grateful to all of you for building up my confidence.

I would also like to express my sincere gratitude to the Department of Mathematical Sciences at Carnegie Mellon for supporting me academically and financially throughout my time as a graduate student. Most importantly, thanks for supporting my visit to the Simons Institute in Fall 2018 and to London in Summer 2019.

My time at Carnegie Mellon has been an incredible journey, in large part due to the many new friends as well as old friends back from my undergraduate years. Special thanks to Prof. Bill Hrusa, Prof. Steven Shreve, Roy Zilin Jiang, Yue Pu, Ryan Yan Xu, Zhi Li, Luxi Wang, Yifan Sun, Mochong Duan, Yufeng Shen, An Xiao and most importantly, Iris Zhipeng Yang and Daze Wu. Thanks for being there for me in good times and bad times.

Finally, I would like to thank my mom Yajun Tian and my grandmothers Yuying Wu and Qingsheng Hao. This thesis is dedicated to my father Rong Shi and my grandfather Meiheng Shi. I hope I will make you proud.

Abstract

This thesis studies risk-sharing equilibria where trading is subject to transaction costs.

In an infinite-horizon model with specific state dynamics and exogenous price volatility but general convex trading costs, we determine equilibrium prices and trading strategies in closed form and show how this allows us to calibrate the model to time-series data for prices and trading volume.

For more general state dynamics and endogenous volatilities, equilibria with transaction costs correspond to fully-coupled systems of nonlinear forward-backward stochastic differential equations. We propose a simulation-based deep-learning algorithm that allows us to approximate the solution of such systems numerically. For quadratic trading costs and specific state dynamics, we complement this with a global wellposedness result. As a byproduct, the latter also yields explicit asymptotic expansions of the equilibrium for small transaction costs.

These small-cost asymptotics formally extend to models with general state dynamics, transaction costs, and endogenous volatilities, leading to explicit asymptotic approximations of equilibrium prices with general trading costs.

Contents

Chapter 1

Introduction

The interplay of *liquidity* and asset prices has been studied extensively in the empirical literature, cf., e.g., Amihud, Mendelson, and Pedersen (2006) and the references therein for an overview. In particular, it is well-documented that asset returns depend on liquidity, but how does liquidity affect the volatility of a financial market? And do "illiquidity discounts" increase when there is substantial *liquidity risk*, i.e., if liquidity fluctuates randomly over time?

To develop a theoretical underpinning for the impact of illiquidity on price levels, asset returns, and volatilities, one needs to study *equilibrium* asset pricing models. This means that price levels, returns, and volatilities are not modelled as exogenous inputs, but determined endogenously by matching supply and demand. This in turn allows us to study how the price characteristics depend on the market's liquidity.

However, the analysis of equilibrium models that account for trading costs is challenging. Indeed, both models with limited liquidity and equilibrium asset pricing are notoriously intractable in their own right. These difficulties are of course only compounded in models where equilibrium asset prices are determined endogenously in the presence of trading frictions. To wit, trading costs severely complicate the agents' individual optimization problems. Moreover, representative agents cannot be used to analyze the impact of trading costs, since they precisely do not account for the trades between the individual market participants.

As a consequence, most of the existing literature on equilibrium asset pricing with transaction costs focuses either on purely numerical approaches, or is restricted to extremely stylized settings in order to obtain analytical results. For example, Heaton and Lucas (1996); Adam, Beutel, Marcet, and Merkel (2015); Buss and Dumas (2019) numerically solve discrete-time, discrete state models. Vayanos (1998) studies models where equilibrium prices are diffusive but trading follows deterministic patterns. Conversely, trading volume fluctuates randomly in the model of Lo et al. (2004), but the corresponding equilibrium prices are constant. Prices and trading volume fluctuate randomly in the models of Garleanu and Pedersen (2016); Sannikov and Skrzypacz (2016);

Bouchard, Fukasawa, Herdegen, and Muhle-Karbe (2018), but the volatilities in these models are exogenous inputs.

This thesis builds on the recent work of Herdegen, Muhle-Karbe, and Possamaï (2019), who establish the existence of risk-sharing equilibria when trading is subject to quadratic transaction costs. The volatility process is determined endogenously in their model, and their setting is general enough to generate diffusive prices and trading volume with random and mean-reverting dynamics. However, existence results are only established under the restrictive assumption that the agents' preferences are sufficiently similar. Moreover, the analysis in Herdegen et al. (2019) crucially exploits the assumption that trading costs are quadratic, which naturally raises the question of robustness with respect to the specification of the trading cost. Finally, tractable results are only obtained in Herdegen et al. (2019) for a specific model with linear state dynamics and constant transaction costs. In contrast, the comparative statics of models with more general state dynamics and the impact of liquidity risk remain largely uncharted territory.

The present thesis contributes to this line of research in a number of ways. Chapter 2, which is based on Gonon, Muhle-Karbe, and Shi (2019), studies the equilibrium implications of general transaction costs in a simple stylized benchmark model. The general trading cost nests quadratic and proportional costs as limiting cases, and also covers trading costs that scale with the 3/2-th power of the the order flow as suggested in various empirical studies. With an infinite time horizon, specific linear state dynamics, and an exogenous volatility process, we show that the equilibrium expected returns and trading strategies can be characterized explicitly in terms of a single scalar ODE that has previously appeared in the partial-equilibrium literature (Guasoni and Weber, 2018; Cayé, Herdegen, and Muhle-Karbe, 2019). These explicit results in turn allow us to calibrate the model to time-series data for prices and trading volume, which provides the basis for the quantitative analysis of this simple model and its extensions that we study in the subsequent chapters.

To wit, in Chapter 3, we extend the baseline model from Chapter 2 to more general state dynamics and endogenous volatilities. In such more general settings, equilibrium asset prices and optimal trading strategies have a natural correspondence to fully-coupled systems of nonlinear forward-backward stochastic differential equations (FBSDEs). These equations do not satisfy the Lipschitz, non-degeneracy or monotonicity assumptions that underlie well-posedness results in the existing literature. Here, as in Gonon, Muhle-Karbe, and Shi (2019), we show how such systems can at least be attacked numerically by adapting the deep-learning algorithm proposed by Han, Jentzen, and E (2018).

Subsequently, in Chapter 4, we rigorously derive global existence results for a specific example with constant quadratic transaction costs and linear state dynamics. In this context, Herdegen et al. (2019) have shown that the general FBSDEs system can be reduced to four coupled ordinary differential equations of Riccati type. However, their existence proof based on Picard iteration again crucially exploits the assumption that the agents' risk aversions are sufficiently similar. We show using direct computations that this assumption is superfluous, in that equilibrium prices with transaction costs exist for arbitrary parameter configurations of the model. The bounds used to derive this global existence result also allow us to obtain rigorous asymptotic expansions of the equilibrium in the limit of small transaction costs.

Starting from the work of Shreve and Soner (1994), this kind of asymptotic analysis has a long history for partial-equilibrium models, see, e.g., the survey Muhle-Karbe, Reppen, and Soner (2017) for an overview. In a specific equilibrium model, formal small-costs first appear in Lo et al. (2004). In Chapter 5, we formally extend the asymptotics from Chapter 4 to much more general models that allow for essentially arbitrary Markovian state dynamics and general transaction costs. Even in this generality, where rigorous existence and uniqueness results are still far out of reach, our formal expansions suggest that explicit formulas still obtain in the small-cost limit. Similar result have been developed in partial-equilibrium contexts by Soner and Touzi (2013); Martin (2014); Kallsen and Muhle-Karbe (2017); Moreau, Muhle-Karbe, and Soner (2017); Cai, Rosenbaum, and Tankov (2017); Cayé, Herdegen, and Muhle-Karbe (2019). Extending these results to a general-equilibrium setting in turn opens the door to future research on the impact of liquidity risk or the interplay of stochastic volatility, liquidity, and trading volume, for example.

Chapter 2

Equilibrium Returns with General Transaction Costs

2.1 Introduction

As discussed in the introduction, much recent progress on understanding the equilibrium effects of transaction costs has been made by focusing on quadratic costs on the agents' trading rates (Garleanu and Pedersen, 2016; Sannikov and Skrzypacz, 2016; Bouchard, Fukasawa, Herdegen, and Muhle-Karbe, 2018; Herdegen, Muhle-Karbe, and Possamaï, 2019). The analysis of these models crucially exploits the linearity of the corresponding first-order conditions, thereby naturally raising the question how delicately the qualitative and quantitative predictions depend on the specific choice of the trading costs. Typical examples of specifications of transaction costs are linear transaction taxes or empirical estimates of actual trading costs that typically correspond to a power of the order flow of around $3/2$ (Lillo, Farmer, and Mantegna, 2003; Almgren, Thum, Hauptmann, and Li, 2005).

In this chapter, we address this challenge by studying risk-sharing equilibria with *general* convex costs levied on the agents' trading rates. This nests quadratic costs as one special case, but also covers proportional costs as another limiting case. We show that in an infinite-horizon model with linear state dynamics and exogenous price volatility, the corresponding equilibrium returns can be characterized *explicitly* up to the solution of a single nonlinear ODE. The latter determines the mean-reverting fluctuations of the frictional equilibrium returns around their frictionless counterparts. If costs are quadratic, this "liquidity premium" is an Ornstein-Uhlenbeck process similarly as in Garleanu and Pedersen (2016); Bouchard, Fukasawa, Herdegen, and Muhle-Karbe (2018); Herdegen, Muhle-Karbe, and Possamaï (2019); for proportional costs it turns out to be a doubly-reflected Brownian motion.

To assess the quantitative differences between the respective equilibrium returns, we calibrate our model to market data. This is challenging, since agents' preferences and endowments are not directly observable. However, we show that this difficulty can be overcome as follows. We first pin down some of the parameters by calibrating the frictionless model to a time series of prices. Then, we fit the additional parameters of our model with proportional transaction costs to trading volume data, by exploiting the fact that the average turnover rate in the model can be computed in closed form. To obtain comparable results for other forms of trading costs, we in turn match the corresponding trading volumes and stationary variances of the liquidity premium.

We find that realistic transaction costs lead to substantial fluctuations around the constant frictionless expected returns if agents' trading targets are calibrated to match the large trading volume observed empirically. In contrast, the differences between the results for proportional, quadratic, and intermediate costs are rather small if the magnitude of these costs is matched appropriately. This provides some justification for the use of quadratic trading costs as a proxy for other less tractable specifications.

Trading volume is given by a nonlinear function of the equilibrium returns in our model, and this transformation magnifies the differences between different cost specifications. Indeed, for quadratic costs, volume follows the absolute value of an Ornstein-Uhlenbeck process, whereas subquadratic costs skew volume towards either zero or infinite rates as observed in the limiting case of proportional costs. The trading volume dynamics implied by our model recapture the main stylized facts observed empirically, such as autocorrelation and mean reversion (Lo and Wang, 2000). However, with realistically small transaction costs, our simple stylized model with constant volatilities and trading needs cannot reproduce the strong persistence observed in real time-series data. Likewise, matching the large average turnover rate observed empirically is tied to excessive fluctuations relative to the data.

The remainder of this chapter is organized as follows. Section 2.2 introduces our frictionless baseline model and derives the corresponding equilibrium returns. In Section 2.3, this model and the equilibrium results are extended to general smooth convex costs on the agents' trading rates. The limiting case of proportional transaction costs is treated separately in Section 2.4. Both models are calibrated to time series data in Section 2.5. For better readability, all proofs are collected in Section 2.6.

Notation. Throughout this chapter, we work on a fixed filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, where the filtration is generated by a one-dimensional standard Brownian motion $W = (W_t)_{t\geq 0}$. For $p \geq 1$, we denote by $\mathbb{H}^p$ the $\mathbb{R}$-valued, progressively measurable processes $X = (X_t)_{t\in[0,T]}$ that satisfy

$$\|X\|_{\mathbb{H}^p} := \left(\mathbb{E}\left[\left(\int_0^T |X_t|^2 dt\right)^{p/2}\right]\right)^{1/p} < \infty.$$

2.2 Frictionless Baseline Model

2.2.1 Risk-Sharing Economy

Randomness is generated by a one-dimensional standard Brownian motion $(W_t)_{t\geq 0}$. We consider two agents indexed by $n = 1, 2$ that receive (cumulative) random endowments

$$d\zeta_t^n = \xi_t^n dW_t, \quad \text{where } \xi_t^n = \xi^n W_t, \quad \xi^n \in \mathbb{R}.$$

To hedge against the fluctuations of their endowment streams, the agents trade a safe and a risky asset. The price of the safe asset is exogenous and normalized to one. The price of the risky asset follows

$$dS_t = \mu_t dt + \sigma dW_t.$$

Here, the constant volatility σ is given exogenously, whereas the expected returns process $\mu \in \mathbb{H}^2$ is to be determined endogenously by matching the agents' demand to the fixed supply $s \in \mathbb{R}$ of the risky asset. See Vayanos (1998); Žitković (2012); Choi and Larsen (2015); Kardaras et al. (2015); Garleanu and Pedersen (2016); Xing and Žitković (2018); Bouchard et al. (2018) for related equilibrium models where the volatility also is a free parameter. Models where the volatility is determined endogenously are discussed in Chapters 3, 4, and 5.

Remark 2.1. Unlike for more general preferences, an additional orthogonal component (and a finite variation drift) of the agents' endowments would not change the optimizers of the simple linear-quadratic goal functionals (2.1), (2.3) that we consider below, compare Bouchard et al. (2018). We therefore focus on the present most parsimonious specification.

The restriction to two agents is made to reduce the dimensionality of the problem. More agents can be treated without difficulties in the frictionless case and, using matrix algebra, also for quadratic costs (Bouchard et al., 2018). For more general transaction costs, however, more than two agents would lead to multidimensional nonlinear differential equations. Therefore, we focus on two (representative) agents for tractability.

Likewise, we restrict ourselves to an extremely specific random endowment in this chapter in order to avoid introducing additional state variables for the optimization problems with transaction costs.

Finally, a constant exogenous volatility is also crucial for obtaining analytical results for general transaction costs in Section 2.3 and 2.4 below.

2.2.2 Frictionless Optimization and Equilibrium

As a reference point, we first consider the frictionless version of the model. Starting from fixed initial positions that clear the market, $\varphi_{0-}^1 + \varphi_{0-}^2 = s$,[1]

[1]Here, the left limits indicate that the agents may immediately change their positions at time $t = 0$. This has no effect on their frictionless goal functional, but initial bulk trades are nonnegligible with proportional costs.

the agents choose their positions $\psi \in \mathbb{H}^2$ in the risky asset to maximize one-period expected returns penalized for the corresponding variances. Without transaction costs, the continuous-time version of this criterion is

$$\begin{aligned}
\bar{J}_T^n(\psi) &= \mathbb{E}\left[\int_0^T (\psi_t dS_t + d\zeta_t^n) - \frac{\gamma^n}{2} d\langle \textstyle\int_0^{\cdot} \psi_u dS_u + \zeta^n \rangle_t\right] \\
&= \mathbb{E}\left[\int_0^T \left(\psi_t \mu_t - \frac{\gamma^n}{2}\left(\psi_t \sigma + \xi_t^n\right)^2\right) dt\right]. \qquad (2.1)
\end{aligned}$$

Put differently, agents trade off expected returns against the tracking error relative to the exogenous target position $-\xi^n/\sigma$ as in Choi et al. (2018); Sannikov and Skrzypacz (2016). The optimal strategy for (2.1) is readily determined by point-wise optimization as

$$\varphi_t^n = \frac{\mu_t}{\gamma^n \sigma^2} - \frac{\xi_t^n}{\sigma}, \quad t \in [0, T].$$

The frictionless *equilibrium* return is in turn pinned down by matching the agents' total demand $\varphi_t^1 + \varphi_t^2$ to the supply s of the risky asset at all times $t \in [0, T]$:

$$\bar{\mu}_t = \bar{\gamma}\left[s\sigma + \xi_t^1 + \xi_t^2\right]\sigma, \quad t \in [0, T], \qquad (2.2)$$

where the aggregate risk aversion is defined as

$$\bar{\gamma} = \frac{\gamma^1 \gamma^2}{\gamma^1 + \gamma^2}.$$

The agents' optimal trading strategies corresponding to this frictionless equilibrium return are

$$\bar{\varphi}_t^1 = \frac{s\gamma^2}{\gamma^1 + \gamma^2} + \frac{\gamma^2 \xi_t^2 - \gamma^1 \xi_t^1}{(\gamma^1 + \gamma^2)\sigma}, \qquad \bar{\varphi}_t^2 = s - \bar{\varphi}_t^1, \qquad t \in [0, T].$$

Note that the frictionless equilibrium return and the corresponding optimal trading strategies are independent of the time horizon T. In particular, the frictionless optimizers also maximize the long-run average performance $\bar{J}_T^n/T$ as $T \to \infty$, in that for all competing admissible strategies ψ,

$$\limsup_{T \to \infty} \frac{1}{T}\left[\bar{J}_T^n(\psi) - \bar{J}_T^n(\bar{\varphi}^n)\right] \le 0.$$

With transaction costs – where the optimizers are no longer independent of the planning horizon – we will directly solve the long-run version of (2.1), see Definitions 2.3 and 2.10 below.

2.3 Equilibrium with Costs on the Trading Rate

2.3.1 Transaction Costs and Strategies

We now take into account transaction costs. A popular class of models originating from the optimal execution literature focuses on absolutely continuous trading strategies, cf. Almgren and Chriss (2001); Almgren (2003),

$$\psi_t = \psi^n_{0-} + \int_0^t \dot{\psi}_u du, \quad t \geq 0,$$

and penalizes the trading rate $\dot{\psi}_t = d\psi_t/dt$ with an *instantaneous trading cost* $\lambda G(\dot{\psi}_t)$, where $\lambda > 0$ is a constant. Portfolio choice problems for the most tractable quadratic specification $\lambda G(x) = \lambda x^2/2$ are analyzed in single-agent models by Garleanu and Pedersen (2016); Almgren and Li (2016); Moreau et al. (2017); Guasoni and Weber (2017); equilibrium returns are determined in Garleanu and Pedersen (2016); Sannikov and Skrzypacz (2016); Bouchard et al. (2018). In Guasoni and Weber (2018); Cayé et al. (2019); Bayraktar et al. (2018), single-agent models are solved for the more general power costs $\lambda G(x) = \lambda |x|^q/q$, $q \in (1, 2]$ proposed by Almgren (2003). Below, we will determine equilibrium returns for general smooth convex cost functions G as studied in the duality theory of Guasoni and Rásonyi (2015):

Assumption 2.2. (i) The trading cost $G : \mathbb{R} \to \mathbb{R}_+$ is convex, symmetric, and strictly increasing on $[0, \infty)$, differentiable on $[0, \infty)$, and satisfies $G(0) = 0$;

(ii) The derivative G' is also strictly increasing and differentiable on $(0, \infty)$ with $G'(0) = 0$;

(iii) There exist constants $C > 0$, $K \geq 2$ and $x_0 > 0$ such that

$$|(G')^{-1}(x)| \leq C(1 + |x|^{K-1}) \text{ for all } x \in \mathbb{R}, \quad G''(x) \leq C \text{ for all } |x| > x_0.$$

One readily verifies that the power functions $G(x) = |x|^q/q$, $q \in (1, 2]$ proposed in Almgren (2003) satisfy all of these requirements, as do linear combinations of these power functions. A relevant example beyond the power class is provided by the empirical estimates of Bucci et al. (2019), who find that impact costs are quadratic for small trades but scale with a power of approximately 3/2 for larger order sizes.

With transaction costs, the analogue of the frictionless mean-variance goal functional (2.1) is

$$J^n_T(\dot{\psi}) = \mathbb{E}\left[\int_0^T \left(\psi_t \mu_t - \frac{\gamma^n}{2}(\psi_t \sigma + \xi^n_t)^2 - \lambda G(\dot{\psi}_t)\right) dt\right]. \tag{2.3}$$

Unlike its frictionless counterpart, this optimization problem is no longer "myopic", since the current position influences future choices in the presence of transaction costs, and since optimal strategies naturally depend on a finite time horizon T here. To simplify the analysis below, we therefore focus on the ergodic limit of (2.3), where the goal is to maximize the long-run average performance $J_T^n(\dot{\psi})/T$ as $T \to \infty$. This criterion has a long history in single-agent problems with transaction costs, cf. Dumas and Luciano (1991); Taksar et al. (1988); De Lataillade et al. (2012); Gerhold et al. (2014); Guasoni and Weber (2017). Here, we show that it also makes the equilibrium analysis of general trading costs tractable. Throughout, we focus on *admissible* strategies

$$\psi_t = \psi_{0-}^n + \int_0^t \dot{\psi}_u du, \quad t \geq 0$$

that satisfy the integrability conditions

$$\mathbb{E}\left[\int_0^T G(\dot{\psi}_t)dt\right] < \infty, \qquad \mathbb{E}\left[\int_0^T \psi_t^2 dt\right] < \infty, \quad \text{for all } T > 0, \tag{2.4}$$

as well as the transversality condition

$$\lim_{T\to\infty} \frac{1}{T^2}\mathbb{E}[\psi_T^2] = 0. \tag{2.5}$$

2.3.2 Equilibrium

For tractability, we focus on long-run Radner equilibria where agents maximize the long-term average of their finite-horizon goal functionals:

Definition 2.3. A process $\mu = (\mu_t)_{t\in[0,T]} \in \mathbb{H}^2$ is *(long-run) equilibrium return* if there exist admissible trading rates $\dot{\varphi}^1$, $\dot{\varphi}^2$ for agents 1 and 2 such that:

(i) *market clearing*: The agents' optimal position clears the the market for the risky asset at all times, i.e. $\varphi_t^1 + \varphi_t^2 = s$, $t \in [0,T]$.

(ii) *individual optimality*: The trading rate $\dot{\varphi}^n$ is optimal for the long-run version of agent n's control problem (2.3) in that,

$$\limsup_{T\to\infty} \frac{1}{T}\left[J_T^n(\dot{\psi}) - J_T^n(\dot{\varphi}^n)\right] \leq 0; \tag{2.6}$$

Remark 2.4. Note that as in, e.g., Lo et al. (2004); Buss and Dumas (2019), our transaction cost is an exogenous deadweight cost and not an output of the trading process in equilibrium.

The construction of the equilibrium return is based on the solution of a nonlinear ODE. For single-agent models with instantaneous trading costs of power form, a corresponding equation has been introduced and studied by

Guasoni and Weber (2018).[2] In Section 2.6.3, we show that their existence and uniqueness proof can be extended to general cost functions satisfying Assumption 2.2.

Lemma 2.5. *Suppose the instantaneous trading cost G satisfies Assumption 2.2. Then the ordinary differential equation*

$$\frac{1}{2}\left(\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}\right)^2 g''(x)+g'(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right)=\frac{(\gamma^1+\gamma^2)\sigma^2}{2}x \tag{2.7}$$

has a unique solution g on $\mathbb{R}$ such that $xg(x)\leq 0$ for all $x\in\mathbb{R}$. Moreover, g is odd, non-increasing on $\mathbb{R}$ and g satisfies the growth conditions

$$\lim_{x\to-\infty}\frac{g(x)}{\lambda(G^*)^{-1}(\frac{(\gamma^1+\gamma^2)\sigma^2}{4\lambda}x^2)}=1,\quad \lim_{x\to+\infty}\frac{g(x)}{\lambda(G^*)^{-1}(\frac{(\gamma^1+\gamma^2)\sigma^2}{4\lambda}x^2)}=-1, \tag{2.8}$$

where G^* is the Legendre transform of G.

Remark 2.6. For power functions $G(x)=|x|^q/q$, $q\in(1,2]$, the Legendre transform is

$$\begin{aligned}G^*(x)&=\sup_y\{xy-G(y)\}=x(G')^{-1}(x)-G\left((G')^{-1}(x)\right)\\&=x\,\mathrm{sgn}(x)|x|^{1/(q-1)}-\frac{1}{q}|x|^{q/(q-1)}=\frac{q-1}{q}|x|^{q/q-1}=|x|^p/p,\end{aligned}$$

where $p=q/(q-1)$ is the conjugate of q.

With the function g from Lemma 2.5, we can now define the ergodic state variable that will drive both the expected returns and optimal trading rates in equilibrium. (In equilibrium, this process describes the deviation of agent 1's actual position φ^1_t from its frictionless counterpart $\bar{\varphi}^1_t$, compare (2.12), motivating our notation.)

Lemma 2.7. *Let g be the solution of the ODE (2.7) from Lemma 2.5. There exists a unique strong solution of the SDE*

$$\begin{aligned}d\Delta\varphi^1_t&=(G')^{-1}\left(\frac{g(\Delta\varphi^1_t)}{\lambda}\right)dt+\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}dW_t,\quad t\geq 0,\\ \Delta\varphi^1_0&=\varphi^1_{0-}-\frac{s\gamma^2}{\gamma^1+\gamma^2}.\end{aligned} \tag{2.9}$$

Moreover, this process is a recurrent diffusion.

Proof. See Section 2.6.1. □

[2]Indeed, if $G(x)=|x|^q/q$, $q\in(1,2]$, then differentiating the first-order ODE (15) in (Guasoni and Weber, 2018, Theorem 4.1) and a change of variables as in Section 2.6.4 lead to the second-order ODE (2.7). The same link to a first-order equation is exploited in our existence proof in Section 2.6.3.

Remark 2.8. If the instantaneous trading cost is quadratic, $G(x) = x^2/2$, then with $(G')^{-1}(x) = x$ and the solution to the ODE (2.7) from Lemma 2.5 is

$$g(x) = -\left(\frac{(\gamma^1+\gamma^2)\sigma^2\lambda}{2}\right)^{1/2} x.$$

Accordingly,

$$\begin{aligned} d\Delta\varphi_t^1 &= -\left(\frac{(\gamma^1+\gamma^2)\sigma^2}{2\lambda}\right)^{1/2} \Delta\varphi_t^1 dt + \frac{\gamma^1\xi^1 - \gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma} dW_t, \quad t \geq 0, \\ \Delta\varphi_0^1 &= \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}. \end{aligned}$$

Whence, $\Delta\varphi_t^1$ is an Ornstein-Uhlenbeck process in this case. In general, the drift rate in (2.9) describes the nonlinear attraction of the process $\Delta\varphi_t^1$ towards its average level zero, where $xg(x) \leq 0$ ensures that the process is indeed mean reverting and in turn converges to an ergodic limit.

We now present our first main result. It identifies the equilibrium return for general smooth, convex cost functions.

Theorem 2.9. *Recall $\bar\gamma = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}$. With the solution $(\Delta\varphi_t^1)_{t\geq 0}$ of (2.9), define*

$$\mu_t = \bar\gamma\left[s\sigma + \xi_t^1 + \xi_t^2\right]\sigma + \frac{(\gamma^1-\gamma^2)\sigma^2}{2}\Delta\varphi_t^1, \quad t \geq 0. \tag{2.10}$$

Then, the trading rates

$$\dot\varphi_t^1 = -\dot\varphi_t^2 = (G')^{-1}\left(\frac{g(\Delta\varphi_t^1)}{\lambda}\right), \quad t \geq 0 \tag{2.11}$$

clear the corresponding market and are individually optimal in the long run. Therefore, $(\mu_t)_{t\geq 0}$ is an equilibrium return.

Proof. See Section 2.6.1. □

The first term in (2.10) is the frictionless equilibrium return from (2.2). Accordingly, the second term describes how the equilibrium return changes due to transaction costs. Evidently, if both agents have the same risk aversion, then the adjustment is zero like for the quadratic costs studied by Bouchard et al. (2018). In this case, both agents are adversely affected by the transaction costs, but the market still clears at the frictionless equilibrium price.

For heterogenous agents, there is a nontrivial liquidity premium depending on the current demand imbalance. Indeed, in equilibrium, the state dynamics $d\Delta\varphi_t$ also describe the evolution of the deviation between agent 1's actual position and its frictionless counterpart,

$$d\Delta\varphi_t^1 = (G')^{-1}\left(\frac{g(\Delta\varphi_t^1)}{\lambda}\right) dt + \frac{\gamma^1\xi^1 - \gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma} dW_t = d(\varphi_t^1 - \bar\varphi_t^1). \tag{2.12}$$

By market clearing, the sign is reversed for agent 2. Accordingly, the liquidity premium is positive if the more risk averse agent sells and negative if the more risk averse agent buys to move closer to the corresponding frictionless allocation. In each case, the return adjustment ensures market clearing by offsetting the more risk averse agent's stronger motive to trade.

For quadratic costs, we recover the Ornstein-Uhlenbeck returns from Corollary 5.5 in Bouchard et al. (2018). For general convex trading costs, these are replaced by processes with nonlinear mean-reversion speeds.

2.4 Equilibrium with Proportional Costs

One important cost specification is not covered by Assumption 2.2: proportional transaction costs. These arise as the limit $p \to 1$ in the model of Almgren (2003). Rather than studying the (singular) limiting behaviour of the corresponding optimal strategies as in Guasoni and Weber (2018), we instead show that the equilibrium with proportional costs can be constructed directly using singular rather than regular stochastic control.

Since proportional costs only penalize trade size but not speed, risky positions are naturally described by general finite-variation processes in this case or, equivalently, by their Jordan-Hahn decompositions into minimal increasing processes – the cumulative numbers of shares purchased and sold:

$$\psi_t = \psi^n_{0-} + \psi^{\uparrow}_t - \psi^{\downarrow}_t.$$

As in Janeček and Shreve (2010); Martin and Schöneborn (2011); De Lataillade et al. (2012); Martin (2014) we assume for simplicity that the (cumulative) costs $\lambda(\psi^{\uparrow}_T + \psi^{\downarrow}_T)$, $\lambda > 0$, are proportional to the number of shares traded (rather than the monetary amount transacted). Agent n's goal functional in turn becomes

$$J^n_T(\psi) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\left(\psi_t\sigma + \xi^n_t\right)^2\right)dt - \lambda(\psi^{\uparrow}_T + \psi^{\downarrow}_T)\right]. \tag{2.13}$$

We again focus on the long-run average performance $J^n_T(\varphi)/T$ as $T \to \infty$ of *admissible strategies* that satisfy the integrability condition

$$\mathbb{E}\left[\int_0^T \psi_t^2 dt\right] < \infty, \qquad \mathbb{E}[\psi^{\uparrow}_T + \psi^{\downarrow}_T] < \infty, \quad \text{for all } T > 0, \tag{2.14}$$

as well as the transversality condition

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[|\psi_T|\right] = 0. \tag{2.15}$$

2.4.1 Equilibrium

We use an analogous notion of Radner equilibrium as in Definition 2.3:

Definition 2.10. A process $\mu \in \mathbb{H}^2$ is a *(long-run) equilibrium return* if there exist admissible strategies φ^1, φ^2 for agents 1 and 2 such that:

(i) *market clearing*: The agents' optimal position clear the the market for the risky asset at all times, i.e. $\varphi_t^1 + \varphi_t^2 = s$, $t \in [0, T]$;

(ii) *individual optimality*: The strategy φ^n is optimal for the long-run version of agent n's control problem (2.13) in that,

$$\limsup_{T\to\infty} \frac{1}{T} \left[J_T^n(\psi) - J_T^n(\varphi^n)\right] \leq 0. \tag{2.16}$$

The construction of the equilibrium return with proportional costs is based on the analogue of the mean-reverting process from Lemma 2.7. This turns out to be a doubly-reflected Brownian motion,

$$d\Delta\varphi_t^1 = \frac{\gamma^1\xi^1 - \gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma} dW_t + dL_t - dU_t. \tag{2.17}$$

where $\Delta\varphi_{0-}^1 = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}$ and L, U are the minimal increasing processes with $L_{0-} = U_{0-} = 0$ that keep $(\Delta\varphi_t^1)_{t\geq 0}$ in the interval $[-l, l]$,[3] whose endpoints have the following explicit expression:

$$l = \left(\frac{3\lambda(\gamma^1\xi^1 - \gamma^2\xi^2)^2}{(\gamma^1+\gamma^2)^3\sigma^4}\right)^{1/3}. \tag{2.18}$$

With the state variable $\Delta\varphi^1$ at hand, we can now formulate our second main result. It shows that the equilibrium return with proportional costs can be expressed in direct analogy to its counterpart for the smooth, superlinear costs treated in Theorem 2.9. The only difference is that the mean-reverting state variable in Theorem 2.9 is replaced by the doubly-reflected Brownian motion from (2.17).

Theorem 2.11. *Recall* $\bar\gamma = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}$. *With the solution* $(\Delta\varphi_t^1)_{t\geq 0}$ *of* (2.17), *define*

$$\mu_t = \bar\gamma\left[s\sigma + \xi_t^1 + \xi_t^2\right]\sigma + \frac{(\gamma^1-\gamma^2)\sigma^2}{2}\Delta\varphi_t^1, \quad t \geq 0. \tag{2.19}$$

Then, the trading strategies

$$\varphi_t^1 = \varphi_{0-}^1 + L_t - U_t, \quad \varphi_t^2 = s - \varphi_{0-}^2 + U_t - L_t, \quad t \geq 0, \tag{2.20}$$

clear the market and are individually optimal in the long run. Therefore $(\mu_t)_{t\geq 0}$ *is an equilibrium return.*

[3] See Kruk et al. (2007) for the pathwise construction of L, U. In particular, there is an initial jump in L or U if the initial value $\Delta\varphi_{0-}^1$ lies below $-l$ or above l, respectively. On $(0, T]$, L and U have continuous paths.

Proof. See Section 2.6.2. □

Note that, in equilibrium, each agent's singular control problem has a fully explicit solution. Similar closed-form expressions for optimal no-trade regions also obtain for the ergodic control of Brownian motion, which underlies the tractability of problems with *small* transaction costs Soner and Touzi (2013); Kallsen and Muhle-Karbe (2017); Cai et al. (2017). Surprisingly, the equilibrium constructed in Theorem 2.11 displays the same tractability, even though the corresponding equilibrium return is not zero but a reflected Brownian motion.

2.5 Calibration

To assess the quantitative properties of our equilibrium returns, we now calibrate the model to price and trading-volume data for the US equity market. More specifically, we consider the 320 current constituents of the S&P500 for which ten years of uninterrupted data are available from January 2, 2009 to January 2, 2019 on the CRSP database.[4] To obtain the price dynamics of a "typical stock", we then compute the capitalization-weighted average of the respective prices. The total number of outstanding shares of this average stock then is the number of shares outstanding for all our stocks. Likewise, the total share turnover is also aggregated across all stocks.

2.5.1 Calibration of the Frictionless Baseline Model

We first consider the frictionless baseline version of the model from Section 2.2.2. The exogenous (absolute) daily volatility σ can be estimated directly from the time series of stock prices, leading to $\sigma = 1.88$ for our dataset.[5] To obtain a simple parsimonious model for the equilibrium returns, we suppose throughout as in Lo et al. (2004) that there is no aggregate endowment ($\xi^1_t = -\xi^2_t$). Then, the frictionless equilibrium expected return from (2.2) is $\bar{\mu} = \bar{\gamma} s \sigma^2$. As the number of shares outstanding is $s = 2.46 \times 10^{11}$, we choose $\bar{\gamma} = 8.31 \times 10^{-14}$ to match this to the average (absolute) daily returns $\bar{\mu}$ of 0.072 in our time series.[6]

2.5.2 Calibration with Transaction Costs

Whereas the frictionless equilibrium price only depends on the aggregate risk aversion $\bar{\gamma} = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}$ and aggregate endowment $\xi^1 + \xi^2$, the individual values of

[4] We do not work with an even longer time series, since the corresponding larger changes in price levels then become problematic for our arithmetic model.

[5] Since the average stock prices are 124.11 in our time series, this corresponds to a Black-Scholes volatility of around 23.8%.

[6] This corresponds to a yearly Black-Scholes return of 14.44% relative to the average price level.

these parameters need to be pinned down to determine equilibria with transaction costs. Moreover, the initial allocations of the agents need to be specified and an appropriate estimate for the respective trading cost is evidently needed.

Proportional Costs For proportional costs, we use the estimate obtained in Novy-Marx and Velikov (2016) for value-weighted trading strategies: 0.25% of the average stock prices, that is $\lambda_1 = 0.31$ for our dataset.[7] Once the aggregate risk aversion $\bar{\gamma}$ is fixed, the individual agents' absolute risk aversions γ^1, γ^2 are free parameters in the present model, which correspond to the agents' sizes relative to each other. If both agents are of the same size, the frictional equilibrium coincides with its frictionless counterpart. To illustrate the effect of heterogeneity, we set $\gamma^2 = 2\gamma^1$, so that the larger agent 2 has twice the risk capacity of agent 1.[8] Then, with $\bar{\gamma} = 8.31 \times 10^{-14}$ we have $\gamma^1 = 1.25 \times 10^{-13}$ and $\gamma^2 = 2.5 \times 10^{-13}$. For the initial allocations, we suppose for simplicity that each agent initially holds a fraction of the total supply equal to their share of the total risk tolerance, $\varphi^1_{0-} = \frac{\gamma^2}{\gamma^1+\gamma^2} s = s - \varphi^2_{0-}$. This minimizes the effect of transaction costs because no initial bulk trades are necessary in this case. But the initial allocation generally only affects the initial conditions of the state variables in our long-run equilibria in Theorems 2.11, so that the effect of different specifications disappears quickly in any case.

Finally, we calibrate the value of the endowment volatilities $\xi^1_1 = -\xi^2_1 = \xi_1$ to time-series data for trading volume. More specifically, given our estimate $\lambda_1 = 0.312$ from the proportional cost, we choose the parameter β_1 to match the average daily share turnover in 2009-2018, which is $\mathrm{ShTu} = 1.84 \times 10^9$ (that is, about 187% of the outstanding shares per year), to the corresponding long-term average value in our model. Using the ergodic theorem, the latter can be calculated as in (Gerhold et al., 2014, Lemma C.2),

$$\mathrm{ShTu} = \lim_{T\to\infty} \frac{1}{T} \int_0^T d|\varphi|_t = \lim_{T\to\infty} \frac{L_T}{T} + \lim_{T\to\infty} \frac{U_T}{T} = \left(\frac{\gamma^1+\gamma^2}{24\lambda_1\sigma^2} \right)^{1/3} \xi_1^{4/3} \quad \text{a.s.}$$

Accordingly, we have

$$\xi_1 = \left(\frac{24\mathrm{ShTu}^3\lambda_1\sigma^2}{\gamma^1+\gamma^2} \right)^{1/4} = 2.57 \times 10^{10}.$$

Superlinear Costs For comparison, we also consider the power transaction costs $\lambda_q G_q(x) = \lambda_q |x|^q/q$, $q \in (1, 2]$. In this case, to choose the endowment volatilities $\xi^1_q = -\xi^2_q = \xi_q$, we apply the ergodic theorem to compute the

[7]Somewhat larger bid-ask spreads of 1% are used by Lynch and Tan (2011); Buss and Dumas (2019), for example.

[8]These specific parameter values are chosen because they lead to realistic levels of liquidity premia in the extended version of the model with endogenous volatilities, see Section 4.4.5.

long-term average of the daily share turnover as

$$\mathrm{ShTu} = \lim_{T\to\infty} \frac{1}{T} \int_0^T |\dot{\varphi}_t^1| dt = \int_{-\infty}^{\infty} \left| (G_q')^{-1} \left(\frac{g_q(x)}{\lambda_q} \right) \right| \nu_q(x) dx \quad \text{a.s.}$$

Here, $\nu_q(x)$ is the invariant density of the stationary law of the state variable X. For quadratic costs $\lambda_2 G_2(x) = \lambda_2 x^2/2$, this is an Ornstein-Uhlenbeck process (cf. Remark 2.8) whose stationary distribution is Gaussian with mean zero and variance $(\lambda_2 \xi_2^4)^{1/2}/\sigma^3(2(\gamma^1+\gamma^2))^{1/2}$. As

$$(G_2')^{-1} \left(\frac{g_2(x)}{\lambda_2} \right) = - \left(\frac{(\gamma^1+\gamma^2)\sigma^2}{2\lambda_2} \right)^{1/2} x,$$

the average turnover per year in turn is proportional to the endowment volatility ξ_2 in this case,

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T |\dot{\varphi}_t^1| dt = \left(\frac{\gamma^1+\gamma^2}{2\pi^2\sigma^2\lambda_2} \right)^{1/4} \xi_2, \quad \text{a.s.}$$

Accordingly, to match the average share turnover for a given quadratic transaction cost λ_2, we need $\xi_2 = \mathrm{ShTu}/(\frac{\gamma^1+\gamma^2}{2\pi^2\sigma^2\lambda_2})^{1/4}$. Whence, it remains to choose an appropriate value for the trading cost parameter λ_2. To make its impact comparable to the proportional cost, we choose it to obtain the same stationary variance of the state variable $\Delta\varphi$ as with proportional costs.

With proportional costs, this process has a uniform stationary law with standard deviation $l/3^{1/2}$. With quadratic costs, the stationary standard deviation of the Ornstein-Uhlenbeck state variable is

$$\xi_2 \left(\frac{\lambda_2 \sigma^2}{4\gamma} \right)^{1/4} = \mathrm{ShTu} \left(\frac{\pi \lambda_2}{2\gamma\sigma^2} \right)^{1/2}.$$

To match this with the stationary standard deviation for proportional costs, we choose $\lambda_2 = 1.08 \times 10^{-10}$. This leads to $\xi_2 = 2.19 \times 10^{10}$. Note that this "equivalent quadratic cost" is of the same order of magnitude as the direct estimate obtained from proprietary trade execution data for S&P500 stocks in (Collin-Dufresne et al., 2020, Table 5).

For general power costs $\lambda_q G_q(x) = \lambda_q |x|^q/q$ the solution g_q of the ODE (2.7) is not known explicitly. However, by exploiting the homotheticity of the power function, a change of variable allows us to reduce (2.7) to an equation that only depends on the elasticity q of the price impact function, but not the parameters λ_q, ξ_q that we are trying to determine here. Accordingly, the values of λ_q, ξ_q that match the average share turnover observed empirically as well as the variance of the state variable for proportional costs can be expressed as integrals of this universal function. For fixed q, these can in turn be computed by using a quadrature formula to integrate the numerical solution of (2.7), cf. Subsection 2.6.4 for more details. For $q = 3/2$, which is in line with empirical estimates of actual trading costs in Almgren et al. (2005); Lillo et al. (2003), this leads to

$$\xi_{3/2} = 2.33 \times 10^{10}, \qquad \lambda_{3/2} = 5.22 \times 10^{-6}.$$

Analogously, for $q = 1.125$ and trading costs close to proportional, we obtain

$$\xi_{1.125} = 2.50 \times 10^{10}, \qquad \lambda_{1.125} = 0.019.$$

Simulations of ten years of daily equilibrium returns (generated with the same Brownian sample path) for these four sets of parameters are shown in Figure 2.1. For our calibrated parameters, the frictional equilibrium returns display substantial deviations around their frictionless counterpart, but the differences between the equilibrium returns for the different cost specifications is much smaller.

Even though these numbers are generated from just one sample path, they in fact quite accurately reflect the stationary distributions of the state variables by the ergodic theorem. This is illustrated in Figure 2.2, where we compare the empirical probability density functions to the stationary normal distribution for $q = 2$. While the empirical distribution clearly does become more spread out for smaller q (it is normal for $q = 2$ but uniform for $q = 1$), the realized distributions are nevertheless quite similar for our calibrated parameters.

The simulated daily share turnover for $q = 2$ and $q = 3/2$ is compared to the historical trading volume data in Figure 2.3. By the calibration, the averages of the simulated trading volumes agree with the empirical data and broadly display the same mean-reverting behavior.

However, for the simple model with constant price volatility and homogenous trading needs, the variances of trading volume are substantially larger than in the data. Moreover, the autocorrelation functions in the model also decay much faster than their empirical counterparts.

2.6 Proofs

To ease notation, define

$$\widetilde{\gamma} = \frac{\gamma^1 + \gamma^2}{2}, \qquad \delta = \frac{\gamma^1 \xi^1 - \gamma^2 \xi^2}{(\gamma^1 + \gamma^2)\sigma}. \tag{2.21}$$

Then the ODE (2.7) in Lemma 2.5 can be rewritten as

$$\frac{\delta^2}{2} g''(x) + g'(x) \left(G'\right)^{-1} \left(\frac{g(x)}{\lambda} \right) = \widetilde{\gamma} \sigma^2 x,$$

and the SDE (2.9) in Lemma 2.7 as

$$d\Delta\varphi_t^1 = \left(G'\right)^{-1} \left(\frac{g(\Delta\varphi_t^1)}{\lambda} \right) dt + \delta dW_t.$$

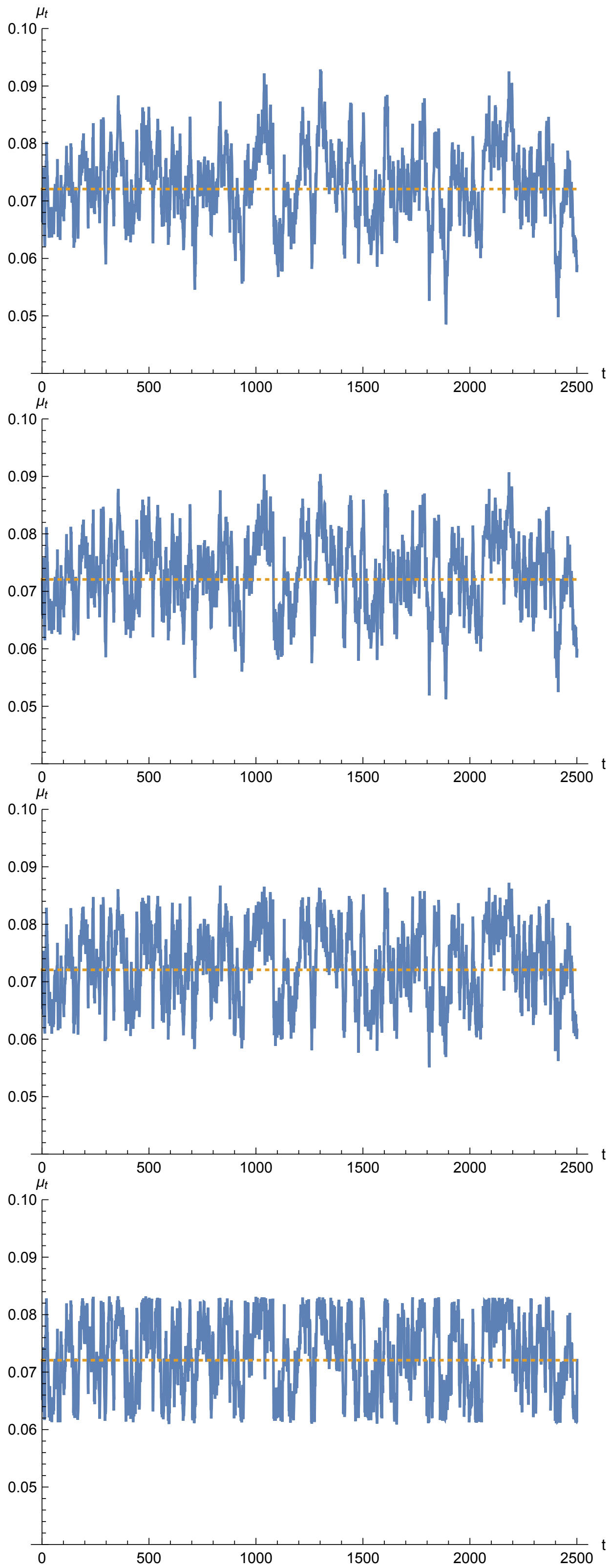

Figure 2.1: Simulated frictional equilibrium returns with calibrated parameters for quadratic trading costs (top panel), costs proportional to the 3/2-th power of the agents' trading rates (second panel), to the 9/8-th power (third panel), and proportional costs (bottom panel). The corresponding (daily) frictionless equilibrium return is constant and equal to 0.072 here.

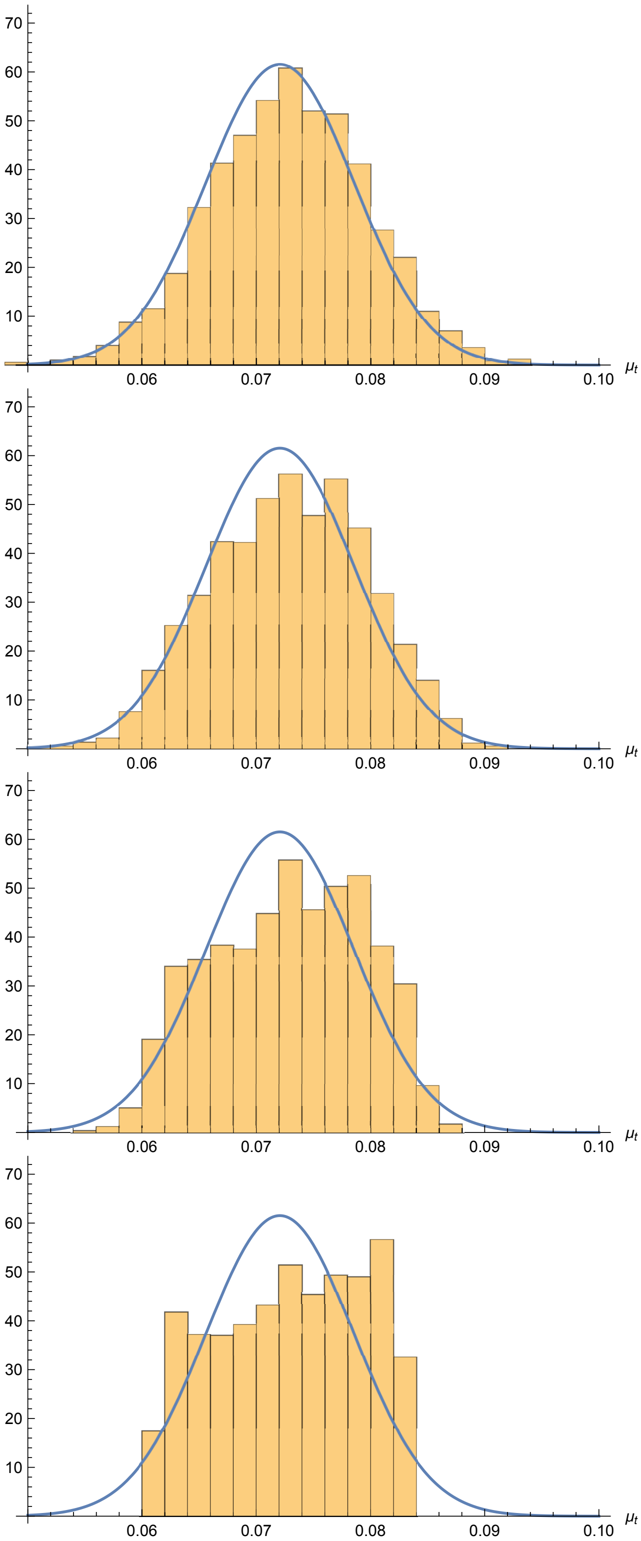

Figure 2.2: Empirical probability density functions for quadratic trading costs (top panel), costs proportional to the 3/2-th power of the agents' trading rates (second panel), to the 9/8-th power (third panel), and proportional costs (bottom panel) compared to the stationary normal distribution for quadratic costs.

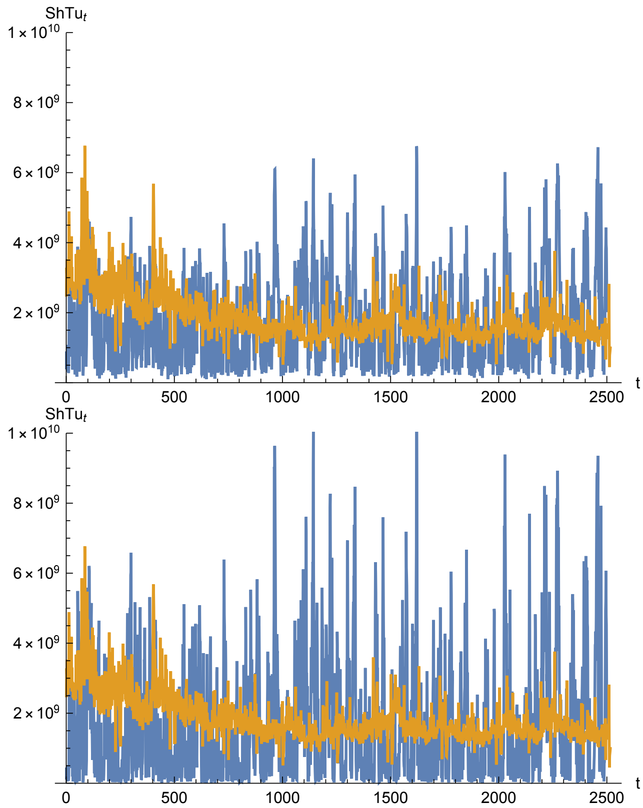

Figure 2.3: Simulated daily share turnover for quadratic trading costs (blue, upper panel) and costs proportional to the 3/2-th power of the agents' trading rates (blue, lower panel), compared to empirical trading volume (orange).

2.6.1 Proofs for Section 2.3

Proof of Lemma 2.7. This proof is based on Lemma 2.5, whose proof is in Section 2.6.3. The (strong) existence and uniqueness of the (fast) mean-reverting processes. follows from results of Veretennikov (1997). Strong existence and uniqueness follow from a standard localization argument, cf. (Cayé et al., 2020, Proof of Proposition 1.1). By Lemma 2.5, we have $g(x) \leq 0$ for $x > 0$ and, in view of Assumption 2.2(iii) there exists $M > 0$ such that

$$|(G')^{-1}(x)| \geq \frac{c}{2}|x| \quad \text{for } |x| \geq |M|.$$

As $(G')^{-1}$ is odd, it follows that, for x such that $|g(x)| \geq \lambda M$,

$$x(G')^{-1}\left(\frac{g(x)}{\lambda}\right) = -|x|(G')^{-1}\left(\frac{|g(x)|}{\lambda}\right) \leq -\frac{c}{2}|x||g(x)|.$$

From the properties of g listed in Lemma 2.5, we can infer that $|g|$ is increasing on $[0, \infty)$ and satisfies $\lim_{|x|\to\infty} |g(x)| = \infty$. Whence, there exists $M_0 > 0$ such that for every $|x| > M_0$, we have

$$g(-|x|) = |g(x)| \geq \lambda M.$$

Hence, for every $r > 0$ and $|x| \geq 2r/c|g(M_0)| + M_0$,

$$\frac{x}{|x|}(G')^{-1}\left(\frac{g(x)}{\lambda}\right) \leq -\frac{c}{2}|g(x)| \leq -\frac{c}{2}|g(M_0)| \leq -\frac{r}{|x|}.$$

Thus, (Veretennikov, 1997, Condition (6)) is satisfied and the (strong) existence and uniqueness result follows. For later use also note that, by (Veretennikov, 1997, Lemma 1), we have the following uniform moment bounds:

$$\sup_{T\geq 0} \mathbb{E}\left[|\Delta\varphi_T^1|^k\right] < \infty, \quad \text{for every } k \in \mathbb{N}. \tag{2.22}$$

The second part of the assertion is established in (Cayé et al., 2020, Appendix D.2). □

Proof of Theorem 2.9. Market clearing evidently holds by definition of the trading rates (2.11). Observe that the corresponding strategy φ^1 is admissible and satisfies the transversality condition (2.5). Moreover,

$$
\begin{aligned}
&\mu_t - \gamma^1(\varphi_t^1\sigma + \xi_t^1)\sigma \\
&= \frac{\gamma^1-\gamma^2}{2}\sigma^2\Delta\varphi_t^1 + \bar{\gamma}\left[s\sigma + \xi_t^1 + \xi_t^2\right]\sigma - \gamma^1(\varphi_t^1\sigma + \xi_t^1)\sigma \\
&= \frac{\gamma^1-\gamma^2}{2}\sigma^2\Delta\varphi_t^1 - \gamma^1\sigma^2\varphi_t^1 + \frac{\gamma^1\sigma}{\gamma^1+\gamma^2}\left[\gamma^2 s\sigma + \gamma^2\xi_t^1 + \gamma^2\xi_t^2 - (\gamma^1+\gamma^2)\xi_t^1\right] \\
&= \frac{\gamma^1-\gamma^2}{2}\sigma^2\Delta\varphi_t^1 - \gamma^1\sigma^2\varphi_t^1 + \frac{\gamma^1\sigma}{\gamma^1+\gamma^2}\left[\gamma^2 s\sigma + \gamma^2\xi_t^2 - \gamma^1\xi_t^1\right] \\
&= \frac{\gamma^1-\gamma^2}{2}\sigma^2\Delta\varphi_t^1 - \gamma^1\sigma^2\varphi_t^1 + \gamma^1\sigma^2\bar{\varphi}_t^1 \\
&= \frac{\gamma^1-\gamma^2}{2}\sigma^2\Delta\varphi_t^1 - \gamma^1\sigma^2\Delta\varphi_t^1 \\
&= -\widetilde{\gamma}\sigma^2\Delta\varphi_t^1.
\end{aligned}
$$

Hence,

$$
\mu_t - \gamma^1(\varphi_t^1\sigma + \xi_t^1)\sigma = -\widetilde{\gamma}\sigma^2\Delta\varphi_t^1. \tag{2.23}
$$

Consider a competing admissible strategy ψ for the first agent and, to ease notation, set

$$
\dot{\theta}_t = \dot{\psi}_t - \dot{\varphi}_t^1, \qquad \text{so that} \quad \theta_t = \int_0^t \left(\dot{\psi}_u - \dot{\varphi}_u^1\right) du = \psi_t - \varphi_t^1.
$$

Identity (2.23) and the convexity of G yield

$$
\begin{aligned}
&J_T^1(\dot{\psi}) - J_T^1(\dot{\varphi}^1) \\
&= \mathbb{E}\left[\int_0^T \theta_t\mu_t - \frac{\gamma^1}{2}\theta_t(\psi_t\sigma + \varphi_t^1\sigma + 2\xi_t^1)\sigma + \lambda\big(G(\dot{\varphi}_t^1) - G(\dot{\psi}_t)\big)\, dt\right] \\
&= \mathbb{E}\left[\int_0^T \theta_t\mu_t - \frac{\gamma^1}{2}\theta_t\big(\psi_t\sigma - \varphi_t^1\sigma + 2(\varphi_t^1\sigma + \xi_t^1)\big)\sigma + \lambda\big(G(\dot{\varphi}_t^1) - G(\dot{\psi}_t)\big)\, dt\right] \\
&\leq \mathbb{E}\left[\int_0^T -\frac{1}{2}\gamma^1\left(\theta_t\sigma\right)^2 + \theta_t\big(\mu_t - \gamma^1(\varphi_t^1\sigma + \xi_t^1)\sigma\big) + \lambda G'(\dot{\varphi}_t^1)\big(\dot{\varphi}_t^1 - \dot{\psi}_t\big)\, dt\right] \\
&= \mathbb{E}\left[\int_0^T -\frac{1}{2}\gamma^1\left(\theta_t\sigma\right)^2 - \gamma\theta_t\sigma^2\Delta\varphi_t^1 - \lambda G'(\dot{\varphi}_t^1)\dot{\theta}_t\, dt\right].
\end{aligned} \tag{2.24}
$$

We now analyze the terms on the right-hand side. The dynamics (2.9) of $\Delta\varphi_t^1$, Itô's formula, and the ODE (2.7) for g imply

$$
\begin{aligned}
dg(\Delta\varphi_t^1) &= \left(\frac{1}{2}\delta^2 g''(\Delta\varphi_t^1) + g'(\Delta\varphi_t^1)(G')^{-1}\left(\frac{g(\Delta\varphi_t^1)}{\lambda}\right)\right) dt + g'(\Delta\varphi_t^1)\delta dW_t \\
&= \widetilde{\gamma}\sigma^2\Delta\varphi_t^1 dt + g'(\Delta\varphi_t^1)\delta dW_t.
\end{aligned} \tag{2.25}
$$

Integration by parts and the dynamics (2.25) in turn yield

$$d\left[\theta_t g(\Delta\varphi_t^1)\right] = \left(\dot{\theta}_t g(\Delta\varphi_t^1) + \gamma\theta_t\sigma^2\Delta\varphi_t^1\right)dt + \theta_t g'(\Delta\varphi_t^1)\delta dW_t. \tag{2.26}$$

Here, the local martingale part is a true martingale. Indeed, by Hölder's inequality, the integrability condition (2.4) and the boundedness of g' established in Lemma 2.18,

$$\mathbb{E}\left[\int_0^t |\theta_u g'(\Delta\varphi_u^1)|^2 du\right] \leq K^2\mathbb{E}\left[\int_0^t \theta_u^2 du\right] < \infty.$$

Also taking into account that

$$G'(\dot{\varphi}_t^1) = G'\left((G')^{-1}\left(\frac{g(\Delta\varphi_t^1)}{\lambda}\right)\right) = \frac{g(\Delta\varphi_t^1)}{\lambda},$$

we can therefore use (2.26) to replace the second and the third terms on the right-hand side of (2.24), obtaining

$$J_T^1(\dot{\psi}) - J_T^1(\dot{\varphi}^1) \leq -\mathbb{E}[g(\Delta\varphi_T^1)\theta_T] - \mathbb{E}\left[\int_0^T \frac{1}{2}\gamma^1\left(\theta_t\sigma\right)^2 dt\right].$$

The Cauchy-Schwartz inequality yields

$$\begin{aligned}\left|\mathbb{E}[g(\Delta\varphi_T^1)\theta_T]\right| &\leq \left(\mathbb{E}[g(\Delta\varphi_T^1)^2]\mathbb{E}[\theta_T^2]\right)^{1/2}\\ &\leq \left(\mathbb{E}[2g(\Delta\varphi_T^1)^2](\mathbb{E}[(\psi_T)^2] + \mathbb{E}[(\varphi_T^1)^2])\right)^{1/2}.\end{aligned}$$

By the polynomial growth of g established in Lemma 2.18 and (2.22), we have $\sup_{T\geq 0}\mathbb{E}[g(\Delta\varphi_T^1)^2] < \infty$. Together with the transversality condition (2.5), it follows that

$$\begin{aligned}0 &\leq \lim_{T\to\infty}\frac{1}{T}\left|\mathbb{E}[g(\Delta\varphi_T^1)\theta_T]\right|\\ &\leq \lim_{T\to\infty}\frac{1}{T}\left(\mathbb{E}[2g(\Delta\varphi_T^1)^2](\mathbb{E}[(\psi_T)^2] + \mathbb{E}[(\varphi_T^1)^2])\right)^{1/2} = 0.\end{aligned}$$

Therefore, the trading rate $\dot{\varphi}^1$ is indeed long-run optimal for agent 1:

$$\begin{aligned}&\limsup_{T\to\infty}\frac{1}{T}\left[J_T^1(\dot{\psi}) - J_T^1(\dot{\varphi}^1)\right]\\ &\leq \limsup_{T\to\infty}\frac{1}{T}\left[-\mathbb{E}[g(\Delta\varphi_T^1)\theta_T] - \mathbb{E}\left[\int_0^T \frac{1}{2}\gamma^1\left(\theta_t\sigma\right)^2 dt\right]\right]\\ &= -\lim_{T\to\infty}\frac{1}{T}\mathbb{E}[g(\Delta\varphi_T^1)\theta_T] + \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[-\int_0^T \frac{1}{2}\gamma^1\left(\theta_t\sigma\right)^2 dt\right] \leq 0.\end{aligned}$$

An analogous argument shows that $\dot{\varphi}^2$ is long-run optimal for agent 2. This completes the proof. □

2.6.2 Proofs for Section 2.4

The following lemma provides the counterpart of the function g from Lemma 2.5 for proportional costs. It is given in closed form; its properties listed here are therefore easily verified by direct calculations:

Lemma 2.12. *With the constant l from (2.18), define*

$$g(x) = \frac{\gamma\sigma^2}{3\delta^2}\left(x^3 - 3l^2 x\right)\mathbb{1}_{[|x|\leq l]} - \lambda \operatorname{sgn}(x)\mathbb{1}_{[|x|>l]}. \tag{2.27}$$

This function has the following properties:

(i) g is an odd, decreasing function;

(ii) $\frac{1}{2}\delta^2 g''(x) = \gamma\sigma^2 x$ for $x \in (-l, l)$;

(iii) g' is continuous on $\mathbb{R}$ and $g'(l) = g'(-l) = 0$;

(iv) For every $x \in [0, l]$, we have $0 \geq g(x) \geq g(l) = -\lambda$.

Lemma 2.13. *The strategies from Theorem 2.11 are admissible and satisfy the transversality condition (2.15). Moreover they clear the market.*

Proof. Let $x = |\varphi^1_{0-}| + |\varphi^2_{0-}| + l + s$. First, note that the initial jump satisfies

$$-l \leq \Delta\varphi^1_0 = L_0 - U_0 + \Delta\varphi^1_{0-} \leq l,$$

and hence

$$\Delta\varphi^1_t = \delta W_t + L_t - U_t + \Delta\varphi^1_{0-}.$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}[|L_T - U_T|] &= \mathbb{E}[|\Delta\varphi^1_T - \delta W_T - \Delta\varphi^1_{0-}|] \\
&\leq \delta\mathbb{E}[|W_T|] + \mathbb{E}[|\Delta\varphi^1_T|] + |\Delta\varphi^1_{0-}| \\
&\leq x + \delta\left(\frac{2T}{\pi}\right)^{1/2},
\end{aligned}$$

so that the transversality condition (2.15) is satisfied.

Notice that

$$\begin{aligned}
|L_t - U_t|^2 &\leq \left(|\Delta\varphi^1_t| + |\Delta\varphi^1_{0-}| + \delta|W_t|\right)^2 \\
&\leq (x + \delta|W_t|)^2 \leq 2x^2 + 2\delta^2|W_t|^2.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
\mathbb{E}\left[\int_0^T (L_t - U_t)^2 dt\right] &\leq \mathbb{E}\left[\int_0^T 2x^2 + 2\delta^2|W_t|^2 dt\right] \\
&= 2x^2 T + 2\delta^2\mathbb{E}\left[\int_0^T |W_t|^2 dt\right] \\
&= 2x^2 T + \delta^2 T^2,
\end{aligned}$$

so that φ^1 satisfies the first integrability condition in (2.14).

Now, apply Itô's formula to $(\Delta\varphi^1_T + l)^2/4l$, obtaining

$$
\begin{aligned}
&\frac{1}{4l}(\Delta\varphi^1_T + l)^2 - \frac{1}{4l}(\Delta\varphi^1_0 + l)^2 \\
&= \int_0^T \frac{\delta}{2l}(\Delta\varphi^1_t + l)dW_t + \int_0^T \frac{\delta^2}{4l}dt + \int_0^T \frac{1}{2l}(-l + l)dL_t - \int_0^T \frac{1}{2l}(l + l)dU_t \\
&= \int_0^T \frac{\delta}{2l}(\Delta\varphi^1_t + l)dW_t + \frac{\delta^2}{4l}T - U_T + U_0.
\end{aligned}
$$

Rearranging, taking expectations, and $0 \leq U_0 \leq |\Delta\varphi^1_0| \leq x$ leads to

$$
\begin{aligned}
&\mathbb{E}[U_T] \\
&= U_0 + \frac{1}{4l}(\Delta\varphi^1_0 + l)^2 + \frac{\delta^2}{4l}T - \mathbb{E}\left[\int_0^T \frac{\delta}{2l}(\Delta\varphi^1_t + l)dW_t\right] - \mathbb{E}\left[\frac{1}{4l}(\Delta\varphi^1_T + l)^2\right] \\
&\leq x + l + \frac{\delta^2}{4l}T.
\end{aligned}
\tag{2.28}
$$

After applying Itô's formula to $(\Delta\varphi^1_T - l)^2/4l$, a symmetric calculation and $0 \leq L_0 \leq |\Delta\varphi^1_0| \leq x$ show

$$
\begin{aligned}
&\mathbb{E}[L_T] \\
&= L_0 + \frac{1}{4l}(\Delta\varphi^1_0 - l)^2 + \frac{\delta^2}{4l}T - \mathbb{E}\left[\int_0^T \frac{\delta}{2l}(\Delta\varphi^1_t - l)dW_t\right] - \mathbb{E}\left[\frac{1}{4l}(\Delta\varphi^1_T - l)^2\right] \\
&\leq x + l + \frac{\delta^2}{4l}T.
\end{aligned}
\tag{2.29}
$$

Combining (2.28) and (2.29) yields the second integrability condition in (2.14); therefore φ^1 is indeed admissible. Market clearing evidently holds by construction; in particular φ^2 is admissible as well. For later use also observe that, by definition,

$$
\varphi^1_t = \Delta\varphi^1_t - \delta W_t + \frac{s\gamma^2}{\gamma^1 + \gamma^2}, \qquad \gamma^1\sigma(\sigma\varphi^1_t + \xi^1_t) - \mu_t = \widetilde{\gamma}\sigma^2\Delta\varphi^1_t. \tag{2.30}
$$

□

Proof of Theorem 2.11. Consider a competing admissible strategy with Jordan-Hahn decomposition $\psi = \varphi^1_{0-} + \psi^\uparrow - \psi^\downarrow$. To ease notation, set

$$
\theta_t = \psi_t - \varphi^1_t, \qquad \text{so that} \quad d\theta_t = d\psi^\uparrow_t - d\psi^\downarrow_t - dL_t + dU_t, \qquad \theta_{0-} = 0.
$$

By properties (i) and (iv) of g from Lemma 2.12, and the fact that L, U only grow on the sets $\{\Delta\varphi^1_t = -l\}$ and $\{\Delta\varphi^1_t = l\}$, respectively, we have

$$\begin{aligned}
\mathbb{1}_{(-l,0)}(\Delta\varphi^1_t)g(\Delta\varphi^1_t)d\theta_t &\le \lambda \mathbb{1}_{(-l,0)}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t + d\psi^\downarrow_t\right] \\
&\le \lambda \mathbb{1}_{(-l,0)}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t + d\psi^\downarrow_t - dL_t - dU_t\right], \qquad (2.31) \\
\mathbb{1}_{(0,l)}(\Delta\varphi^1_t)g(\Delta\varphi^1_t)d\theta_t &\le \lambda \mathbb{1}_{(0,l)}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t + d\psi^\downarrow_t\right] \\
&\le \lambda \mathbb{1}_{(0,l)}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t + d\psi^\downarrow_t - dL_t - dU_t\right]. \qquad (2.32)
\end{aligned}$$

Properties (i) and (iv) of g from Lemma 2.12 and (2.31-2.32) show that

$$\begin{aligned}
&\int_0^T g(\Delta\varphi^1_t)d\theta_t \\
&= \lambda \int_0^T \mathbb{1}_{\{-l\}}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t - d\psi^\downarrow_t - dL_t\right] - \mathbb{1}_{\{l\}}(\Delta\varphi^1_t)\left[d\psi^\uparrow_t - d\psi^\downarrow_t + dU_t\right] \\
&\quad + \lambda \int_0^T \mathbb{1}_{(-l,l)}(\Delta\varphi^1_t)g(\Delta\varphi^1_t)d\theta_t \\
&\le \lambda \int_0^T \left(\mathbb{1}_{\{-l\}}(\Delta\varphi^1_t) + \mathbb{1}_{(-l,l)}(\Delta\varphi^1_t) + \mathbb{1}_{\{l\}}(\Delta\varphi^1_t)\right)\left[d\psi^\uparrow_t + d\psi^\downarrow_t - dL_t - dU_t\right] \\
&= \lambda\left[\psi^\uparrow_T + \psi^\downarrow_T - L_T - U_T\right] - \lambda\left[\psi^\uparrow_{0-} + \psi^\downarrow_{0-} - L_{0-} - U_{0-}\right] \\
&= \lambda\left[\psi^\uparrow_T + \psi^\downarrow_T - L_T - U_T\right].
\end{aligned}$$

Together with (2.30), it follows that

$$\begin{aligned}
&J^1_T(\psi) - J^1_T(\varphi^1) \\
&= \mathbb{E}\left[\int_0^T \left((\psi_t - \varphi^1_t)\mu_t - \frac{\gamma^1}{2}\sigma(\psi_t - \varphi^1_t)(\sigma\psi_t + \sigma\varphi^1_t + 2\xi^1_t)\right)dt\right] \\
&\quad - \lambda\mathbb{E}\left[\psi^\uparrow_T + \psi^\downarrow_T - L_T - U_T\right] \\
&= \mathbb{E}\left[\int_0^T \left((\psi_t - \varphi^1_t)\mu_t - \frac{\gamma^1}{2}\sigma(\psi_t - \varphi^1_t)\left(\sigma\psi_t - \sigma\varphi^1_t + 2(\sigma\varphi^1_t + \xi^1_t)\right)\right)dt\right] \\
&\quad - \lambda\mathbb{E}\left[\psi^\uparrow_T + \psi^\downarrow_T - L_T - U_T\right] \\
&= \mathbb{E}\left[\int_0^T -\left(\frac{1}{2}\gamma^1\sigma^2(\psi_t - \varphi^1_t)^2 + \gamma\sigma^2\Delta\varphi^1_t(\psi_t - \varphi^1_t)\right)dt\right] \\
&\quad - \lambda\mathbb{E}\left[\psi^\uparrow_T + \psi^\downarrow_T - L_T - U_T\right] \\
&\le -\mathbb{E}\left[\int_0^T \frac{1}{2}\gamma^1\sigma^2\theta_t^2 dt\right] - \mathbb{E}\left[\int_0^T \gamma\sigma^2\Delta\varphi^1_t\theta_t dt + \int_0^T g(\Delta\varphi^1_t)d\theta_t\right]. \qquad (2.33)
\end{aligned}$$

To simplify this expression, use Itô's formula, the dynamics (2.17) of the doubly-reflected Brownian motion $\Delta\varphi^1$, the fact that L, U only grow on the sets $\{\Delta\varphi_t^1 = -l\}$ and $\{\Delta\varphi_t^1 = l\}$ respectively, and the ODE for g from Lemma 2.12(ii) to compute

$$\begin{aligned} dg(\Delta\varphi_t^1) &= \frac{1}{2}\delta^2 g''(\Delta\varphi_t^1)dt + g'(\Delta\varphi_t^1)\left[dL_t - dU_t\right] + \delta g'(\Delta\varphi_t^1)dW_t \\ &= \widetilde{\gamma}\sigma^2\Delta\varphi_t^1 dt + \delta g'(\Delta\varphi_t^1)dW_t. \end{aligned}$$

Integration by parts in turn yields

$$d\left[g(\Delta\varphi_t^1)\theta_t\right] = g(\Delta\varphi_t^1)d\theta_t + \widetilde{\gamma}\sigma^2\theta_t\Delta\varphi_t^1 dt + \delta\theta_t g'(\Delta\varphi_t^1)dW_t.$$

Since g' is bounded, the integrability condition (2.14) implies that the local martingale part in this decomposition is a true martingale, so that

$$\begin{aligned} \mathbb{E}\left[\int_0^T \widetilde{\gamma}\sigma^2\Delta\varphi_t^1\theta_t dt + \int_0^T g(\Delta\varphi_t^1)d\theta_t\right] &= \mathbb{E}\left[g(\Delta\varphi_T^1)\theta_T\right] - \mathbb{E}\left[g(\Delta\varphi_{0-}^1)\theta_{0-}\right] \\ &= \mathbb{E}\left[g(\Delta\varphi_T^1)\theta_T\right]. \end{aligned} \tag{2.34}$$

Now, the long-run optimality of φ^1 for agent 1 follows from (2.33) and (2.34) by taking into account that property (iv) of g and the transversality condition (2.15) imply

$$\begin{aligned} \lim_{T\to\infty}\frac{1}{T}\left|\mathbb{E}\left[g(\Delta\varphi_T^1)\theta_T\right]\right| &\leq \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[|g(\Delta\varphi_T^1)\theta_T|\right] \\ &\leq \lim_{T\to\infty}\frac{\lambda}{T}\mathbb{E}\left[|\theta_T|\right] \\ &\leq \lim_{T\to\infty}\frac{\lambda}{T}\mathbb{E}\left[|\psi_T| + |\varphi_T^1|\right] = 0. \end{aligned}$$

An analogous argument shows that φ^2 is optimal for agent 2, thereby completing the proof. □

2.6.3 Proof of Lemma 2.5

In this subsection, we establish existence, uniqueness, and properties for the second-order nonlinear ODE (2.7) from Lemma 2.5. To this end, we introduce the following first-order nonlinear ODE:

$$y'(x) = f(x, y(x)) := -ax^2 + b + F(y(x)), \tag{2.35}$$

and extend the ideas of Guasoni and Weber (2018) to general functions $F : \mathbb{R} \to \mathbb{R}$ which satisfy Assumption 2.14 below. That is, in Lemma 2.17, we establish that for suitable functions F, and any choice of $a > 0$ and $b \in \mathbb{R}$, (2.35) has a unique positive solution on its maximal domain, which contains $[\sqrt{\max\{b,0\}/a}, \infty)$. Then, for the first-order ODE:

$$g'(x) = ax^2 - b - F(g(x)), \tag{2.36}$$

Lemma 2.18 shows that there is a unique value of b that guarantees there is a solution on $\mathbb{R}$ such that $xg(x) \leq 0$, and the solution is unique. Moreover, Lemma 2.19 proves that this solution to (2.36) is also the unique solution of the second-order ODE:

$$g''(x) = 2ax - F'(g(x))g'(x). \tag{2.37}$$

Finally, with the help of Lemma 2.20 pointing out the relationship between Assumption 2.2 and Assumption 2.14, we establish the proof of Lemma 2.5 with F chosen to be proportional to the Legendre transform of the trading cost function G.

To carry out this program, we first introduce the assumptions on F that are needed to generalize the argument developed for power functions by Guasoni and Weber (2018). Subsequently, in Remark 2.15 and Lemma 2.16, we derive a number of consequences, which are crucial tools for the analysis.

Assumption 2.14. (i) F is convex, differentiable, even, and strictly increasing on $[0, \infty)$ with $F(0) = 0$;

(ii) F' is also differentiable and strictly increasing on $[0, \infty)$ with $F'(0) = 0$;

(iii) There exists a constant K such that $F(x) \leq K(1 + |x|^p)$ for some $p \geq 2$;

(iv) There exist constants $\tilde{C} > 0$ and $x_0 > 0$ such that $F''(x) > \tilde{C}$ for every $|x| > x_0$.

Remark 2.15. Some immediate consequences of Assumption 2.14 are as follows:

(i) F' is increasing on the whole real line, since it is an odd function (as F is even) and F' is strictly increasing on $[0, \infty)$;

(ii) Assumption (iv) implies that there is some $\hat{a} > 0$ such that $F(x) > \hat{a}x^2$ for large $x > 0$. This is why $p \geq 2$ in Assumption 2.14(iii) is without loss of generality.

Lemma 2.16. *Suppose F satisfies Assumption 2.14. Then:*

(i) F^{-1} exists and is concave on $[0, \infty)$;

(ii) For every $x \geq 0$ and every $\alpha \geq 1$:

$$\alpha F(x) \leq F(\alpha x), \qquad F^{-1}(\alpha x) \leq \alpha F^{-1}(x);$$

(iii) For $x, y \geq 0$:

$$F(x + y) \geq F(x) + F(y), \qquad F^{-1}(x) + F^{-1}(y) \geq F^{-1}(x + y);$$

(iv) On $(0, \infty)$, F^{-1} is strictly increasing but $(F^{-1})'$ is strictly decreasing;

(v) *There exists constant* $C > 0$ *that* $F^{-1}(x^2) \leq C|x|$ *and* $2x(F^{-1})'(x^2) \leq 2C$ *for every* $|x| > x_0$.

Proof. (i): Convexity of F implies that, for $x, y \geq 0$ and $0 < a < 1$,

$$ax + (1-a)y = aF(F^{-1}(x)) + (1-a)F(F^{-1}(y)) \geq F(aF^{-1}(x) + (1-a)F^{-1}(y)).$$

As F is increasing, F^{-1} is increasing as well. Applying F^{-1} on both sides of the above estimate in turn yields the concavity of F^{-1}.

(ii): Recall that $F(0) = 0$ and again use convexity of F to obtain, for every $x \geq 0$ and $\alpha \geq 1$,

$$F(\alpha x) = \alpha \left[\frac{1}{\alpha} F(\alpha x) + \left(1 - \frac{1}{\alpha}\right) F(0) \right] \geq \alpha F\left(\frac{1}{\alpha} \alpha x\right) = \alpha F(x).$$

Analogously, the concavity of F^{-1} yields $F^{-1}(\alpha x) \leq \alpha F^{-1}(x)$.

(iii): Since F' is increasing we have $F'(x+y) - F'(x) \geq 0$ for every $x, y > 0$. As a consequence, $F(x+y) - F(x) \geq F(0+y) - F(0) = F(y)$ as asserted. It implies that

$$F(F^{-1}(x) + F^{-1}(y)) \geq F(F^{-1}(x)) + F(F^{-1}(y)) = x + y.$$

By the strictly increasing property of F, we can infer

$$F^{-1}(x) + F^{-1}(y) \geq F^{-1}(x+y).$$

(iv): Since F is convex and F and F' are strictly increasing on $[0, \infty)$, then $F' \geq 0$, $F'' \geq 0$ and they are both not equal to zero on any interval, hence

$$(F^{-1})'(x) = \frac{1}{F'(F^{-1}(x))} \geq 0, \qquad (F^{-1})''(x) = -\frac{F''(F^{-1}(x))}{\left(F'(F^{-1}(x))\right)^3} \leq 0,$$

and they are both not zero on any interval. So F^{-1} is strictly increasing on $[0, \infty)$ but $(F^{-1})'$ is strictly decreasing on $(0, \infty)$ as asserted.

(v): By directly integrating the inequality in Assumption 2.14 (iv) and choosing C large enough, together with (ii) in 2.16, it's easy to see that the first statement holds. For the second statement, by Assumption 2.14(ii),

$$\frac{d}{dx}\left[xF'(x) - F(x)\right] = xF''(x) + F'(x) - F'(x) = xF''(x) \geq 0.$$

By strictly increasing property of F^{-1}, for $x > 0$, $F^{-1}(x^2) > 0$, hence

$$F'(F^{-1}(x^2)) \geq \frac{F(F^{-1}(x^2))}{F^{-1}(x^2)} = \frac{x^2}{F^{-1}(x^2)}.$$

Together Assumption 2.14 (iv) it follows that, for $x \geq |x_0|$,

$$\frac{d}{dx}F^{-1}(x^2) = 2x(F^{-1})'(x^2) = \frac{2x}{F'(F^{-1}(x^2))} \leq \frac{2xF^{-1}(x^2)}{x^2} \leq \frac{2Cx^2}{x^2} = 2C,$$

which yields the desired result. $\square$

Now we address the existence and uniqueness of the positive solution to (2.35) on $[\sqrt{\max\{b,0\}/a}, \infty)$.

Lemma 2.17. *Let F be a function satisfying Assumption 2.14 and $a > 0$, $b \in \mathbb{R}$. Then there exists a unique solution y of*

$$y'(x) = f(x, y(x)) := -ax^2 + b + F(y(x)), \tag{2.35}$$

such that $[\sqrt{\max\{b,0\}/a}, \infty)$ is contained in its maximal interval of existence, and $y(x) \geq 0$ for every $x \geq \sqrt{\max\{b,0\}/a}$. Moreover y is increasing on $[\sqrt{\max\{b,0\}/a}, \infty)$, and satisfies the growth condition

$$\lim_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} = 1. \tag{2.38}$$

Further, in (2.35), if we rewrite $y(x) = y(x;b)$, then for every $x \in [0, \infty)$, $b \in \mathbb{R}$,

$$\frac{\partial y(x;b)}{\partial b} < 0. \tag{2.39}$$

Proof. Let $b_+ = \max\{b, 0\}$. On $[\sqrt{b_+/a}, +\infty)$, we define the function $h(x) = F^{-1}(ax^2 - b)$. Notice that by definition of $h(x)$ we have $f(x, h(x)) = 0$ and h is strictly increasing on $[\sqrt{b_+/a}, +\infty)$. Thus we can infer that h is a supersolution on $(\sqrt{b_+/a}, \infty)$ in that $h'(x) \geq f(x, h(x)) = 0$.

Notice that $f(x, y)$ is locally Lipschitz, so that local existence and uniqueness hold for the initial-value problem (2.35) with initial condition (x_0, y_0). For every $\bar{x} > \sqrt{b_+/a}$, let $y(x; \bar{x}, h(\bar{x}))$ denote the unique solution to (2.35) with initial condition $(\bar{x}, h(\bar{x}))$ on its maximal interval of existence (T^-, T^+). The first step is to show that the following inequalities hold for every $\bar{x} > \sqrt{b_+/a}$:

$$\text{on } [\sqrt{b_+/a}, \bar{x}),\ y(x; \bar{x}, h(\bar{x})) > h(x) \geq 0,\ y(\cdot; \bar{x}, h(\bar{x})) \text{ is increasing}, \tag{2.40}$$
$$\text{on } [\bar{x}, T^+), \qquad y(x; \bar{x}, h(\bar{x})) \leq h(x). \tag{2.41}$$

First, by directly calculating the first-order derivative, we find that for every $\bar{x} > \sqrt{b_+/a}$,

$$y'(\bar{x}; \bar{x}, h(\bar{x})) = f(\bar{x}, y(\bar{x}; \bar{x}, h(\bar{x}))) = f(\bar{x}, h(\bar{x})) = 0 < h'(\bar{x}).$$

Therefore, there exists $\epsilon^- \in (0, \bar{x} - \sqrt{b_+/a} \vee T^-)$ such that $y(x; \bar{x}, h(\bar{x})) > h(x)$ for $x \in (\bar{x} - \epsilon^-, \bar{x})$. Define

$$x_0 = \inf\{x \in [\sqrt{b_+/a}, \bar{x}) \cap (T^-, T^+) : y(\cdot; \bar{x}, h(\bar{x})) > h \text{ on } (x, \bar{x}]\}.$$

It is easy to see that on $(x_0, \bar{x})$, $y(x; \bar{x}, h(\bar{x}))$ is increasing through

$$y'(x; \bar{x}, h(\bar{x})) = f(x, y(x; \bar{x}, h(\bar{x}))) > f(x, h(x)) = 0,$$

and since $y(x_0;\bar{x},h(\bar{x}))$ is between $h(x_0)>0$ and $h(\bar{x})<\infty$, we conclude $x_0\in(T^-,T^+)$. Suppose $x_0>\sqrt{b_+/a}$. The definition of x_0 yields $y(x_0;\bar{x},h(\bar{x}))=h(x_0)$ and $y'(x_0;\bar{x},h(\bar{x}))\geq h'(x_0)>0$; but plugging $y(x_0;\bar{x},h(\bar{x}))=h(x_0)$ into (2.35) gives $y'(x_0;\bar{x},h(\bar{x}))=f(x_0,h(x_0))=0$, a contradiction. Therefore, $y(\cdot;\bar{x},h(\bar{x}))$ is increasing on $(x_0,\bar{x})$ and $x_0=\sqrt{b_+/a}$, hence (2.40) holds.
To show (2.41), we calculate the second-order derivative,

$$y''(x;\bar{x},h(\bar{x}))=-2ax+F'(y(x;\bar{x},h(\bar{x})))y'(x;\bar{x},h(\bar{x})), \tag{2.42}$$

which implies that $y''(\bar{x};\bar{x},h(\bar{x}))<0$ and there exists $\epsilon^+>0$ such that $y(x;\bar{x},h(\bar{x}))<h(x)$ for $x\in(\bar{x},\bar{x}+\epsilon^+)$. Define

$$x_1:=\sup\{x\in[\bar{x},T^+):y(\cdot;\bar{x},h(\bar{x}))<h\text{ on }[\bar{x},x)\}.$$

Suppose $x_1<T^+$, $y(x_1;\bar{x},h(\bar{x}))=h(x_1)$, and $y'(x_1;\bar{x},h(\bar{x}))\geq h'(x_1)>0$; but on the other hand, $y'(x_1;\bar{x},h(\bar{x}))=f(x_1,h(x_1))=0$. We have a contradiction, which implies (2.41) holds.
Define

$$x_2:=\sup\{x\geq\bar{x}:y(\cdot;\bar{x},h(\bar{x}))\geq 0,\ \text{or } y'(\cdot;\bar{x},h(\bar{x}))\leq 0\text{ on }[x,\bar{x}]\}. \tag{2.43}$$

From (2.42), we see that $y''(x;\bar{x},h(\bar{x}))\leq -2ax$ for $x\in[\bar{x},x_2)$, so $y(\cdot;\bar{x},h(\bar{x}))$ is strictly decreasing and strictly concave on $[\bar{x},x_2)$. This implies $x_2<+\infty$, and by continuity we have $y(x_2;\bar{x},h(\bar{x}))=0$, $y'(x_2;\bar{x},h(\bar{x}))<0$, and in addition that $y(\cdot;\bar{x},h(\bar{x}))<0$ in a right-neighbourhood of x_2. For $x>x_2$, we claim that $y(x;\bar{x},h(\bar{x}))\leq 0$, because in order to become positive again, $y(\cdot;\bar{x},h(\bar{x}))$ would need to cross zero, but $y(x;\bar{x},h(\bar{x}))=0$ implies $y'(x;\bar{x},h(\bar{x}))=-ax^2+b<0$. Therefore, we can conclude that either $T^+=\infty$ or $\lim_{x\uparrow T^+}y(x;\bar{x},h(\bar{x}))=-\infty$. For the latter case, define

$$y(x;\bar{x},h(\bar{x}))=-\infty,\qquad\text{for } x\in[T^+,\infty).$$

Now consider the relationship between $y(x;\bar{x}_1,h(\bar{x}_1))$ and $y(x;\bar{x}_2,h(\bar{x}_2))$ for $\bar{x}_2>\bar{x}_1>\sqrt{b_+/a}$. By (2.40), at $\bar{x}_1$, $y(\bar{x}_1;\bar{x}_1,h(\bar{x}_1))=h(\bar{x}_1)<y(\bar{x}_1;\bar{x}_2,h(\bar{x}_2))$. By (local) uniqueness of the initial value problems associated with (2.35), there cannot exist x such that $y(x;\bar{x}_1,h(\bar{x}_1))=y(x;\bar{x}_2,h(\bar{x}_2))>-\infty$, thus the graph of $y(x;\bar{x}_1,h(\bar{x}_1))$ lies strictly below the graph of $y(x;\bar{x}_2,h(\bar{x}_2))$ except when they both take the value $-\infty$. In summary

$$\text{on }(\sqrt{b_+/a},\infty),\quad y(x;\cdot,h(\cdot))\text{ is increasing.} \tag{2.44}$$

Next, we show that any solution y of (2.35) such that $[0,\infty)$ is contained in its maximum interval of existence with $y(x)\geq 0$ for every $x\geq\sqrt{b_+/a}$, automatically satisfies the growth condition (2.38). From the above argument concerning the relationship between $h(x)$ and $y(x;\bar{x},h(\bar{x}))$, an important observation is for every $x>\sqrt{b_+/a}$ and every $\bar{x}>x$, we need to have $y(x)>y(x;\bar{x},h(\bar{x}))\geq h(x)$; otherwise the solution y will not stay positive. We summarize the properties of y as follows:

i) $y(x) > h(x) \geq 0$, $y'(x) = -ax^2 + b + F(y(x)) > -ax^2 + b + F(h(x)) = 0$, which means y is strictly increasing on $(\sqrt{b_+/a}, +\infty)$;

ii) $[\sqrt{b_+/a}, +\infty) \subset D$, where D is the maximal interval of existence of $y(x)$.

From Property i) and Lemma 2.16 (iii,iv), it follows that

$$\begin{aligned}1 &= \lim_{x\to\infty} \frac{F^{-1}(ax^2) - F^{-1}(b_+)}{F^{-1}(ax^2)} \\ &\leq \liminf_{x\to\infty} \frac{h(x)}{F^{-1}(ax^2)} \\ &\leq \limsup_{x\to\infty} \frac{h(x)}{F^{-1}(ax^2)} \\ &\leq \lim_{x\to\infty} \frac{F^{-1}(ax^2)}{F^{-1}(ax^2)} = 1,\end{aligned}$$

and in turn

$$\liminf_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} \geq 1.$$

Next we show that $L = \lim_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)}$ exists and $L = 1$. To this end, set $M = \limsup_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)}$ and notice that $1 \leq M \leq \infty$. If $M = 1$ then we can conclude that $L = 1$.

Assume $1 < M < \infty$. We first want to show $M = L$. There exists a sequence $(x_n)_{n\geq 0} \to \infty$ such that

$$\lim_{n\to\infty} \frac{y(x_n)}{F^{-1}(ax_n^2)} = M.$$

In particular, for any $\delta \in (0, M-1)$ there exists $N_\delta \in \mathbb{N}$ such that for every $n \geq N_\delta$ we have

$$y(x_n) \geq (M-\delta)F^{-1}(ax_n^2).$$

For large x, we claim that the function $s(x) = (M-\delta)F^{-1}(ax^2)$ is a subsolution of (2.35). By Lemma 2.16 (v), we know that for $x \geq |x_0|/\sqrt{a}$,

$$0 < s'(x) = (M-\delta)(F^{-1})'(ax^2)2ax \leq 4\sqrt{a}(M-\delta)C.$$

Since $M - \delta > 1$, there exists $\bar{x}$ such that for $x \geq \bar{x}$, we have $(M-\delta)ax^2 - ax^2 + b \geq 4\sqrt{a}(M-\delta)C$. As a consequence,

$$\begin{aligned}s'(x) &\leq 4\sqrt{a}(M-\delta)C \\ &\leq -ax^2 + b + (M-\delta)ax^2 \\ &= -ax^2 + b + F(F^{-1}((M-\delta)(ax^2))) \\ &\leq -ax^2 + b + F((M-\delta)F^{-1}(ax^2)) \\ &= f(x, s(x)).\end{aligned}$$

On the other hand, notice that $y(x_n) \geq s(x_n)$ for every $n \geq N_\delta$. Thus by the comparison lemma, for every $\delta \in (0, M-1)$ and some large x_N, from $y(x_N) \geq (M-\delta)F^{-1}(ax_N^2) = s(x_N)$ we can conclude $y(x) \geq s(x) = (M-\delta)F^{-1}(ax^2)$ for $x \geq x_N$. In particular, for every small δ,

$$\liminf_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} \geq M - \delta,$$

and therefore

$$\liminf_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} = M = \limsup_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)}.$$

If $M = \infty$, we substitute $M - \delta$ with $N \in \mathbb{N}$ and then infer with the same argument that $\liminf_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} = \infty$. In other words, the limit L exists and $L = M \in [1, \infty]$.

Next, we show $L = 1$. First, assume to the contrary $1 < L < \infty$. Since $\lim_{x\to\infty} \frac{y(x)}{F^{-1}(ax^2)} = L$, by Lemma 2.16 (v), there exists a constant $K > 0$ such that $y(x) \leq Kx$ for large $x > 0$. Moreover, for every $\delta \in (0, L-1)$ and large x, by Lemma 2.16 (ii),

$$y(x) \geq (L-\delta)F^{-1}(ax^2) \geq F^{-1}((L-\delta)ax^2).$$

As a consequence,

$$\liminf_{x\to\infty} \frac{F(y(x))}{ax^2} \geq L - \delta.$$

On the other hand, (2.35) implies

$$\liminf_{x\to\infty} \frac{y'(x)}{ax^2} \geq L - \delta - 1 > 0,$$

so that $y'(x)$ grows at least quadratically, leading to a contradiction.

Now assume that $L = +\infty$. With a similar argument as above, for every $L' > 0$ and x sufficiently large,

$$F(y(x)) \geq F(F^{-1}(L'ax^2)) = L'ax^2,$$

and it follows that

$$\lim_{x\to\infty} \frac{ax^2}{F(y(x))} = 0.$$

From (2.35) it follows that

$$\lim_{x\to\infty} \frac{y'(x)}{F(y(x))} = \lim_{x\to\infty} \frac{-ax^2 + b + F(y(x))}{F(y(x))} = 1.$$

Notice that for large x such that $y(x) > Cx_0$, Lemma 2.16 (v) yields

$$\frac{y(x)^2}{C^2} = F\left(F^{-1}\left(\frac{y(x)^2}{C^2}\right)\right) \leq F\left(C\frac{y(x)}{C}\right) = F(y(x)).$$

Thus, for small δ and sufficiently large $\bar{x}$, for all $x > \bar{x}$ we have

$$\frac{1-\delta}{C^2} y^2(x) \leq (1-\delta)F(y(x)) \leq y'(x).$$

Hence for sufficiently large $\xi > \bar{x}$,

$$\frac{y'(\xi)}{y^2(\xi)} \geq \frac{1-\delta}{C^2} > 0.$$

Integrating this inequality from $\xi = \bar{x}$ to $\xi = x$, we obtain

$$y(x) \geq \frac{1}{\frac{1}{y(\bar{x})} - \frac{1-\delta}{C^2}(x-\bar{x})}.$$

In particular, $y(x)$ has a vertical asymptote, contradicting Property ii). In summary, $L = 1$.

We now establish the uniqueness of $y(x)$. Suppose there exists another solution y_2 of (2.35) such that $[0, \infty)$ is contained in its maximal domain, and $y_2(x) \geq 0$ for every $x \geq \sqrt{b_+/a}$, and there exists $\bar{x}$, $\delta > 0$ such that $y_2(\bar{x}) \geq y(\bar{x}) + \delta$. Then, on $[\bar{x}, \infty)$, the graph of y_2 always lies above y; otherwise it will violate the local uniqueness of the initial value problems associated with (2.35). Moreover, for $x \geq \bar{x}$,

$$y_2'(x) - y'(x) = F(y_2(x)) - F(y(x)) \geq 0,$$

which means $y_2 - y$ is increasing. As a result,

$$\begin{aligned}
y_2'(x) - y'(x) &= F(y_2(x)) - F(y(x)) \\
&\geq F(y_2(x) - y(x)) \\
&\geq F(y_2(\bar{x}) - y(\bar{x})) \\
&\geq F(\delta) > 0,
\end{aligned}$$

which implies that, for every $x > \bar{x}$,

$$y_2(x) - y(x) \geq \delta + (x - \bar{x})F(\delta).$$

But y_2 also satisfies (2.38), and for large x we have

$$F^{-1}(ax^2) \leq \sqrt{a}Cx.$$

Whence,

$$0 = \lim_{x\to\infty} \frac{y_2(x) - y(x)}{F^{-1}(ax^2)} \geq \lim_{x\to\infty} \frac{\delta + (x-\bar{x})F(\delta)}{\sqrt{a}Cx} = \frac{F(\delta)}{\sqrt{a}C} > 0,$$

which leads to contradiction. A symmetric argument yields the same results for the case where there exists $\bar{x}$ and $\delta > 0$ such that $y_2(\bar{x}) \leq y(\bar{x}) - \delta$. This establishes uniqueness.

We now establish the existence of $y(x)$ by constructing it. To this end, fix $x \geq \sqrt{b_+/a}$ and define

$$y_*(x) = \sup\{y(x; \bar{x}, h(\bar{x})) : \bar{x} > x\}.$$

Let $x_0 > 0$ and $C > 0$ be the constant in Lemma 2.16 (v). For every $x_1 \geq \sqrt{b_+/a}$, we can choose a large $y_1 > F^{-1}(ax_1^2 + 2\sqrt{a}C + x_0^2 + |b|)$, and for $x \geq x_1$ define

$$\tilde{y}(x) = F^{-1}(F(y_1) + a(x^2 - x_1^2)).$$

Then by $F(y_1) - ax_1^2 + b > 0$, $\tilde{y}(x) > h(x)$ for every $x \geq x_1$. Moreover, from the fact that $F(y_1) - ax_1^2 + b > 2\sqrt{a}C + x_0^2 + |b| + b > x_0^2$, we can infer

$$\begin{aligned} 0 \leq \tilde{y}'(x) &= 2ax(F^{-1})'(F(y_1) + a(x^2 - x_1^2)) \\ &\leq 2\sqrt{a}\sqrt{F(y_1) + a(x^2 - x_1^2)}(F^{-1})'(F(y_1) + a(x^2 - x_1^2)) \\ &\leq 2\sqrt{a}C \\ &< F(y_1) - ax_1^2 + b = f(x_1, \tilde{y}(x_1)). \end{aligned}$$

In particular, the unique local solution $y(x; x_1, y_1)$ to (2.35) with initial condition (x_1, y_1) satisfies

$$\tilde{y}'(x_1) < f(x_1, \tilde{y}(x_1)) = f(x_1, y_1) = y'(x_1; x_1, y_1).$$

Thus for every $\bar{x} > x_1$, $y(\bar{x}; x_1, y_1) > \tilde{y}(\bar{x}) > h(\bar{x}) = y(\bar{x}; \bar{x}, h(\bar{x}))$. The local uniqueness of the initial value problems associated with (2.35) implies that $y(\cdot; x_1, y_1)$ and $y(\cdot; \bar{x}, h(\bar{x}))$ cannot cross, so the graph of $y(\cdot; x_1, y_1)$ lies above $y(\cdot; \bar{x}, h(\bar{x}))$ and, in particular, $y_1 = y(x_1; x_1, y_1) > y(x_1; \bar{x}, h(\bar{x}))$. Taking the supremum over $\bar{x}$ yields that $y_*(x_1) \leq y_1 < +\infty$. In summary y_* is defined pointwise on $[\sqrt{b_+/a}, \infty)$, and $y_* < \infty$ is guaranteed.

Next we study the continuity and differentiability of y_*. Notice that by (2.40), we know that $y_*(x) \geq h(x) \geq 0$ and is increasing on $[\sqrt{b/a}, \infty)$. Therefore, for every $x \in [\sqrt{b_+/a}, +\infty)$, we have $y_*(x+) = \lim_{\epsilon \to 0^+} y_*(x + \epsilon)$ exists; and for every $x \in (\sqrt{b_+/a}, +\infty)$, $y_*(x-) = \lim_{\epsilon \to 0^+} y_*(x - \epsilon)$ exists as well. In particular, $f(x, y_*(x))$ is locally integrable on $[\sqrt{b_+/a}, \infty)$.

For $x_2 > x_1 \geq \sqrt{b_+/a}$, we want to estimate $y_*(x_2) - y_*(x_1)$. By the monotonicity of $y(x; \bar{x}, h(\bar{x}))$ in $\bar{x}$ established in (2.44), we know that

$$y_*(x) = \sup\{y(x; \bar{x}, h(\bar{x})) : \bar{x} > \sqrt{b_+/a}\}.$$

Together with $y_* \geq 0$ and since F is increasing on $[0, \infty)$, it follows that

$$\begin{aligned}
y_*(x_2) - y_*(x_1) &= \sup\{y(x_2; \bar{x}, h(\bar{x})) - y_*(x_1) : \bar{x} > \sqrt{b_+/a}\} \\
&\leq \sup\{y(x_2; \bar{x}, h(\bar{x})) - y(x_1; \bar{x}, h(\bar{x})) : \bar{x} > \sqrt{b_+/a}\} \\
&= \sup\left\{\int_{x_1}^{x_2} y'(\xi; \bar{x}, h(\bar{x}))\, d\xi : \bar{x} > \sqrt{b_+/a}\right\} \\
&\leq \int_{x_1}^{x_2} \sup\{y'(\xi; \bar{x}, h(\bar{x})) : \bar{x} > \sqrt{b/a}\}\, d\xi \\
&= \int_{x_1}^{x_2} -a\xi^2 + b + \sup\{F(y(\xi; \bar{x}, h(\bar{x}))) : \bar{x} > \sqrt{b_+/a}\}\, d\xi \\
&= \int_{x_1}^{x_2} -a\xi^2 + b + F(y_*(\xi))\, d\xi \\
&= \int_{x_1}^{x_2} f(\xi, y_*(\xi)) d\xi. \qquad (2.45)
\end{aligned}$$

For every $\delta > 0$, there exists $\bar{x}$ such that $y(x_1; \bar{x}, h(\bar{x})) + \delta > y_*(x_2)$. By (2.44), without loss of generality, we can assume that $\bar{x} > x_2$, and therefore $y(\xi; \bar{x}, h(\bar{x}))$ is increasing in ξ on the interval $[x_1, x_2]$ by (2.40). Thus for every $\delta > 0$ and for every $\xi \in [x_1, x_2]$, the monotonicity of F on $[0, \infty)$ yields $F(y(\xi; \bar{x}, h(\bar{x}))) \geq F(y(x_1; \bar{x}, h(\bar{x}))) \geq F(y_*(x_1) - \delta)$. Therefore,

$$\begin{aligned}
y_*(x_2) - y_*(x_1) &\geq y_*(x_2) - (y(x_1; \bar{x}, h(\bar{x})) + \delta) \\
&\geq y(x_2; \bar{x}, h(\bar{x})) - y(x_1; \bar{x}, h(\bar{x})) - \delta \\
&= \int_{x_1}^{x_2} y'(\xi; \bar{x}, h(\bar{x}))\, d\xi - \delta \\
&= \int_{x_1}^{x_2} -a\xi^2 + b + F(y(\xi; \bar{x}, h(\bar{x}))) d\xi - \delta \\
&\geq (x_2 - x_1) F(y(x_1; \bar{x}, h(\bar{x}))) - \delta + \int_{x_1}^{x_2} (-a\xi^2 + b)\, d\xi \\
&\geq (x_2 - x_1) F(y_*(x_1) - \delta) - \delta + \int_{x_1}^{x_2} (-a\xi^2 + b)\, d\xi.
\end{aligned}$$

As this holds for arbitrary small $\delta > 0$, it follows from the continuity of F that

$$y_*(x_2) - y_*(x_1) \geq (x_2 - x_1) F(y_*(x_1)) + \int_{x_1}^{x_2} (-a\xi^2 + b) d\xi. \qquad (2.46)$$

By (2.45) and (2.46), we can conclude the continuity of y_* on $[\sqrt{b_+/a}, +\infty)$. We then establish the differentiability of y_*. First for $x \in [\sqrt{b_+/a}, +\infty)$, by (2.45) and the continuity of y_*,

$$\limsup_{\epsilon \to 0^+} \frac{y_*(x + \epsilon) - y_*(x)}{\epsilon} \leq \limsup_{\epsilon \to 0^+} \frac{1}{\epsilon} \int_x^{x+\epsilon} f(\xi, y_*(\xi)) d\xi = f(x, y_*(x)),$$

and by (2.46).

$$\begin{aligned}\liminf_{\epsilon\to 0^+}\frac{y_*(x+\epsilon)-y_*(x)}{\epsilon} &\geq \liminf_{\epsilon\to 0^+}\frac{1}{\epsilon}\left(\epsilon F(y_*(x))+\int_x^{x+\epsilon}(-a\xi^2+b)d\xi\right)\\ &= -ax^2+b+F(y_*(x))\\ &= f(x,y_*(x)).\end{aligned}$$

In addition, for $x\in(\sqrt{b_+/a},+\infty)$, by (2.45), (2.46) and the continuity of y_*,

$$\begin{aligned}f(x,y_*(x)) &= \limsup_{\epsilon\to 0^+}\frac{1}{\epsilon}\int_{x-\epsilon}^{x} f(\xi,y_*(\xi))d\xi\\ &\geq \limsup_{\epsilon\to 0^+}\frac{y_*(x)-y_*(x-\epsilon)}{\epsilon}\\ &\geq \liminf_{\epsilon\to 0^+}\frac{y_*(x)-y_*(x-\epsilon)}{\epsilon}\\ &\geq \liminf_{\epsilon\to 0^+}\frac{1}{\epsilon}\left(\epsilon F(y_*(x-\epsilon))+\int_{x-\epsilon}^{x}(-a\xi^2+b)d\xi\right)\\ &= -ax^2+b+F(y_*(x))\\ &= f(x,y_*(x)).\end{aligned}$$

Hence we can conclude that y_*' exists and

$$y_*'(x)=f(x,y_*(x)),\ \text{for all } x\in[\sqrt{b_+/a},+\infty).$$

In summary, for $b\in\mathbb{R}$, the function y_* therefore is a solution of (2.35) that satisfies properties i), ii) and hence satisfies also the growth condition (2.38).

We only need to show (2.39) holds. When $b>0$, where $b_+=b$, $[0,\sqrt{b/a})$ is contained in the maximal interval of existence of y_* is a side product in the proof of (2.39).

We rewrite $h(x;b)=F^{-1}(ax^2-b)$, and notice that h' is strictly decreasing by Lemma 2.16 (iv). Let $y(x;\bar{x},h(\bar{x};b))$ denote the solution to (2.35) with constant b and initial condition $(\bar{x},h(\bar{x};b))$ with $\bar{x}>\sqrt{b/a}$. Then by Proposition 2.76 in Chicone (1999), we know that

$$\frac{\partial}{\partial b}y(x;\bar{x},h(\bar{x};b))=\Phi(x;\bar{x},b),\tag{2.47}$$

where $\Phi(x;\bar{x},b)$ is the solution to the following initial value problem on its maximal interval of existence,

$$\begin{cases}\Phi'(x;\bar{x},b)=F'\left(y(x;\bar{x},h(\bar{x};b))\right)\Phi(x;\bar{x},b)+1,\\ \Phi(\bar{x};\bar{x},b)=\dfrac{\partial}{\partial b}h(\bar{x};b)=-\left(F^{-1}\right)'(a\bar{x}^2-b).\end{cases}\tag{2.48}$$

To wit, for $\bar{x}>x\geq\sqrt{b_+/a}$, $y(x;\bar{x},h(\bar{x};b))\geq h(x;b)\geq 0$,

$$\begin{aligned}\Phi(x;\bar{x},b)=&-\left(F^{-1}\right)'(a\bar{x}^2-b)e^{-\int_x^{\bar{x}}F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\ &-\int_x^{\bar{x}}e^{-\int_x^{u}F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}du\leq 0.\end{aligned}\tag{2.49}$$

Notice that $F(F^{-1}(x)) = x$, then differentiate on both sides yields

$$F'(F^{-1}(x))\left(F^{-1}\right)'(x) = 1,$$

thus

$$F'(F^{-1}(a\bar{x}^2 - b))\left(F^{-1}\right)'(a\bar{x}^2 - b) = 1,$$

In order to see whether $\Phi(x;\bar{x},b)$ is increasing in $\bar{x}$ or not, we calculate its partial derivatives with respect to $\bar{x}$, and using the property (2.44), $F' \geq 0$, $(F^{-1})' \geq 0$, $(F^{-1})'' \leq 0$ to estimate the lower bound:

$$\begin{aligned}
&\frac{\partial}{\partial \bar{x}}\Phi(x;\bar{x},b)\\
&= \left(F^{-1}\right)'(a\bar{x}^2 - b)e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\int_x^{\bar{x}} F'\left(y(\xi;\bar{x},h(\bar{x};b))\right)\frac{\partial}{\partial \bar{x}}y(\xi;\bar{x},h(\bar{x};b))d\xi\\
&\quad + \left(F^{-1}\right)'(a\bar{x}^2 - b)F'\left(y(\bar{x};\bar{x},h(\bar{x};b))\right)e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\
&\quad - 2a\bar{x}\left(F^{-1}\right)''(a\bar{x}^2 - b)e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\
&\quad - e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\
&\quad + \int_x^{\bar{x}}\left(\int_x^u F''\left(y(\xi;\bar{x},h(\bar{x};b))\right)\frac{\partial}{\partial \bar{x}}y(\xi;\bar{x},h(\bar{x};b))\,d\xi\right)e^{-\int_x^u F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}du\\
&\geq \left(F^{-1}\right)'(a\bar{x}^2 - b)F'\left(h(\bar{x};b)\right)e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi} - e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\
&= \left(F^{-1}\right)'(a\bar{x}^2 - b)F'\left(F^{-1}(a\bar{x}^2 - b)\right)e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi} - e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi}\\
&= e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi} - e^{-\int_x^{\bar{x}} F'(y(\xi;\bar{x},h(\bar{x};b)))\,d\xi} = 0,
\end{aligned}$$

hence, we conclude that $\Phi(x;\bar{x},b)$ is increasing in $\bar{x}$ on $(x,+\infty)$.
We rewrite $y(x;b)$ as the unique solution to (2.35) with constant b such that $[\sqrt{b_+/a},+\infty)$ is contained in its maximal interval of existence D_b, such that $y(x;b) \geq 0$ for every $x \geq \sqrt{b_+/a}$. For $x \geq \sqrt{b_+/a}$, we claim $\frac{\partial}{\partial b}y(x;b)$ exists and

$$\frac{\partial}{\partial b}y(x;b) = \sup\left\{\Phi(x;\bar{x},b) : \bar{x} > x\right\} \leq 0. \tag{2.50}$$

For $b_2 > b_1 \geq 0$, fix $x \geq \sqrt{b_{2+}/a}$, we first see that

$$\begin{aligned}
y(x;b_2) - y(x;b_1) &= \sup\left\{y(x;\bar{x},h(\bar{x};b_2)) - y(x;b_1) : \bar{x} > x\right\}\\
&\leq \sup\left\{y(x;\bar{x},h(\bar{x};b_2)) - y(x;\bar{x},h(\bar{x};b_1)) : \bar{x} > x\right\}\\
&= \sup\left\{\int_{b_1}^{b_2}\frac{\partial}{\partial b}y(x;\bar{x},h(\bar{x};b))\,db : \bar{x} > x\right\}\\
&\leq \int_{b_1}^{b_2}\sup\left\{\frac{\partial}{\partial b}y(x;\bar{x},h(\bar{x};b)) : \bar{x} > x\right\}\,db\\
&= \int_{b_1}^{b_2}\sup\left\{\Phi(x;\bar{x},b) : \bar{x} > x\right\}\,db.
\end{aligned}\tag{2.51}$$

On the other hand, for every $\delta > 0$, there exists $\bar{x}_\delta$ such that

$$y(x; \bar{x}_\delta, h(\bar{x}_\delta; b_1)) + \delta > y(x; b_1).$$

By (2.44), we can assume without loss of generality that $\bar{x}_\delta > x$ and for every $\bar{x} > \bar{x}_\delta$

$$y(x; \bar{x}, h(\bar{x}; b_1)) + \delta > y(x; b_1).$$

Therefore, for every $\bar{x} > \bar{x}_\delta$,

$$\begin{aligned} y(x; b_2) - y(x; b_1) &\geq y(x; b_2) - y(x; \bar{x}, h(\bar{x}; b_1)) - \delta \\ &\geq y(x; \bar{x}, h(\bar{x}; b_2)) - y(x; \bar{x}, h(\bar{x}; b_1)) - \delta \\ &= \int_{b_1}^{b_2} \frac{\partial}{\partial b} y(x; \bar{x}, h(\bar{x}; b))\, db - \delta \\ &= \int_{b_1}^{b_2} \Phi(x; \bar{x}, b)\, db - \delta, \end{aligned}$$

By the fact that $\Phi(x; \bar{x}, b)$ is increasing in $\bar{x}$ on $(x, +\infty)$, we know that

$$\begin{aligned} y(x; b_2) - y(x; b_1) &\geq \sup \left\{ \int_{b_1}^{b_2} \Phi(x; \bar{x}, b)\, db : \bar{x} > \bar{x}_\delta \right\} - \delta \\ &= \int_{b_1}^{b_2} \sup \left\{ \Phi(x; \bar{x}, b) : \bar{x} > \bar{x}_\delta \right\}\, db - \delta \\ &= \int_{b_1}^{b_2} \sup \left\{ \Phi(x; \bar{x}, b) : \bar{x} > x \right\}\, db - \delta, \end{aligned}$$

since the above inequality holds for every $\delta > 0$, we conclude that

$$y(x; b_2) - y(x; b_1) \geq \int_{b_1}^{b_2} \sup \left\{ \Phi(x; \bar{x}, b) : \bar{x} > x \right\}\, db. \tag{2.52}$$

By (2.51) and (2.52), $y(x; b)$ is continuous and differentiable with respect to b, and we can infer our claim (2.50) holds from (2.49), hence (2.39) holds. By Theorem 2.77 in Chicone (1999), we know $[0, \infty)$ is contained in the maximal interval of existence of $y(\cdot; b)$ for every $b \in \mathbb{R}$. □

In Lemma 2.17, we have shown that for every $b \geq 0$, there exists a non-negative solution y_r to (2.35) on $[0, \infty)$. A symmetric argument yields that for every $b \geq 0$, there exists a non-positive solution y_l to (2.35) on $(-\infty, 0]$. Then, as we shall see in the proof of Lemma 2.18 below, by the monotonicity of $y(0; b)$ with respect to b, there exists a unique choice of the constant b in (2.35) that allows us to smoothly paste together the solution y_l and y_r at 0, thereby obtaining a solution of (2.35) on the whole real line.

Lemma 2.18. *Let F be a function satisfying Assumption 2.14. Then there exists a unique constant $b_F > 0$ such that when $b = b_F$, the ODE*

$$g'(x) = ax^2 - b - F(g(x)), \tag{2.36}$$

has a solution g on $\mathbb{R}$ such that $xg(x) \le 0$. Moreover, g is unique, and it is odd and decreasing and satisfies the following growth conditions:

$$\lim_{x\to-\infty} \frac{g(x)}{F^{-1}(ax^2)} = 1, \qquad \lim_{x\to+\infty} \frac{g(x)}{F^{-1}(ax^2)} = -1. \tag{2.53}$$

Further, there exists $K > 0$, such that for $x \in \mathbb{R}$,

$$|g(x)| \le K(1 + |x|), \qquad |g'(x)| \le K.$$

Proof. From Lemma 2.17, we know that for every parameter $b \ge 0$ there exists a unique solution $y_r(x;b)$ on its maximal domain D_b of existence and $y_r(x;b) \ge 0$ for every $x \ge \sqrt{b/a}$, and $y_r(x;b)$ is unique and satisfies

$$\lim_{x\to+\infty} \frac{y_r(x;b)}{F^{-1}(ax^2)} = 1.$$

By Lemma 2.17, we have $[0,\infty) \subset D_b$. Define $y_l(x;b) = -y_r(-x;b)$ on $(-\infty, 0]$. Then

$$\lim_{x\to-\infty} \frac{y_l(x;b)}{F^{-1}(ax^2)} = -\lim_{x\to-\infty} \frac{y_r(-x;b)}{F^{-1}(ax^2)} = -\lim_{x\to\infty} \frac{y_r(x;b)}{F^{-1}(ax^2)} = -1.$$

Moreover, since F is even, for $x \le 0$,

$$\begin{aligned} {y_l}'(x;b) = {y_r}'(-x;b) &= -a(-x)^2 + b + F(y_r(-x;b)) \\ &= -ax^2 + b + F(-y_r(-x;b)) \\ &= -ax^2 + b + F(y_l(x;b)). \end{aligned} \tag{2.54}$$

That is, $y_l(x;b)$ also satisfies (2.35) on $(-\infty, 0]$.

For $b = 0$, by i) in the proof of Lemma 2.17, $y_r(x;0) > F^{-1}(ax^2)$, hence

$$y_r(0;0) > F^{-1}(0) = 0 > -y_r(0;0) = y_l(0;0).$$

By (2.39) in Lemma 2.17, for $x \ge 0$, $y_r(x;b)$ is strictly decreasing in b and thus $y_r(x;b) \le y_r(x;0) < \infty$ for all $b \ge 0$. In addition, we claim that as $b \to +\infty$, $y_r(0;b)$ goes to $-\infty$. Suppose not. Then there exists

$$\delta_1 := \lim_{b\to+\infty} y_r(0;b) > -\infty.$$

As a result,

$$\begin{aligned} y_r(1;b) &= y_r(0;b) + \int_0^1 -ax^2 + b + F(y(x;b))\, dx \\ &\ge y_r(0;b) + \int_0^1 (-a+b) dx \\ &\ge \delta_1 + b - a, \end{aligned}$$

and, for $b \to +\infty$,

$$y_r(1;0) \geq \lim_{b\to+\infty} y_r(1;b) \geq \lim_{b\to+\infty} \delta_1 + b - a = +\infty,$$

which leads to contradiction. Hence as $b \to +\infty$, $y_r(0;b)$ goes to $-\infty$, and $y_l(0;b) = -y_r(0;b)$ goes to $+\infty$. Thus, for some constant b_F we have 0 is contained in D_{b_F} and

$$y_r(0;b_F) = 0 = y_l(0;b_F). \tag{2.55}$$

As $y_r(x;b)$ is decreasing in b, the constant b_F is unique.

Now we use $y_r(\cdot;b_F)$ and $y_l(\cdot;b_F)$ to construct the solution for (2.36):

$$g(x) := -y_r(x;b_F)\mathbb{1}_{[x\geq 0]} - y_l(x;b_F)\mathbb{1}_{[x<0]}.$$

It's easy to see that g is defined on $\mathbb{R}$ and satisfies the growth conditions (2.53). We now show that g is indeed a solution of (2.36) with $b = b_F$. Using (2.55), we can see that g is continuous and equal to zero at $x = 0$. Therefore,

$$\begin{aligned} g(x) &= -y_r(x;b_F)\mathbb{1}_{[x\geq 0]} + y_r(-x;b_F)\mathbb{1}_{[x<0]} \\ &= -y_r(x;b_F)\mathbb{1}_{[x>0]} + y_r(-x;b_F)\mathbb{1}_{[x\leq 0]} = -g(-x), \end{aligned}$$

which implies that g is odd. Furthermore, as y_r is increasing on $[\sqrt{b_F/a}, \infty)$, and for $x \in [0, \sqrt{b_F/a}]$,

$$y_r'(x;b_F) = -ax^2 + b_F + F(y_r(x;b_F)) \geq -ax^2 + b_F \geq 0,$$

y_r is increasing on $[0,\infty)$, and we infer that g is decreasing. Since F is even, we have

$$F(g(x)) = F(-y_r(x;b_F)) = F(-y_l(-x;b_F)) = F(g(-x)), \quad \text{for } x \geq 0.$$

Therefore we can conclude that

$$g'(x) = -{y_r}'(x) = ax^2 - b_F - F(y_r(x;b_F)) = ax^2 - b_F - F(g(x)), \quad \text{for } x > 0.$$

Likewise,

$$g'(x) = -{y_l}'(x) = ax^2 - b_F - F(y_l(x;b_F)) = ax^2 - b_F - F(g(x)), \quad \text{for } x < 0.$$

Moreover, the continuity of g' is guaranteed at $x = 0$ since

$$\lim_{x\to 0^+} g'(x) = -{y_r}'(0;b_F) = -b_F = -{y_l}'(0;b_F) = \lim_{x\to 0^-} g'(x).$$

In summary, g therefore is indeed a solution of (2.36) with $b = b_F$.

Next, we show that g is unique. Let g, defined and continuously differentiable on $\mathbb{R}$, satisfy (2.36) for some $b \geq 0$ and also satisfy $xg(x) \leq 0$ for $x \in \mathbb{R}$. Then $-g$ is the unique function $y(\cdot;b)$ in Lemma 2.17. Because F is even and

g satisfies (2.36), we know $g(-x)$ also satisfies the conditions of Lemma 2.17. Hence $g(-x) = y(x;b)$ for x in the maximal interval of existence of $y(\cdot;b)$. Therefore,

$$-g(0) = y(0;b) = g(0),$$

which implies $y(0;b) = 0$. This forces b to be equal to b_F, and g to be the function constructed above.

The growth condition (2.53) and Lemma 2.16 (v) imply that there exist $x_0 > 0$ and $\hat{c} > 0$ such that, for every $|x| > x_0$,

$$|g(x)| \leq 2|F^{-1}(ax^2)| \leq 2\hat{c}|x|.$$

Therefore, for all x, and since $-g$ is increasing,

$$|g(x)| \leq |g(x_0)| + 2\hat{c}|x|. \tag{2.56}$$

Now we would like to show the boundedness of g', which follows the same idea as Bayraktar et al. (2018). Since g is odd, we only need to show that for $x > 0$, g' is bounded from below. From (2.56), we can infer that as $x \to \infty$, g' cannot go to $-\infty$. Therefore, there exists $M > 0$ and an increasing sequence $\{x_n\}_{n=1}^{\infty}$ such that $x_n \to \infty$ and $-M \leq g'(x_n) \leq 0$. Now suppose g' is not bounded from below, which means that for every integer $n > M$, there exists $z_n > x_n$ such that $g'(z_n) \leq -n$. For each $n > M$, let $m(n) > n$ denote the first integer such that $x_n < z_n < x_{m(n)}$. Then from

$$g'(z_n) \leq -n < -M \leq \min\{g'(x_n), g'(x_{m(n)})\},$$

we can infer that there exists a local minimum of g' on $[x_n, x_{m(n)}]$ for every integer $n > M$, denoted by ξ_n. Therefore, for every integer $n > M$, $g''(\xi_n) = 0$, and

$$\begin{aligned} 0 \leq g'''(\xi_n) &= 2a - F''(g(\xi_n))\big(g'(\xi_n)\big)^2 - F'(g(\xi_n))g''(\xi_n) \\ &= 2a - F''(g(\xi_n))\big(g'(\xi_n)\big)^2. \end{aligned}$$

Together with Assumption 2.14 (iv), we know that $F''(g(\xi_n)) > 0$ for n large enough, and hence

$$n^2 \leq \big(g'(\xi_n)\big)^2 \leq \frac{2a}{F''(g(\xi_n))} \leq \frac{2a}{\tilde{C}},$$

which leads to a contradiction. Without loss of generality, we choose $M > 0$ large enough so that $|g'(x)| < M$ for every $|x| > x_0$.

Now choose $K > M + |g(x_0)| + 2\hat{c}$, we have

$$|g(x)| \leq K(1 + |x|), \qquad |g'(x)| \leq K$$

as asserted. This completes the proof. $\square$

Next, we show that with $b = b_F$, the solution to the first-order ODE (2.36) on $\mathbb{R}$ with $xg(x) \leq 0$ is also the unique solution on $\mathbb{R}$ to the second-order ODE (2.37) with $xg(x) \leq 0$.

Lemma 2.19. *Let F be a function satisfying Assumption (2.14). Then the unique solution g on $\mathbb{R}$ to (2.36) such that $xg(x) \leq 0$ is also the unique solution on $\mathbb{R}$ of the second-order ODE*

$$g''(x) = 2ax - F'(g(x))g'(x) \tag{2.37}$$

such that $xg(x) \leq 0$.

Proof. In view of the first-order ODE (2.36) satisfied by g, its derivative is also differentiable. Differentiating the ODE for g in turn shows that g also satisfies the second-order ODE (2.37).

Now suppose $\tilde{g}$ is a solution of the second-order ODE (2.37) satisfying $x\tilde{g}(x) \leq 0$, hence we can infer that $\tilde{g}$ is non-increasing at zero. As

$$(F(\tilde{g}(x)))' = F'(\tilde{g}(x))\tilde{g}'(x),$$

integrating both sides of (2.37) gives

$$\tilde{g}'(x) = \tilde{g}'(0) + \int_0^x \Big(2a\xi - F'(\tilde{g}(\xi))\tilde{g}'(\xi)\Big) d\xi = ax^2 - \tilde{b} - F(\tilde{g}(x)),$$

for some constant $\tilde{b} = F(\tilde{g}(0)) - \tilde{g}'(0)$. By Lemma 2.18 we know $b_F > 0$ is the unique constant such that (2.36) has a solution on $\mathbb{R}$ with $xg(x) \leq 0$. Thus, $\tilde{b} = b_F$ and, by the uniqueness of g, we have $\tilde{g} = g$. This completes the proof. $\square$

We introduce one more lemma before the proof of Lemma 2.5.

Lemma 2.20. *Suppose the general cost function G satisfies Assumption 2.2. Then G^*, the Legendre transform of G, satisfies Assumption 2.14, and so does $cG^*(\frac{x}{c})$, where $c > 0$ is a constant.*

Proof. Observe that the Legendre transformation of the cost function $G(x)$ is

$$G^*(x) = x(G')^{-1}(x) - G((G')^{-1}(x)).$$

Since the instantaneous cost G is even, G' and in turn $(G')^{-1}$ are odd, so that the function G^* is even. Moreover, $G(0) = G'(0) = 0$ imply $G^*(0) = 0$. As both G and $(G')^{-1}$ are differentiable,

$$(G^*)'(x) = (G')^{-1}(x) > 0.$$

In particular, $(G^*)^{-1}$ exists on $[0, \infty)$ and is differentiable. Moreover, by the convexity and twice differentiability of G,

$$(G^*)''(x) = ((G')^{-1})'(x) > 0.$$

It follows that G^* is convex and $(G^*)'$ is strictly increasing, so that Assumptions 2.14 (i,ii) are satisfied. By Assumption 2.2, $|(G')^{-1}(x)| \leq C(1+|x|^{k-1})$ for $C>0$ and $k\geq 2$. Whence, there exists a constant $K>0$ such that

$$G^*(x) = |G^*(x)| \leq |x(G')^{-1}(x)| \leq C(|x|+|x|^k) \leq K(1+|x|^k).$$

Therefore, Assumption 2.14(iii) is also satisfied. Again by Assumption 2.2, $(G')^{-1}$ is increasing, and there exists $C>0$ and $x_0>0$, such that for large $x>x_0$, $(G')^{-1}(x)$ is large and by Assumption 2.2 (iii)

$$(G^*)''(x) = ((G')^{-1})'(x) = \frac{1}{G''\left((G')^{-1}(x)\right)} \geq \frac{1}{C}.$$

Thus, Assumption 2.14(iv) holds as well. □

We now turn to the proof of Lemma 2.5.

Proof of Lemma 2.5. Let G^* denote the Legendre transform of G, and define

$$a = \frac{\widetilde{\gamma}\sigma^2}{\delta^2}, \quad F(x) = \frac{2\lambda}{\delta^2} G^*\left(\frac{x}{\lambda}\right),$$

where $\widetilde{\gamma}$ and δ are defined as in (2.21). By Lemma 2.20, G^* and in turn F satisfy Assumption 2.14. For the above choices of a and F, Lemma 2.17 and Lemma 2.18 therefore yield the existence and uniqueness of the constant b_F and the solution g on $\mathbb{R}$ to the first-order ODE (2.36) such that $xg(x)\leq 0$ for every $x\in\mathbb{R}$. In view of the first-order ODE (2.36) satisfied by g,

$$\begin{aligned} g'(x) &= \frac{\widetilde{\gamma}\sigma^2}{\delta^2}x^2 - F(g(x)) - b_F \\ &= \frac{\widetilde{\gamma}\sigma^2}{\delta^2}x^2 - \frac{2\lambda}{\delta^2}\left[\frac{g(x)}{\lambda}(G')^{-1}\left(\frac{g(x)}{\lambda}\right) - G\left((G')^{-1}\left(\frac{g(x)}{\lambda}\right)\right)\right] - b_F, \end{aligned}$$

multiplying the above equation by $\frac{1}{2}\delta^2$ and differentiating on both sides show that g is also the unique solution to the ODE (2.7) from Lemma 2.5:

$$\begin{aligned} \frac{1}{2}\delta^2 g''(x) &= -\left[g(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right) - \lambda G\left((G')^{-1}\left(\frac{g(x)}{\lambda}\right)\right)\right]' + \widetilde{\gamma}\sigma^2 x \\ &= -g'(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right) + \widetilde{\gamma}\sigma^2 x. \end{aligned}$$

Here, we have used in the last step that

$$\left[g(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right)\right]' = g'(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right) + g(x)((G')^{-1})'\left(\frac{g(x)}{\lambda}\right)\frac{g'(x)}{\lambda},$$

and

$$\left[\lambda G\left((G')^{-1}\left(\frac{g(x)}{\lambda}\right)\right)\right]' = \lambda G'\left((G')^{-1}\left(\frac{g(x)}{\lambda}\right)\right)((G')^{-1})'\left(\frac{g(x)}{\lambda}\right)\frac{g'(x)}{\lambda}$$
$$= g(x)((G')^{-1})'\left(\frac{g(x)}{\lambda}\right)\frac{g'(x)}{\lambda}.$$

To complete the proof, notice that

$$F^{-1}(ax^2) = F^{-1}\left(\frac{\gamma\sigma^2}{\delta^2}x^2\right) = (G^*)^{-1}\left(\frac{\delta^2}{2\lambda}\frac{\widetilde{\gamma}\sigma^2}{\delta^2}x^2\right) = \lambda(G^*)^{-1}\left(\frac{\widetilde{\gamma}\sigma^2}{2\lambda}x^2\right),$$

which yields the analogue of the growth conditions 2.53 that are exactly (2.8):

$$\lim_{x\to-\infty}\frac{g(x)}{\lambda(G^*)^{-1}(\frac{\widetilde{\gamma}\sigma^2}{2\lambda}x^2)} = 1, \qquad \lim_{x\to+\infty}\frac{g(x)}{\lambda(G^*)^{-1}(\frac{\widetilde{\gamma}\sigma^2}{2\lambda}x^2)} = -1. \tag{2.8}$$

□

2.6.4 Calibration Details

In this section, we provide some additional details concerning the calibration of the model with costs of general power form at the end of Section 2.5.2. If $G_q(x) = |x|^q/q$ with cost parameter $\lambda_q > 0$, $q \in (1, 2]$, then the nonlinear ODE (2.7) from Lemma 2.5 can be simplified by rescaling. Indeed, the solution then can be written as

$$g_q(x) = \left(\frac{\lambda_q}{q}\right)^{\frac{3}{q+2}}\left(\frac{\widetilde{\gamma}\sigma^2\delta_q^4}{8}\right)^{\frac{q-1}{q+2}}\tilde{g}_q\left(2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}}x\right), \tag{2.57}$$

where $\tilde{g}_q$ is the unique solution on $\mathbb{R}$ of[9]

$$\tilde{g}_q''(x) + \tilde{g}_q'(x)\,\mathrm{sgn}(\tilde{g}_q(x))\left|\frac{\tilde{g}_q(x)}{q}\right|^{\frac{1}{q-1}} = 2x. \tag{2.58}$$

This rescaled ODE only depends on the elasticity q of the trading cost but not on the other primitives of the model. As a consequence, the rescaled ODE only needs to be solved numerically once for each q to determine the transaction cost λ_q and δ_q that match the variance of the state variable for proportional costs and the average share turnover observed empirically. To this end, first notice that

$$(G_q')^{-1}\left(\frac{g_q(x)}{\lambda_q}\right) = -\,\mathrm{sgn}(x)\left|\frac{g_q(x)}{\lambda_q}\right|^{\frac{1}{q-1}}$$
$$= -\left(\frac{q\widetilde{\gamma}\sigma^2\delta_q^4}{8\lambda_q}\right)^{\frac{1}{q+2}}\mathrm{sgn}(x)\left|\frac{\tilde{g}_q\left(2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}}x\right)}{q}\right|^{\frac{1}{q-1}}. \tag{2.59}$$

[9]As shown in Lemma 2.19, $\tilde{g}_q$ is in fact the solution to the first-order equation (17) in (Guasoni and Weber, 2018, Theorem 6), with $q = \alpha + 1$.

For power costs $G_q(x) = |x|^q/q$, the dynamics of the state-variable $\Delta\varphi_t^1$ from Lemma 2.7 in turn are given by

$$d\Delta\varphi_t^1 = -\left(\frac{q\widetilde{\gamma}\sigma^2\delta_q^4}{8\lambda_q}\right)^{\frac{1}{q+2}} \operatorname{sgn}(\Delta\varphi_t^1) \left| \frac{\tilde{g}_q\left(2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}}\Delta\varphi_t^1\right)}{q} \right|^{\frac{1}{q-1}} dt$$
$$+ \delta_q dW_t. \tag{2.60}$$

The stationary density (c.f. (Karatzas and Shreve, 1998, Chapter 5)) of the $\Delta\varphi_t^1$ therefore can therefore be computed via the normalized speed measure as[10]

$$\nu_q(x) = \frac{2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}} \exp\left(-\int_0^{2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}}x} \left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}} dy\right)}{2\int_0^\infty \exp\left(-\int_0^x \left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}} dy\right) dx}.$$

The goal now is to choose the model parameters λ_q and δ_q to match the share turnover in the model to its empirical level and the stationary variance of the state variable to its counterpart for proportional costs. To this end, define

$$\tilde{c}_q = \left[2\int_0^\infty \exp\left(-\int_0^x \left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}} dy\right) dx\right]^{-1},$$
$$\tilde{v}_q = 2\tilde{c}_q \int_0^\infty x^2 \exp\left(-\int_0^x \left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}} dy\right) dx.$$

To match the total share turnover, we then need

$$\text{ShTu} = \int_{-\infty}^\infty \left|\frac{g_q(x)}{\lambda_q}\right|^{\frac{1}{q-1}} \nu_q(x)dx = 2^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{1}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{2q}{q+2}} \tilde{c}_q\delta_q^2, \tag{2.61}$$

which is satisfied if we choose

$$\delta_q = \left(\frac{\lambda_q}{2^{q-1}q\widetilde{\gamma}\sigma^2}\left(\frac{\text{ShTu}}{\tilde{c}_q}\right)^{q+2}\right)^{1/4}. \tag{2.62}$$

After matching the average share turnover, we now choose the size λ_q of the trading cost to match the stationary variance of the state variable to its counterpart for proportional costs λ_1. For power costs with elasticity q, the stationary mean of the state variable is zero by the symmetry of ν_q, and the stationary

[10] For quadratic costs $q = 2$, where the state variable has Ornstein-Uhlenbeck dynamics, this reduces to the density of the normal distribution.

variance can in turn be computed by integrating against the stationary density,

$$2\int_0^\infty x^2\nu_q(x)dx = \frac{\tilde{v}_q}{4^{\frac{q-1}{q+2}}\left(\frac{q\widetilde{\gamma}\sigma^2}{\lambda_q}\right)^{\frac{2}{q+2}}\left(\frac{1}{\delta_q}\right)^{\frac{4q}{q+2}}} = \frac{\tilde{v}_q\lambda_q}{2^{q-1}q\widetilde{\gamma}\sigma^2}\left(\frac{\mathrm{ShTu}}{\tilde{c}_q}\right)^q, \qquad (2.63)$$

where we have inserted (2.62) in the second step. For proportional costs the state variable is a doubly reflected Brownian motion, whose stationary law is the uniform distribution on $[-l, l]$ (which has variance $l^2/3$). Recall from Section 2.5.2 that for proportional costs λ_1, we need

$$\delta_1 = \frac{\xi_1}{\sigma} = \left(\frac{12\mathrm{ShTu}^3\lambda_1}{\widetilde{\gamma}\sigma^2}\right)^{1/4},$$

to match the average share turnover ShTu. After inserting this into Formula (2.18) for the trading boundary l, it follows that the stationary variance $l^2/3$ of the state variable for proportional costs is given by

$$\frac{l^2}{3} = \frac{1}{3}\left(\frac{3\lambda_1\delta_1^2}{2\widetilde{\gamma}\sigma^2}\right)^{2/3} = \frac{\lambda_1\mathrm{ShTu}}{\widetilde{\gamma}\sigma^2}.$$

To match this with the corresponding stationary variance (2.63) for power costs with elasticity q, we therefore need to choose

$$\lambda_q = \frac{q\tilde{c}_q}{\tilde{v}_q}\left(\frac{2\tilde{c}_q}{\mathrm{ShTu}}\right)^{q-1}\lambda_1. \qquad (2.64)$$

In summary, for a given value of q, the solution $\tilde{g}_q$ of (2.58) therefore needs to be computed numerically on a fine grid once. Then, we can use numerical integration to determine $\tilde{c}_q$, $\tilde{v}_q$. This finally allows us to compute the λ_q corresponding to the proportional trading cost λ_1 via (2.64), and pins down the corresponding δ_q through (2.62).

Chapter 3

Endogenous Volatilities and Nonlinear FBSDEs

3.1 Introduction

In the previous chapter, we have studied a simple, tractable equilibrium model with transaction costs. There, very specific assumptions on the exogenous inputs of the model have allowed us to obtain explicit solutions that facilitate a calibration of the model to time-series data. However, this very stylized model suffers from a number of limitations. We have already discussed discrepancies between trading volume in the model and the empirical data in Chapter 2. Two other key restriction of the benchmark model are that liquidity premia are zero on average and volatilities are given exogenously. The first property is at odds with a large empirical literature that documents that less liquid securities exhibit higher average expected returns (Amihud and Mendelson, 1986; Brennan and Subrahmanyam, 1996; Pástor and Stambaugh, 2003). The second limitation rules out studies of the effects changes in market liquidity have on market volatility (compare, e.g., the numerical analyses of the impact of transaction taxes in Adam, Beutel, Marcet, and Merkel (2015); Buss, Dumas, Uppal, and Vilkov (2016)).

In order to address these limitations, it is natural to extend our baseline model to more general state dynamics (where volatilities are mean-reverting stochastic processes, for example, that lead to richer trading-volume dynamics) and to determine the volatility process endogenously by matching an exogenous terminal dividend (which allows us to study how liquidity influences volatility). As a byproduct, models of this kind can also generate systematic liquidity premia as demonstrated in a model with quadratic costs by Herdegen, Muhle-Karbe, and Possamaï (2019). However, the analysis of models with more general state dynamics and endogenous volatilities is substantially more involved, in that it leads to fully-coupled systems of nonlinear forward-backward stochastic differential equations (FBSDEs). Indeed, the optimal positions evolve forward from the agents' initial allocations. In contrast, the initial optimal trading rates need to be determined as part of the solution, taking into account that trading stops at the terminal time. Likewise, the stock dynamics also need to be derived from the terminal dividend. For quadratic trading costs, wellposedness of this multidimensional and fully-coupled system has recently been established by Herdegen et al. (2019) for agents with sufficiently similar risk aversions.

In this chapter, we show how to derive similar FBSDE systems for models with general transaction costs. If trading costs are not quadratic, wellposedness of the system becomes an even more challenging and completely open problem. Moreover, simplifications to systems of coupled Riccati equations as in Herdegen et al. (2019) are not possible even for the simplest linear state dynamics. In order to nevertheless shed some first light on the behaviour of such models, we demonstrate in this chapter how to approximate their solutions numerically by adapting the simulation-based deep learning approach of Han, Jentzen, and E (2018) if the time horizon is not too long. Here, the idea is to use a deep neural network to parametrize the "decoupling field" that describes the backward components as a function of the forward variables. For each choice of the decoupling field, the corresponding forward dynamics of the system can in turn be simulated by a standard Euler scheme, so that it remains to keep updating the initial guess for the decoupling field using stochastic gradient descent until the simulation matches the terminal condition of the equation sufficiently well.

We verify that our algorithm produces accurate results by comparing it to the Riccati system that describes the equilibrium in a benchmark example with quadratic costs and linear state dynamics. With minor adjustments, the same algorithm is also able to deal with other trading cost specifications. As in our baseline model from Chapter 2, the numerical results we report here suggest that the specification of the trading cost has only a minor effect on the equilibrium price dynamics for our calibrated parameters. Complementing these numerical results with a rigorous verification theorem is an important direction for future research.[1]

This chapter is organized as follows: we first introduce the general setting in Section 3.2. In Section 3.3, we then define the (Radner) equilibria with endogenous volatilities and discuss the frictionless baseline model. The relationship between frictional equilibrium with general transaction costs on the trading rate and a FBSDE system is derived in Section 3.4. We then describe the deep-learning approach in Section 3.5. Finally, equilibria in models with proportional transaction costs and their relation to FBSDEs with reflection are discussed in Section 3.6.

Notation. We fix a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, \mathbb{P})$ with finite time horizon $T > 0$, where the filtration is generated by a d-dimensional stan-

[1]For the model with quadratic costs and linear state dynamics from Herdegen et al. (2019), we extend their local results to *global* existence in Chapter 4.

dard Brownian motion $W = (W_t)_{t\in[0,T]}$. Throughout, let $\|\cdot\|$ be the 2-norm of a real-valued vector. We denote by FV the set of finite-variation processes and, for $p \geq 1$, write $\mathbb{H}^p$ for the $\mathbb{R}$-valued, progressively measurable processes $X = (X_t)_{t\in[0,T]}$ that satisfy

$$\|X\|_{\mathbb{H}^p} := \left(\mathbb{E}\left[\left(\int_0^T |X_t|^2 dt\right)^{p/2}\right]\right)^{1/p} < \infty.$$

3.2 Market

In this chapter, we consider more general state dynamics than in Sections 2.2.2, 2.3, and 2.4. To wit, randomness in the market is now generated by a d-dimensional Brownian motion $(W_t)_{t\in[0,T]}$. The cumulative random endowment of agent $n = 1, 2$ now is of the general form

$$d\zeta_t^n = \xi_t^n dW_t, \quad \text{for a general } d\text{-dim process } \xi^n \in \mathbb{H}^2,$$

and the price of the risky asset has dynamics

$$dS_t = \mu_t dt + \sigma_t dW_t. \tag{3.1}$$

In contrast to Chapter 2, not just the equilibrium return process $(\mu_t)_{t\in[0,T]}$ but also also the initial price $S_0 \in \mathbb{R}$ and the volatility process σ are to be determined in equilibrium by matching the agents' demand to the supply $s \in \mathbb{R}$ of the risky asset. To pin down these additional quantities, we assume as in Herdegen et al. (2019) that the terminal stock price is given by an exogenous $\mathscr{F}_T$-measurable random variable:

$$S_T = \mathfrak{S}. \tag{3.2}$$

This can be interpreted as a fundamental value as in Kyle (1985) or as a terminal dividend as in Grossman and Stiglitz (1980). In order to ensure that the agents' mean-variance optimization problems are well-defined for a sufficiently large class of trading strategies, we require the volatility process to belong to $\mathbb{H}^2$ and the corresponding return process to satisfy the no-arbitrage condition $\mu = \sigma\kappa$ for an $\mathbb{R}^d$-valued market price of risk process $\kappa \in \mathbb{H}^2$.

3.3 Frictionless Optimization and Equilibrium

3.3.1 Individual Optimization

As a reference point, we once again first consider the frictionless version of the model. As in Section 2.2, the agents maximize expected returns penalized for the corresponding quadratic variations. That is, without transaction costs, they optimize

$$\bar{J}_T^n(\psi) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\|\psi_t\sigma_t + \xi_t^n\|^2\right)dt\right], \tag{3.3}$$

over *admissible* strategies satisfying the integrability condition $\psi\sigma \in \mathbb{H}^2$ with respect to the given price process S with dynamics (3.1).

The frictionless results from Section 2.2 readily adapt to this more general setting. Indeed, also for a general stochastic volatility process, pointwise maximization of the goal functional (3.3) still yields the agents' individually optimal strategies,

$$\varphi_t^n = \frac{\mu_t}{\gamma^n\sigma_t\sigma_t^\top} - \frac{\xi_t^n\sigma_t^\top}{\sigma_t\sigma_t^\top}, \qquad t \in [0,T]. \tag{3.4}$$

Remark 3.1. Recall that we require the expected return process to satisfy the no-arbitrage condition $\mu = \sigma\kappa$ for a market price of risk $\kappa \in \mathbb{H}^2$. By the Cauchy-Schwarz inequality, the admissibility condition $\psi\sigma \in \mathbb{H}^2$ therefore guarantees that

$$\mathbb{E}\left[\int_0^T |\psi_t\mu_t| dt\right] \leq \frac{1}{2}\left[\mathbb{E}\left[\int_0^T \|\psi_t\sigma_t\|^2 dt\right] + \mathbb{E}\left[\int_0^T \|\kappa_t\|^2 dt\right]\right].$$

Also observe that the individually optimal strategy φ^n from (3.4) satisfies $\varphi^n\sigma \in \mathbb{H}^2$ automatically:

$$\begin{aligned}\mathbb{E}\left[\int_0^T \|\varphi_t^n\sigma_t\|^2 dt\right] &= \mathbb{E}\left[\int_0^T \left\|\frac{\mu_t\sigma_t}{\gamma^n\sigma_t\sigma_t^\top} - \frac{\xi_t^n\sigma_t^\top}{\sigma_t\sigma_t^\top}\sigma_t\right\|^2 dt\right] \\ &\leq 2\mathbb{E}\left[\int_0^T \left\|\frac{\kappa_t}{\gamma^n}\right\|^2 + \|\xi_t^n\|^2 dt\right] < \infty.\end{aligned}$$

3.3.2 Equilibrium

With the agents' individually optimal strategies at hand, we now turn to the corresponding equilibrium prices. In the present context with endogenous volatility process, the notion of Radner equilibrium from Definition 2.3 naturally generalizes as follows:

Definition 3.2. A price process S for the risky asset with initial asset price $S_0 \in \mathbb{R}$, expected returns process $\mu = (\mu_t)_{t\in[0,T]}$, as well as volatility process $\sigma = (\sigma_t)_{t\in[0,T]} \in \mathbb{H}^2$ is called a *(Radner) equilibrium* if:

(i) the price process satisfies the no-arbitrage condition $\mu = \sigma\kappa$ for $\kappa \in \mathbb{H}^2$;

(ii) the terminal condition $S_T = \mathfrak{S}$ is satisfied;

(iii) *individual optimality*: there exists admissible strategies φ^n for agents $n = 1, 2$ maximizing their goal functionals $\bar{J}^n_T$ with respect to the given price process S;

(iv) *market clearing*: the agents' optimal strategies clear the the market for the risky asset at all times: $\varphi^1_t + \varphi^2_t = s$, $t \in [0, T]$.

As in Section 2.2, the equilibrium return is pinned down by matching the agents' total demand $\varphi^1_t + \varphi^2_t$ to the supply s of the risky asset at all times $t \in [0, T]$:

$$\bar{\mu}_t = \bar{\gamma}\left[s\bar{\sigma}_t + \xi^1_t + \xi^2_t\right]\bar{\sigma}_t^\top, \quad t \in [0, T], \quad \text{where recall that } \bar{\gamma} = \frac{\gamma^1\gamma^2}{\gamma^1 + \gamma^2}. \tag{3.5}$$

Now, however, we also need to determine the corresponding initial price of the risky asset and its volatility. To this end, insert (3.5) into (3.1) and recall the terminal condition (3.2). This leads the following scalar quadratic BSDE:

$$d\bar{S}_t = \bar{\gamma}\left[s\bar{\sigma}_t + \xi^1_t + \xi^2_t\right]\bar{\sigma}_t^\top dt + \bar{\sigma}_t dW_t, \qquad S_T = \mathfrak{S}, \tag{3.6}$$

As is well known, the solution of this equation can be expressed in terms of the Laplace transform of the terminal condition, leading to explicit solutions in many concrete examples. We sumarize these results in Proposition 3.3; a detailed discussion can be found in (Herdegen et al., 2019, Section 4.1).

Proposition 3.3. *Suppose the following assumptions hold:*

(i) the aggregate random endowment satisfies $\xi^1 + \xi^2 \in \mathbb{H}^2$;

(ii) the local martingale $Z^\xi := \exp\left(-\int_0^\cdot \bar{\gamma}(\xi^1_t + \xi^2_t)dW_t\right)$ *is a true martingale;*

(iii) $\mathbb{E}\left[\left(Z^\xi_T\right)^{\frac{\widehat{p}}{\widehat{p}-1}} + \left(Z^\xi_T\right)^{-\frac{(1+p)\tilde{p}}{\tilde{p}-1}} + e^{4\frac{(1+p)\widehat{p}}{p}\bar{\gamma}s\mathfrak{S}} + e^{-4(1+p)\tilde{p}\bar{\gamma}s\mathfrak{S}}\right] < \infty$ *for some constants* $p > 1$, $\widehat{p} > 1$, *and* $\tilde{p} > 1$.

Then there exists a solution $(\bar{S}, \bar{\sigma}) \in (\mathbb{H}^2)^2$ *of* (3.6), *and* $\bar{S}$ *is an equilibrium price in the frictionless market.*

Remark 3.4. In fact, the uniqueness of frictionless equilibrium can also be established among a large class of processes, see Proposition 4.3 in Herdegen et al. (2019). We will therefore henceforth refer to *the* frictionless equilibrium price (3.6).

The agents' optimal trading strategies (3.4) corresponding to the frictionless equilibrium price (3.6) are

$$\bar{\varphi}^1_t = \frac{s\gamma^2}{\gamma^1 + \gamma^2} + \frac{(\gamma^2\xi^2_t - \gamma^1\xi^1_t)\bar{\sigma}_t^\top}{(\gamma^1 + \gamma^2)\bar{\sigma}_t\bar{\sigma}_t^\top} \qquad \bar{\varphi}^2_t = s - \bar{\varphi}^1_t, \qquad t \in [0, T]. \tag{3.7}$$

Using (3.7), the frictionless equilibrium return (3.5) can be rewritten as follows:

$$\bar{\mu}_t = \frac{1}{2}\left[(\gamma^2 s + (\gamma^1 - \gamma^2)\bar{\varphi}_t^1)\bar{\sigma}_t + (\gamma^1\xi_t^1 + \gamma^2\xi_t^2)\right]\bar{\sigma}_t^\top, \quad t \in [0,T]. \tag{3.8}$$

If the aggregate endowment is zero and the terminal condition is a linear function of the driving Brownian motion, then the frictionless equilibrium prices have Bachelier dynamics just as in the model with exogenous volatility studied in Chapter 2:

Example 3.5. Suppose the state variable is a one-dimensional Brownian motion ($d = 1$), the aggregate endowment is zero as in Lo et al. (2004) ($\xi^1 = -\xi^2$), and the terminal dividend is a linear function of the driving Brownian motion ($\mathfrak{S} = \beta T + \alpha W_T$). Then, the frictionless equilibrium price $\bar{S}$ is a Bachelier model with constant expected returns and volatilities:

$$d\bar{S}_t = \bar{\gamma} s\alpha^2 dt + \alpha dW_t, \qquad \bar{S}_0 = (\beta - \bar{\gamma} s\alpha^2)T.$$

To wit, the volatility of the asset is inherited from uncertainty about the terminal dividend here. The corresponding expected return in turn compensates the agents for exposure to this risk.

Agent $n = 1, 2$'s optimal trading strategies in this frictionless equilibrium are also Brownian motions:

$$\bar{\varphi}_t^n = \frac{s\bar{\gamma}}{\gamma^n} - \frac{\xi_t^n}{\alpha}, \quad t \in [0,T]. \tag{3.9}$$

3.4 Frictional Optimization and Equilibrium

3.4.1 Individual Optimization

With transaction costs, both individual optimization and the corresponding equilibria become significantly more involved, leading to systems of fully-coupled nonlinear FBSDEs. As in Section 2.3, we focus on absolutely continuous trading strategies here, and penalize the trading rate $\dot{\psi}_t = d\psi_t/dt$ with an *instantaneous trading cost* $G_t(\dot{\psi}_t) = \lambda_t G(\dot{\psi}_t)$, where the smooth convex function G satisfies Assumption 2.2. The singular limiting case of proportional costs is in turn discussed separately in Section 3.6.

Like in the partial-equilibrium model of Moreau et al. (2017), we allow the transaction cost to fluctuate randomly over time:

$$\lambda_t = \lambda\Lambda_t, \qquad t \in [0,T]. \tag{3.10}$$

Here, the constant $\lambda > 0$ modulates the magnitude of the cost (this scaling parameter will be sent to zero in the small-cost asymptotics we consider in Chapters 4 and 5 below). The *strictly positive* stochastic processes $(\Lambda_t)_{t\in[0,T]}$ describes the fluctuations of liquidity over time, and allows to model "liqudity risk" as in Acharya and Pedersen (2005), for example.

Let us now consider the agents' individual optimization problems for a given initial asset price $S_0 \in \mathbb{R}$, expected returns process $(\mu_t)_{t\in[0,T]}$ and volatility process $(\sigma_t)_{t\in[0,T]}$. The frictional analogue of the mean-variance goal functional (3.3) is

$$J_T^n(\psi) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\|\sigma_t\psi_t + \xi_t^n\|^2 - \lambda_t G\left(\dot{\psi}_t\right)\right)dt\right]. \tag{3.11}$$

To make sure all terms are well defined, we focus on *admissible* strategies that satisfy the same integrability condition $\psi\sigma \in \mathbb{H}^2$ as in the frictionless case, and whose expected trading costs are finite:

$$E\left[\int_0^T \lambda_t G\left(\dot{\psi}_t\right)dt\right] < \infty. \tag{3.12}$$

By strict convexity of the goal functional (3.3), optimality of a trading rate $\dot{\varphi}^n$ for agent n is equivalent to the first-order condition that the Gateaux derivative $\lim_{\rho\to 0}(J_T^n(\dot{\varphi}^n + \rho\dot{\eta}) - J_T^n(\dot{\varphi}^n))/\rho$ vanishes for *any* admissible perturbation η, cf. Ekeland and Temam (1999):

$$0 = \mathbb{E}_t\left[\int_0^T \left(\mu_t \int_0^t \dot{\eta}_u du - \gamma^n(\sigma_u\varphi_u^n + \xi_u^n)\sigma_u^\top \int_0^t \dot{\eta}_u du - \lambda_t G'(\dot{\varphi}_t^n)\dot{\eta}_t\right)dt\right].$$

As in Bank et al. (2017), this can be rewritten using Fubini's theorem as

$$0 = \mathbb{E}_t\left[\int_0^T \left(\int_t^T \left(\mu_u - \gamma^n(\sigma_u\varphi_u^n + \xi_u^n)\sigma_u^\top\right)du - \lambda_t G'(\dot{\varphi}_t^n)\right)\dot{\eta}_t dt\right].$$

Since this has to hold for *any* perturbation $\dot{\eta}_t$, the tower property of conditional expectation yields

$$\begin{aligned}\lambda_t G'\left(\dot{\varphi}_t^n\right) &= \mathbb{E}_t\Big[\int_t^T \left(\mu_u - \gamma^n(\sigma_u\varphi_u^n + \xi_u^n)\sigma_u^\top\right)du\Big] \\ &= M_t^n + \int_0^t \left(\gamma^n(\sigma_u\varphi_u^n + \xi_u^n)\sigma_u^\top - \mu_u\right)du, \end{aligned} \tag{3.13}$$

for a martingale $dM_t^n = Z_t^n dW_t$ that needs to be determined as part of the solution. Solving for the dynamics of the agents' optimal trading rates would introduce the dynamics of the trading costs. Accordingly, it is preferable to instead work with the marginal trading cost as the backward process that describes the agents' optimal controls:

$$Y_t^n := \lambda_t G'\left(\dot{\varphi}_t^n\right), \tag{3.14}$$

and from (3.13) we can easily infer that $Y_T^n = 0$. With this notation, the corresponding trading rates are

$$\dot{\varphi}_t^n = (G')^{-1}\left(\frac{Y_t^n}{\lambda_t}\right).$$

Agent n's optimal position φ^n and the corresponding marginal trading costs Y^n in turn solve the nonlinear FBSDE

$$d\varphi_t^n = (G')^{-1}\left(\frac{Y_t^n}{\lambda_t}\right)dt, \qquad \varphi_0^n = \varphi_{0-}^n, \tag{3.15}$$

$$dY_t^n = \big(\gamma^n(\sigma_t\varphi_t^n + \xi_t^n)\sigma_t^\top - \mu_t\big)dt + Z_t^n dW_t, \qquad Y_T^n = 0. \tag{3.16}$$

For constant quadratic costs $\lambda x^2/2$ and constant volatility σ, this FBSDE becomes linear and can in turn be solved by reducing it to some standard Riccati equations (Bank et al., 2017; Bouchard et al., 2018). For volatilities and quadratic costs that fluctuate randomly, these ODEs are replaced by a backward *stochastic* Riccati equation, compare Kohlmann and Tang (2002); Ankirchner and Kruse (2015). With nonlinear costs, no such simplifications are possible. In fact, the wellposedness of the system is generally unclear even for short time horizons since no Lipschitz condition for $(G')^{-1}$ is satisfied for costs of power form $G(x) = |x|^q/q$, $q \in (1,2)$, for example.

3.4.2 Equilibrium

Despite these difficulties, formally solving for the corresponding equilibrium return is – surprisingly – not more difficult than for quadratic costs. To see this, first observe that symmetry of the trading cost G implies that the marginal cost G' and in turn its inverse $(G')^{-1}$ are antisymmetric. As a consequence, the market clearing condition $\dot\varphi^1 = -\dot\varphi^2$ implies that $(G')^{-1}(\dot\varphi_t^1) = -(G')^{-1}(\dot\varphi_t^2)$ and in turn

$$Y_t^1 + Y_t^2 = 0, \qquad \text{for all } t \in [0,T]. \tag{3.17}$$

After summing the corresponding backward equations (3.16), we see that the frictional equilibrium return has to satisfy

$$0 = \mu_t - \gamma^1(\sigma_t\varphi_t^1 + \xi_t^1)\sigma_t^\top + \mu_t - \gamma^2(\sigma_t\varphi_t^2 + \xi_t^2)\sigma_t^\top.$$

Together with the market clearing condition $\varphi_t^1 + \varphi_t^2 = s$, it follows that the frictional equilibrium return has the same relationship to the frictional volatility and the agents' optimal positions as in the frictionless case (3.8):

$$\mu_t = \frac{1}{2}\left[(\gamma^2 s + (\gamma^1 - \gamma^2)\varphi_t^1)\sigma_t + (\gamma^1\xi_t^1 + \gamma^2\xi_t^2)\right]\sigma_t^\top. \tag{3.18}$$

Plugging expression (3.18) back into agent 1's optimality condition (3.16) in turn yields a backward equation that is linear in the optimal position, like for quadratic costs:[2]

$$dY_t^1 = \left(\widetilde{\gamma}\sigma_t\varphi_t^1 - \frac{\gamma^2 s}{2}\sigma_t - \frac{1}{2}(\gamma^2\xi_t^2 - \gamma^1\xi_t^1)\right)\sigma_t^\top dt + Z_t^1 dW_t, \quad Y_T^1 = 0, \tag{3.19}$$

[2]Note that these linear dynamics obtain here if this equation is expressed in terms of the marginal cost $Y_t^1 = \lambda_t G'(\dot\varphi_t^1)$ rather than the trading rate $\dot\varphi_t^1$.

where recall that

$$\widetilde{\gamma} = \frac{\gamma^1 + \gamma^2}{2}.$$

All nonlinearities are absorbed into the corresponding forward component,

$$d\varphi_t^1 = (G')^{-1}\left(\frac{Y_t^1}{\lambda_t}\right) dt, \quad \varphi_0^1 = \varphi_{0-}^1. \tag{3.20}$$

If the volatility process σ of the risky asset is not given exogenously as in Chapter 2, it needs to be determined from the terminal condition $\mathfrak{S}$. By plugging expression (3.18) for the equilibrium return into the price dynamics (3.1), we obtain the following BSDE, which is coupled to the forward-backward system (3.19-3.20):

$$\begin{aligned} dS_t &= \frac{1}{2}\left[(\gamma^2 s + (\gamma^1 - \gamma^2)\varphi_t^1)\sigma_t + (\gamma^1 \xi_t^1 + \gamma^2 \xi_t^2)\right]\sigma_t^\top dt + \sigma_t dW_t, \\ S_T &= \mathfrak{S}. \end{aligned} \tag{3.21}$$

This is again the same equation as for quadratic costs (Herdegen et al., 2019). In particular, if both agents' risk aversions coincide ($\gamma^1 = \gamma^2 = 2\bar{\gamma}$), then (3.18) becomes

$$\mu_t = \frac{1}{2}\left[\gamma^2 s \sigma_t + (\gamma^1 \xi_t^1 + \gamma^2 \xi_t^2)\right]\sigma_t^\top = \bar{\gamma}\left[s\sigma_t + \xi_t^1 + \xi_t^2\right]\sigma_t^\top.$$

Therefore, the price process decouples from the forward-backward system (3.19-3.20) and leads to the same equilibrium dynamics as without transaction costs. For heterogenous but sufficiently similar risk aversions $\gamma^1 \approx \gamma^2$ and quadratic costs, it is shown in Herdegen et al. (2019) that a solution of (3.19-3.21) exists and identifies an equilibrium with transaction costs. However, the proof crucially exploits the fact that, with quadratic costs, the forward-backward system (3.19-3.20) for a given volatility process $(\sigma_t)_{t\in[0,T]}$ can be studied by means of the stochastic Riccati equation from Kohlmann and Tang (2002). Establishing such results for more general trading costs – where such tools are not available – is a challenging direction for further research.

3.4.3 Reparametrization

In applications, we are mainly interested in the changes induced by transaction costs relative to the frictionless version of the model. To facilitate both the numerical and the analytical analysis of the system in Section 3.4 and Chapter 4, respectively, we subtract the BSDE (3.6) for the frictionless equilibrium price from its frictional counterpart (3.21). This in turn leads to a backward equation with zero terminal condition for the change of the equilibrium price due to transaction costs:

$$d(\Delta S_t) := \Delta\mu_t dt + \Delta\sigma_t dW_t, \qquad \Delta S_T = 0, \tag{3.22}$$

where

$$\Delta\mu_t := \mu_t - \bar{\mu}_t, \qquad \Delta\sigma_t := \sigma_t - \bar{\sigma}_t. \tag{3.23}$$

To identify the terms in these dynamics that will be small for small transaction costs, write

$$\Delta\varphi_t^1 := \varphi_t^1 - \bar{\varphi}_t^1 \tag{3.24}$$

for the difference between the frictional and frictionless equilibrium positions of agent 1, which we expect to vanish as the transaction costs tend to zero. In view of the forward equation (3.20) and the strategy $\bar{\varphi}^1$ of the frictionless equilibrium strategy, this process has dynamics

$$d(\Delta\varphi_t^1) = (G')^{-1}\left(\frac{Y_t^1}{\lambda_t}\right)dt - d\bar{\varphi}_t^1, \quad \Delta\varphi_0^1 = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}. \tag{3.25}$$

With this notation and by plugging in the frictionless equilibrium return (3.8) and the equilibrium position (3.7), the drift term $\Delta\mu_t$ of ΔS_t from (3.22) can be rewritten as follows after some algebraic manipulations:

$$\begin{aligned}\Delta\mu_t &= \frac{1}{2}\left(2\Delta\sigma_t\bar{\sigma}_t^\top + \Delta\sigma_t\Delta\sigma_t^\top\right)\left(2\bar{\gamma}s + \frac{\gamma^1-\gamma^2}{\gamma^1+\gamma^2}\frac{(\gamma^2\xi_t^2-\gamma^1\xi_t^1)\bar{\sigma}_t^\top}{\bar{\sigma}_t\bar{\sigma}_t^\top}\right)\\ &\quad + \frac{1}{2}(\gamma^1\xi_t^1+\gamma^2\xi_t^2)\Delta\sigma_t^\top + \frac{\gamma^1-\gamma^2}{2}\sigma_t\sigma_t^\top\Delta\varphi_t^1.\end{aligned}$$

Next, observe that the drift rate $\mu_t^{Y^1}$ of the marginal trading cost Y_t^1 from (3.19) can be rewritten as

$$\begin{aligned}\mu_t^{Y^1} &= \frac{\gamma^1+\gamma^2}{2}\sigma_t\sigma_t^\top\Delta\varphi_t^1 + \frac{\gamma^2\xi_t^2-\gamma^1\xi_t^1}{2}\left(\frac{\sigma_t\sigma_t^\top}{\bar{\sigma}_t\bar{\sigma}_t^\top}\bar{\sigma}_t^\top - \sigma_t^\top\right)\\ &= \gamma\sigma_t\sigma_t^\top\Delta\varphi_t^1 + \frac{\gamma^2\xi_t^2-\gamma^1\xi_t^1}{2}\left(\frac{2\bar{\sigma}_t\Delta\sigma_t^\top}{\bar{\sigma}_t\bar{\sigma}_t^\top}\bar{\sigma}_t^\top + \frac{\Delta\sigma_t\Delta\sigma_t^\top}{\bar{\sigma}_t\bar{\sigma}_t^\top}\bar{\sigma}_t^\top - \Delta\sigma_t^\top\right).\end{aligned} \tag{3.26}$$

After further rearrangement, it follows that

$$\Delta\mu_t = \frac{\gamma^1-\gamma^2}{\gamma^1+\gamma^2}\mu_t^{Y^1} + \bar{\gamma}\left(2s\bar{\sigma}_t + s\Delta\sigma_t + \xi_t^1 + \xi_t^2\right)\Delta\sigma_t^\top.$$

Accordingly, the process

$$\mathcal{Y}_t := \Delta S_t - \widehat{\gamma}Y_t^1, \quad \mathcal{Z}_t := \Delta\sigma_t - \widehat{\gamma}Z_t^1, \quad \text{with } \widehat{\gamma} := \frac{\gamma^1-\gamma^2}{\gamma^1+\gamma^2} \tag{3.27}$$

solves the BSDE

$$\begin{aligned}d\mathcal{Y}_t &= \left[\bar{\gamma}\left(2s\bar{\sigma}_t + s\Delta\sigma_t + \xi_t^1 + \xi_t^2\right)\Delta\sigma_t^\top\right]dt + \left(\Delta\sigma_t - \widehat{\gamma}Z_t^1\right)dW_t\\ &=: \left[\bar{\gamma}\left(2s\bar{\sigma}_t + s\left(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\right) + \xi_t^1 + \xi_t^2\right)\left(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\right)^\top\right]dt + \mathcal{Z}_t dW_t,\\ \mathcal{Y}_T &= 0.\end{aligned} \tag{3.28}$$

In view of (3.26), the corresponding BSDE (3.19) for the marginal trading cost Y^1_t of agent 1 can in turn be rewritten as

$$
\begin{aligned}
dY^1_t = \Bigg[& \frac{\gamma^2\xi^2_t - \gamma^1\xi^1_t}{2}\left(\frac{(2\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z^1_t)(\mathcal{Z}_t + \widehat{\gamma}Z^1_t)^\top \bar{\sigma}_t}{\bar{\sigma}_t\bar{\sigma}_t^\top} - \mathcal{Z}_t - \widehat{\gamma}Z^1_t\right)^\top \\
& + \gamma(\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z^1_t)(\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z^1_t)^\top \Delta\varphi^1_t\Bigg] dt + Z^1_t dW_t, \\
Y^1_T = 0. &
\end{aligned} \tag{3.29}
$$

The two backward equations (3.28), (3.29) then again form an autonomous forward-backward system together with the forward equation (3.25) for the deviation $\Delta\varphi^1_t$ of agent 1's position from its frictionless counterpart.

In the following proposition, we summarize the above discussion leading to the correspondence between solutions of the FBSDE system and Radner equilibria with transaction costs:

Proposition 3.6. *Suppose the assumptions in Proposition 3.3 are satisfied and let $(\bar{S}, \bar{\sigma})$ as well as $\bar{\varphi}^1$ be the corresponding frictionless (Radner) equilibrium price and frictionless optimal strategy for agent 1. Suppose there exists a solution $(\Delta\varphi^1, \mathcal{Y}, Y^1, \mathcal{Z}, Z^1) \in (\mathbb{H}^2)^5$ of the FBSDE system (3.25), (3.28), (3.29) such that $(\bar{\varphi}^1 + \Delta\varphi^1)(\bar{\sigma} + \mathcal{Z} + \widehat{\gamma}Z^1) \in \mathbb{H}^2$, and*

$$
\mathbb{E}\left[\int_0^T \lambda_t G\left((G')^{-1}\left(\frac{Y^1_t}{\lambda_t}\right)\right)\right] < \infty.
$$

Then

$$
S_t = \bar{S}_t + \mathcal{Y}_t + \widehat{\gamma}Y^1_t, \quad t \in [0, T],
$$

is an equilibrium price with transaction costs. The agents' optimal trading rates are

$$
\dot{\varphi}^1_t = -\dot{\varphi}^2_t = (G')^{-1}\left(\frac{Y^1_t}{\lambda_t}\right), \quad t \in [0, T], \tag{3.30}
$$

and the corresponding optimal positions are given by

$$
\varphi^1_t = s - \varphi^2_t = \bar{\varphi}^1_t + \Delta\varphi^1_t, \quad t \in [0, T].
$$

As already emphasized above, a general existence proof for the FBSDE system (3.25), (3.28), (3.29) remains a challenging open problem. For a specific model with quadratic costs, we report some first global existence results in Chapter 4. Here, let us just briefly sketch how the nonlinear FBSDE (3.25), (3.28), (3.29) reduces to a nonlinear ODE in the context of Section 2.3, where:

(i) the volatilities $\bar{\sigma} = \sigma > 0$ in the models with and without transaction costs are given by the same exogenous constant (and no terminal condition is imposed);

(ii) the endowment volatilities $\xi_t^n = \xi^n W_t, n = 1,2$ follow Brownian motions;

(iii) the cost parameter is constant ($\Lambda = 1$).

When the volatility process is exogenous and the same with and without trading costs, we have $\Delta\sigma = 0$ in (3.22). As a consequence, (3.28) directly yields $\mathcal{Y}_t = 0$ and $\mathcal{Z}_t = 0$ for $t \in [0,T]$. In view of (3.26), the drift rate of the marginal costs Y^1 for agent 1 is

$$\mu_t^Y = \frac{(\gamma_1 + \gamma_2)\sigma^2}{2}\Delta\varphi_t^1.$$

For the specific endowment volatilities we consider here, the frictionless strategy for agent 1 is a Brownian motion, hence

$$d\bar{\varphi}_t^1 = \frac{\gamma^2\xi^2 - \gamma^1\xi^1}{(\gamma^1+\gamma^2)\sigma}dW_t.$$

The forward-backward system (3.25), (3.29) in turn becomes autonomous,

$$\begin{aligned} d\Delta\varphi_t^1 &= (G')^{-1}\left(\frac{Y_t^1}{\lambda}\right)dt + \frac{\gamma^1\xi^1 - \gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}dW_t, \qquad && \Delta\varphi_0^1 = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \\ dY_t^1 &= \frac{(\gamma_1+\gamma_2)\sigma^2}{2}\Delta\varphi_t^1 dt + Z_t^1 dW_t, && Y_T^1 = 0. \end{aligned} \tag{3.31}$$

Now use the standard ansatz that the backward component Y_t^1 should be a function $g(t,\Delta\varphi_t^1)$ of time and the forward component. Itô's formula and the dynamics of the forward component in turn yield [3]:

$$\begin{aligned} &dY_t^1 \\ &= \left(\partial_t + (G')^{-1}\left(\frac{g}{\lambda}\right)\partial_x + \frac{1}{2}\left(\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}\right)^2\partial_{xx}\right)g dt + \partial_x g\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}dW_t. \end{aligned}$$

Comparing the drift rate to the BSDE (3.31), we therefore obtain the following semilinear PDE for the function $g(t,x)$:

$$\begin{cases} \partial_t g + \partial_x g\ (G')^{-1}\left(\frac{g}{\lambda}\right) + \frac{1}{2}\left(\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}\right)^2\partial_{xx}g = \frac{(\gamma_1+\gamma_2)\sigma^2}{2}x, \\ g(T,x) = 0. \end{cases} \tag{3.32}$$

If the time horizon T is large, then the solution should become stationary ($\partial_t g(t,x) \approx 0$). Such a stationary solution should in turn solve the nonlinear ODE from Lemma 2.5 in Chapter 2:

$$g'(x)(G')^{-1}\left(\frac{g(x)}{\lambda}\right) + \frac{1}{2}\left(\frac{\gamma^1\xi^1-\gamma^2\xi^2}{(\gamma^1+\gamma^2)\sigma}\right)^2 g''(x) = \frac{(\gamma^1+\gamma^2)\sigma^2}{2}x. \tag{2.7}$$

[3] we drop the argument $(t,\Delta\varphi_t^1)$ to ease notation

For finite time horizons T, the PDE (3.32) cannot be reduced to an ODE. Far from the terminal time T, it is natural to expect that the correct solution is still identified by the same growth condition in the space variable as for the ODE 2.5 in Chapter 2. For the numerical solution of the ODE, this growth condition also serves as a boundary condition that is approximately correct for large values of the space variable. For the PDE, however, it becomes incompatible with the zero terminal condition at maturity, which describes that trading slows down and eventually stops near the terminal time. For more general versions of the model, even the stationary boundary conditions in the space dimensions are not readily available and it is not clear how to paste them together with the terminal condition. Accordingly, it is not straightforward to solve the PDE (3.32) and its extensions using finite-difference schemes.

As a remedy, in the next section we therefore propose a numerical algorithm in the spirit of Han et al. (2018). It solves the FBSDE by simulation and therefore bypasses the need to identify the correct boundary conditions. The algorithm approximates the dependence of the backward component on the forward components by a deep neural network. Whence, it is also able to handle higher-dimensional settings, e.g., with endogenous volatilities or random and time-varying transaction costs, provided that the time horizon of the model is sufficiently short.

3.5 Numerics

We now present a numerical algorithm to solve the FBSDEs from Section 3.4. The algorithm is then tested for the calibrated parameters from Section 2.5. To wit, we compare our numerical results to the analytical solutions from Chapter 2 as well as the ones from Herdegen et al. (2019) and Chapter 4. Finally, we also apply our numerical algorithm to a model with nonlinear costs and endogenous volatility, for which no benchmarks are available.

3.5.1 Deep-Learning Algorithm

Overview Solving the FBSDE system (3.25), (3.28), (3.29) is challenging because it is multidimensional and the forward and backward components are fully coupled. Nevertheless, it is amenable to the simulation-based approach of Han et al. (2018), which approximates the solution by a deep neural network. In Han et al. (2018) the focus lies on BSDEs, but the approach can readily be extended to FBSDEs, compare Gonon et al. (2019); Han and Long (2018).

We now briefly sketch the idea underlying this approach. The main observation is the following: the dynamics of FBSDE system (3.25) (3.28) (3.29) are pinned down if the volatility $(\mathcal{Z}, Z^1)$ of the backward component $(\mathcal{Y}, Y^1)$ is specified. Together with a guess for the initial values $(\mathcal{Y}_0, Y_0^1)$ of the backward components, we can then simulate the system with a standard forward scheme and check whether it satisfies the terminal conditions. Searching for the correct initial values is relatively straightforward. The more difficult challenge is how to parametrize the controls $(\mathcal{Z}, Z^1)$ – which are functions of time and the forward processes – and update the corresponding initial guesses until the terminal condition is matched sufficiently well. To this end, Han et al. (2018) consider an equidistant partition of the time interval into N subintervals. At each time point $t_m = mT/N$, they then parametrize $(\mathcal{Z}_{t_m}, Z^1_{t_m})$ with a (shallow) network structure F^{θ_m}, where the time t_m value of the forward components are the inputs for the network. Together with the guess for the initial values $(\mathcal{Y}_0, Y^1_0)$ and a time-discretization of the FBSDE system (3.25), (3.28), (3.29) via the Euler scheme (for example), the system can then be simulated forward in time. In this way, the shallow network structures are concatenated over time and becomes a deep neural network architecture, since the number of layer of the architecture and the number of parameters in the networks grows linearly in the number of discretization steps N. In other words, we input simulated Brownian paths into the deep network structure, and it outputs the terminal values of the backward components.

Our task is to update the parameters $\{\mathcal{Y}_0, Y^1_0, \theta_m, m = 0, \dots, N\}$ until the terminal conditions $\mathcal{Y}_T = 0 = Y^1_T$ are matched sufficiently well. This iterative update of the network parameters is the essence of *machine learning* tasks. State-of-the-art performance is achieved through backpropagation and stochastic gradient descent-type algorithms, see Goodfellow et al. (2016). [4]

Algorithm Let us now describe the machine learning algorithm in more detail. We fix a time partition $0 = t_0 < \ldots < t_N = T$, where $t_m = mT/N$ and $\Delta t = T/N$. Let $(\Delta W_m)_{m=0}^N$ be iid normally distributed random variables with mean zero and variance Δt. At time t_m, we have $\xi^1_{t_m}, \xi^2_{t_m}$ from the market, and the frictionless volatility $\bar{\sigma}_{t_m}$ and the increment of the frictionless strategy $\Delta\bar{\varphi}^1_{t_m}$ of agent 1 available. Together with $(\Delta W_{t_m}, W_{t_m}, \Delta\varphi^1_{t_m}, \mathcal{Z}_{t_m}, Z^1_{t_m})$ as input, the discrete-time analogue of the forward update rule for the FBSDE system (3.25), (3.28), (3.29) is:

$$\Delta\sigma_{t_m} = \mathcal{Z}_{t_m} + \widehat{\gamma} Z^1_{t_m}, \tag{3.33}$$

$$\mathcal{Y}_{t_{m+1}} = \mathcal{Y}_{t_m} + \bar{\gamma}\left(2s\bar{\sigma}_{t_m} + s\Delta\sigma_{t_m} + \xi^1_{t_m} + \xi^2_{t_m}\right)\Delta\sigma^\top_{t_m}\Delta t + \mathcal{Z}_{t_m}\Delta W_{t_m}, \tag{3.34}$$

$$\begin{aligned} Y^1_{t_{m+1}} = Y^1_{t_m} &+ \frac{\gamma^2\xi^2_{t_m} - \gamma^1\xi^1_{t_m}}{2}\left(\frac{(2\bar{\sigma}_{t_m} + \Delta\sigma_{t_m})\Delta\sigma^\top_{t_m}\bar{\sigma}_{t_m}}{\bar{\sigma}_{t_m}\bar{\sigma}^\top_{t_m}} - \Delta\sigma_{t_m}\right)^\top \Delta t \\ &+ \gamma(\bar{\sigma}_t + \Delta\sigma_{t_m})(\bar{\sigma}_t + \Delta\sigma_{t_m})^\top\Delta\varphi^1_{t_m}\Delta t + Z^1_{t_m}\Delta W_{t_m}, \end{aligned} \tag{3.35}$$

$$\Delta\varphi^1_{t_{m+1}} = \Delta\varphi^1_{t_m} + (G')^{-1}\left(\frac{Y^1_{t_m}}{\lambda_{t_m}}\right)\Delta t - \Delta\bar{\varphi}^1_{t_m}, \tag{3.36}$$

$$W_{t_{m+1}} = W_{t_m} + \Delta W_{t_m}. \tag{3.37}$$

[4]Recently, Raghu et al. (2017); Allen-Zhu et al. (2018); Du et al. (2018) show that if the network is deep enough (which in our case if the time discretization is fine enough), first-order optimization can find global minima under some regularity conditions.

Remark 3.7. Recall that there is no assumption on the dynamic of the market. Here, however, we need to mildly assume that we can simulate the endowment volatilities ξ^1 and ξ^2.

Now, we focus on the parametrization of $(\mathcal{Z}_{t_m}, Z^1_{t_m})$, $m = 0, \ldots, N$ within a function class $\{F^\theta : \theta \in \Theta\}$. A very popular class of functions in the machine learning community is the "singular activation function Rectified Linear Unit" ($ReLU$): $ReLu(x) = \max\{x, 0\}$. A popular class of approximation function F is the convolution of linear functions with $ReLu$ activations. For the numerical experiments in Section 3.5.2, at each time t_m, the F^{θ_m} we use is a neural network with one hidden layer and $N = 15$ hidden units, that is

$$F^{\theta_m}(x) = w^2_{\theta_m}\left(ReLu\left(w^1_{\theta_m}x + b^1_{\theta_m}\right)\right) + b^2_{\theta_m}.$$

Recall that the Brownian motion W is a d-dim process, the forward component $\Delta\varphi$ is 1-dim, $\mathcal{Z}_{t_m}$ and $Z^1_{t_m}$ are both d-dim vectors. $w^1_{\theta_m} \in \mathbb{R}^{N\times(d+1)}$ and $b^1_{\theta_m} \in \mathbb{R}^N$ are called the input *weights* and *biases* of the network, whereas $w^1_{\theta_m} \in \mathbb{R}^{2d\times N}$ and $b^1_{\theta_m} \in \mathbb{R}^{2d}$ the output *weights* and *biases* of the network. We let y_0 denote the initial guess for $y_0 = (\mathcal{Y}_0, Y^1_0)$, $\theta_m = (w^1_{\theta_m}, b^1_{\theta_m}, w^2_{\theta_m}, b^2_{\theta_m})$, and summarize the forward update procedure in Algorithm 1:

Algorithm 1: Forward Update

Data: $(x_0, \Delta\varphi^1_{0-})$, `batch_size` sample paths $\Delta W = (\Delta W_m)_{m=0}^N$;

Initialization: $k = 0$, $W = 0$, $\Delta{\varphi^1}^\theta = \Delta\varphi^1_{0-}$, $(\mathcal{Y}^\theta, {Y^1}^\theta) = y_0$;

while $k \leq N$ **do**

 $\Delta W = \Delta W_k$;$(\mathcal{Z}^\theta, {Z^1}^\theta) = F^{\theta_k}(W, \Delta{\varphi^1}^\theta)$;

 $\mathcal{Y}^\theta, {Y^1}^\theta, \Delta{\varphi^1}^\theta, W =$ update rule (3.33-3.37) $(\Delta W, W, \Delta{\varphi^1}^\theta, \mathcal{Z}^\theta, {Z^1}^\theta)$;

 $k{+}{+}$;

end

return: $(\mathcal{Y}^\theta_T, {Y^1}^\theta_T)$

For an arbitrary choice of the parameters $\theta := \{y_0, \theta_m, m = 0, \ldots, N\}$, the output $(\mathcal{Y}^\theta_T, {Y^1}^\theta_T)$ of the algorithm may not be anywhere close to the terminal condition $(0, 0)$. In order to find a set of parameters $\{y_0, \theta_m, m = 0, \ldots, N\}$ for which the terminal condition is matched sufficiently well, we introduce the loss function `Loss` and re-formulate the task into the following optimization problem:

$$\min_{\{y_0, \theta_m, m=0,\ldots,N\}} \texttt{Loss} := \frac{1}{\texttt{batch\_size}}\left[\|\mathcal{Y}^\theta_T\|^2 + \|{Y^1}^\theta_T\|^2\right]. \tag{3.38}$$

The minimization problem (3.38) can be tackled using "`Adam`" (Kingma and Ba, 2014), which is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. In our case, the "stochastic" comes from the random sampling in the nature of our learning task. The first-order gradient-based optimization is an extension of classical gradient descent algorithm to non-differentiable functions with subgradient, cf. Shor (2012) for more details. Starting from an initial guess $\theta^{(0)}$, the algorithm iteratively updates our target functional `Loss`

$$\theta^{(j+1)} = \theta^{(j)} - \eta_j \partial \texttt{Loss}(\theta^{(j)}), \tag{3.39}$$

where ∂`Loss` represent the subgradient of `Loss`. To optimize the performance, we adaptively choose a sequence of decreasing learning rates $\eta_j > 0$ based on adaptive estimates of lower-order moments of $\theta^{(j)}$. However, as is apparent from (3.34 - 3.37), the dependence of the solution on the parameter θ is complex, since the state variables and parametric functions are iteratively added, multiplied and composed. For instance, ${Y^1}^\theta_{t_{m+1}}$ depends not only on θ_m, but also (via $\Delta{\varphi^1}^\theta_{t_m}$) on $y_0, \theta_0, \ldots, \theta_{m-1}$. This makes the computational solution of (3.38) by classical numerical techniques highly challenging.

Thanks to the compositional structure of neural networks, one can instead apply the chain rule to obtain the subgradient $\partial\texttt{Loss}(\theta)$. The parameters can then be updated efficiently using the so-called backpropagation algorithm, see, e.g., Goodfellow et al. (2016). Then, $(\Delta{\varphi^1}^\theta_{t_m}, \mathcal{Y}^\theta_{t_m}, {Y^1}^\theta_{t_m}, \mathcal{Z}^\theta_{t_m}, {Z^1}^\theta_{t_m})_{m=0}^N$ can be viewed as the outputs of a deep neural network with random input $(\Delta W_{t_m})_{m=0}^N$. Finally, all of this can be implemented with 10 lines of code in the computational graph structure employed in Python libraries such as Torch or TensorFlow.

In summary, the learning algorithm iteratively updates the network parameters as follows until a desired `error_bound` is reached:

Algorithm 2: Training Procedure

Data: $(x_0, \Delta\varphi^1_{0-})$, and $(\mathcal{Y}_T, Y^1_T) = (0, 0)$;
Initialization: $k = 0$, initialization of parameters $\{y_0, \theta_k, k = 0, \ldots, N\}$;
while *Loss* $\geq$ *error_bound* **do**
  sample `batch_size` of iid ΔW;
  $(\mathcal{Y}^\theta_T, {Y^1}^\theta_T)$ = output of Algorithm 1 with input $(x_0, \Delta\varphi^1_{0-})$, ΔW;
  `Loss` $= \left[\|\mathcal{Y}_T\|^2 + \|Y^1_T\|^2\right]$ /`batch_size`;
  Calculate the gradient of `Loss`;
  Back propagate updates for $\{y_0, \theta_m, m = 0, \ldots, N\}$ via `Adam`;
end
return: (local) optimizer $\theta^* = \{y^*_0, \theta^*_m, m = 0, \ldots, N\}$.

3.5.2 Numerical results

The algorithm introduced in Section 3.5.1 is now applied to solve the FBSDE system corresponding to equilibria with transaction costs. Here we are using 252 (trading) days for a year, so we have 21 (trading) days for a month.

We first consider the simplest version (3.31) of the model where the volatility is exogenous as a sanity check. In this setting, we compare the numerical solution to the nonlinear ODE from Chapter 2 that describes the exact solution of the infinite-horizon version of the model.

Subsequently, we consider the model with endogenous volatility. In order to test the performance of the learning algorithm in this case, we compare its results to the semi-explicit solution in terms of Riccati equations obtained for quadratic costs in Herdegen et al. (2019) and Chapter 4.

Exogenous volatility As a sanity check, we first consider the finite-horizon version of the model from Section 2.3 with power transaction costs of 3/2, i.e. $\lambda_{3/2}G_{3/2}(x) = 2\lambda_{3/2}|x|^{3/2}/3$, initial positions $\Delta\varphi_0^1 = 0 = \Delta\varphi_0^2$ and model parameters as calibrated in Section 2.5. The algorithm described in Section 3.5.1 can be readily adapted to the forward-backward system (3.31) with one less equation and exogenous volatility, i.e. $\Delta\sigma = 0$. The output of the network is $(\Delta{\varphi^1}^{\theta^*}_{t_m}, {Y^1}^{\theta^*}_{t_m}, {Z^1}^{\theta^*}_{t_m})_{m=0}^N$.

We sample 10000 Brownian paths ΔW and input these into the deep neural network with learned parameter θ^*, and compare the output paths $(\Delta{\varphi^1}^{\theta^*}_{t_m}, {Y^1}^{\theta^*}_{t_m}, {Z^1}^{\theta^*}_{t_m})_{m=0}^N$ to the long-run optimal trading rate from Theorem 2.9, where g is given by the solution of the nonlinear ODE from Lemma 2.5. Figure 3.1 shows the graph of both functions at 30 days before, 10 days before and 1 day before maturity T, i.e., the scatter plot of $(\Delta{\varphi^1}^{\theta^*}_{t_m}, (G_q')^{-1}({Y^1}^{\theta^*}_{t_m}/\lambda_q))$. We observe that the numerical-solution of the finite-horizon problem is already close to the long-run optimum, even for a time horizon as short as a month and a half.

Endogenous volatility We now turn to the model with endogenous volatility from Section 3.4, where the terminal liquidating condition as well as the frictionless equilibrium benchmark are given as in Example 3.5. We consider $\lambda_q G_q(x) = \lambda_q|x|^q/q$ both for $q = 2$ (quadratic costs) and $q = 3/2$ (power costs). For λ_q, γ_1, γ_2, and $\xi^1 = -\xi^2 = \xi_q$ we use the same parameter values as for the model with exogenous volatility (cf. Section 2.5) and we also again set $\Delta\varphi_0^1 = 0 = \Delta\varphi_0^2$. The additional parameters α and β are calibrated to the frictionless equilibrium from Section 3.3. To wit, α is estimated from the time series (resulting in the same value as for σ in Section 2.5.1) and β is chosen so that $\bar{S}_0 = (\beta - s\bar{\gamma}\alpha^2)T$ matches the current stock price. We focus on a short time horizon for $T = 21$ trading days discretized into $n = 168$ time steps. To produce Figure 3.2 and Figure 3.3, we simulate a Brownian sample path ΔW, and put the same Brownian increments ΔW into the FBSDE system with the learned parameters θ_2^* and $\theta_{3/2}^*$ for quadratic costs and 3/2-power costs respectively.

The deep-learning algorithm from Section 3.5.1 in turn yields an approximate solution of the FBSDE system (3.25), (3.28), (3.29). To assess the effect of different transaction costs, we compare the equilibrium price and volatility to the respective quantity in the frictionless equilibrium, i.e. we examine (sample paths of) the price difference $\Delta S^{\theta^*} = \mathcal{Y}^{\theta^*} + \widehat{\gamma}{Y^1}^{\theta^*}$ and the volatility

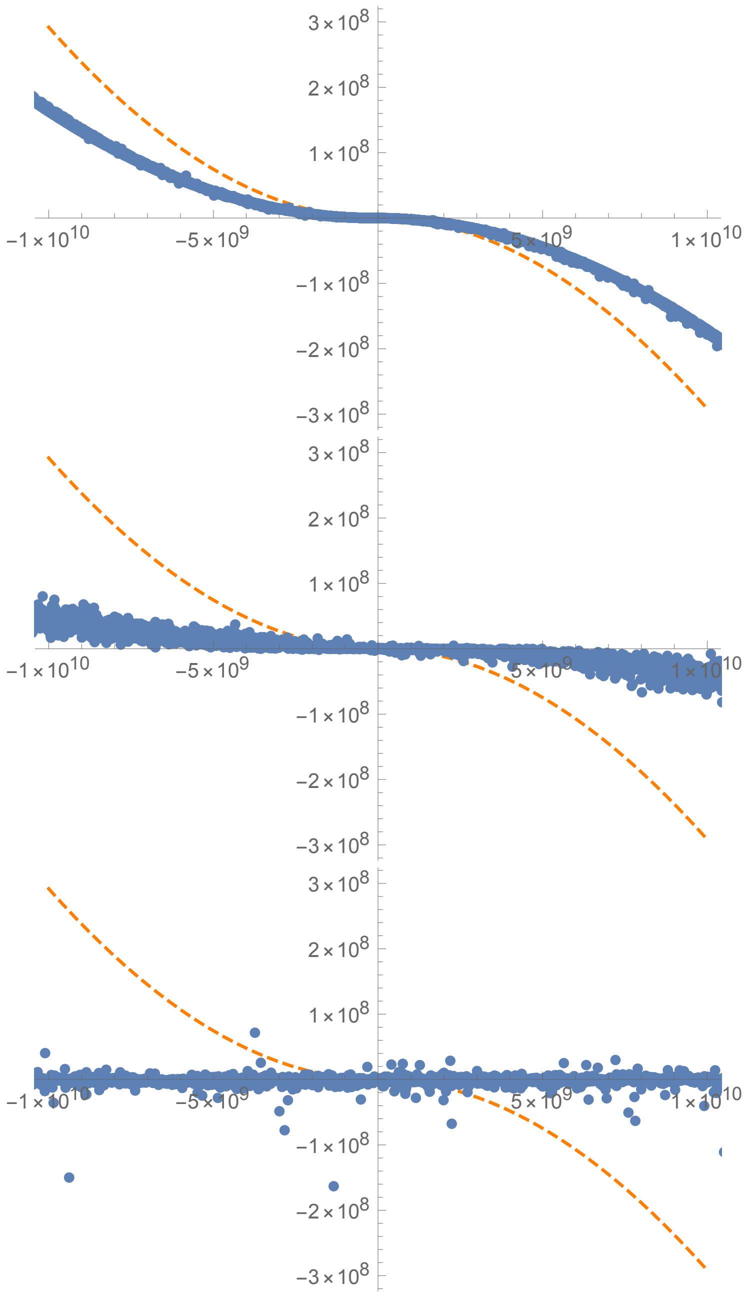

Figure 3.1: Comparison of the long-run optimal trading rate to the neural-network approximation of its finite horizon counterpart for power costs with $q = 1.5$ and 30-day before maturity (upper panel), 10-day before maturity and 1-day before maturity (lower panel).

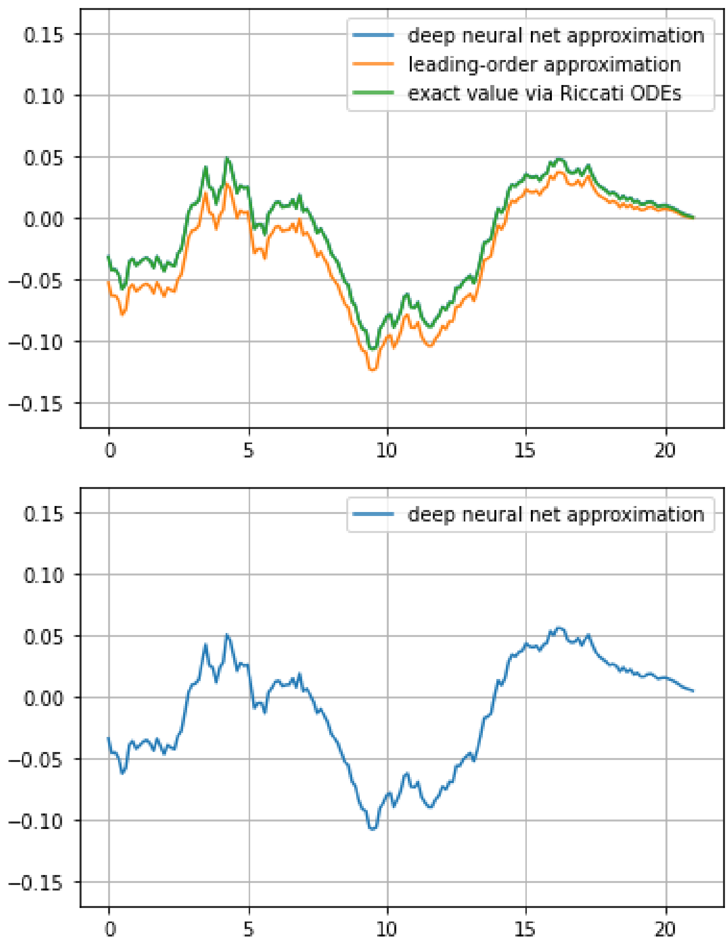

Figure 3.2: Difference between simulated frictional and frictionless equilibrium prices changes with calibrated parameters for quadratic costs (upper panel) and 3/2-costs (lower panel).

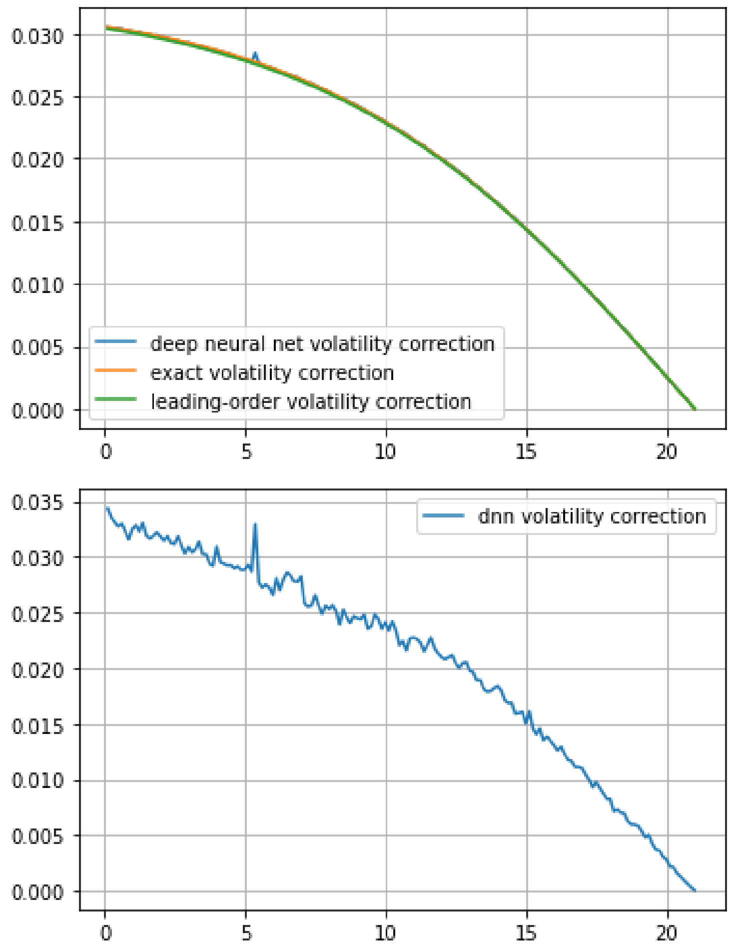

Figure 3.3: Difference between simulated frictional and frictionless equilibrium volatilities with calibrated parameters for quadratic costs (upper panel) and 3/2-costs (lower panel).

difference $\Delta\sigma^{\theta^*} = \mathcal{Z}^{\theta^*} + \widehat{\gamma} Z^{1^{\theta^*}}$ over time. For quadratic costs it has been shown in Herdegen et al. (2019) that $\Delta\varphi^1$, $\mathcal{Y}$ and Y^1 can be described in terms of a system of coupled Riccati ODEs; in Chapter 4, we show that a solution exists for all parameter configurations of the model. This provides a benchmark in the case of quadratic costs. The upper panels in Figure 3.2 and Figure 3.3 show one sample path of the price and volatility differences for quadratic costs calculated by both methods, i.e., by applying the neural network based algorithm described above and by solving the system of ODEs in Section 4.3 with a standard ODE solver. The neural-network based method provides an accurate approximation of the equilibrium quantities for the short time horizons considered here.

The analogous plots for power costs with $q = 3/2$ are shown in the lower panels of Figure 3.2 and Figure 3.3 (in order to compare these to the corresponding results for quadratic costs, we use the same Brownian noise in each case). Note that no benchmark is available in this case. The equilibrium prices for the two cost specifications turn out to be quite similar. This corroborates the findings from Section 2.5 and suggests that quadratic costs can also serve as useful proxies for other less tractable costs specifications in settings with endogenous volatilities.

3.6 Equilibrium with Proportional Costs

The discussion above has focused on convex trading costs on the agents trading rates. Similarly as in the model with exogenous volatilities studied in Chapter 2, proportional costs can viewed as a singular limit of the corresponding regular control problems. Alternatively, they can be analyzed directly using techniques from singular control. Here, we briefly outline how this formally leads to a correspondence between equilibria with transaction costs and FBSDEs with instantaneous reflection.

3.6.1 Model

As in Section 2.4, in models with proportional costs it is natural to consider general finite-variation strategies, which can be decomposed in the cumulative numbers of shares purchased and sold:

$$\psi_t = \varphi^n_{0-} + \psi^{\uparrow}_t - \psi^{\downarrow}_t.$$

With proportional costs on the number of shares transacted, the mean-variance goal functional for agent n then is

$$J^n_T(\psi) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\|\psi_t\sigma_t + \xi^n_t\|^2\right) dt - \lambda_t(d\psi^{\uparrow}_t + d\psi^{\downarrow}_t)\right] \qquad (3.40)$$

To make this well defined, we focus on *admissible* strategies that satisfy the usual integrability condition $\psi\sigma \in \mathbb{H}^2$ with respect to the given price process

S, and have finite expected trading costs:

$$E\left[\int_0^T \lambda_t d\psi_t^{\uparrow} + \lambda_t d\psi_t^{\downarrow}\right] < \infty. \tag{3.41}$$

As in Section 3.4.2 the proportional cost $\lambda_t = \lambda\Lambda_t$ can be time-dependent and random here.

3.6.2 Individual Optimization

If a trading strategy φ^n is optimal for agent n, then for *any* perturbation η $J_T^n(\varphi^n + \eta) - J_T^n(\varphi^n) \le 0$ is necessarily nonpositive. Now let $\eta^{\perp}$ and $\eta^{\ll}$ be the mutually singular as well as absolutely continuous decomposition of η with respect to $\check{\varphi}^n = \varphi^{n,\uparrow} + \varphi^{n,\downarrow}$.

We first focus on absolutely continuous perturbation $\eta^{\ll}$. The optimally condition is equivalent to $\lim_{\rho\to 0}\left(J_T^n(\varphi^n + \rho\eta^{\ll}) - J_T^n(\varphi^n)\right)/\rho = 0$, i.e. the Gateaux derivative equal to 0, with the help of the following Lemma 3.8:

Lemma 3.8.

$$\lim_{\rho\to 0}\frac{1}{\rho}\left[\left(\varphi^{n,\uparrow} - \varphi^{n,\downarrow} + \rho\eta^{\ll}\right)_T^{\uparrow} - \varphi_T^{n,\uparrow}\right] = \int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\, d\varphi_t^{n,\uparrow},$$
$$\lim_{\rho\to 0}\frac{1}{\rho}\left[\left(\varphi^{n,\uparrow} - \varphi^{n,\downarrow} + \rho\eta^{\ll}\right)_T^{\downarrow} - \varphi_T^{n,\downarrow}\right] = -\int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\, d\varphi_t^{n,\downarrow}.$$

Proof. First notice

$$\eta_T^{\ll} = \int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\, d\check{\varphi}_t^n = \int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\, d\varphi_t^{n,\uparrow} + \int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\, d\varphi_t^{n,\downarrow}.$$

For $|\rho| < \varphi_T^{n,\uparrow}/(1+\check{\eta}_T^{\ll})$, notice

$$|\rho|\int_0^T \left|\frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\right| d\varphi_t^{n,\uparrow} = |\rho|\int_0^T \frac{d\check{\eta}_t^{\ll}}{d\check{\varphi}_t^n} d\varphi_t^{n,\uparrow} \le |\rho|\check{\eta}_T^{\ll} \le \varphi_T^{n,\uparrow},$$

and symmetrically for $|\rho| < \varphi_T^{n,\downarrow}/(1+\check{\eta}_T^{\ll})$,

$$|\rho|\int_0^T \left|\frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n}\right| d\varphi_t^{n,\downarrow} = |\rho|\int_0^T \frac{d\check{\eta}_t^{\ll}}{d\check{\varphi}_t^n} d\varphi_t^{n,\downarrow} \le |\rho|\check{\eta}_T^{\ll} \le \varphi_T^{n,\downarrow}.$$

Hence

$$\left(\varphi^{n,\uparrow} - \varphi^{n,\downarrow} + \rho\eta^{\ll}\right)_T^{\uparrow} = \varphi^{n,\uparrow} + \rho\int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n} d\varphi_t^{n,\uparrow},$$
$$\left(\varphi^{n,\uparrow} - \varphi^{n,\downarrow} + \rho\eta^{\ll}\right)_T^{\downarrow} = \left(\varphi^{n,\downarrow} - \varphi^{n,\uparrow} - \rho\eta^{\ll}\right)_T^{\uparrow} = \varphi^{n,\downarrow} - \rho\int_0^T \frac{d\eta_t^{\ll}}{d\check{\varphi}_t^n} d\varphi_t^{n,\downarrow}.$$

□

From Lemma 3.8, we have

$$
\begin{aligned}
&\lim_{\rho\to 0}\frac{1}{\rho}\left[\left(\varphi^{n,\uparrow}-\varphi^{n,\downarrow}+\rho\eta^{\ll}\right)^{\uparrow}_T+\left(\varphi^{n,\uparrow}-\varphi^{n,\downarrow}+\rho\eta^{\ll}\right)^{\downarrow}_T-\varphi^{n,\uparrow}_T-\varphi^{n,\downarrow}_T\right]\\
&=\int_0^T\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\varphi^{n,\uparrow}_t-\int_0^T\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\varphi^{n,\downarrow}_t\\
&=\int_0^T\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\varphi^{n}_t.
\end{aligned}
$$

We define the "margin utility" Y^n for agent n as:

$$
Y^n_t := \mathbb{E}_t\left[\int_t^T\left(\mu_u-\gamma^n\sigma_u(\sigma_u\varphi^n_u+\xi^n_u)^\top\right)du\right]. \tag{3.42}
$$

The Gateaux derivative for absolutely continuous perturbation $\eta^{\ll}$ is therefore:

$$
\begin{aligned}
0&=\lim_{\rho\to 0}\frac{1}{\rho}(J^n_T(\varphi^n+\rho\eta^{\ll})-J^n_T(\varphi^n))\\
&=\mathbb{E}\left[\int_0^T\left(\mu_t-\gamma^n\sigma_t(\sigma_t\varphi^n_t+\xi^n_t)^\top\right)\eta^{\ll}_t dt-\lambda_t\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\varphi^n_t\right]\\
&=\mathbb{E}\left[\int_0^T\int_t^T\left(\mu_u-\gamma^n\sigma_u(\sigma_u\varphi^n_u+\xi^n_u)^\top\right)du\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\check{\varphi}^n_t-\lambda_t\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,d\varphi^n_t\right]\\
&=\mathbb{E}\left[\int_0^T Y^n_t\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,\left(d\varphi^{n,\uparrow}_t+d\varphi^{n,\downarrow}_t\right)-\lambda_t\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\,\left(d\varphi^{n,\uparrow}_t-d\varphi^{n,\downarrow}_t\right)\right]\\
&=\mathbb{E}\left[\int_0^T\frac{d\eta^{\ll}_t}{d\check{\varphi}^n_t}\left[(Y^n_t-\lambda_t)\,d\varphi^{n,\uparrow}_t+(Y^n_t+\lambda_t)\,d\varphi^{n,\downarrow}_t\right]\right].
\end{aligned}
$$

Since the above has to hold for any absolutely continuous perturbation $\eta^{\ll}$, it follows that purchases and sales of the optimal strategy only happen on the sets $\{Y^n_t=\lambda_t\}$ and $\{Y^n_t=-\lambda_t\}$, respectively.

Now we look at the mutually singular perturbation $\eta^{\perp}$. Unfortunately, the Gateaux derivative does not exists, since we can calculate directly that

$$
\left(\varphi+\rho\eta^{\perp}\right)^{\uparrow}_T+\left(\varphi+\rho\eta^{\perp}\right)^{\downarrow}_T=\check{\varphi}_T+|\rho|\check{\eta}^{\perp}_T,
$$

and the derivative with respect to ρ does not exists. However, we can still get the bounds for Y^n, through direct calculation:

$$
\begin{aligned}
&J^n_T(\varphi^n+\eta^{\perp})-J^n_T(\varphi^n)\\
&=\mathbb{E}\left[\int_0^T\left(\mu_t-\gamma^n\sigma_t(\sigma_t\varphi^n_t+\xi^n_t+\frac{1}{2}\sigma_t\eta^{\perp}_t)^\top\right)\eta^{\perp}_t dt-\lambda_t d\check{\eta}^{\perp}_t\right]\\
&=\mathbb{E}\left[\int_0^T\int_t^T\left(\mu_u-\gamma^n\sigma_u(\sigma_u\varphi^n_u+\xi^n_u)^\top\right)du\,d\eta^{\perp}_t-\lambda_t d\check{\eta}^{\perp}_t\right]-\frac{1}{2}\mathbb{E}\left[\int_0^T\sigma_t\sigma_t^\top\left(\eta^{\perp}_t\right)^2dt\right]\\
&=\mathbb{E}\left[\int_0^T Y^n_t\,d\eta^{\perp}_t-\lambda_t d\check{\eta}^{\perp}_t\right]-\frac{1}{2}\mathbb{E}\left[\int_0^T\sigma_t\sigma_t^\top\left(\eta^{\perp}_t\right)^2dt\right]\\
&=\mathbb{E}\left[\int_0^T(Y^n_t-\lambda_t)\,d\eta^{\perp,\uparrow}_t-(Y^n_t+\lambda_t)d\eta^{\perp,\downarrow}_t\right]-\frac{1}{2}\mathbb{E}\left[\int_0^T\sigma_t\sigma_t^\top\left(\eta^{\perp}_t\right)^2dt\right].
\end{aligned}
$$

Since this inequality needs to hold for arbitrary perturbation $\eta^\perp$, we must have $Y^n_t \in [-\lambda_t, \lambda_t]$.

Together with the fact that purchases and sales of the optimal strategy only happen on the sets $\{Y^n_t = \lambda_t\}$ and $\{Y^n_t = -\lambda_t\}$, respectively,[5] we have

$$\int_0^T \mathbb{1}_{\{Y^n_t < \lambda_t\}} d\varphi^{n,\uparrow}_t = 0 = \int_0^T \mathbb{1}_{\{Y^n_t > -\lambda_t\}} d\varphi^{n,\downarrow}_t, \qquad \text{a.s.} \tag{3.43}$$

In summary, this discussion shows that the optimal strategy $\varphi^n = \varphi^n_{0-} + \varphi^{n,\uparrow} - \varphi^{n,\downarrow}$ and the process Y^n solve the Skorokhod problem:

$$\varphi^n_t = \varphi^n_{0-} + \int_0^T \mathbb{1}_{\{Y^n_t = \lambda_t\}} d\varphi^{n,\uparrow}_t - \int_0^T \mathbb{1}_{\{Y^n_t = -\lambda_t\}} d\varphi^{n,\downarrow}_t, \tag{3.44}$$

$$Y^n_t = -\int_t^T \left(\gamma^n(\sigma_t\varphi^n_t + \xi^n_t)\sigma^\top_t - \mu_t\right) dt + \int_t^T Z^n_t dW_t \in [-\lambda_t, \lambda_t], \tag{3.45}$$

where the martingale part Z^n need to be determined as part of the solution.

3.6.3 Equilibrium

We now concatenate the agents' individually optimal trading strategies to an equilibrium. To achieve market clearing, we impose that, in analogy to (3.17),

$$Y^1_t + Y^2_t = 0, \quad t \in [0, T]. \tag{3.46}$$

Indeed, together with (3.43), it then follows that $s - \varphi^1_t = \varphi^2_{0-} + \varphi^{1,\downarrow}_t - \varphi^{1,\uparrow}_t$ and Y^2 solve the Skorokhod problem characterizing agent 2's optimal strategy. Whence, the market clears and the clearing condition $\varphi^1_t + \varphi^2_t = s$ implies that the frictional equilibrium return has the same relationship to the frictional volatility and the agents' optimal positions as for superlinear costs in (3.18) and in the frictionless case (3.8):

$$\mu_t = \frac{1}{2}\left[(\gamma^2 s + (\gamma^1 - \gamma^2)\varphi^1_t)\sigma_t + (\gamma^1\xi^1_t + \gamma^2\xi^2_t)\right]\sigma^\top_t.$$

Plugging this expression back into agent 1's optimality condition (3.45) and the price dynamics (3.1) in turn yields the same backward equations as in (3.19) and (3.21). The difference is that the forward equation (3.44) is completelty different from its counterpart (3.20) for superlinear costs, in that the agents' equilibrium positions with proportional costs are not adjusted at an absolutely continuous rate but only on a singular set. With the same reparametrizations

[5] The process Y^n_t describes the marginal benefit of future expected returns and risk reductions, and trades happen when this process becomes large enough to match the (constant) marginal cost – the direct analogue of the classical gradient constraints in dynamic programming equations for singular control problems as in Davis and Norman (1990); Shreve and Soner (1994).

as in Section 3.4.3, this FBSDE system with reflection can be rewritten as follows:

$$
\begin{aligned}
d\Delta\varphi_t^1 &= \mathbb{1}_{\{Y_t^1=\lambda_t\}} d\varphi_t^{1,\uparrow} - \mathbb{1}_{\{Y_t^1=-\lambda_t\}} d\varphi_t^{1,\downarrow} - d\bar{\varphi}_t^1, \\
&\qquad \Delta\varphi_{0-}^1 = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \qquad (3.47) \\
d\mathcal{Y}_t &= \left[\bar{\gamma}\left(2s\bar{\sigma}_t + s\left(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\right) + \xi_t^1 + \xi_t^2\right)\left(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\right)^\top\right] dt + \mathcal{Z}_t dW_t, \\
&\qquad \mathcal{Y}_T = 0. \qquad (3.28) \\
dY_t^1 &= \left[\frac{\gamma^2\xi_t^2 - \gamma^1\xi_t^1}{2}\left(\frac{(2\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z_t^1)(\mathcal{Z}_t + \widehat{\gamma}Z_t^1)^\top\bar{\sigma}_t}{\bar{\sigma}_t\bar{\sigma}_t^\top} - \mathcal{Z}_t - \widehat{\gamma}Z_t^1\right)^\top \right. \\
&\qquad \left. + \gamma(\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z_t^1)(\bar{\sigma}_t + \mathcal{Z}_t + \widehat{\gamma}Z_t^1)^\top\Delta\varphi_t^1\right] dt + Z_t^1 dW_t, \\
&\qquad Y_T^1 = 0. \qquad (3.29)
\end{aligned}
$$

In summary, we have established the following correspondence between solutions of this FBSDE with reflection and equilibria with proportional transaction costs:

Proposition 3.9. *Suppose the assumptions in Proposition 3.3 are satisfied and let $(\bar{S}, \bar{\sigma})$ and $\bar{\varphi}^1$ be the corresponding frictionless equilibrium price and optimal strategy for agent 1. Suppose $(\varphi^{1,\uparrow}, \varphi^{1,\downarrow}, \Delta\varphi, \mathcal{Y}, Y^1, \mathcal{Z}, Z^1) \in (FV)^2 \times (\mathbb{H}^2)^5$ is a solution to the system (3.47), (3.28), (3.29) such that*

$$
(\bar{\varphi}^1 + \Delta\varphi^1)(\bar{\sigma} + \mathcal{Z} + \widehat{\gamma}Z^1) \in \mathbb{H}^2 \quad \textit{and} \quad \mathbb{E}\left[\int_0^T \lambda_t d\varphi_t^{1,\uparrow} + \lambda_t d\varphi_t^{1,\downarrow}\right] < \infty.
$$

Then

$$
S = \bar{S} + \mathcal{Y} + \widehat{\gamma}Y^1
$$

is an equilibrium price with transaction costs and the agents' optimal trading strategies are given by

$$
\varphi_t^1 = \varphi_{0-}^1 + \varphi_t^{1,\uparrow} - \varphi_t^{1,\downarrow}, \qquad \varphi_t^2 = s - \varphi_{0-}^1 + \varphi_t^{1,\downarrow} - \varphi_t^{1,\uparrow}. \qquad (3.48)
$$

Like for its counterpart for superlinear trading costs, existence and uniqueness results for the above fully-coupled FBSDE system with reflection are wide open problems. However, this description of the equilibrium allows to apply the deep-learning algorithm proposed in Section 3.5, since at each time discretization t_m, instead of $(\mathcal{Z}, Z^1)$, we can parametrize $(d\varphi^{1,\uparrow}, d\varphi^{1,\downarrow})$ and adapt the loss function `Loss` with the reflection condition at each time discretization, i.e.

$$
\texttt{Loss} = \frac{1}{\texttt{batch\_size}}\left[\|\mathcal{Y}_T^\theta\|^2 + C\sum_{m=0}^N \left\|\left(|{Y^1}_{t_m}^\theta| - \lambda_{t_m}\right)\mathbb{1}_{\{|{Y^1}_{t_m}^\theta| > \lambda_{t_m}\}}\right\|^2\right],
$$

where the constant C is the penalizing constant. Moreover, the above FBSDE formulation provides the starting point for the formal asymptotic analysis of small transaction costs that we carry out in Chapter 5.

Chapter 4

Global Existence for a Model with Linear State Dynamics

4.1 Introduction

In Chapter 3, we have seen that equilibrium asset prices and trading strategies with transaction costs generally correspond to coupled systems of forward-backward stochastic differential equations (FBSDEs). These equations are not Lipschitz, have a multidimensional backward component, and display full coupling between the (degenerate) forward and backward components. Accordingly, these FBSDEs fall outside the scope of known existence results, see (Herdegen, Muhle-Karbe, and Possamaï, 2019, Section 4.2) for a detailed discussion of the related literature.

In Herdegen et al. (2019), a *local* wellposedness result is established for quadratic transaction costs and agents whose risk aversion parameters are sufficiently similar. This approach is based on the observation that, for agents with the same risk aversion, the backward equation for the equilibrium price decouples from the forward-backward system for the optimal position and trading rate, and in fact reduces to the BSDE for the frictionless equilibrium price. As a consequence, transaction costs affect the agents' trading behaviour (and welfare) in this case, but the same equilibrium prices still clear the market. Herdegen et al. (2019) in turn establish a local existence result around this "expansion point". To wit, they show that a Picard iteration produces an equilibrium close to the frictionless one provided that the agents' risk aversion parameters are sufficiently similar. This raises the natural question whether this "smallness condition" is indeed necessary, or whether equilibrium asset prices with transaction costs exist in general, irrespective of the heterogeneity of the agents' preferences.

In this chapter, we show that this is indeed the case for the simplest example with linear state dynamics. In this setting, there is no aggregate endowment and both endowment volatilities (and, in turn, frictionless target strategies) as well as the asset's terminal condition are linear functions of the driving Brownian motion. In this context, the frictionless equilibrium price has Bachelier dynamics and the FBSDE describing its frictional counterpart has been reduced to a system of four coupled Riccati ODEs in Herdegen et al. (2019). However, existence is again only established by Picard iteration for sufficiently similar risk aversions. Here, we show how to establish global existence for this system using direct computations. To wit, we first apply a number of changes of variables, which allow us to reduce the system to two coupled ODEs. Variation of constants in turn allows us to futher reduce this to a single path-dependent but scalar ODE, for which we can establish wellposedness by elementary comparison arguments.

As a byproduct, the estimates derived in the proof allow us to derive in a rigorous manner explicit asymptotic approximations of the equilibrium in the limiting regime of *small* transaction costs. We test these approximations against the numerical solutions of the Riccati ODEs and find that they provide excellent approximations for the calibrated parameters from Sections 2.5 and 3.5. The closed-form approximations provide a lot of additional qualitative insight into the properties of the equilibrium already in this simplest setting (where the ODEs determining the exact equilibrium are straightforward to solve numerically). Moreover, the scalings and structure of the asymptotic expansions that we derive rigorously here also motivate the formal asymptotic analysis of the general models that we carry out in Chapter 5.

In the present model with linear state dynamics, the frictionless equilibrium price has constant expected returns and volatilities. With small transaction costs, expected returns and trading volumes follow Ornstein-Uhlenbeck processes as in the model with exogenous volatility studied in Chapter 2. However, when the volatility is determined endogenously, price levels and average expected returns can change due to the introduction of the trading costs. In our model, the empirically relevant case of positive illiquidity discounts as in Amihud and Mendelson (1986) (or, equivalently, liquidity premia as in Brennan and Subrahmanyam (1996); Pástor and Stambaugh (2003)) necessarily corresponds to a positive relationship between transaction costs and volatility. This is in line with empirical studies that test this relationship using natural experiments such as the introduction of transaction taxes (Umlauf, 1993; Jones and Seguin, 1997; Hau, 2006). The positive link between trading costs and volatility is also corroborated by numerical results for risk-sharing equilibria (Adam, Beutel, Marcet, and Merkel, 2015; Buss and Dumas, 2019) as well as by models with asymmetric information (Danilova and Julliard, 2019) and heterogenous beliefs (Muhle-Karbe, Nutz, and Tan, 2020).

In our model, the magnitude of these effects is modulated by the heterogeneity of the agents' preferences. If both agents' risk aversions are the same, then trading costs have no impact on equilibrium dynamics. However, realistic liquidity premia of around 0.5% per annum obtain for the parameters calibrated to S&P500 time series in Sections 2.5 and 3.5 if one of the agents has twice the risk aversion of the other. In this case, the volatility increases by about 1.7% relative to its frictionless value. The (yearly) Sharpe ratio of the risky asset in turn increases from 60.77% to 61.81% when transaction costs are taken into consideration.

The remainder of this chapter is organized as follows. The exogenous inputs of the model, the frictionless equilibrium, and the FBSDE system describing its frictional counterpart are recalled in Section 4.2. Subsequently, in Section 4.3, we reduce this probabilistic representation to a system of Riccati equations and in turn a path-dependent but scalar ODE. We then establish the existence of a global solution and show how it allows us to construct a solution of the FBSDE system. Next, Section 4.4 provides explicit approximation formulas in the limit for small transaction costs. For better readability, all proofs are collected in Section 4.5.

4.2 Model

In this chapter, we focus on the simplest specification of the model with endogenous volatility from Chapter 3:

(i) the exogenous state process $X_t = W_t$ is a scalar Brownian motion;

(ii) the volatilities

$$\xi^1(X_t) = -\xi^2(X_t) = \xi X_t, \quad \xi \in (0,\infty)$$

of agents' endowments (and, in turn, their frictionless positions) sum to zero and are linear functions of the state variable;

(iii) the terminal condition is a linear function of the driving Brownian motion, too:

$$\mathfrak{S} = \beta T + \alpha X_T, \qquad \beta \in \mathbb{R},\ \alpha \in (0,\infty).$$

For these primitives, Example 3.5 shows that the frictionless volatility and expected return are constant:

$$\bar{\sigma}_t = \bar{\sigma} = \alpha, \quad \bar{\mu}_t = \bar{\mu} = \bar{\gamma} s \alpha^2.$$

The frictionless equilibrium price in turn has Bachelier dynamics:

$$d\bar{S}_t = \bar{\gamma} s \alpha^2 dt + \alpha dW_t, \quad \bar{S}_0 = \beta T - \bar{\gamma} s \alpha^2 T.$$

Regarding the convex trading costs $\lambda_t G(\dot{\varphi}_t^n)$ on the agents' trading rate, we follow Herdegen et al. (2019) and use the most tractable specification:

(iv) the trading costs are quadratic and of the same magnitude for all times and states:

$$\lambda_t = \lambda \in (0,\infty), \qquad G(x) = \frac{1}{2}x^2.$$

With these specifications, the FBSDE system (**??**), (3.25), (3.28), (3.29) from Chapter 3 that describes the frictional equilibrium simplifies to

$$
\begin{aligned}
dX_t &= dW_t, & X_0 &= 0, \qquad (4.1)\\
d\Delta\varphi_t^1 &= \frac{Y_t^1}{\lambda}dt + \frac{\xi}{\alpha}dW_t, & \Delta\varphi_0^1 &= \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \qquad (4.2)\\
d\mathcal{Y}_t &= \bar{\gamma}s\left(2\alpha + \mathcal{Z}_t + \widehat{\gamma}Z_t^1\right)\left(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\right)dt + \mathcal{Z}_t dW_t, & \mathcal{Y}_T &= 0, \qquad (4.3)\\
dY_t^1 &= \widetilde{\gamma}\left[(\alpha + \mathcal{Z}_t + \widehat{\gamma}Z_t^1)^2\Delta\varphi_t^1 - \frac{\xi}{\alpha}W_t\big(\mathcal{Z}_t + \widehat{\gamma}Z_t^1\big)\big(\alpha + \mathcal{Z}_t + \widehat{\gamma}Z_t^1\big)\right]dt\\
&\quad + Z_t^1 dW_t, & Y_T^1 &= 0. \qquad (4.4)
\end{aligned}
$$

Here, we recall from Chapter 3 that

$$\widetilde{\gamma} = \frac{\gamma^1+\gamma^2}{2}, \quad \bar{\gamma} = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}, \quad \widehat{\gamma} = \frac{\gamma^1-\gamma^2}{\gamma^1+\gamma^2}.$$

If a solution of the system (4.1)-(4.4) exists then, by Proposition 3.6, an equilibrium price with transaction costs is given by

$$S_t = \bar{S}_t + \mathcal{Y}_t + \widehat{\gamma}Y_t^1.$$

Moreover, the agents' corresponding equilibrium trading rates and positions then are

$$\dot{\varphi}_t^1 = -\dot{\varphi}_t^2 = \frac{Y_t^1}{\lambda}, \quad \varphi_t^1 = s - \varphi_t^2 = \bar{\varphi}_t^1 + \Delta\varphi_t^1.$$

To wit, the first forward components of the FBSDE system (4.1)-(4.4) is the exogenous and uncoupled state variable that drives the agents' random endowments and the terminal dividend. The second forward component describes the deviation of agent 1's actual position from its frictionless counterpart, which is naturally coupled to the second backward component that corresponds to the marginal trading cost $\lambda G'(\dot{\varphi}_t^1)$ of agent 1's optimal trading rate. Finally, a linear combination of both backward components determines the price adjustment due to transaction costs.

We now address the question whether a solution of the FBSDE system (4.2)-(4.4) indeed exists for arbitrary parameters of the model.

4.3 Linear Ansatz and Riccati System

In general, the backward components of Markovian FBSDEs are deterministic functions of time and the forward components. Here, motivated by the linearity of the forward dynamics (4.1)-(4.2), we make the ansatz that the functions determining the backward components $\mathcal{Y}_t$ and Y_t^1 are also linear:[1]

$$\mathcal{Y}_t = A_\lambda(t) + B_\lambda(t)W_t + C_\lambda(t)\Delta\varphi_t^{1,\lambda}, \tag{4.5}$$
$$Y_t^1 = D_\lambda(T-t) + E_\lambda(T-t)W_t - F_\lambda(T-t)\Delta\varphi_t^{1,\lambda}, \tag{4.6}$$

for deterministic functions $A_\lambda, B_\lambda, C_\lambda, D_\lambda, E_\lambda, F_\lambda : [0,T] \to \mathbb{R}$ satisfying

$$A_\lambda(T) = B_\lambda(T) = C_\lambda(T) = 0, \quad D_\lambda(0) = E_\lambda(0) = F_\lambda(0) = 0. \tag{4.7}$$

Itô's formula and the dynamics (4.2) of $\Delta\varphi_t^1$ in turn yield

$$\begin{aligned}
d\mathcal{Y}_t &= [A'_\lambda + B'_\lambda W_t + C'_\lambda \Delta\varphi_t]\,dt + B_\lambda dW_t + C_\lambda d\Delta\varphi_t^{1,\lambda} \\
&= \left[A'_\lambda + B'_\lambda W_t + C'_\lambda\Delta\varphi_t^{1,\lambda} + \frac{C_\lambda}{\lambda}\Big(D_\lambda + E_\lambda W_t - F_\lambda \Delta\varphi_t^{1,\lambda}\Big)\right] dt \\
&\quad + \left(B_\lambda + \frac{\xi}{\alpha}C_\lambda\right) dW_t \\
&= \left[A'_\lambda + \frac{C_\lambda D_\lambda}{\lambda} + \Big(B'_\lambda + \frac{C_\lambda E_\lambda}{\lambda}\Big)W_t + \Big(C'_\lambda - \frac{C_\lambda F_\lambda}{\lambda}\Big)\Delta\varphi_t^{1,\lambda}\right] dt \\
&\quad + \left(B_\lambda + \frac{\xi}{\alpha}C_\lambda\right) dW_t,
\end{aligned}$$

as well as

$$\begin{aligned}
dY_t^1 &= [-D'_\lambda - E'_\lambda W_t + F'_\lambda \Delta\varphi_t]\,dt + E_\lambda dW_t - F_\lambda d\Delta\varphi_t^{1,\lambda} \\
&= \left[-D'_\lambda - E'_\lambda W_t + F'_\lambda \Delta\varphi_t - \frac{F_\lambda}{\lambda}(D_\lambda + E_\lambda W_t - F_\lambda \Delta\varphi_t^{1,\lambda})\right] dt \\
&\quad + \left(E_\lambda - \frac{\xi}{\alpha}F_\lambda\right) dW_t \\
&= -\left[D'_\lambda + \frac{F_\lambda D_\lambda}{\lambda} + \Big(E'_\lambda + \frac{F_\lambda E_\lambda}{\lambda}\Big)W_t - \Big(F'_\lambda + \frac{F_\lambda^2}{\lambda}\Big)\Delta\varphi_t^{1,\lambda}\right] dt \\
&\quad + \left(E_\lambda - \frac{\xi}{\alpha}F_\lambda\right) dW_t.
\end{aligned}$$

To match the respective volatility terms in the backward equations (4.3) and (4.4), we therefore need

$$\mathcal{Z}_t = B_\lambda(t) + \frac{\xi}{\alpha}C_\lambda(t), \qquad Z_t^1 = E_\lambda(T-t) - \frac{\xi}{\alpha}F_\lambda(T-t). \tag{4.8}$$

Matching the drift terms in dY_t^1 and $d\mathcal{Y}_t$ and comparison of coefficients in

[1]Here, the minus sign in front of F ensures that large deviations from the frictionless equilibrium position are reduced for positive F.

turn leads to the following system of coupled Riccati ODEs:

$$\begin{aligned}
F_\lambda' &= \widetilde{\gamma}\left(\alpha + B_\lambda + \frac{\xi}{\alpha}C_\lambda + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)^2 - \frac{F_\lambda^2}{\lambda},\\
E_\lambda' &= \frac{\widetilde{\gamma}\xi}{\alpha}\left(\alpha + B_\lambda + \frac{\xi}{\alpha}C_\lambda + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)\left(B_\lambda + \frac{\xi}{\alpha}C_\lambda + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right) - \frac{F_\lambda E_\lambda}{\lambda},\\
D_\lambda' &= -\frac{F_\lambda D_\lambda}{\lambda},\\
C_\lambda' &= \frac{F_\lambda C_\lambda}{\lambda},\\
B_\lambda' &= -\frac{E_\lambda C_\lambda}{\lambda},\\
A_\lambda' &= \bar{\gamma}s\left(2\alpha + B_\lambda + \frac{\xi}{\alpha}C_\lambda + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)\left(B_\lambda + \frac{\xi}{\alpha}C_\lambda + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F\right) - \frac{D_\lambda C_\lambda}{\lambda}.
\end{aligned}$$

Together with the vanishing boundary conditions (4.7), it follows that

$$B_\lambda = C_\lambda = D_\lambda = 0, \tag{4.9}$$

and in turn

$$A_\lambda(t) = -\widehat{\gamma}\bar{\gamma}s\int_t^T \left(2\alpha + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)\left(E_\lambda - \frac{\xi}{\alpha}F_\lambda\right)dr. \tag{4.10}$$

It therefore remains to determine the functions F_λ, E_λ as solutions of the following initial-value problem:

$$\begin{cases} F_\lambda' = \widetilde{\gamma}\left(\alpha + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)^2 - \frac{F_\lambda^2}{\lambda}, \quad F_\lambda(0) = E_\lambda(0) = 0,\\ E_\lambda' = \widetilde{\gamma}\widehat{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}E_\lambda - \frac{\widehat{\gamma}\xi}{\alpha}F_\lambda\right)\left(E_\lambda - \frac{\xi}{\alpha}F_\lambda\right) - \frac{F_\lambda E_\lambda}{\lambda}. \end{cases} \tag{4.11}$$

As the right-hand side of (4.11) is continuously differentiable in F_λ, E_λ, uniqueness and *local* existence for the system (4.11) follow from standard arguments as in (Chicone, 1999, Theorem 1.261), for example.

Lemma 4.1. *There exists at most one solution of* (4.11)*, and there is a unique local solution* (F_λ, E_λ) *of* (4.11) *on a maximum interval of existence* $[0, T_{\max})$.

Proof. See Section 4.5.2. □

Comparing to the analysis in Section 5 of Herdegen et al. (2019), the reparametrization from Section 3.4.3 has therefore already reduced the dimension of the Riccati system from four equations to two. However, this still does not allow the application apply comparison arguments in a straightforward manner to establish global existence. Fortunately, some further transformations allow a reduction of the system (4.11) to a path-dependent, but *scalar* ODE. To wit, use the local solution to define

$$H_\lambda = E_\lambda - \frac{\xi}{\alpha} F_\lambda, \quad \text{such that } H_\lambda(0) = 0. \tag{4.12}$$

Then, the Riccati equations for E_λ and F_λ show that on $[0, T_{\max})$,

$$\begin{aligned} H_\lambda' &= \widetilde{\gamma}\widehat{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}H_\lambda\right) H_\lambda - \frac{F_\lambda E_\lambda}{\lambda} - \widetilde{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}H_\lambda\right)^2 + \frac{\xi}{\alpha}\frac{F_\lambda^2}{\lambda} \\ &= \widetilde{\gamma}\widehat{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}H_\lambda\right) H_\lambda - \widetilde{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}H_\lambda\right)^2 - \frac{F_\lambda H_\lambda}{\lambda} \\ &= -\widetilde{\gamma}\xi\alpha - \left(\widetilde{\gamma}\widehat{\gamma}\xi + \frac{F_\lambda}{\lambda}\right) H_\lambda. \end{aligned}$$

Whence, variation of constants gives

$$H_\lambda(\tau) = -\widetilde{\gamma}\xi\alpha \int_0^\tau e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)} dr, \quad \tau \in [0, T], \tag{4.13}$$

and in turn

$$E_\lambda(\tau) = \frac{\xi}{\alpha} F_\lambda(\tau) - \widetilde{\gamma}\xi\alpha \int_0^\tau e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)} dr, \quad \tau \in [0, T]. \tag{4.14}$$

After inserting (4.14) into the first Riccati equation from (4.11), we find that F_λ is a local solution of the following path-dependent but scalar ODE:

$$\begin{cases} F_\lambda'(\tau) = \widetilde{\gamma}\alpha^2 \left(1 - \widetilde{\gamma}\widehat{\gamma}\xi \displaystyle\int_0^\tau e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)} dr\right)^2 - \dfrac{1}{\lambda} F_\lambda(\tau)^2, \\ F_\lambda(0) = 0. \end{cases} \tag{4.15}$$

Global existence for this one-dimensional equation and in turn the original system (4.11) can now be established by elementary comparison arguments:

Theorem 4.2. *There exists a unique solution F_λ of (4.15) on $[0, T]$. It is increasing and satisfies*

$$\frac{\widetilde{\gamma}^{1/2}\alpha}{e^{\widetilde{\gamma}\xi T}} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}}\tau\right) \le \frac{F_\lambda(\tau)}{\lambda^{1/2}} \le \widetilde{\gamma}^{1/2}\alpha e^{\widetilde{\gamma}\xi T} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha e^{\widetilde{\gamma}\xi T}}{\lambda^{1/2}}\tau\right). \tag{4.16}$$

The unique global solution of (4.11) is then given by F_λ and E_λ from (4.14).

Proof. See Section 4.5.2. □

With the solution F_λ, E_λ of the Ricatti system (4.11) at hand and A_λ defined via (4.10), it is now straightforward to construct a solution of the FBSDE system (4.2)-(4.4):

Corollary 4.3. *Let F_λ, E_λ be the solutions of the Ricatti system (4.11). Then, a solution of the FBSDE system (4.2)-(4.4) is given by the solution of the linear SDE*

$$d\Delta\varphi_t^{1,\lambda} = \frac{F_\lambda(T-t)}{\lambda}\left(\frac{E_\lambda(T-t)}{F_\lambda(T-t)}W_t - \Delta\varphi_t^{1,\lambda}\right)dt + \frac{\xi}{\alpha}dW_t,$$
$$\Delta\varphi_0^{1,\lambda} = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \tag{4.17}$$

as well as

$$\mathcal{Y}_t^\lambda = A_\lambda(t) = -\widehat{\gamma}\bar{\gamma}s\int_t^T \left(2\alpha + \widehat{\gamma}H_\lambda(T-u)\right)H_\lambda(T-u)\,du, \tag{4.18}$$
$$Y_t^{1,\lambda} = E_\lambda(T-t)W_t - F_\lambda(T-t)\Delta\varphi_t^{1,\lambda}. \tag{4.19}$$

In view of Proposition 3.6, this solution of the FBSDE in turn leads to the following equilibrium prices, optimal trading rates, and optimal positions:

$$S_t = \bar{S}_t + \mathcal{Y}_t^\lambda + \widehat{\gamma}Y_t^{1,\lambda}, \quad \dot{\varphi}_t^1 = -\dot{\varphi}_t^2 = \frac{Y_t^{1,\lambda}}{\lambda}, \quad \varphi_t^1 = s - \varphi_t^2 = \bar{\varphi}_t^1 + \Delta\varphi_t^{1,\lambda}.$$

Proof. See Section 4.5.2. □

Corollary 4.3 guarantees global existence of a Radner equilibrium with transaction costs for arbitrary parameter configurations of the model. In particular, the agents' risk aversions can be arbitrary here, whereas they need to be sufficiently similar for the local existence result established using Picard iteration in (Herdegen et al., 2019, Theorem 5.2).

Let us briefly comment on the economic implications of this result. Without transaction costs, the driving Brownian motion is the only state variable of the equilibrium. With trading costs, equilibrium prices and trading rates are additionally influenced by the deviation $\Delta\varphi_t^1$ of agent 1's position from its frictionless counterpart. This process has mean-reverting dynamics driven by agent 1's frictionless strategy, reflecting the fact that with transaction costs these deviations can only be reduced gradually but not immediately. The same mean-reverting process also appears in the price correction due to transaction costs, leading to a mean-reverting return component but a deterministic volatility adjustment. In addition, the price correction also includes an additional deterministic drift $\mathcal{Y}_t$ here.

Compared to the model with exogenous volatility studied in Chapter 2, the mean-reverting return components are a common feature of both models. In contrast, the volatility was fixed to the same value with and without transaction costs in Chapter 2, whereas it can change here. Finally, whereas the expected return adjusted was always zero on average in Chapter 2, it can potentially include a systematic "liquidity premium" here. In order to say more about the qualitative and quantitative properties of all these effects, a useful analytical tool is to consider their leading-order asymptotics for *small transaction costs*, to which we turn next.

4.4 Small-Costs Asymptotics

The Riccati system (4.11) can be solved numerically in a straighforward manner. In order to shed more lights on the structure of the solution and its implications for equilibrium asset prices and trading strategies, it is nevertheless instructive to expand the solutions in the practically relevant limiting regime of *small* transaction costs $\lambda \approx 0$. The estimates derived in the proof of Theorem 4.2 allow us to derive rigorous error bounds for these approximations of the Riccati equations (4.11) and in turn the FBSDE system (4.17)-(4.19). The scalings and structures of these approximation in turn form the starting point for the formal asymptotic analysis of more general models that we carry out in Chapter 5.

4.4.1 Approximation of the Riccati Equations

We start by deriving the leading-order asymptotics of the solutions F_λ, E_λ of the Riccati system (4.11) and the function A_λ from (4.10):

Lemma 4.4. *Define*

$$F(\tau) = \widetilde{\gamma}^{1/2}\alpha \tanh\left(\widetilde{\gamma}^{1/2}\alpha\, \tau\right), \tag{4.20}$$

$$E(\tau) = \widetilde{\gamma}\widehat{\gamma}\frac{\xi^2}{\alpha}\left(\operatorname{sech}\left(\widetilde{\gamma}^{1/2}\alpha\, \tau\right) - 1\right), \tag{4.21}$$

$$A(t) = 2\bar{\gamma}\widehat{\gamma}\widetilde{\gamma}^{1/2}\xi\alpha s(T-t). \tag{4.22}$$

Then, for small transaction costs $\lambda \downarrow 0$, the functions $\lambda^{1/2}F(\tau/\lambda^{1/2})$, $\lambda E(\tau/\lambda^{1/2})$, $\lambda^{1/2}A(t)$ approximate F_λ, E_λ, A_λ uniformly at the leading orders in that[2]

$$\left\|F_\lambda(\tau) - \lambda^{1/2}F(\tau/\lambda^{1/2})\right\|_\infty = O(\lambda), \tag{4.23}$$

$$\left\|E_\lambda(\tau) - \lambda E(\tau/\lambda^{1/2})\right\|_\infty = O(\lambda^{3/2}), \tag{4.24}$$

$$\left\|A_\lambda(t) - \lambda^{1/2}A(t)\right\|_\infty = O(\lambda). \tag{4.25}$$

Proof. See Section 4.5.3. □

In order to assess the practical relevance of these approximations, we compare them to the numerical solution of the Riccati equations (4.11), (4.10) in Figure 4.1. As parameters, we use the values calibrated to S&P500 time series data in Sections 2.5 and 3.5:

$$\alpha = 1.88, \quad s = 2.46 \times 10^{11}, \quad \gamma^2 = 2\gamma^1 = 1.25 \times 10^{-13},$$

as well as

$$\lambda = 1.08 \times 10^{10}, \quad \xi = 2.19 \times 10^{10}, \quad \varphi^1_{0-} = s - \varphi^2_{0-} = \frac{\gamma^2}{\gamma^1+\gamma^2}s.$$

[2] These are indeed the relevant orders since $\|F\|_\infty = \|E\|_\infty = \|A\|_\infty = O(1)$.

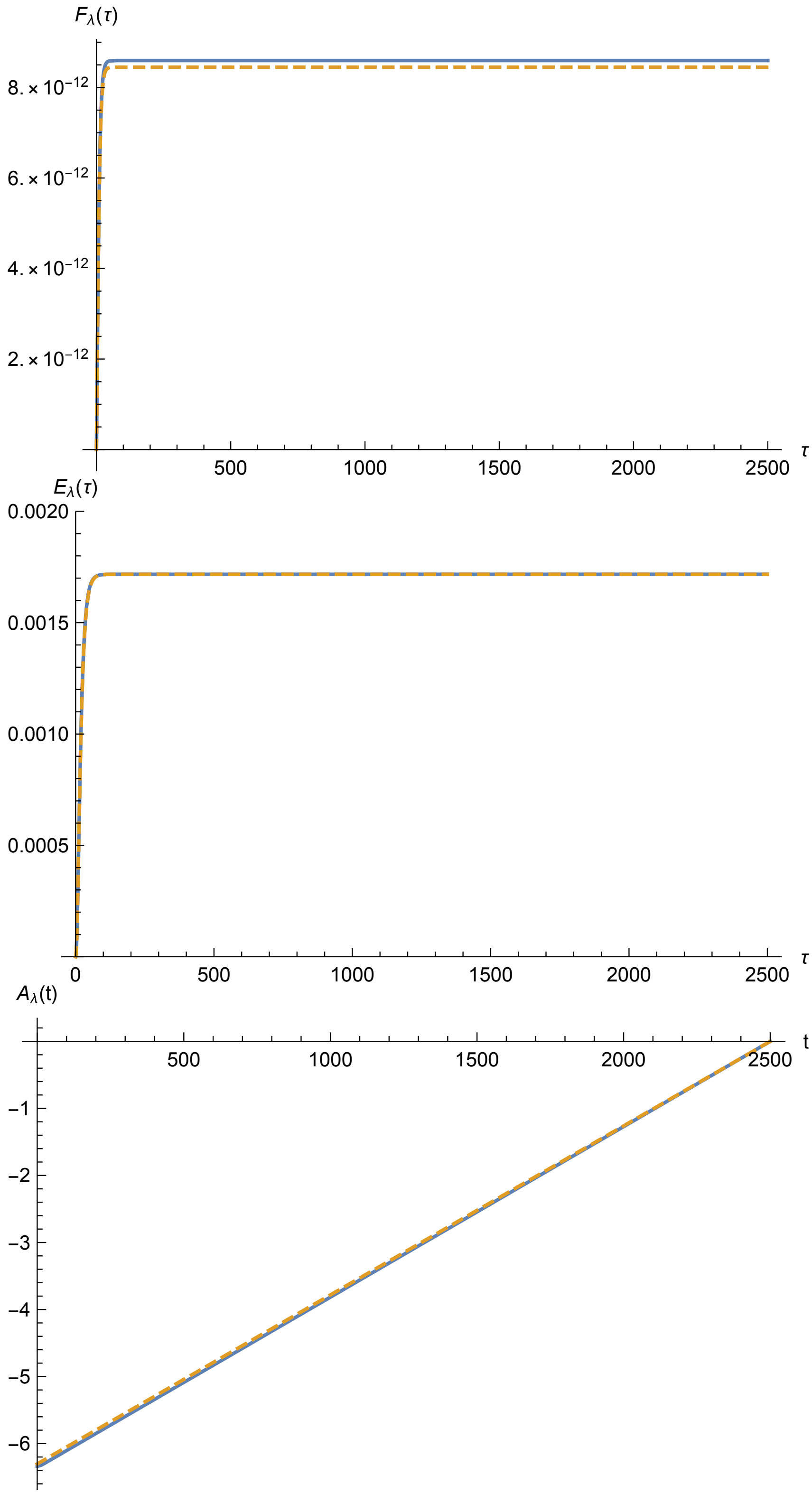

Figure 4.1: Numerical solution (solid, blue) and approximations from Lemma 4.4 (dashed, yellow) for the for the function F_λ (top panel), E_λ (middle panel), and A_λ (lower panel).

We find that the exact solutions and their leading-order approximations are virtually indistinguishable even for a time horizon as long as ten years ($T = 2500$ trading days). This provides some first evidence that the small-cost regime is indeed the practically relevant case.

4.4.2 Approximation of the FBSDE System

Next, we turn to the small-cost asymptotics of the FBSDE system (4.2)-(4.4). Using the expansions of the functions F_λ and A_λ we have determined above,[3] we obtain closed-form expressions that match – at the respective leading orders for small λ – the forward process $\Delta\varphi_t^{1,\lambda}$ from (4.17) as well as the backward processes $\mathcal{Y}_t^\lambda$ and $Y_t^{1,\lambda}$ from (4.18) and (4.19). Since all these processes are driven by Brownian motions, the respective error bounds of course only hold in expectation here, rather than uniformly as in Lemma 4.4.

Theorem 4.5. *Fix $p \geq 1$ and define the functions F and A as in Lemma 4.4. Then, for small transaction costs $\lambda \downarrow 0$, the solution of the linear SDE*

$$d\Delta_t^\lambda = -\frac{1}{\lambda^{1/2}} F\left(\frac{T-t}{\lambda^{1/2}}\right)\Delta_t^\lambda dt + \frac{\xi}{\alpha} dW_t, \qquad \Delta_0^\lambda = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \tag{4.26}$$

yields the following closed-form approximations of the forward process $\Delta\varphi_t^{1,\lambda}$, and the backward processes $\mathcal{Y}_t^\lambda$, $Y_t^{1,\lambda}$ in the FBSDE system (4.2)-(4.4):[4]

$$\begin{aligned}
\left\| \Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda \right\|_p &= O(\lambda^{1/2}), && t \in [0,T],\\
\|\mathcal{Y}_t^\lambda - \lambda^{1/2} A(t)\|_\infty &= O(\lambda), && t \in [0,T],\\
\left\| Y_t^{1,\lambda} + \lambda^{1/2} F\left(\frac{T-t}{\lambda^{1/2}}\right)\Delta_t^\lambda \right\|_p &= O(\lambda), && t \in [0,T].
\end{aligned}$$

Proof. See Section 4.5.3. □

Theorem 4.5 in turn allows us to approximate the equilibrium price and its drift and volatility coefficients as follows:

Corollary 4.6. *Fix $p \geq 1$. For small transaction costs $\lambda \downarrow 0$, the frictional equilibrium price from Corollary 4.3 has the following leading-order approximation:*

$$\|S_t^\lambda - \bar{S}_t - \lambda^{1/2} A(t)\|_p = O(\lambda^{3/4}), \quad t \in [0,T].$$

[3] In contrast, the function E_λ does not contribute to the leading-order terms.

[4] These are indeed the relevant asymptotic orders since, by Lemma 4.14, (4.18), (4.19), and Corollary 4.13, the L_p-norms of $\Delta\varphi_t^\lambda$, $\mathcal{Y}_t^\lambda$, $Y_t^{1,\lambda}$ are of the orders $O(\lambda^{1/4})$, $O(\lambda^{1/2})$, and $O(\lambda^{3/4})$, respectively.

The leading-order asymptotics of its drift, average drift, and volatility are

$$\begin{aligned} \|\mu_t^\lambda - \bar{\mu} - \widetilde{\gamma}\widehat{\gamma}\alpha^2\Delta_t^\lambda\|_p &= O(\lambda^{1/2}), && t \in [0,T], \\ \|\mathbb{E}[\mu_t^\lambda] - \bar{\mu} - \lambda^{1/2}A'(t)\|_\infty &= O(\lambda), && t \in [0,T], \\ \left\| \sigma_t^\lambda - \alpha + \frac{\lambda^{1/2}\widehat{\gamma}\xi}{\alpha} F\left((T-t)/\lambda^{1/2}\right) \right\|_p &= O(\lambda), && t \in [0,T]. \end{aligned}$$

Proof. See Section 4.5.3. □

4.4.3 Time-Averaged Aysmptotics

The hyperbolic functions in Lemma 4.4 are crucial to approximate the solutions of the Riccati equations (4.11) and the FBSDE (4.2)-(4.4) pointwise in time. However, for small transaction costs λ, this time inhomogeneity only affects the solution just before the terminal time T. Accordingly, if one drops the hyperbolic functions to obtain even simpler formulas, then the asymptotic order of the approximation error remains unchanged if the error is averaged across both time and states in the $\mathbb{H}^p$-norm.[5] The function F_λ can in fact be approximated by a constant:

Lemma 4.7. *Fix $p \geq 1$. The function F_λ from Theorem 4.2 satisfies*

$$\int_0^T \left| F_\lambda(\tau) - (\widetilde{\gamma}\alpha^2\lambda)^{1/2} \right| d\tau = O(\lambda). \tag{4.27}$$

Proof. See Section 4.5.3. □

The solution of the FBSDE system (4.2)-(4.4) in turn has the following closed-form approximation:

Theorem 4.8. *Fix $p \geq 1$, recall the explicit linear function $A(t)$ from (4.22), and define the Ornstein-Uhlenbeck process*

$$d\widehat{\Delta}_t^\lambda = -\left(\frac{\widetilde{\gamma}\alpha^2}{\lambda}\right)^{1/2} \widehat{\Delta}_t^\lambda dt + \frac{\xi}{\alpha} dW_t, \qquad \widehat{\Delta}_0^\lambda = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1 + \gamma^2}. \tag{4.28}$$

The forward process $\Delta\varphi^{1,\lambda}$ and the backward processes $\mathcal{Y}^\lambda$, $Y^{1,\lambda}$ in the FBSDE system (4.2)-(4.4) have the following leading-order approximations:

$$\begin{aligned} \|\Delta\varphi^{1,\lambda} - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p} &= O(\lambda^{1/2}), \\ \|\mathcal{Y}^\lambda - \lambda^{1/2}A\|_\infty &= O(\lambda), \\ \|Y^{1,\lambda} + (\lambda\widetilde{\gamma}\alpha^2)^{1/2}\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} &= O(\lambda). \end{aligned}$$

Proof. See Section 4.5.3. □

[5] Recall that the $\mathbb{H}^p$-norm of a process X is $\|X\|_{\mathbb{H}^p} := (\mathbb{E}[(\int_0^T |X_t|^2 dt)^{p/2}])^{1/p}$.

For the frictional equilibrium price from Corollary 4.3, this finally yields the following simple approximations. We express them in terms of the primitives of the model to facilitate their interpretation in Section 4.4.5 below.

Corollary 4.9. *Fix $p \geq 1$ and recall the Ornstein-Uhlenbeck process $\widehat{\Delta}_t^\lambda$ from Theorem 4.8. The equilibrium price for small transaction costs $\lambda \downarrow 0$ has the following closed-from approximation:*

$$\left\| S_t - \bar{S}_t - \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s \lambda^{1/2}(T-t) \right\|_{\mathbb{H}^p} = O(\lambda^{3/4}). \tag{4.29}$$

The leading-order asymptotics of its drift, average drift, and volatility are

$$\left\| \mu^\lambda - \bar{\mu} - \frac{\gamma^1-\gamma^2}{2}\alpha^2\widehat{\Delta}^\lambda \right\|_{\mathbb{H}^p} = O(\lambda^{1/2}), \tag{4.30}$$

$$\left\| \mathbb{E}[\mu^\lambda] - \bar{\mu} + \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s \lambda^{1/2} \right\|_{\infty} = O(\lambda), \tag{4.31}$$

$$\left\| \sigma^\lambda - \bar{\sigma} + \frac{\gamma^1-\gamma^2}{(2(\gamma^1+\gamma^2))^{1/2}}\xi\lambda^{1/2} \right\|_{\mathbb{H}^p} = O(\lambda). \tag{4.32}$$

The agents' equilibrium trading rates $\dot{\varphi}_t^1 = -\dot{\varphi}_t^2$ have the following leading-order asymptotics:

$$\left\| \dot{\varphi}^1 + \left(\frac{\gamma^1+\gamma^2}{2\lambda}\right)^{1/2}\alpha\widehat{\Delta}^\lambda \right\|_{\mathbb{H}^p} = O(1). \tag{4.33}$$

Proof. See Section 4.5.3. □

4.4.4 Accuracy of the Approximations

For the calibrated parameters from Section 2.5, we now assess the accuracy of the time-averaged approximations of the equilibrium price from Corollary 4.9. In Figure 4.2 we compare a simulated sample path of the price adjustment $S_t - \bar{S}_t$ to its deterministic approximation from (4.29).[6] While the price correction clearly does have a nonzero volatility in line with (4.32), the linear approximation from Corollary 4.9 nevertheless provides a very good fit.

[6]These and all other simulations in this section are generated from the same Brownian sample path simulated on a grid with one hundred time steps per day.

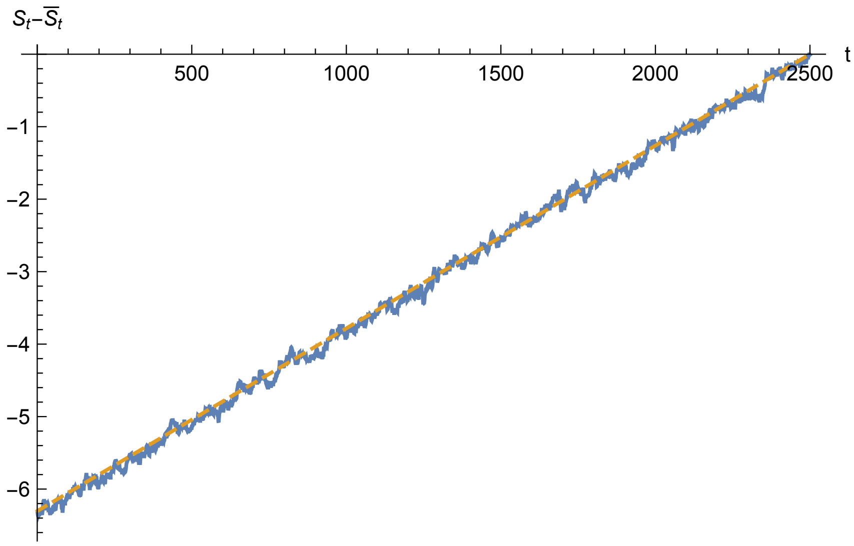

Figure 4.2: Sample path of the price adjustment $S_t - \bar{S}_t$ (solid, blue) and its linear approximation (4.29) (dashed, yellow).

In Figure 4.3, we in turn compare simulated sample paths of the exact adjustment $\mu_t - \bar{\mu}_t$ of the expected return and its approximation (4.30). The overall behavior is fit quite well, but the approximation displays a noticeable negative bias. This is more clearly visible in Figure 4.4, where we plot the corresponding approximation error.

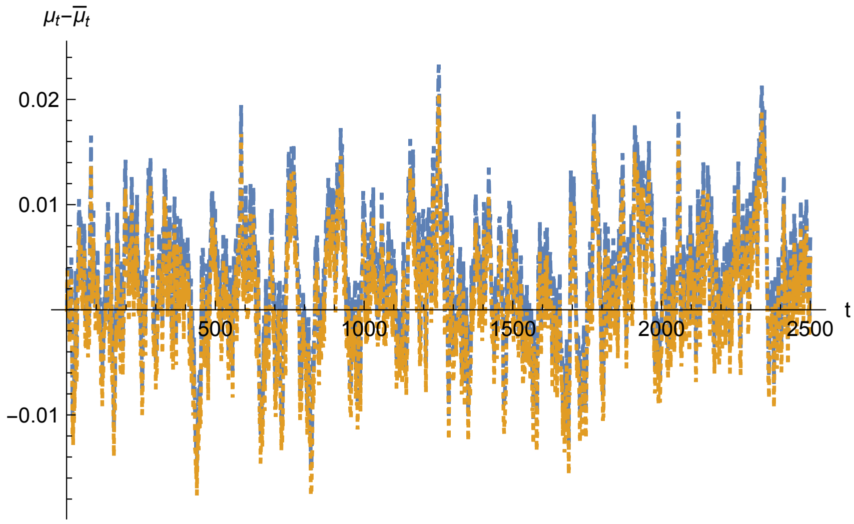

Figure 4.3: Sample path of the expected return adjustment $\mu_t - \bar{\mu}_t$ (solid, blue) and its Ornstein-Uhlenbeck approximation (4.29) (dashed, yellow).

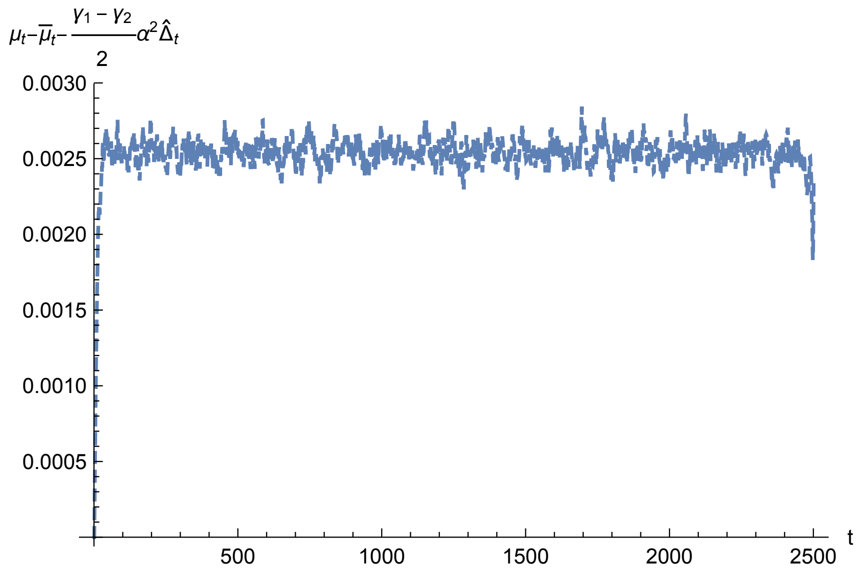

Figure 4.4: Error of the leading-order approximation (4.30) along the sample path from Figure 4.3.

The negative bias almost exactly corresponds to the nonzero average expected return from (4.31), which is given by 0.0025 for the parameters considered here. If one adjusts the Ornstein-Uhlenbeck process (4.30) by this constant as depicted in Figure 4.5, then the average absolute approximation error drops to less than 3% of this constant.

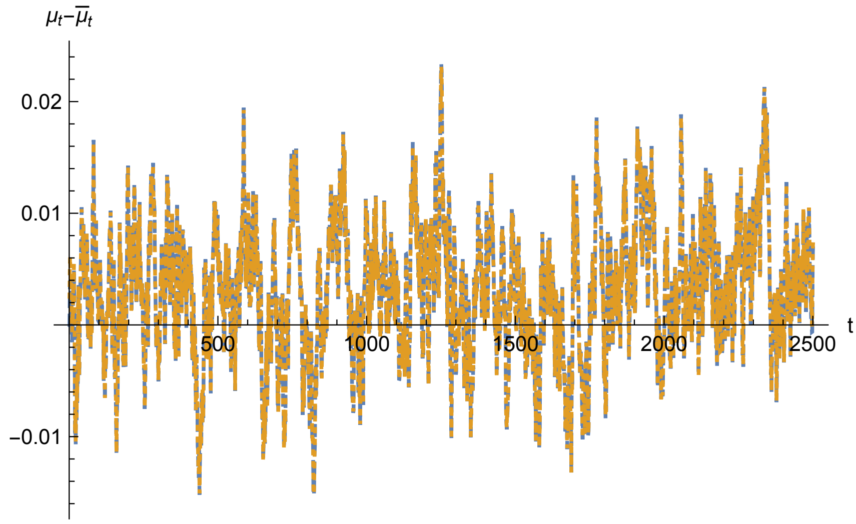

Figure 4.5: Sample path of the expected return adjustment $\mu_t - \bar{\mu}_t$ (solid, blue) and its Ornstein-Uhlenbeck approximation (4.29) shifted by the approximation (4.31) of the average expected return adjustment (dashed, yellow).

Next, the exact volatility correction from Corollary 4.3 is compared to the closed-form approximation (4.32) in Figure 4.6. The exact value and the asymptotic approximation are virtually the same over a time horizon of ten years – differences only become visible about one month before the terminal time is reached.

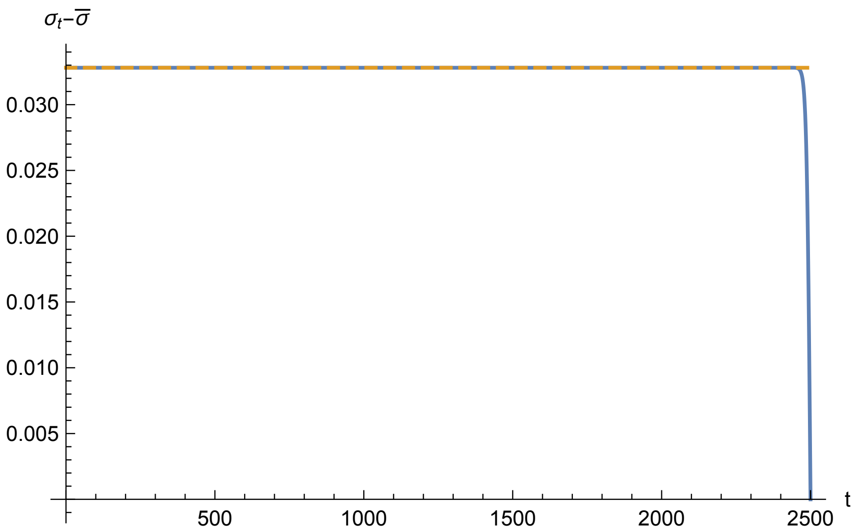

Figure 4.6: Numerical solution (solid, blue) and approximation from (4.32) (dashed, yellow) of the volatility correction due to small transaction costs.

Finally, in Figure 4.7, we compare the daily share turnover in the exact model to the approximation (4.33). Here, the approximation with an Ornstein-Uhlenbeck process already performs very well – the average absolute approximation error is only about 1.4% of the average trading volume in the exact model.

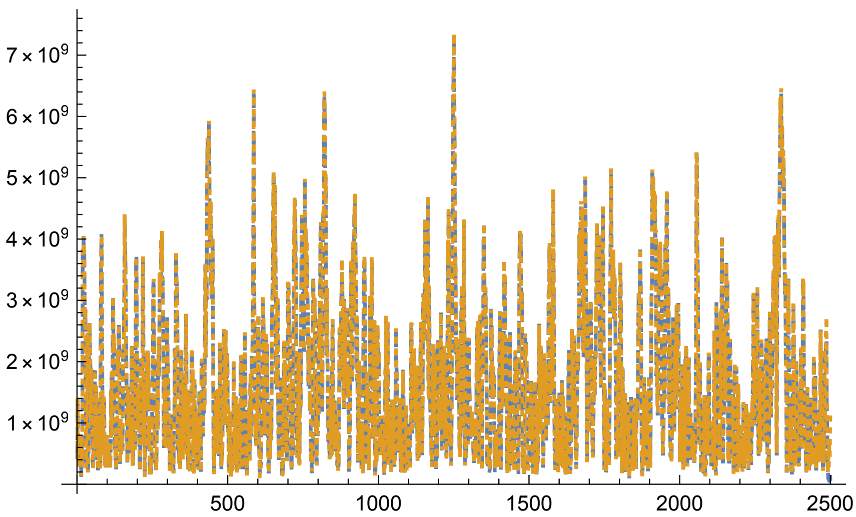

Figure 4.7: Daily share turnover in the numerical solution of the exact equilibrium (solid, blue) and its approximation from (4.33) (dashed, yellow).

In summary, our comparison of the exact equilibrium prices and trading volume to the approximations from Corollary 4.9 suggest the following:

(i) The level of the equilibrium-price adjustment $S_t - \bar{S}_t$ due to transaction costs is well approximated by the linear function from (4.29).

(ii) The adjustment $\mu_t - \bar{\mu}_t$ of the equilibrium expected returns is well approximated by the Ornstein-Uhlenbeck process from (4.30), shifted by the constant from (4.31).

(iii) The volatility adjustment is well approximated by the constant from (4.32).

(iv) Trading volume is well approximated by the Ornstein-Uhlenbeck process from (4.33).

We now discuss some of the economic insights that can be gleaned from these closed-form approximations.

4.4.5 Discussion of the Results

After discussing the accuracy of the small-cost asymptotics from Corollary 4.9 in the previous section, we now turn to their qualitative and quantitative predictions, and compare them to the empirical literature. We begin our discussion with the impact of transaction costs on the price level, cf. (4.29):

$$S_t - \bar{S}_t \approx \lambda^{1/2} A(0) = \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s\lambda^{1/2}(T-t). \tag{4.34}$$

This formula shows that, as in the overlapping-generations model with proportional costs studied by Vayanos (1998), asset prices can be either increased or decreased by the transaction costs. In our model, the sign of this effect is determined by whether agent 1 or agent 2 is more risk averse. In the empirically relevant case of an "illiquidity discount" (Amihud and Mendelson, 1986), the price adjustment (4.34) is concave in the size of the transaction costs in line with empirical results of Amihud and Mendelson (1986). More specifically, it scales with the square root of the trading cost here, like the effect of small transaction costs on the value function in partial-equilibrium models (Guasoni and Weber, 2017; Moreau et al., 2017). The other terms in (4.34) show that the illiquidity discount is large if the asset or the agents' trading targets are volatile, again paralleling similar results for value functions of partial equilibrium models (Moreau et al., 2017). The intuition is that both higher risk and more wildly fluctuating trading targets lead to stronger trading needs, and in turn bigger effects of transaction costs. Here, the magnitude of these terms is multiplied by the total supply (which scales up the trading costs for the agents' total portfolio) and by the agents' remaining time horizon (which scales the trading costs accumulated over time).

In order to assess the magnitude of the price correction, it is therefore preferable to consider the "liquidity premium", that is, the difference $\mu_t - \bar{\mu}_t$ between expected returns with and without transaction costs. In view of (4.31), the average liquidity premia are given by almost the same formula as the illiquidity discount (4.34), but are independent of the agents' planning horizon. For $\gamma^2 = 2\gamma^1$ and the remaining parameters calibrated to a capitalization-weighted time series of S&P500 stocks as in Sections 2.5 and 3.5, we obtain a daily liquidity premium of about 0.0025, that is, about 0.5% in yearly terms relative to the average stock prices. For trading volume calibrated to the large levels observed empirically, the model therefore produces realistic levels of liquidity premia without extreme heterogeneity between the agents preferences. Indeed, we have chosen $\gamma^2 = 2\gamma^1$ to obtain a ratio of the average liquidity premium and the (equivalent) proportional cost from Section 2.5 that is of a similar size as in Amihud and Mendelson (1986) – 2 in our model, compared to about 1.9 in their empirical analysis.

After the sign and size of the liquidity premium are calibrated to empirical data, the sign of the volatility correction is fixed in our model and therefore provides a testable implication. To wit, the expansion (4.32) shows that the sign of the volatility correction due to small transaction costs always has the same sign as the average liquidity premium in (4.31). Accordingly, our model predicts that the positive liquidity premia observed empirically necessarily correspond to a positive relationship between trading costs and volatility. This is in line with empirical studies that test this relationship using natural experiments such as the introduction of transaction taxes (Umlauf, 1993; Jones and Seguin, 1997; Hau, 2006). The positive link between trading costs and volatility is also corroborated by numerical results for risk-sharing equilibria (Adam et al., 2015; Buss and Dumas, 2019) as well as by models with asymmetric information (Danilova and Julliard, 2019) and heterogenous beliefs (Muhle-Karbe et al., 2020). For our parameters calibrated to liquid stocks, the magnitude of this effect turn out to be relatively modest, in that volatility increases by abut 1.7% of its frictionless value. Accordingly, the average (yearly) Sharpe ratio of the risky asset increases from 60.77% to 61.81% for our calibration when transaction costs are taken into account.

In addition to shifting average expected return approximately by a constant, transaction costs also introduce a mean-reverting stochastic liquidity premium. In view of (4.30), the latter has Ornstein-Uhlenbeck dynamics for small transaction costs, and is in fact of exactly the same form as in the model with exogenous volatility studied in Chapter 2. Likewise, trading volume for small transaction costs is described by the same Ornstein-Uhlenbeck process as in Chapter 2. In summary, with linear state dynamics, the model with endogenous volatility therefore makes the same predictions for mean-reverting returns and trading volume as its counterpart with exogenous volatility, but complements these by systematic liquidity premia and a positive link between trading costs and volatility in line with the empirical literature.

For the simple model we consider in this section, the findings reported here parallel the asymptotics for similar risk aversions studied in Herdegen et al. (2019). However, the small-cost asymptotics considered have the advantage that they formally extend to tractable formulas for much more general models, that we derive in Chapter 5.

4.5 Proofs

4.5.1 Elementary Integrals

We first discuss three elementary integrals that are used repeatedly in the proofs for this chapter.

Lemma 4.10. *For every positive constant $c > 0$, we have*

$$\int_0^\tau \int_0^v e^{-\frac{1}{c}\int_r^\tau \tanh\left(\frac{u}{c}\right)du} dr dv = c^2 \left(1 - \operatorname{sech}\left(\frac{\tau}{c}\right)\right).$$

Proof. Two elementary integrations yield

$$\begin{aligned}
\int_0^\tau \int_0^v e^{-\frac{1}{c}\int_r^\tau \tanh\left(\frac{u}{c}\right)du} dr dv &= \int_0^\tau \int_0^v e^{-\log \frac{\cosh\left(\frac{\tau}{c}\right)}{\cosh\left(\frac{r}{c}\right)}} dr dv \\
&= \int_0^\tau \int_0^v \frac{\cosh\left(\frac{r}{c}\right)}{\cosh\left(\frac{\tau}{c}\right)} dr dv \\
&= c \int_0^\tau \frac{\sinh\left(\frac{v}{c}\right)}{\cosh\left(\frac{\tau}{c}\right)} dv \\
&= c^2 \frac{\cosh\left(\frac{\tau}{c}\right) - 1}{\cosh\left(\frac{\tau}{c}\right)} = c^2 \left(1 - \operatorname{sech}\left(\frac{\tau}{c}\right)\right).
\end{aligned}$$

□

Lemma 4.11. *For every constant $c > 0$, and $0 < t < T$, we have*

$$\int_0^t e^{-\frac{1}{c}\int_{T-t}^{T-r} \tanh\left(\frac{u}{c}\right)du} dr \leq 2c.$$

Proof. Two elementary integrations and the definition of the hyperbolic cosine show that

$$
\begin{aligned}
\int_0^t e^{-\frac{1}{c}\int_{T-t}^{T-r}\tanh\left(\frac{u}{c}\right)du}dr &= \int_0^t e^{-\log\frac{\cosh\left(\frac{T-r}{c}\right)}{\cosh\left(\frac{T-t}{c}\right)}}dr \\
&= \int_0^t \frac{\cosh\left(\frac{T-t}{c}\right)}{\cosh\left(\frac{T-r}{c}\right)}dr \\
&= \cosh\left(\frac{T-t}{c}\right)\int_0^t \frac{2}{e^{\frac{T-r}{c}}+e^{-\frac{T-r}{c}}}dr \\
&\le 2\cosh\left(\frac{T-t}{c}\right)\int_0^t e^{-\frac{T-r}{c}}dr \\
&= 2c\cosh\left(\frac{T-t}{c}\right)e^{-\frac{T}{c}}\left(e^{\frac{t}{c}}-1\right) \\
&\le 2c\cosh\left(\frac{T-t}{c}\right)e^{-\frac{T-t}{c}} \\
&\le 2c.
\end{aligned}
$$

□

Lemma 4.12. *For every constant $c > 0$, and $T > 0$, we have*

$$
\int_0^T \left(1-\tanh\left(\frac{\tau}{c}\right)\right)^2 d\tau \le \int_0^T 1-\tanh\left(\frac{\tau}{c}\right)d\tau \le c.
$$

Proof. Since $\tanh \in [0,1]$ on the positive real line, we have

$$
\left(1-\tanh\left(\frac{\tau}{c}\right)\right)^2 \le 1-\tanh\left(\frac{\tau}{c}\right), \quad \tau \in [0,T],
$$

and in turn

$$
\int_0^T 1-\tanh\left(\frac{\tau}{c}\right)d\tau = \int_0^T \frac{2e^{-\frac{\tau}{c}}}{e^{\frac{\tau}{c}}+e^{-\frac{\tau}{c}}}d\tau \le \int_0^T 2e^{-\frac{2}{c}\tau}d\tau \le c.
$$

□

4.5.2 Proofs for Section 4.3

Proof of Lemma 4.1. As the right hand side of (4.11) is continuously differentiable with respect to (F_λ, E_λ), the local existence and uniqueness follow from standard arguments as in (Chicone, 1999, Theorem 1.261), for example. Moreover, the existence holds on a maximal interval of existence $I_{\max} := (T_{\min}, T_{\max})$, with $0 \in I_{\max}$. Since we only care about $t \ge 0$, with a little abuse of notation, we can conclude that there exists a unique local solution on a maximal interval of existence $[0, T_{\max})$. □

Proof of Theorem 4.2. Let (F_λ, E_λ) be the unique local solution of (4.11). In view of (4.12) and (4.14), F is also a local solution of (4.15). Conversely, any solution of this equation yields a local solution of (4.11) via (4.14), so that uniqueness for (4.15) is inherited from (4.11). First, notice that we can re-write the path-dependent ODE (4.15) as

$$F_\lambda'(\tau) = \widetilde{\gamma}\left(\alpha + \widehat{\gamma} H_\lambda(\tau)\right)^2 - \frac{1}{\lambda} F_\lambda(\tau)^2. \tag{4.35}$$

Recall from (4.13) that

$$H_\lambda(\tau) = -\widetilde{\gamma}\xi\alpha \int_0^\tau e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(\tau-r)} dr \le 0.$$

This gives us an estimation for the first term of (4.35) on the right-hand side.

To establish global existence, it remains to show that the local solution F_λ of (4.15) remains bounded on $[0,T]$. To this end, we first we show that F_λ is nonnegative. On $[0, T_{\max})$, the ODE (4.15) for F_λ yields

$$\begin{aligned}\left(F_\lambda(\tau) e^{\frac{1}{\lambda}\int_0^\tau F_\lambda(u)du}\right)' &= \left(F_\lambda'(\tau) + \frac{1}{\lambda} F_\lambda^2(\tau)\right) e^{\frac{1}{\lambda}\int_0^\tau F_\lambda(u)du} \\ &= \widetilde{\gamma}\left(\alpha + \widehat{\gamma} H_\lambda(\tau)\right)^2 e^{\frac{1}{\lambda}\int_0^\tau F_\lambda(u)du}.\end{aligned}$$

Together with the initial condition $F_\lambda(0) = 0$ it follows that

$$F_\lambda(\tau) = \widetilde{\gamma}\int_0^\tau \left(\alpha + \widehat{\gamma} H_\lambda(v)\right)^2 e^{-\frac{1}{\lambda}\int_v^\tau F_\lambda(u)du} dv.$$

As a consequence, F_λ is nonnegative on $[0, T_{\max})$.

We now use this initial estimate to establish the bounds (4.16) which show, in particular, that the local solution in fact remains bounded on $[0,T]$. To this end, we distinguish two cases. First suppose that $1 \ge \widehat{\gamma} \ge 0$. Since F_λ is nonnegative we then have, for $\tau \in [0, T_{\max}) \cap [0,T]$,

$$\alpha \ge \alpha + \widehat{\gamma} H_\lambda(\tau) \ge \alpha\left(1 - \widetilde{\gamma}\widehat{\gamma}\xi \int_0^\tau e^{-\widetilde{\gamma}\widehat{\gamma}\xi(\tau-r)} dr\right) \ge \alpha e^{-\gamma\xi T} > 0. \tag{4.36}$$

Whence, we can estimate the right-hand side of (4.15) from above and below, and in turn bound the local solution F_λ of (4.15) by the explicit global solutions of the corresonding standard Riccati equations.[7] Thus for $\tau \in [0, T_{\max}) \cap [0,T]$,

$$\frac{\widetilde{\gamma}^{1/2}\alpha}{e^{\widetilde{\gamma}\xi T}} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}}\tau\right) \le \frac{F_\lambda(\tau)}{\lambda^{1/2}} \le \widetilde{\gamma}^{1/2}\alpha \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right); \tag{4.37}$$

[7] For example, let $\tilde{F}(\tau) = \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha \tanh(\widetilde{\gamma}^{1/2}\alpha\tau/\lambda^{1/2})$, and notice it is the global solution of $\tilde{F}' = \widetilde{\gamma}\alpha^2 - \tilde{F}^2/\lambda$. Then, by (4.36), we have $\lambda(F - \tilde{F})' \le -(F + \tilde{F})(F - \tilde{F})$. Gronwall's lemma and the initial conditions $F(0) = \tilde{F}(0) = 0$ in turn show that $F - \tilde{F} \le 0$, first on any compact subset of $[0, T_{\max})$ and in turn on $[0, T_{\max})$. The lower bound follows analogously.

in particular, $T_{\max} > T$. Similarly, if $-1 \leq \widehat{\gamma} \leq 0$, then for $\tau \in [0, T_{\max}) \cap [0, T]$,

$$\alpha \leq \alpha + \widehat{\gamma} H_\lambda(\tau) \leq \alpha \left(1 - \widetilde{\gamma}\widehat{\gamma}\xi \int_0^\tau e^{-\widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)} dr\right) \leq \alpha e^{\gamma\xi T}. \tag{4.38}$$

The local solution F_λ of (4.15) can therefore again be bounded from above and below by the explicit global solutions of standard Ricatti equations:

$$\widetilde{\gamma}^{1/2}\alpha \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right) \leq \frac{F_\lambda(\tau)}{\lambda^{1/2}} \leq \frac{\widetilde{\gamma}^{1/2}\alpha}{e^{-\widetilde{\gamma}\xi T}} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{-\widetilde{\gamma}\xi T}}\tau\right). \tag{4.39}$$

Together, (4.37), (4.39) yield the estimate (4.16) and thus global existence for (4.15) and (4.11).

Moreover, when $-1 \leq \widehat{\gamma} \leq 0$, if $\widetilde{\gamma}\xi < \widetilde{\gamma}^{1/2}\alpha/\lambda^{1/2}$, i.e. $\lambda < \alpha^2/\widetilde{\gamma}\xi^2$ we can further infer that

$$\begin{aligned}
\alpha + \widehat{\gamma} H_\lambda(\tau) &\leq \alpha \left(1 - \widetilde{\gamma}\widehat{\gamma}\xi \int_0^\tau \frac{\cosh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}} r\right)}{\cosh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right)} e^{-\widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)} dr\right) \\
&\leq \alpha \left(1 - \frac{2\widetilde{\gamma}\widehat{\gamma}\xi e^{-\widetilde{\gamma}\widehat{\gamma}\xi\tau}}{\cosh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right)} \int_0^\tau e^{\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}} r} e^{\widetilde{\gamma}\widehat{\gamma}\xi r} dr\right) \\
&\leq \alpha \left(1 - \frac{2\widetilde{\gamma}\widehat{\gamma}\xi e^{-\widetilde{\gamma}\widehat{\gamma}\xi\tau}}{\cosh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right)} \frac{e^{\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau} e^{\widetilde{\gamma}\widehat{\gamma}\xi\tau}}{\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}} + \widetilde{\gamma}\widehat{\gamma}\xi}\right) \\
&\leq \alpha(1 + 2) = 3\alpha.
\end{aligned}$$

Hence as $\lambda \downarrow 0$, we have a tighter upper bound for F_λ:

$$\frac{\widetilde{\gamma}^{1/2}\alpha}{e^{\widetilde{\gamma}\xi T}} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}}\tau\right) \leq \frac{F_\lambda(\tau)}{\lambda^{1/2}} \leq 3\widetilde{\gamma}^{1/2}\alpha \tanh\left(3\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\tau\right). \tag{4.40}$$

□

Corollary 4.13. *Fix $T, \xi, \widetilde{\gamma} > 0$ and $\widehat{\gamma} \in [-1, 1]$. Then, the solutions F_λ, E_λ of (4.11) for transaction costs $\lambda \downarrow 0$ satisfy the following uniform bounds:*

$$\|F_\lambda\|_\infty \leq 3\lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha = O(\lambda^{1/2}), \quad \|E_\lambda\|_\infty \leq 4\frac{|\widehat{\gamma}|\widetilde{\gamma}\xi^2 e^{4\widetilde{\gamma}\xi T}}{\alpha}\lambda = O(\lambda). \tag{4.41}$$

Proof. (4.40) and $\|\tanh\|_\infty = 1$ directly yield the first bound in (4.41). To derive the corresponding estimate for E_λ, first recall (4.36) (4.38) from the proof of Theorem 4.2 that for every $\tau \in [0, T]$,

$$\alpha e^{-\widetilde{\gamma}\xi T} \leq \alpha + \widehat{\gamma} H_\lambda(\tau) \leq \alpha e^{\widetilde{\gamma}\xi T}. \tag{4.42}$$

Next, observe that the ODE (4.11) for E_λ and (4.14) show that

$$
\begin{aligned}
E'_\lambda(\tau) &= \widetilde{\gamma}\widehat{\gamma}\frac{\xi}{\alpha}\left(\alpha + \widehat{\gamma}H_\lambda(\tau)\right)H_\lambda(\tau) - \frac{1}{\lambda}F_\lambda E_\lambda \\
&= -\widehat{\gamma}\widetilde{\gamma}^2\xi^2\left(\alpha + \widehat{\gamma}H_\lambda(\tau)\right)\int_0^\tau e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(\tau - r)}dr - \frac{1}{\lambda}F_\lambda E_\lambda.
\end{aligned}
$$

Together with $E(0) = 0$, it follows that

$$
\begin{aligned}
E_\lambda(\tau) &= \widetilde{\gamma}\widehat{\gamma}\frac{\xi}{\alpha}\int_0^\tau \left(\alpha + \widehat{\gamma}H_\lambda(v)\right)H_\lambda(v)e^{-\frac{1}{\lambda}\int_v^\tau F_\lambda(u)du}dv \\
&= -\widehat{\gamma}\widetilde{\gamma}^2\xi^2\int_0^\tau \left(\left(\alpha + \widehat{\gamma}H_\lambda(v)\right)\int_0^v e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(v - r)}dr\right)dv. \qquad (4.43)
\end{aligned}
$$

Together, (4.43), (4.42), and (4.16) show that

$$
\begin{aligned}
|E_\lambda(\tau)| &\leq |\widehat{\gamma}|\widetilde{\gamma}^2\xi^2\int_0^\tau \left(\alpha + \widehat{\gamma}H_\lambda(\tau)\right)\int_0^v e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(v - r)}dr\ dv \\
&\leq |\widehat{\gamma}|\widetilde{\gamma}^2\xi^2\alpha\int_0^\tau e^{\widetilde{\gamma}\xi T}\int_0^v e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du + \widetilde{\gamma}\xi T}drdv \\
&\leq |\widehat{\gamma}|\widetilde{\gamma}^2\xi^2\alpha e^{2\widetilde{\gamma}\xi T}\int_0^\tau\int_0^v e^{-\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}e^{\widetilde{\gamma}\xi T}}\int_r^\tau \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}e^{\widetilde{\gamma}\xi T}}u\right)du}drdv.
\end{aligned}
$$

The second bound in (4.41) in turn follows from Lemma 4.10. □

4.5.3 Proofs for Section 4.4

Proof of Lemma 4.4. Differentiation shows that F solves the Riccati equation $F = \widetilde{\gamma}\alpha^2 - F^2$ with initial condition $F(0) = 0$. Combining this with the ODE (4.15) for F_λ, we obtain

$$
\begin{aligned}
&\frac{d}{d\tau}\left(F_\lambda(\tau) - \lambda^{1/2}F\left(\frac{\tau}{\lambda^{1/2}}\right)\right) \\
&= \widehat{\gamma}\widetilde{\gamma}\left(2\alpha + \widehat{\gamma}H_\lambda\right)H_\lambda - \frac{1}{\lambda}\left(F_\lambda^2(\tau) - \lambda F^2\left(\frac{\tau}{\lambda^{1/2}}\right)\right) \\
&= -\frac{1}{\lambda}\left(F_\lambda(\tau) - \lambda^{1/2}F\left(\frac{\tau}{\lambda^{1/2}}\right)\right)\left(F_\lambda(\tau) + \lambda^{1/2}F\left(\frac{\tau}{\lambda^{1/2}}\right)\right) \\
&\quad + \widehat{\gamma}\widetilde{\gamma}\left(2\alpha + \widehat{\gamma}H_\lambda(\tau)\right)H_\lambda(\tau).
\end{aligned}
$$

Variation of constants in turn yields

$$
\begin{aligned}
&F_\lambda(\tau) - \lambda^{1/2}F\left(\frac{\tau}{\lambda^{1/2}}\right) \\
&= -\widehat{\gamma}\widetilde{\gamma}^2\xi\alpha\int_0^\tau \left(2\alpha + \widehat{\gamma}H_\lambda(v)\right)\int_0^v e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du - \int_{\frac{v}{\lambda^{1/2}}}^{\frac{\tau}{\lambda^{1/2}}} F(u)du - \widetilde{\gamma}\widehat{\gamma}\xi(v - r)}drdv.
\end{aligned}
$$

Together with (4.42), $F_\lambda, F \geq 0$, and the lower bound for F_λ from (4.16), it follows that

$$\begin{aligned}
&\left|F_\lambda(\tau) - \lambda^{1/2} F\left(\frac{\tau}{\lambda^{1/2}}\right)\right| \\
&\leq |\widehat{\gamma}|\widetilde{\gamma}^2 \xi\alpha \int_0^\tau \left(\alpha + \alpha e^{\widetilde{\gamma}\xi T}\right) \int_0^v e^{-\frac{1}{\lambda}\int_r^\tau F_\lambda(u)du} e^{\widetilde{\gamma}\xi T} dr\, dv \\
&\leq 2|\widehat{\gamma}|\widetilde{\gamma}^2 \alpha^2 \xi e^{2\widetilde{\gamma}\xi T} \int_0^\tau \int_0^v e^{-\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}} \int_r^\tau \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}}(T-u)\right) du} dr dv.
\end{aligned}$$

Whence, (4.23) is a consequence of Lemma 4.10. $\square$

As a preparation for the proof of Theorem 4.5, we first establish the asymptotic order of $\Delta\varphi^{1,\lambda}$ for small transaction costs λ. In order to write down the formulas without changing the page to landscape setting, let us assume $\varphi^1_{0-} = \frac{s\gamma^2}{\gamma^1+\gamma^2}$ without loss of generality. For the cases where $\varphi^1_{0-} \neq \frac{s\gamma^2}{\gamma^1+\gamma^2}$, we need to introduce an extra term $\exp\{-\frac{1}{\lambda}\int_{T-t}^T F_\lambda(u)du\}|\varphi^1_{0-} - \frac{s\gamma^2}{\gamma^1+\gamma^2}|$ and $\exp\{-\frac{\gamma^{1/2}\alpha}{\lambda^{1/2}}(T-t)\}|\varphi^1_{0-} - \frac{s\gamma^2}{\gamma^1+\gamma^2}|$ in the estimation for $|\Delta\varphi_t|$ and $|\widehat{\Delta}_t|$, respectively. With the triangle inequality and elementary integrations we can still achieve the same order for Theorem 4.5 and Theorem 4.8.

Lemma 4.14. *For $t \in [0,T]$ and $p \geq 1$, we have*

$$\mathbb{E}\left[\left|\Delta\varphi^{1,\lambda}_t\right|^p\right]^{1/p} = O(\lambda^{1/4}). \tag{4.44}$$

Proof. The linear SDE (4.17) for $\Delta\varphi^{1,\lambda}_t$ has the solution

$$\Delta\varphi^{1,\lambda}_t = \frac{1}{\lambda}\int_0^t \left(e^{-\frac{1}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} E_\lambda(T-r)W_r\right) dr + \frac{\xi}{\alpha}\int_0^t e^{-\frac{1}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} dW_r.$$

Notice that, for every $t \in [0,T]$,

$$\Delta\varphi^{1,\lambda}_t \sim N\left(0, \frac{\xi^2}{\alpha^2}\int_0^t e^{-\frac{2}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} dr\right).$$

The lower bound for F_λ from (4.16) and Lemma 4.11 yield

$$\begin{aligned}
\int_0^t e^{-\frac{2}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} dr &\leq \int_0^t e^{-\frac{1}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} dr \\
&\leq \int_0^t e^{-\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}}\int_{T-t}^{T-r} \tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2} e^{\widetilde{\gamma}\xi T}} u\right) du} dr \\
&\leq 2\lambda^{1/2}\frac{e^{\widetilde{\gamma}\xi T}}{\widetilde{\gamma}^{1/2}\alpha}.
\end{aligned} \tag{4.45}$$

By the formula for the moments of centred Gaussian distribution, we have

$$\mathbb{E}\left[\left|\Delta\varphi^{1,\lambda}_t\right|^p\right]^{1/p} = \frac{\xi}{\alpha}\left(2\int_0^t e^{-\frac{2}{\lambda}\int_{T-t}^{T-r} F_\lambda(u)du} dr\right)^{1/2} \left(\frac{\Gamma(\frac{p+1}{2})}{\pi^{1/2}}\right)^{1/p} = O(\lambda^{1/4}).$$

This is the desired estimate. $\square$

Proof of Theorem 4.5. The definitions of $\mathcal{Y}^\lambda$ in (4.18), an elementary integration as in Lemma 4.12, the estimates for F_λ, E_λ from Corollary 4.13 and for $F_\lambda(\tau) - \lambda^{1/2} F\left(\frac{\tau}{\lambda^{1/2}}\right)$ from Lemma 4.4 yield

$$
\begin{aligned}
&\left|\mathcal{Y}_t^\lambda - \lambda^{1/2} A(t)\right| \\
&\leq |\widehat{\gamma}|\bar{\gamma} s \left| \int_t^T 2\xi \lambda^{1/2} \widetilde{\gamma}^{1/2} \alpha - \left(2\alpha + \widehat{\gamma} H_\lambda(T-u)\right) H_\lambda(T-u) du \right| \\
&\leq |\widehat{\gamma}|\bar{\gamma} s \int_t^T \left[2\alpha \|E_\lambda\|_\infty + |\widehat{\gamma}| \|H_\lambda\|_\infty^2 + 2\xi |\lambda^{1/2} \widetilde{\gamma}^{1/2} \alpha - F_\lambda(T-u)|\right] du \\
&= O(\lambda) + 2\xi \lambda^{1/2} \bar{\gamma} s \int_0^{T-t} \left| \widetilde{\gamma}^{1/2} \alpha - F\left(\frac{\tau}{\lambda^{1/2}}\right) \right| + \left| \frac{F_\lambda(\tau)}{\lambda^{1/2}} - F\left(\frac{\tau}{\lambda^{1/2}}\right) \right| d\tau \\
&= O(\lambda) + 2\xi \lambda^{1/2} \widetilde{\gamma}^{1/2} \alpha \bar{\gamma} s \int_0^{T-t} 1 - \tanh\left(\frac{\widetilde{\gamma}^{1/2} \alpha}{\lambda^{1/2}} \tau\right) d\tau \\
&\leq O(\lambda) + 2\xi \lambda^{1/2} \widetilde{\gamma}^{1/2} \alpha \bar{\gamma} s \int_0^{T} 1 - \tanh\left(\frac{\widetilde{\gamma}^{1/2} \alpha}{\lambda^{1/2}} \tau\right) d\tau \\
&= O(\lambda).
\end{aligned}
$$

Next observe that, by the respective definitions in (4.17) and (4.26),

$$
d\left(\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda\right) = \frac{E_\lambda(T-t) W_t + \lambda^{1/2} F\left(\frac{T-t}{\lambda^{1/2}}\right) \Delta_t^\lambda - F_\lambda(T-t) \Delta\varphi_t^{1,\lambda}}{\lambda} dt.
$$

Whence, variation of constants shows that the solution of this linear (random) ODE is given by

$$
\begin{aligned}
&\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda \\
&= \int_0^t \frac{E_\lambda(T-r) W_r + \left(\lambda^{1/2} F\left(\frac{T-r}{\lambda^{1/2}}\right) - F_\lambda(T-r)\right) \Delta\varphi_r^{1,\lambda}}{\lambda e^{\frac{1}{\lambda^{1/2}} \int_{T-t}^{T-r} F\left(\frac{u}{\lambda^{1/2}}\right) du}} dr.
\end{aligned}
$$

By Minkowski's integral inequality (Hardy et al., 1952, Theorem 202),

$$
\begin{aligned}
&\mathbb{E}\left[\left|\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda\right|^p\right]^{1/p} \\
&\leq \int_0^t \frac{\mathbb{E}\left[\left|E_\lambda(T-r) W_r + \left(\lambda^{1/2} F\left(\frac{T-r}{\lambda^{1/2}}\right) - F_\lambda(T-r)\right) \Delta\varphi_r^{1,\lambda}\right|^p\right]^{1/p}}{\lambda e^{\frac{1}{\lambda^{1/2}} \int_{T-t}^{T-r} F\left(\frac{u}{\lambda^{1/2}}\right) du}} dr \\
&\leq \int_0^t \frac{\left[O(\lambda^p) \mathbb{E}[|W_r|^p] + O(\lambda^p) \mathbb{E}\left[\left|\Delta\varphi_r^{1,\lambda}\right|^p\right]\right]^{1/p}}{\lambda} e^{-\frac{1}{\lambda^{1/2}} \int_{T-t}^{T-r} F\left(\frac{u}{\lambda^{1/2}}\right) du} dr \\
&= \int_0^t \left[\mathbb{E}[|W_r|^p] + \mathbb{E}\left[\left|\Delta\varphi_r^{1,\lambda}\right|^p\right]\right]^{1/p} e^{-\frac{1}{\lambda^{1/2}} \int_{T-t}^{T-r} F\left(\frac{u}{\lambda^{1/2}}\right) du} dr.
\end{aligned}
$$

Now, notice that

$$
\sup_{0 \leq r \leq T} \mathbb{E}[|W_r|^p] = O(1), \qquad \sup_{0 \leq r \leq T} \mathbb{E}\left[\left|\Delta\varphi_r^\lambda\right|^p\right] = O(\lambda^{p/4}).
$$

These two estimates, Corollary 4.13, Lemma 4.4, and Lemma 4.11 in turn show that $\|\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda\|_p = O(\lambda^{1/2})$ as asserted.

Finally, this estimate together with the definitions of $Y_t^{1,\lambda}$ in (4.19), the estimates for F_λ, E_λ from Corollary 4.13 and for $F_\lambda(\tau) - \lambda^{1/2}F\left(\frac{\tau}{\lambda^{1/2}}\right)$ from Lemma 4.4 yield

$$\begin{aligned}
&\mathbb{E}\left[\left|Y_t^{1,\lambda} + \lambda^{1/2}F\left(\frac{T-t}{\lambda^{1/2}}\right)\Delta_t^\lambda\right|^p\right] \\
&= \mathbb{E}\left[\left|E_\lambda(T-t)W_t + \lambda^{1/2}F\left(\frac{T-t}{\lambda^{1/2}}\right)\Delta_t^\lambda - F_\lambda(T-t)\Delta\varphi_t^{1,\lambda}\right|^p\right] \\
&\leq 3^{p-1}\left[O(\lambda^p)\|W_t\|_p^p + O(\lambda^p)\|\Delta\varphi_t^{1,\lambda}\|_p^p + \lambda^{p/2}\|F\|_\infty^p\ \|\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda\|_p^p\right] \\
&= O(\lambda^p) + O(\lambda^{5p/4}) + O(\lambda^p).
\end{aligned}$$

This establishes the last asserted estimate and thereby completes the proof. □

Proof of Corollary 4.6. The approximation of the frictional equilibrium price from Corollary 4.3 follows directly from Theorem 4.5 and the Minkowski inequality. For the approximation of its drift rate, plug (4.8) and (4.9) into the BSDEs (4.3), (4.4). The stated approximation then follows from Lemma 4.4 and Theorem 4.5. To obtain the expansion of the expected drift, take expectations and use the fact that the W_t and $\Delta\varphi_t^{1,\lambda}$ both have expectation zero; the assertion then also follows from Lemma 4.4. Finally, for the approximation of the volatility, recall from (4.8) and (4.9) that the volatility of the frictional equilibrium price from Corollary 4.3 is

$$\alpha + \widehat{\gamma}(E_\lambda(T-t) - F_\lambda(T-t)\xi/\alpha).$$

Whence, the started expansion follows from Lemma 4.4, and the E_λ term is dropped due to $\|E_\lambda\|_\infty/\|F_\lambda\|_\infty = O(\lambda^{1/2})$. □

Proof of Lemma 4.7. By Lemma 4.12, we have

$$\int_0^T \left|F_\lambda(\tau) - \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\right| d\tau \leq \int_0^T \lambda^{1/2}\left|\widetilde{\gamma}^{1/2}\alpha - F\left(\frac{\tau}{\lambda^{1/2}}\right)\right| + O(\lambda)du = O(\lambda)$$

as asserted. □

Proof of Theorem 4.8. In view of Theorem 4.5, we only need to show that $\|\widehat{\Delta}^\lambda - \Delta^\lambda\|_{\mathbb{H}^p}$ is of order $O(\lambda^{1/2})$. By (4.28),

$$\widehat{\Delta}_t^\lambda \sim N\left(0,\ \frac{\xi^2\lambda^{1/2}}{2\alpha^3\widetilde{\gamma}^{1/2}}\left(1 - e^{-2\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}t}\right)\right), \quad t \in [0,T].$$

As a consequence,

$$\mathrm{Var}[\Delta_t^\lambda] \leq \frac{\xi^2\lambda^{1/2}}{2\alpha^3\widetilde{\gamma}^{1/2}}, \quad t \in [0,T].$$

The formula for the p-th centred moment of normally distributed random variables in turn yields

$$\sup_{0\le r\le T} \mathbb{E}\left[\left|\widehat{\Delta}_r\right|^p\right] \le \left(\frac{\xi^2\lambda^{1/2}}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{p/2} \frac{\Gamma(\frac{p+1}{2})}{\pi^{1/2}} \le \lambda^{p/4}\left(\frac{p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{p/2}.$$

Combining the dynamics of (4.26) and (4.28), we obtain

$$d\left(\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right) = -\frac{F\left(\frac{T-t}{\lambda^{1/2}}\right)}{\lambda^{1/2}}\left(\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right)dt + \frac{\widehat{\Delta}_t^\lambda}{\lambda^{1/2}}\left(\widetilde{\gamma}^{1/2}\alpha - F\left(\frac{T-t}{\lambda^{1/2}}\right)\right)dt.$$

Whence, variation of constants shows that the solution of this linear (random) ODE is given by

$$\Delta_t^\lambda - \widehat{\Delta}_t^\lambda = \int_0^t \frac{e^{-\frac{1}{\lambda^{1/2}}\int_{T-t}^{T-r}F\left(\frac{u}{\lambda^{1/2}}\right)du}}{\lambda^{1/2}}\left(\widetilde{\gamma}^{1/2}\alpha - F\left(\frac{T-r}{\lambda^{1/2}}\right)\right)\widehat{\Delta}_r^\lambda dr.$$

By Minkowski's integral inequality, Lemma 4.11, and the monotonicity of the hyperbolic tangent, it follows that

$$\begin{aligned}
&\left(\mathbb{E}\left[\left|\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right|^{2p}\right]\right)^{1/2p} \\
&\le \int_0^t \frac{e^{-\frac{1}{\lambda^{1/2}}\int_{T-t}^{T-r}F\left(\frac{u}{\lambda^{1/2}}\right)du}}{\lambda^{1/2}}\left(\widetilde{\gamma}^{1/2}\alpha - F\left(\frac{T-r}{\lambda^{1/2}}\right)\right)\mathbb{E}\left[\left|\widehat{\Delta}_r^\lambda\right|^{2p}\right]^{1/2p} dr \\
&\le \lambda^{1/4}\left(\frac{2p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{1/2}\int_0^t \frac{e^{-\frac{1}{\lambda^{1/2}}\int_{T-t}^{T-r}F\left(\frac{u}{\lambda^{1/2}}\right)du}}{\lambda^{1/2}}\left(\widetilde{\gamma}^{1/2}\alpha - F\left(\frac{T-r}{\lambda^{1/2}}\right)\right)dr \\
&\le \lambda^{1/4}\left(\frac{2p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{1/2}\int_0^t \frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}\left(1-\tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}(T-r)\right)\right)e^{-\frac{1}{\lambda^{1/2}}\int_{T-t}^{T-r}F\left(\frac{u}{\lambda^{1/2}}\right)du}dr \\
&\le \lambda^{1/4}\left(\frac{2p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{1/2}\left(1-\tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}(T-t)\right)\right)\int_0^t \frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}e^{-\frac{1}{\lambda^{1/2}}\int_{T-t}^{T-r}F\left(\frac{u}{\lambda^{1/2}}\right)du}dr \\
&\le 2\lambda^{1/4}\left(\frac{2p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\right)^{1/2}\left(1-\tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}(T-t)\right)\right).
\end{aligned}$$

Together with Lemma 4.12 and Minkowski's integral inequality, this shows that

$$
\begin{aligned}
\|\Delta^\lambda - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p} &= \mathbb{E}\left[\left(\int_0^T \left|\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right|^2 dt\right)^{p/2}\right]^{1/p} \\
&\leq \mathbb{E}\left[\left(\int_0^T \left|\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right|^2 dt\right)^{p}\right]^{1/2p} \\
&= \left(\mathbb{E}\left[\left(\int_0^T \left|\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right|^2 dt\right)^{p}\right]^{1/p}\right)^{1/2} \\
&\leq \left(\int_0^T \mathbb{E}\left[\left|\Delta_t^\lambda - \widehat{\Delta}_t^\lambda\right|^{2p}\right]^{1/p} dt\right)^{1/2} \\
&\leq \left(4\lambda^{1/2}\frac{2p\xi^2}{\alpha^3\widetilde{\gamma}^{1/2}}\int_0^T \left(1-\tanh\left(\frac{\widetilde{\gamma}^{1/2}\alpha}{\lambda^{1/2}}(T-t)\right)\right)^2 dt\right)^{1/2} \\
&\leq \left(\frac{8p\xi^2}{\alpha^4\widetilde{\gamma}}\lambda\right)^{1/2} = O(\lambda^{1/2}).
\end{aligned}
$$

Therefore, again by the Minkowski's integral inequality and triangle inequality for the $\mathbb{H}^p$ norm, we have that

$$
\begin{aligned}
\|\Delta\varphi^{1,\lambda} - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p} &\leq \|\Delta\varphi^{1,\lambda} - \Delta^\lambda\|_{\mathbb{H}^p} + \|\Delta^\lambda - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p} \\
&\leq \left(\int_0^T \mathbb{E}\left[\left|\Delta\varphi_t^{1,\lambda} - \Delta_t^\lambda\right|^{2p}\right]^{1/p} dt\right)^{1/2} + \|\Delta^\lambda - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p} \\
&= \left(\int_0^T O(\lambda)dt\right)^{1/2} + O(\lambda^{1/2}).
\end{aligned}
$$

Similarily,

$$
\|Y^{1,\lambda} + \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} \leq \|Y^{1,\lambda} + \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\Delta^\lambda\|_{\mathbb{H}^p} + \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\|\Delta^\lambda - \widehat{\Delta}^\lambda\|_{\mathbb{H}^p}.
$$

□

Proof of Corollary 4.9. The approximation for the frictional equilibrium price, the equilibrium return, expected return as well as the equilibrium volatility follows similar argument as in the proof of Corollary 4.6. More specifically, recalling that

$$
\widetilde{\gamma} = \frac{\gamma^1+\gamma^2}{2}, \quad \bar{\gamma} = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}, \quad \widehat{\gamma} = \frac{\gamma^1-\gamma^2}{\gamma^1+\gamma^2},
$$

from (4.22) we have

$$
\begin{aligned}
A(t) &= \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s\lambda^{1/2}(T-t), \\
A'(t) &= -\frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s\lambda^{1/2}.
\end{aligned}
$$

Since

$$S_t = \bar{S}_t + \mathcal{Y}_t + \widehat{\gamma} Y_t^1,$$

from Theorem 4.8 we have

$$\begin{aligned}
&\|S_t - \bar{S}_t - \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}}\xi\alpha s\lambda^{1/2}(T-t)\|_{\mathbb{H}^p} \\
&= \|S_t - \bar{S}_t - A(t)\|_{\mathbb{H}^p} \\
&\leq \|S - \bar{S} - \mathcal{Y}\|_{\mathbb{H}^p} + \|\mathcal{Y} - A(t)\|_{\infty} \\
&\leq |\widehat{\gamma}|\|Y^{1,\lambda} + \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} + |\widehat{\gamma}|\lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\|\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} + O(\lambda) \\
&= O(\lambda) + O(\lambda^{3/4}) + O(\lambda).
\end{aligned}$$

The estimation for the drift rate and the volatility follows similarly as in Corollary 4.6 with the help of Lemma 4.7. For the trading rate, recall that

$$\dot{\varphi}_t^1 = \frac{Y_t^{1,\lambda}}{\lambda}.$$

Hence Theorem 4.8 yields the asserted approximation of the turnover rates:

$$\|\dot{\varphi}^1 + \lambda^{-1/2}\widetilde{\gamma}^{1/2}\alpha\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} = \frac{1}{\lambda}\|Y^{1,\lambda} + \lambda^{1/2}\widetilde{\gamma}^{1/2}\alpha\widehat{\Delta}^\lambda\|_{\mathbb{H}^p} = \frac{1}{\lambda}O(\lambda) = O(1).$$

□

Chapter 5

General Formal Small-Cost Asymptotics

5.1 Introduction

In Chapter 4, we have considered the simplest version of the equilibrium models with transaction costs from Chapter 3. For constant quadratic trading costs, zero aggregate endowment, and linear state dynamics, the FBSDE system from Chapter 3 could then be reduced to a system of Riccati ODEs and finally a path-dependent but scalar ODE. Existence for this equation could in turn be established using elementary comparison arguments. The bounds established in this argument then allowed us to derive a rigorous asymptotic expansion of the solution for small transaction costs. Starting from this approximation of the ODEs, we finally constructed a leading-order approximation of the original FBSDE system and, in turn, the equilibrium with transaction costs.

For more general state dynamics and/or non-constant or non-quadratic costs, the reduction to a system of Riccati equations is no longer possible. In Chapter 3, we have proposed a deep learning algorithm that presents one possible approach to dealing with the FBSDE system in such more general settings. It has the advantage of being very flexible, but its implementation is nontrivial. Indeed, even with substantial tuning of hyper-parameters and computation on a GPU unit, the algorithm only produces stable results for relatively short time horizons of a few months.

Buoyed by the accuracy of the asymptotic expansions form Section 4.4 for the model with linear state dynamics, we therefore now formally extend these results to the general setting from Chapter 3. To this end, we make a scaling ansatz motivated by the structure of the rigorous expansion in Chapter 4 and in turn determine the leading-order coefficients by matching the leading-order terms in the FBSDEs.

Given that the lack of even basic existence and uniqueness results for the FBSDE systems appearing here, this analysis is necessarily partially heuristic, in that rigorous convergence proofs remain a challenging direction for future research. However, our formal results suggest that tractable asymptotic approximations are available even in surprisingly general settings. In partial equilibrium models with exogenous price dynamics, such general asymptotic results have been obtained by Soner and Touzi (2013); Martin (2014); Kallsen and Muhle-Karbe (2015, 2017); Ahrens (2015); Moreau, Muhle-Karbe, and Soner (2017); Cai, Rosenbaum, and Tankov (2017); Herdegen and Muhle-Karbe (2018); Cayé, Herdegen, and Muhle-Karbe (2019), for example. A specific equilibrium model with fixed costs is studied using formal perturbation arguments by Lo et al. (2004). The analysis in this chapter suggests that a general probabilistic version of their approach leads to tractable results even for time-dependent and random transaction costs and arbitrary state dynamics like in corresponding results for partial equilibrium models (Moreau, Muhle-Karbe, and Soner, 2017; Herdegen and Muhle-Karbe, 2018). In the present equilibrium context, this allows to shed light on the impact of "liquidity risk" on asset prices. This is studied by Acharya and Pedersen (2005) in a model with one-shot investors; our asymptotic analysis makes it possible to extend this analysis to settings where agents optimize dynamically over time.

Let us briefly comment on the difficulties in making these formal results rigorous. Proofs of the partial-equilbrium expansions are usually either based on stability results for viscosity solutions of partial differential equations (cf., e.g., Soner and Touzi (2013); Moreau et al. (2017)) or convex duality (Ahrens, 2015; Herdegen and Muhle-Karbe, 2018; Cayé, Herdegen, and Muhle-Karbe, 2019). It is not clear how to generalize either approach to equilibrium models. Indeed, the equilibrium price is not linked to a dynamic programming principle and not the solution of a convex optimization problem. Accordingly, existence needs to be established using methods tailored to the specific structure of the model at hand, which is generally highly nontrivial already for frictionless equilibrium models, compare, e.g., Kardaras, Xing, and Žitković (2015); Xing and Žitković (2018); Kramkov and Pulido (2016a). With wellposedness results at hand, asymptotic expansions would then need to be established in a second step, see, e.g., Kramkov and Pulido (2016b) for the expansion of an incomplete frictionless equilibrium model around its complete counterpart. Developing such results in a general model with trading costs is an important but challenging direction for future research.

The remainder of this chapter is organized as follows. For the convenience of the reader, the model with general transaction costs and state dynamics is recalled from Chapter 3 in Section 5.2. Subsequently, in Section 5.3, we discuss the derivation of the asymptotic approximation with general state dynamics for quadratic costs, and show that our results indeed lead to an extension of Theorem 4.8 in Chapter 4. In Section 5.4, we give four examples to illustrate the effect of liquidity risk. Finally, in Section 5.5 and 5.6 we extend the asymptotic approximations for general power costs and proportional costs, respectively.

Notation. We fix a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ with finite time horizon $T > 0$, where the filtration is generated by a d-dimensional standard Brownian motion $W = (W_t)_{t\in[0,T]}$. Throughout, let $\|\cdot\|$ be the 2-norm of a real-valued vector. For $p \geq 1$, we denote by $\mathbb{H}^p$ the $\mathbb{R}$-valued, progressively measurable processes $X = (X_t)_{t\in[0,T]}$ that satisfy

$$\|X\|_{\mathbb{H}^p} := \left(\mathbb{E}\left[\left(\int_0^T |X_t|^2 dt\right)^{p/2}\right]\right)^{1/p} < \infty.$$

5.2 Model

In this chapter, we study the small-cost asymptotics of the general model from Chapter 3. For the convenience of the reader, we first briefly recall the exogenous inputs of the model, the agents' optimization problems, and the FBSDEs describing asset prices and the corresponding optimal trading strategies.

Throughout this chapter, we fix a finite time horizon $T > 0$. Randomness is generated by a d-dimensional standard Brownian motion $(W_t)_{t\in[0,T]}$, which drives a k-dimensional Markovian state variable $(X_t)_{t\in[0,T]}$ with dynamics

$$dX_t = b(X_t)dt + a(X_t)dW_t, \qquad X_0 = x_0. \tag{??}$$

Here, the drift rate b and diffusion matrix a take values in $\mathbb{R}^k$ and $\mathbb{R}^{k\times d}$, respectively. The corresponding infinitesimal generator is denoted by

$$\mathcal{L}^X := b^\top \nabla + \frac{1}{2}\mathrm{tr}\left(aa^\top \nabla^2\right). \tag{5.1}$$

We consider two agents indexed by $n = 1, 2$ that receive (cumulative) random endowments

$$d\zeta_t^n = \xi^n(X_t)dW_t, \quad \text{where } \xi^n : \mathbb{R}^k \to \mathbb{R}^d. \tag{5.2}$$

To hedge against the fluctuations of these endowment streams, the agents trade a safe and a risky asset. The price of the safe asset is exogenous and normalized to one. The price of the risky asset has dynamics

$$dS_t = \mu_t dt + \sigma_t dW_t \tag{5.3}$$

and matches an exogenous liquidating dividend at the terminal time:

$$S_T = \mathfrak{S}(X_T). \tag{5.4}$$

Our goal is to determine the (scalar) expected returns process μ and the ($\mathbb{R}^d$-valued) volatility process σ that match the agents' demand to the fixed supply $s \geq 0$ of the risky asset.

5.2.1 Frictionless Optimization and Equilibrium

Without transaction costs, the agents choose their positions $(\psi_t)_{t\in[0,T]}$ in the risky asset to maximize expected returns penalized for the corresponding quadratic variations as in the previous chapters:

$$\bar{J}^n_T(\psi) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\|\psi_t\sigma_t + \xi^n_t\|^2\right)dt\right]. \tag{5.5}$$

For given price dynamics, each agents' optimal strategy (3.4) for (5.5) is then readily determined by pointwise optimization. Matching their sum to the supply of the risky asset in turn pins down the frictionless equilibrium return 3.5. Plugging the latter into the dynamics (5.3) and taking into account the terminal condition (5.4), it follows that the frictionless equilibrium price of the risky asset and its volatility are determined by the following scalar quadratic backward stochastic differential equation (BSDE):

$$d\bar{S}_t = \bar{\gamma}\left[\bar{\sigma}_t s + \xi^1_t + \xi^2_t\right]\bar{\sigma}_t^\top dt + \bar{\sigma}_t dW_t, \quad S_T = \mathfrak{S}(X_T), \tag{5.6}$$

where recall that

$$\bar{\gamma} = \frac{\gamma^1\gamma^2}{\gamma^1+\gamma^2}.$$

The agents' frictionless equilibrium strategies corresponding to the frictionless equilibrium price (5.6) are

$$\bar{\varphi}^1_t = \frac{s\gamma^2}{\gamma^1+\gamma^2} + \frac{(\gamma^2\xi^2_t - \gamma^1\xi^1_t)\bar{\sigma}_t^\top}{(\gamma^1+\gamma^2)\bar{\sigma}_t\bar{\sigma}_t^\top} \qquad \bar{\varphi}^2_t = s - \bar{\varphi}^1_t, \qquad t\in[0,T]. \tag{5.7}$$

Like the equilibrium returns $\bar{\mu}_t = \bar{\mu}(t,X_t)$, and volatilities $\bar{\sigma}_t = \bar{\sigma}(t,X_t)$, these equilibrium positions are also functions of time and the state variable only, $\bar{\varphi}^1_t = \bar{\varphi}^1(t,X_t)$. In particular, using Itô's formula, we can write

$$d\bar{\varphi}^1_t = \bar{b}_t dt + \bar{a}_t dW_t, \tag{5.8}$$

where the coefficients $\bar{b}_t = \bar{b}(t,X_t)$ and $\bar{a}_t = \bar{a}(t,X_t)$ take values in $\mathbb{R}$ and $\mathbb{R}^d$, respectively, and given explicitly in terms of the derivatives of the primitives $\xi^1(x)$, $\xi^2(x)$, $b(x)$, $a(x)$, and the frictionless equilibrium volatility $\bar{\sigma}(t,x)$.

5.2.2 Optimization and Equilibrium with Power Costs

For

$$\lambda_t = \lambda\Lambda(X_t),$$

we consider transaction costs of power form,[1]

$$\lambda_t G_q(\dot{\varphi}^n_t), \quad \text{where } G_q(x) = |x|^q/q \text{ for } q\in(1,2],$$

[1]As in the partial equilibrium models of Guasoni and Weber (2018); Cayé et al. (2019) our small-cost asymptotics below crucially exploit the homotheticity of the cost function; therefore we do not consider the most general convex specification from Chapter 3 here.

imposed on each agent n's turnover rate. The frictional version of (5.5) in turn is

$$J_T^n(\dot{\psi}) = \mathbb{E}\left[\int_0^T \left(\psi_t\mu_t - \frac{\gamma^n}{2}\|\sigma_t\varphi_t + \xi_t^n\|^2 - \lambda_t G_q(\dot{\varphi}_t)\right)dt\right]. \tag{5.9}$$

With trading frictions, the agents' optimal strategies are no longer myopic. Instead, as discussed in Chapter 3, the optimal position φ_t^n and the corresponding trading rate $\dot{\varphi}_t^n$ of agent n solve a FBSDE for each price process of the risky asset. The expected return of the latter in turn needs to be chosen so that the agents' total demand matches the supply of the risky asset. After inserting this equilibrium return into the asset dynamics (5.3), the equilibrium volatility is in turn pinned down by the BSDE obtained by matching the terminal dividend. Unlike its counterpart (5.6) in the frictionless case, this BSDE is not autonomous, but coupled to the FBSDE determining agent 1's corresponding optimal position and trading rate (their counterparts for agent 2 are then fixed by market clearing). After the reparametrizations from Subsection 3.4.3, the resulting FBSDE system then reads as follows:

$$dX_t = b(X_t)dt + a(X_t)dW_t, \qquad X_0 = x_0, \tag{5.10}$$

$$d(\Delta\varphi_t^{1,\lambda}) = ((G_q')^{-1}\left(\frac{Y_t^{1,\lambda}}{\lambda_t}\right) - \bar{b}_t)dt - \bar{a}_t dW_t, \Delta\varphi_0^{1,\lambda} = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}, \tag{5.11}$$

$$d\mathcal{Y}_t^\lambda = \left[\bar{\gamma}\left(2s\bar{\sigma}_t + s\left(\mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda}\right) + \xi_t^1 + \xi_t^2\right)\left(\mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda}\right)^\top\right]dt + \mathcal{Z}_t^\lambda dW_t,$$
$$\mathcal{Y}_T^\lambda = 0, \tag{5.12}$$

$$\begin{aligned} dY_t^{1,\lambda} = \Bigg[&\frac{\gamma^2\xi_t^2 - \gamma^1\xi_t^1}{2}\left(\frac{(2\bar{\sigma}_t + \mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda})(\mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda})^\top\bar{\sigma}_t}{\bar{\sigma}_t\bar{\sigma}_t^\top} - \mathcal{Z}_t^\lambda - \widehat{\gamma}Z_t^{1,\lambda}\right)^\top \\ &+ \gamma(\bar{\sigma}_t + \mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda})(\bar{\sigma}_t + \mathcal{Z}_t^\lambda + \widehat{\gamma}Z_t^{1,\lambda})^\top\Delta\varphi_t^{1,\lambda}\Bigg]dt + Z_t^{1,\lambda}dW_t, \end{aligned}$$
$$Y_T^{1,\lambda} = 0. \tag{5.13}$$

Here, the exogenous forward process X_t is autonomous. In contrast, the other forward process $\Delta\varphi_t^{1,\lambda} = \varphi_t^{1,\lambda} - \bar{\varphi}_t^{1,\lambda}$ describes the difference between agent 1 frictional and frictionless equilibrium positions. It is therefore naturally coupled to the backward component $Y_t^{1,\lambda} = \lambda_t G_q'(\dot{\varphi}_t^{1,\lambda})$, which corresponds to the marginal trading cost of agent 1's optimal trading rate (and therefore in turn depends on the deviation $\Delta\varphi_t^{1,\lambda}$ from the frictionless allocation). Finally, the two backward components are also coupled together and, together, describe the adjustment of the equilibrium price due to transaction costs:

$$S_t^\lambda - \bar{S}_t = \mathcal{Y}_t^\lambda + \widehat{\gamma}Y_t^{1,\lambda}, \quad \text{where } \widehat{\gamma} = \frac{\gamma^1 - \gamma^2}{\gamma^1 + \gamma^2}.$$

5.3 Small-Cost Asymptotics

We now derive small-cost asymptotics for models with general state dynamics. In order to make the derivation easier to follow, we first consider the simplest case of quadratic costs ($q = 2$), and show how to generalize the asymptotics for the model with linear state dynamics from Theorem 4.8 in this case. Subsequently, we then turn to models with transaction costs of a general power form, before concluding with a discussion of the important limiting case of proportional costs.

5.3.1 Asymptotics for Quadratic Costs

The backward processes $\mathcal{Y}_t^\lambda$ and $Y_t^{1,\lambda}$ in the Markovian FBSDE system (5.11)-(5.13) are generally functions of time t as well as the forward processes X_t and $\Delta\varphi_t^{1,\lambda}$. For small transaction costs, we now make the following ansatz inspired by the asymptotics from Theorem 4.8 of the FBSDE system for the model with linear state dynamics that we have studied in Subsection 4.4.3. Given the volatilities $\bar{\sigma}_t$, $\bar{a}_t$ of the frictionless equilibrium price and trading strategy of agent 1, consider the linear SDE

$$d\widehat{\Delta}_t^\lambda = -\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{1/2}\widehat{\Delta}_t^\lambda dt - \bar{a}_t dW_t, \qquad \widehat{\Delta}_0^\lambda = \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}. \tag{5.14}$$

We then propose the following approximations for (5.11)-(5.13):

$$\widehat{\mathcal{Y}}_t^\lambda := \lambda^{1/2}A(t, X_t), \tag{5.15}$$

$$\widehat{Y}_t^{1,\lambda} := -\left(\lambda\Lambda_t\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{1/2}\widehat{\Delta}_t^\lambda. \tag{5.16}$$

As in Theorem 4.8, $\widehat{\Delta}_t^\lambda$ approximates the deviation $\Delta\varphi_t^{1,\lambda} = \varphi_t^{1,\lambda} - \bar{\varphi}_t^1$ of agent 1's frictional position from its frictionless counterpart. The drift term in (5.14) therefore corresponds to the approximation (5.16) of agent 1's trading rate $Y_t^{1,\lambda}/\lambda_t$. Compared to Theorem 4.8, we have replaced the constant frictionless volatility and trading cost by their random and time-varying values $\bar{\sigma}_t$ and $\lambda\Lambda_t$ in the present more general model. This is motivated by the results of Moreau et al. (2017), who show that updating the (relative) trading speed in this myopic manner is asymptotically optimal in partial equilibrium models. The diffusion part in (5.14) is induced by the frictionless equilibrium strategy, so that we replace the Brownian motion from Theorem 4.8 by its generalization from (5.8) here. Like in Theorem 4.8, the corresponding drift of the frictionless equilibrium strategy can be neglected for the leading-order approximation we construct here, since it is only of order $O(1)$whereas the trading rate becomes large for small transaction costs. (More precisely, it is of order $O(\lambda^{-1/4})$.)

The approximation (5.15) of the backward process $\mathcal{Y}_t^\lambda$ is a direct generalization of its counterpart in Theorem 4.8. The deterministic function that was sufficient for the leading-order approximation there depends on the frictionless volatility and trading costs. Since these quantities can depend on the driving state variable here, we naturally allow the approximation to be a function of these state processes as well.

We now argue that, for a suitable choice of the function $A(t,x)$, the approximations (5.14)-(5.16) indeed allow us to approximate the FBSDE system (5.10)-(5.13) at the leading order for small λ. To this end, first consider the approximation of the forward component (5.14). Its dynamics (5.14) and the approximation (5.16) show that the forward equation (5.11) is indeed satisfied up to terms of the next-to-leading order $O(1)$.

To proceed, observe that since the mean-reversion speed for $\widehat{\Delta}^\lambda$ is of order $O(\lambda^{-1/2})$, the asymptotic order of (the standard deviations of) $\widehat{\Delta}^\lambda_t$ is $O(\lambda^{1/4})$ similarly as in Lemma 4.14. Itô's formula applied to (5.16) in turn yields

$$\begin{aligned} d\widehat{Y}^{1,\lambda}_t &= -\left(\lambda\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\Lambda_t\right)^{1/2} d\widehat{\Delta}^\lambda_t + O(\lambda^{3/4}) \\ &= \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\widehat{\Delta}^\lambda_t + O(\lambda^{3/4})\right)dt + \left(\left(\lambda\Lambda_t\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{1/2}\bar{a}_t + O(\lambda^{3/4})\right)dW_t. \end{aligned} \tag{5.17}$$

Whence, the corresponding diffusion part has the approximation

$$\widehat{Z}^{1,\lambda}_t = \left(\lambda\Lambda_t\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{1/2}\bar{a}_t + O(\lambda^{3/4}). \tag{5.18}$$

Similarly, Itô's formula applied to (5.15) shows

$$d\widehat{\mathcal{Y}}^\lambda_t = \lambda^{1/2}\left((\partial_t A + \mathcal{L}^X A)dt + \nabla_x A^\top a_t dW_t\right). \tag{5.19}$$

In particular, the corresponding diffusion part is

$$\widehat{\mathcal{Z}}^\lambda_t = \lambda^{1/2}\nabla_x A^\top a_t. \tag{5.20}$$

In view of (5.18) and (5.20), the drift of $\widehat{Y}^{1,\lambda}_t$ in (5.17) indeed matches its counterpart in the backward equation (5.12) at the leading order $O(\lambda^{1/4})$.

We now turn to the other backward equation (5.12). By (5.18) and (5.20), its leading-order drift term is of order $O(\lambda^{1/2})$ and given by

$$\bar{\gamma}\left((2s\bar{\sigma}+\xi^1+\xi^2)\left(a^\top\nabla_x A + \widehat{\gamma}(\Lambda\gamma\bar{\sigma}\bar{\sigma}^\top)^{1/2}\right)\right)\lambda^{1/2}dt.$$

To match this leading-order term with (5.19), the function $A(t,x)$ needs to solve the following linear PDE with source term:

$$\partial_t A + \mathcal{L}^X A - \bar{\gamma}(2s\bar{\sigma}+\xi^1+\xi^2)a^\top\nabla_x A = \widehat{\gamma}\bar{\gamma}\left(\gamma\Lambda\bar{\sigma}\bar{\sigma}^\top\right)^{1/2}(2s\bar{\sigma}+\xi^1+\xi^2)\bar{a}^\top.$$

Together with the terminal condition $A(T,x)=0$,[2] it follows that $A(t,x)$ has the Feynman-Kac representation

$$A(t,x) = -\widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T \left(\gamma\bar{\sigma}_r\bar{\sigma}_r^\top\Lambda_r\right)^{1/2}\left((2s\bar{\sigma}_r+\xi^1_r+\xi^2_r)^\top\bar{a}_r\right)\,dr\right]. \tag{5.21}$$

[2] In contrast, we do not match the terminal condition for $\mathcal{Y}^1_t$ as in the time-averaged asymptotics of Subsection 4.4.3, where this does not affect the order of the accuracy if the approximation errors are averaged across times and states.

Here, the expectation is taken under the measure $\widehat{\mathbb{P}}$ for which the state variable has dynamics

$$dX_t = [b_t - \bar{\gamma} a_t (2s\bar{\sigma}_t + \xi_t^1 + \xi_t^2)^\top]dt + a_t d\widehat{W}_t.$$

In summary, if the function A is chosen appropriately as the solution of a linear PDE with source term, then the ansatz (5.14)-(5.16) indeed allows us to match the leading-order terms in the FBSDE system (5.11)-(5.13). The corresponding approximation of the equilibrium price adjustment due to small transaction costs in turn is

$$S_t - \bar{S}_t = \mathcal{Y}_t + \widehat{\gamma} Y_t^1 \approx \widehat{\mathcal{Y}}_t^\lambda + \widehat{\gamma} \widehat{Y}_t^{1,\lambda} = \lambda^{1/2} A(t, X_t). \tag{5.22}$$

5.4 Examples

We now illustrate the implications the small-cost approximation (5.22) in a number of examples. As a sanity check, we first reconsider the setting with linear state dynamics from Chapter 4.

Example 5.1. Suppose as in Example 3.5 that

$$\xi_t^1 = -\xi_t^2 = \xi W_t \quad \text{and} \quad \mathfrak{S} = \beta T + \alpha W_T,$$

for $\xi, \alpha > 0$ and $\beta \in \mathbb{R}$, so that $\bar{\sigma} = \alpha$ and $\bar{a} = -\xi/\alpha$. For constant transaction costs ($\lambda_t = \lambda > 0$ and $\Lambda(x) = 1$), (5.34) then shows that the leading-order price adjustment (5.22) is given by

$$\lambda^{1/2} A(t,x) = \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1 - \gamma^2)}{(\gamma^1+\gamma^2)^{3/2}} \xi \alpha s \lambda^{1/2}(T-t).$$

This is exactly the leading-order approximation (4.29) from Corollary 4.9. In particular, we need to choose $\gamma^2 > \gamma^1$ to reproduce the illiquidity discounts observed empirically by Amihud and Mendelson (1986).

Next, we consider transaction costs driven by a Brownian motion independent from the one driving endowment and dividend shocks.

Example 5.2. Let $d = k = 2$ and suppose $X = W$ is a bivariate standard Brownian motion. Endowments volatilties and the terminal condition for the risky asset are functions of the first component only,

$$\xi_t^1 = -\xi_t^2 = \xi W_t^{(1)}, \quad \mathfrak{S} = \beta T + \alpha W_T^{(1)}.$$

Accordingly, the frictionless equilibrium is of exactly the same form as in Example 5.1 above. In particular, $\bar{\sigma} = (\alpha, 0)^\top$ and $\bar{a}_t = (-\xi/\alpha, 0)^\top$. The second component of the Brownian motion drives the transaction cost $\lambda_t = \lambda\Lambda(W_t^{(2)})$. By Girsanov's theorem, the dynamics of the second component $X^{(2)} = W^{(2)}$

then remain unchanged by the passage from the original probability to $\tilde{\mathbb{P}}$. Jensen's inequality in turn shows that

$$\widehat{\mathbb{E}}_{t,x}\left[\sqrt{\Lambda(X_r)}\right] = \mathbb{E}_{t,x}\left[\sqrt{\Lambda(X_r)}\right] \leq \sqrt{\mathbb{E}_{t,x}\left[\Lambda(X_r)\right]}.$$

As a consequence, with $\gamma^2 > \gamma^1$ as above, the illiquidity discount

$$\begin{aligned} 0 \geq \lambda^{1/2} A(t,x) &= \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}} \xi\alpha s \lambda^{1/2} \int_t^T \mathbb{E}_{t,x}\left[\sqrt{\Lambda_r}\right] dr \\ &\geq \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1-\gamma^2)}{(\gamma^1+\gamma^2)^{3/2}} \xi\alpha s \lambda^{1/2} \int_t^T \sqrt{\mathbb{E}_{t,x}\left[\Lambda_r\right]} dr.. \end{aligned}$$

with *independent* stochastic transaction costs is *always smaller* than for constant costs of the same average magnitude.

Example 5.2 is reminiscent of a finding for partial equilibrium models in (Moreau et al., 2017, Equation 1.2), where the leading-order correction of the frictionless value function due to small transaction costs also turns out to scale with the square root of the trading cost, suitably average across time and states. Whence, if the volatilities of the market and the frictionless target strategy are constant (like in the present setting) or, more generally, independent of the fluctuations of the transaction costs, then the welfare impact is smaller for stochastic costs than in an otherwise equivalent market where the cost is replaced by its average. The intuitions for this seems to be that by "timing the market" across times and states (i.e., trading more aggressively when trading is comparatively cheap), the welfare impact of transaction costs is reduced. Analogously, the effect on equilibrium prices is reduced in the current setting.

In contrast, Acharya and Pedersen (2005) find that independent liquidity risk is priced as an additional risk factor and in turn increases the liquidity discount. The explanation for this apparent contradiction is that the agents in the overlapping generations model of Acharya and Pedersen (2005) only trade once and therefore cannot time the market. Accordingly, the additional risk of fluctuating trading costs is the main effect for them. In contrast, when agents can shift trades over time to periods of high liquidity, (5.2) shows time-variations in liquidity can in fact be beneficial for agents sophisticated enough to exploit them.

However, this result can be reversed if the fluctuations of trading needs and transaction costs covary, as documented empirically by Acharya and Pedersen (2005); Collin-Dufresne et al. (2020). This is illustrated by the following concrete example.

Example 5.3. Suppose $d = 1$, $k = 2$ with

$$b(x) = \begin{pmatrix} 0 \\ -\frac{\nu^2}{2}x^{(2)} \end{pmatrix}, \quad a(x) = \begin{pmatrix} 1 \\ \nu \end{pmatrix}$$

This means that, as before, the first component $X_t^{(1)}$ of the state variable is a standard Brownian motion. It drives the terminal dividend $\mathfrak{S} = \beta T + \alpha X_T^{(1)}$, so that the frictionless equilibrium price remains a Bachelier model with constant volatility $\bar{\sigma}_t = (\alpha, 0)^\top$ as long as the aggregate endowment is zero ($\xi_t^1 + \xi_t^2 = 0$).

The second component of the state variable is an Ornstein-Uhlenbeck process (with unit variance) that drives both the volatilities of the agents' endowments (and in turn their trading needs) as well as the transaction costs. To wit,

$$d\xi_t^1 = -d\xi_t^2 = \xi |X_t^{(2)}| dW_t,$$

so that the volatility of agent 1's frictionless equilibrium strategy (5.7) is

$$\bar{a}_t = -\frac{\xi}{\alpha} |X_t^{(2)}|.$$

In order to ensure that the trading cost remains positive (and to obtain tractable formulas below), we model it by the square of the OU process,

$$\lambda_t = \lambda |X_t^{(2)}|^2.$$

By Itô's formula, the trading cost has the CIR dynamics and, in particular, long-term mean λ:

$$d\lambda_t = \nu^2 (\lambda - \lambda_t)\, dt + \sqrt{4\nu^2\lambda}\sqrt{\lambda_t} d\tilde{W}_t,$$

for the standard Brownian motion

$$d\tilde{W}_t = \frac{X_t^{(2)}}{|X_t^{(2)}|} \mathbb{1}_{\{X_t^{(2)} \neq 0\}} dW_t + \mathbb{1}_{\{X_t^{(2)} \neq 0\}} dW_t^\perp,$$

where $W^\perp$ is a standard Brownian motion that is independent of W.

With these specifications, the leading-order price adjustment (5.34) due to small transaction costs is

$$\lambda^{1/2} A(0, x) = \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1 - \gamma^2)}{(\gamma^1 + \gamma^2)^{3/2}} \xi \alpha s \lambda^{1/2} \widehat{\mathbb{E}}_{0,x} \left[\int_0^T |X_t^{(2)}|^2 dt \right].$$

If, instead, the stochastic transaction cost is replaced by its long-term average λ, then the corresponding price adjustment is

$$\lambda^{1/2} \bar{A}(0, x) = \frac{2^{1/2}\gamma^1\gamma^2(\gamma^1 - \gamma^2)}{(\gamma^1 + \gamma^2)^{3/2}} \xi \alpha s \lambda^{1/2} \widehat{\mathbb{E}}_{0,x} \left[\int_0^T |X_t^{(2)}| dt \right].$$

By Girsanov's theorem, the state process $X_t^{(2)}$ still is an OU process under the measure $\widehat{\mathbb{P}}$:

$$dX_t^{(2)} = \frac{\nu^2}{2} \left(-\frac{4\alpha s \bar{\gamma}}{\nu} - X_t^{(2)} \right) dt + \nu d\widehat{W}_t^{(2)}.$$

Accordingly, under the measure $\widehat{\mathbb{P}}$, $X_t^{(2)}$, $t \in [0,T]$ is normally distributed with mean $e^{-\frac{1}{2}\nu^2 t}x - (1 - e^{-\frac{1}{2}\nu^2 t})\frac{4\alpha s\bar{\gamma}}{\nu}$ and variance $1 - e^{-\nu^2 t}$. For sufficiently large t, these converge to the stationary values

$$\widehat{m}_X = -\frac{4\alpha s\bar{\gamma}}{\nu} \quad \text{and} \quad s_X^2 = 1.$$

As a consequence, the ergodic theorem allows us to approximate

$$\widehat{\mathbb{E}}_{0,x}\left[\frac{1}{T}\int_0^T |X_t^{(2)}|^2 dt\right] \approx 1 + \widehat{m}_X^2, \tag{5.23}$$

For constant transaction costs, the ergodic theorem and integration against the probability density of the normal distribution show that for sufficiently long time horizon T,

$$\widehat{\mathbb{E}}_{0,x}\left[\frac{1}{T}\int_0^T |X_t^{(2)}| dt\right] \approx e^{-\frac{\widehat{m}_X^2}{2}}\sqrt{\frac{2}{\pi}} + \widehat{m}_X \mathrm{Erf}\left(\frac{\widehat{m}_X}{\sqrt{2}}\right).$$

The derivative of the right-hand side with respect to μ_X is $\mathrm{Erf}(\widehat{m}_X/\sqrt{2}) < 1$. As a consequence, the multiplier for constant transaction costs is smaller than $1 + \widehat{m}_X^2$, the multiplier for stochastic transaction costs. As a result, replacing the transaction costs by their long-run mean reduces the size of the illiquidity discount here – for all parameter values, given that the time horizon is sufficiently long. The ratio of the illiquidity discounts with stochastic and constant costs only depends on the parameter $\widehat{m}_X$. On the negative halfline, this function is decreasing, so that fluctuations of liquidity have a big effect here if they are persistent for small ν.

The previous two examples can of course be combined to obtain models where the impact of liquidity risk depends on the relative magnitude of trading needs independent from and perfectly correlated with the trading cost.

Finally, we have a look at an example with stochastic volatility. In this context, the impact of liquidity risk on illiquidity discounts turns out to be ambiguous.

Example 5.4. Suppose the state variable is a one dimenional CIR process with dynamics

$$dX_t = \kappa(\bar{X} - X_t)dt - \nu\sqrt{X_t}dW_t.$$

The terminal condition for the risky asset is

$$\mathfrak{S} = S_0 + \int_0^T \bar{\gamma} s X_t dt + \int_0^T \sqrt{X_t}dW_t,$$

so that – for a zero aggregate endowment ($\xi_t^1 + \xi_t^2 = 0$) – the frictionless equilibrium price has the (arithmetic) Heston and Nandi (1998) dynamics

$$d\bar{S}_t = \bar{\gamma} s X_t dt + \sqrt{X_t}dW_t.$$

In particular, the frictionless equilibrium volatility is

$$\bar{\sigma}_t = \sqrt{X_t},$$

and the corresponding frictionless optimal trading strategy of agent is

$$d\bar{\varphi}_t^1 = -d\left(\frac{\xi_t^1}{\sqrt{X_t}}\right).$$

For $\xi_t^1 = -\xi_t^2 = \xi X_t$, this is proportional to $\sqrt{X_t}$ which, by Itô's formula, has constant volatility $\nu/2$. In summary, the volatility of the frictionless strategy then is also constant:

$$\bar{a}_t = \frac{\xi\nu}{2}.$$

The last ingredient for the frictional price adjustment are the dynamics of the trading costs themselves. If these are constant ($\lambda_t = \lambda$), then the leading-order price adjustment (5.34) due to small transaction costs is

$$\lambda^{1/2}\bar{A}(0,x) = \frac{\gamma^1\gamma^2(\gamma^1-\gamma^2)}{2^{1/2}(\gamma^1+\gamma^2)^{3/2}}\xi\nu s\lambda^{1/2}\widehat{\mathbb{E}}_{0,x}\left[\int_0^T X_t dt\right]. \tag{5.24}$$

Let us now compare this to stochastic transaction costs $\lambda\Lambda(X_t)$. Then X_t in (5.24) is replaced by $\sqrt{\Lambda(X_t)}X_t$. For concreteness, let us assume that the transaction costs are a multiple c of the infinitesimal variance of the risky asset as implied by (Garleanu and Pedersen, 2016, Section 3.2). Then X_t is replaced by $c^{1/2}X_t^{3/2}$ in (5.24). To ensure that the stochastic and the constant costs have the same long-term means, we choose $c = 1/\bar{X}$. Then, the price correction with stochastic costs is

$$\lambda^{1/2}A(0,x) = \frac{\gamma^1\gamma^2(\gamma^1-\gamma^2)}{2^{1/2}(\gamma^1+\gamma^2)^{3/2}}\xi\nu s\lambda^{1/2}\widehat{\mathbb{E}}_{0,x}\left[\int_0^T \frac{X_t^{3/2}}{\bar{X}^{1/2}} dt\right]. \tag{5.25}$$

Under the measure $\widehat{\mathbb{P}}$, the state process X_t still is an CIR process:

$$dX_t = \widehat{\kappa}\left(\frac{\kappa}{\widehat{\kappa}}\bar{X} - X_t\right)dt + \nu\sqrt{X_t}d\widehat{W}_t, \quad \text{where } \widehat{\kappa} = \kappa + 2\bar{\gamma}s\nu.$$

Whence, the ergodic theorem shows that for large T, the expectation in the price correction for constant costs is approximately

$$\widehat{\mathbb{E}}_{0,x}\left[\int_0^T X_t dt\right] \approx T\bar{X}\frac{\kappa}{\widehat{\kappa}}.$$

To compute the expectation in the leading-order price correction (5.25) with stochastic costs for large T, again use the ergodic theorem. Since the CIR process has a stationary Gamma$(\frac{2\bar{X}\kappa}{\nu^2}, \frac{2\widehat{\kappa}}{\nu^2})$ law under $\widehat{\mathbb{P}}$, this leads to

$$\begin{aligned}\widehat{\mathbb{E}}_{0,x}\left[\int_0^T \frac{X_t^{3/2}}{\bar{X}^{1/2}} dt\right] &\approx T\frac{1}{\bar{X}^{1/2}}\left(\frac{\nu^2}{2\widehat{\kappa}}\right)^{3/2}\frac{\Gamma(\frac{3}{2}+\frac{2\kappa\bar{X}}{\nu^2})}{\Gamma(\frac{2\kappa\bar{X}}{\nu^2})}\\ &= T\bar{X}\left(\frac{\kappa}{\widehat{\kappa}}\right)^{3/2}\left(\frac{2\kappa\bar{X}}{\nu^2}\right)^{-3/2}\frac{\Gamma(\frac{3}{2}+\frac{2\kappa\bar{X}}{\nu^2})}{\Gamma(\frac{2\kappa\bar{X}}{\nu^2})}.\end{aligned}$$

As a result, the illiquidity discount with stochastic transaction costs is bigger than with their constant counterpart if

$$\left(\frac{2\kappa\bar{X}}{\nu^2}\right)^{-3/2}\frac{\Gamma(\frac{3}{2}+\frac{2\kappa\bar{X}}{\nu^2})}{\Gamma(\frac{2\kappa\bar{X}}{\nu^2})} > \left(1+2\bar{\gamma}s\frac{\nu}{\kappa}\right)^{1/2}.$$

The function $y \mapsto y^{-3/2}\Gamma(\frac{3}{2}+y)/\Gamma(y)$ is strictly decreasing and maps $(0,\infty)$ to $(1,\infty)$. Accordingly, liquidity risk increases the liquidity discount here if $\bar{X}$ is small enough, but decreases it if $\bar{X}$ is large enough relative to the other model parameters.

The models considered here only serve as an illustration for the wide scope of the asymptotic expansions. Calibrating such more general models to data and identifying parsimonious yet flexible specifications is an important direction for further research.

5.5 Asymptotics for Power Costs

We now extend the formal asymptotics from Section 5.3.1 to transaction costs of the general power form $\lambda|x|^q/q$, $q \in (1,2]$. To this end, we generalize the ansatz (5.14) - (5.16).

For the approximation $\widehat{\mathcal{Y}}_t^\lambda$ of the backward process $\mathcal{Y}_t$, we expect that it still is a function of time and the state variable as in (5.15), but that the scaling in the transaction cost parameter depends on the elasticity q of the trading cost:

$$\widehat{\mathcal{Y}}_t^\lambda \approx \lambda^{2/q+2}A(t,X_t). \tag{5.26}$$

Remark 5.5. The choice of the power is motivated by the following consideration. The partial-equilibrium results of Guasoni and Weber (2018); Cayé et al. (2019) show that the same power governs the leading-order reductions of the value functions in their settings. For quadratic costs, we have seen above that the adjustment of the equilibrium price is of the same order as the adjustment of the value function. Accordingly, we make the same ansatz here.

Next, consider the approximation of the other backward component $Y_t^{1,\lambda}$. Recall that the equilibrium trading rate of agent 1 is $(G_q')^{-1}(Y_t^{1,\lambda}/\lambda\Lambda_t)$. In view of the small-cost asymptotics for partial equilibrium models developed in Cayé et al. (2019), we expect this to be a function of the state variable and the deviation $\Delta\varphi_t^1$ of the frictional position from its frictionless counterpart. More specifically, inspired by the long-run equilibrium given by $Y_t^1 = g_q(\Delta\varphi_t^1)$ in Chapter 2 and the rescaling for power costs from Section 2.6.4,[3] we assume

[3] As in Section 5.3.1 and in Cayé et al. (2019), the constant volatilities and trading costs are updated myopically over time by plugging in the current values.

that

$$\widehat{Y}_t^{1,\lambda} = \left(\frac{\lambda\Lambda_t}{q}\right)^{\frac{3}{q+2}} \left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top \left(\bar{a}_t\bar{a}_t^\top\right)^2}{8}\right)^{\frac{q-1}{q+2}} \times \tilde{g}_q\left(2^{\frac{q-1}{q+2}}\left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{\frac{1}{q+2}}\left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{q}{q+2}}\widehat{\Delta}_t^\lambda\right). \tag{5.27}$$

Here, we recall that $\tilde{g}_q$ is the solution of the canonical ODE (2.58),

$$\tilde{g}_q''(x) + \tilde{g}_q'(x)\,\mathrm{sgn}(\tilde{g}_q(x))\left|\frac{\tilde{g}_q(x)}{q}\right|^{\frac{1}{q-1}} = 2x,$$

By definition of $\tilde{g}_q$, $\widehat{Y}_t^{1,\lambda}$ and $\widehat{\Delta}_t^\lambda$ then have opposite signs.

Remark 5.6. Here, the rescalings of $\widehat{\Delta}_t^\lambda$ and the dynamic are again inspired by the partial equilibrium results of Guasoni and Weber (2018); Cayé et al. (2019). For quadratic costs, the function $\tilde{g}_2$ was linear in the second variable (compare (5.14)), so that the inner and outer rescalings simplified to the single term $\lambda^{1/2}$. For general power costs, the function $\tilde{g}_q$ is not homothetic, so that both inner and outer rescalings are necessary.

The approximation (5.27) in turn suggests that the dynamics of the deviation $\Delta\varphi_t^1 = \varphi_t^1 - \bar{\varphi}_t^1$ can be approximated by the solution of the following SDE:[4]

$$\begin{aligned} d\widehat{\Delta}_t^\lambda &= -\left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\left(\bar{a}_t\bar{a}_t^\top\right)^2}{8\lambda\Lambda_t}\right)^{\frac{1}{q+2}} \mathrm{sgn}(\widehat{\Delta}_t^\lambda) \\ &\quad \times \left|\frac{\tilde{g}_q\left(2^{\frac{q-1}{q+2}}\left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{\frac{1}{q+2}}\left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{q}{q+2}}\widehat{\Delta}_t^\lambda\right)}{q}\right|^{\frac{1}{q-1}} dt - \bar{a}_t dW_t, \\ \widehat{\Delta}_0^\lambda &= \varphi_{0-}^1 - \frac{s\gamma^2}{\gamma^1+\gamma^2}. \end{aligned} \tag{5.28}$$

By definition of $\tilde{g}_q$, the corresponding deviation $\widehat{\Delta}_t^\lambda$ quickly mean reverts around zero for small transaction costs. (More specifically, it is of order $O(\lambda^{1/(q+2)})$, compare Cayé et al. (2019) and (2.60).) Together with the canon-

[4] Like for quadratic costs in Section 5.3.1, the drift rate of the frictionless equilibrium strategy is again negligible here relative to the large trading speed for small transaction costs.

ical ODE (2.58), Itô's formula applied to (5.27) in turn yields

$$
\begin{aligned}
d\widehat{Y}_t^{1,\lambda} = &\left(\frac{\lambda\Lambda_t}{q}\right)^{\frac{3}{q+2}} \left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\left(\bar{a}_t\bar{a}_t^\top\right)^2}{8}\right)^{\frac{q-1}{q+2}} 2^{\frac{q-1}{q+2}} \left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{\frac{1}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{q}{q+2}} \\
&\times \Bigg[\left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\left(\bar{a}_t\bar{a}_t^\top\right)^2}{8\lambda\Lambda_t}\right)^{\frac{1}{q+2}} \tilde{g}_q'' \left(2^{\frac{q-1}{q+2}} \left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{\frac{1}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{q}{q+2}} \widehat{\Delta}_t^\lambda\right) dt \\
&+ \tilde{g}_q' \left(2^{\frac{q-1}{q+2}} \left(\frac{q\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t}\right)^{\frac{1}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{q}{q+2}} \widehat{\Delta}_t^\lambda\right) d\widehat{\Delta}_t^\lambda\Bigg] + O(\lambda^{3/q+2}) \\
= &\left(\frac{\lambda\Lambda_t}{q}\right)^{\frac{1}{q+2}} \left(\frac{1}{2}\right)^{\frac{2q+1}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{\frac{q+1}{q+2}} \left(\bar{a}_t\bar{a}_t^\top\right)^{\frac{q}{q+2}} \left(\tilde{g}_q'' + \tilde{g}_q' \operatorname{sgn}(\tilde{g}_q)\left|\frac{\tilde{g}_q}{q}\right|^{\frac{1}{q-1}}\right) dt \\
&- \left(\frac{\lambda\Lambda_t}{2^{q-1}q\bar{a}_t\bar{a}_t^\top}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\bar{a}_t\bar{a}_t^\top\right)^{\frac{q}{q+2}} \tilde{g}_q'\bar{a}_t dW_t + O(\lambda^{3/q+2}) \\
= &-\left(\frac{\lambda\Lambda_t}{2^{q-1}q}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{\frac{q}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{2-q}{q+2}} \tilde{g}_q'\bar{a}_t dW_t \\
&+ \gamma\bar{\sigma}_t\bar{\sigma}_t^\top\widehat{\Delta}_t^\lambda dt + O(\lambda^{3/q+2}).
\end{aligned} \tag{5.29}
$$

Comparing the diffusion term to its counterpart in (5.13), we find

$$
\widehat{Z}_t^{1,\lambda} = -\lambda^{2/q+2} \left(\frac{\Lambda_t}{2^{q-1}q}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{\frac{q}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{2-q}{q+2}} \tilde{g}_q'\bar{a}_t + O(\lambda^{3/q+2}). \tag{5.30}
$$

Similarly, Itô's formula applied to (5.26) shows

$$
\begin{aligned}
d\widehat{\mathcal{Y}}_t^\lambda = \lambda^{2/q+2} &\left[\left(\partial_t A(t,X_t) + \mathcal{L}^X A(t,X_t)\right) dt + \nabla_x A(t,X_t)^\top a_t dW_t\right] \\
&+ O(\lambda^{3/q+2}),
\end{aligned} \tag{5.31}
$$

so that comparison with the diffusion part in (5.12) gives

$$
\widehat{\mathcal{Z}}_t^\lambda = \lambda^{2/q+2}\nabla_x A(t,X_t)^\top a_t + O(\lambda^{3/q+2}). \tag{5.32}
$$

Now, compare the leading-order drift rate of order $O(\lambda^{1/q+2})$ in (5.29) to its counterpart in (5.12), also taking into account (5.30) and (5.32). Accordingly, A satisfies a linear PDE with source term,

$$
\begin{aligned}
&\partial_t A + \mathcal{L}^X A - \bar{\gamma}(2s\bar{\sigma} + \xi^1 + \xi^2)a^\top\nabla_x A \\
= &-\widehat{\gamma}\bar{\gamma}\left(\frac{\Lambda}{2^{q-1}q}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}\bar{\sigma}^\top\right)^{\frac{q}{q+2}} \left(\frac{1}{\bar{a}\bar{a}^\top}\right)^{\frac{2-q}{q+2}} (2s\bar{\sigma} + \xi^1 + \xi^2)\bar{a}^\top \\
&\times \tilde{g}_q' \left(2^{\frac{q-1}{q+2}} \left(\frac{q\gamma\bar{\sigma}\bar{\sigma}^\top}{\Lambda}\right)^{\frac{1}{q+2}} \left(\frac{1}{\bar{a}\bar{a}^\top}\right)^{\frac{q}{q+2}} \frac{\widehat{\Delta}}{\lambda^{1/q+2}}\right), \qquad A(T,x) = 0.
\end{aligned} \tag{5.33}
$$

Together with the terminal condition $A(T,x)=0$, this leads to the Feyman-Kac representation

$$
\begin{aligned}
A(t,x) = \widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\Big[\int_t^T \left(\frac{\Lambda_t}{2^{q-1}q}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{\frac{q}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{2-q}{q+2}} (2s\bar{\sigma}_u+\xi_u^1+\xi_u^2)\bar{a}_u^\top \\
\times \tilde{g}_q'\left(2^{\frac{q-1}{q+2}}\left(\frac{q\gamma\bar{\sigma}_u\bar{\sigma}_u^\top}{\Lambda_u}\right)^{\frac{1}{q+2}}\left(\frac{1}{\bar{a}_u\bar{a}_u^\top}\right)^{\frac{q}{q+2}}\frac{\widehat{\Delta}_u^\lambda}{\lambda^{1/q+2}}\right)\, du\Big],
\end{aligned}
\tag{5.34}
$$

where $\widehat{\mathbb{P}}$ is defined by

$$
d\widehat{W}_t = dW_t + \bar{\gamma}(2s\bar{\sigma}_t+\xi_t^1+\xi_t^2)dt. \tag{5.35}
$$

In order to simplify this formula for $A(t,x)$, first integrate both sides of (2.58) and notice that the canonical function $\tilde{g}_q$ also satisfies

$$
\tilde{g}_q'(x) + (q-1)\left|\frac{\tilde{g}_q(x)}{q}\right|^{\frac{1}{q-1}} = x^2 + \tilde{g}_q'(0).
$$

We then introduce the following useful identity for the canonical function $\tilde{g}_q$, cf. the proof of Proposition 4.2 in Cayé et al. (2019)

$$
\begin{aligned}
0 &= \frac{q-1}{q}\tilde{g}_q(x)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\Big|_{-\infty}^{+\infty} \\
&= (q-1)\int_{-\infty}^{+\infty}\left(\frac{\tilde{g}_q'(x)}{q} - \frac{\tilde{g}_q(x)}{q}\left|\frac{\tilde{g}_q(x)}{q}\right|^{\frac{1}{q-1}}\right)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\,dx \\
&= \int_{-\infty}^{+\infty}\left(\frac{q-1}{q}\tilde{g}_q'(x) + (q-1)\left|\frac{\tilde{g}_q(x)}{q}\right|^{\frac{q}{q-1}}\right)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\,dx \\
&= \int_{-\infty}^{+\infty}\left(\frac{q-1}{q}\tilde{g}_q'(x) + \left(x^2+\tilde{g}_q'(0)-\tilde{g}_q'(x)\right)\right)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\,dx \\
&= \int_{-\infty}^{+\infty}\left(x^2+\tilde{g}_q'(0)-\frac{\tilde{g}_q'(x)}{q}\right)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\,dx, \\
&= \frac{\tilde{v}_q+\tilde{g}_q'(0)}{\tilde{c}_q} - \frac{1}{q}\int_{-\infty}^{\infty}\tilde{g}_q'(x)e^{-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy}\,dx.
\end{aligned}
$$

Here, we recall that

$$
\begin{aligned}
\tilde{c}_q &= \left[2\int_0^\infty \exp\left(-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy\right)dx\right]^{-1}, \\
\tilde{v}_q &= \frac{\int_0^\infty x^2\exp\left(-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy\right)dx}{\int_0^\infty \exp\left(-\int_0^x\left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}}dy\right)dx},
\end{aligned}
$$

and in turn

$$\tilde{c}_q \int_{-\infty}^{\infty} \tilde{g}_q'(x) e^{-\int_0^x \left|\frac{\tilde{g}_q(y)}{q}\right|^{\frac{1}{q-1}} dy}\, dx = q(\tilde{v}_q + \tilde{g}_q'(0)). \tag{5.36}$$

Now, by (Cayé et al., 2020, Theorem 1.2), as $\lambda \downarrow 0$, we can approximate $A(t,x)$ with (5.36) as

$$\begin{aligned}
A(t,x) &\approx \widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\Big[\int_t^T \left(\frac{\Lambda_t}{2^{q-1}q}\right)^{\frac{2}{q+2}} \left(\gamma\bar{\sigma}_t\bar{\sigma}_t^\top\right)^{\frac{q}{q+2}} \left(\frac{1}{\bar{a}_t\bar{a}_t^\top}\right)^{\frac{2-q}{q+2}} \\
&\qquad \times (2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top q(\tilde{v}_q + \tilde{g}_q'(0))du\Big] \\
&= \left(\frac{q\gamma}{4}\right)^{\frac{q}{q+2}} (\tilde{g}_q'(0) + \tilde{v}_q) \\
&\qquad \times \widehat{\mathbb{E}}_{t,x}\left[\int_t^T \left(\frac{2\Lambda_u\left(\bar{\sigma}_u\bar{\sigma}_u^\top\right)^{\frac{q}{2}}}{\left(\bar{a}_u\bar{a}_u^\top\right)^{\frac{2-q}{2}}}\right)^{\frac{2}{q+2}} (2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\, dr\right].
\end{aligned}$$

Therefore, the leading order asymptotics for illiquidity discount is

$$\widehat{\mathfrak{Y}}^\lambda = \lambda^{2/q+2} A(t,x),$$

where

$$\begin{aligned}
A(t,x) &\approx \widehat{\gamma}\bar{\gamma}\left(\frac{q\gamma}{4}\right)^{\frac{q}{q+2}} (\tilde{g}_q'(0) + \tilde{v}_q) \\
&\qquad \times \widehat{\mathbb{E}}_{t,x}\left[\int_t^T \left(\frac{2\Lambda_u\left(\bar{\sigma}_u\bar{\sigma}_u^\top\right)^{\frac{q}{2}}}{\left(\bar{a}_u\bar{a}_u^\top\right)^{\frac{2-q}{2}}}\right)^{\frac{2}{q+2}} (2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\, dr\right].
\end{aligned} \tag{5.37}$$

For quadratic costs, the asymptotic price adjustment (5.37) indeed reduces to its counterpart derived directly above.

Remark 5.7. When $q = 2$, the solution to the canonical ODE is

$$\tilde{g}_2(x) = -2x, \tag{5.38}$$

then $\tilde{g}_2'(0) = -2$ and

$$\tilde{v}_2 = \frac{\int_0^\infty x^2 \exp\left(-\int_0^x y dy\right) dx}{\int_0^\infty \exp\left(-\int_0^x y dy\right) dx} = 1. \tag{5.39}$$

Now suppose the dimension of the Brownian Motion is $d = 1$, transaction costs are constant ($\Lambda_t = 1$), and the state dynamics are linear as in Chapter 4: $\xi_t^1 = -\xi_t^2 = \xi W_t$, and $\mathfrak{S} = \beta T + \alpha W_T$ so that $\bar{a}_t = -\xi/\alpha$. Then:

$$\begin{aligned}
&\widehat{\gamma}\bar{\gamma}\left(\frac{2\gamma}{4}\right)^{\frac{2}{2+2}} (\tilde{g}_2'(0) + \tilde{v}_2)\widehat{\mathbb{E}}_{t,x}\left[\int_t^T \left(\frac{2\left(\bar{\sigma}_u\bar{\sigma}_u^\top\right)^{\frac{2}{2}}}{\left(\bar{a}_u\bar{a}_u^\top\right)^{\frac{2-2}{2}}}\right)^{\frac{2}{2+2}} (2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\, dr\right] \\
&= -\widehat{\gamma}\bar{\gamma}\left(\frac{2\gamma}{4}\right)^{\frac{2}{2+2}} \sqrt{2\alpha^2}(T-t)2s\alpha\left(-\frac{\xi}{\alpha}\right) \\
&= 2\bar{\gamma}\widehat{\gamma}\sqrt{\gamma}\xi\alpha s(T-t).
\end{aligned} \tag{5.40}$$

This is exactly the approximation $A(t)$ we have derived directly for quadratic costs in (4.22). For the volatility correction due to small transaction costs, from $g_2'(x) = -2$, implies that

$$\begin{aligned}\mathcal{Z}^\lambda + \widehat{\gamma} Z_t^{1,\lambda} &\approx 0 - \widehat{\gamma}\lambda^{2/4}\left(\frac{1}{2^{2-1}2}\right)^{\frac{2}{4}}(\gamma\alpha^2)^{\frac{2}{4}}\left(\frac{\alpha^2}{\xi^2}\right)^{\frac{2-2}{4}}\tilde{g}_2'\left(-\frac{\xi}{\alpha}\right)\\ &= -\widehat{\gamma}\frac{1}{2}\lambda^{1/2}\gamma^{1/2}\alpha\xi 2\frac{\xi}{\alpha}\\ &= -\widehat{\gamma}\gamma^{1/2}\xi\lambda^{1/2}, \end{aligned} \tag{5.41}$$

which is also the same as the one derived directly.

Remark 5.8. When $q = 3/2$, the solution to the canonical ODE is no longer available explicitly. However, we can still numerically compute $\tilde{g}_{3/2}'(0) = -1.771$ and $\tilde{v}_{3/2} = 0.759$.

Now suppose the dimension of the Brownian Motion is $d = 1$, transaction costs are constant ($\Lambda_t = 1$), and the state dynamics are linear as in Chapter 4: $\xi_t^1 = -\xi_t^2 = \xi W_t$, and $\mathfrak{S} = \beta T + \alpha W_T$ so that $\bar{a}_t = -\xi/\alpha$ and $\bar{a}_t = -\xi/\alpha$. Then:

$$\widehat{\gamma}\bar{\gamma}\left(\frac{3\gamma}{8}\right)^{\frac{3}{3+4}}(\tilde{g}_{3/2}'(0) + \tilde{v}_{3/2})\widehat{\mathbb{E}}_{t,x}\left[\int_t^T \left(\frac{2\left(\bar{\sigma}_u\bar{\sigma}_u^\top\right)^{\frac{3}{4}}}{\left(\bar{a}_u\bar{a}_u^\top\right)^{\frac{4-3}{4}}}\right)^{\frac{4}{3+4}}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\, dr\right]$$

$$\begin{aligned}&= -(-1.771 + 0.759)2\bar{\gamma}\widehat{\gamma}\xi^{5/7}\left(\frac{\alpha}{2}\right)^{8/7}s(6\gamma)^{3/7}(T-t)\\ &= 1.976\bar{\gamma}\widehat{\gamma}\xi^{5/7}\alpha^{8/7}\gamma^{3/7}(T-t),\end{aligned}$$

Therefore, the price adjustment due to transaction costs is

$$S_t - \bar{S}_t = 1.976\bar{\gamma}\widehat{\gamma}\gamma^{3/7}\xi^{5/7}\alpha^{8/7}s\lambda^{4/7}(T-t),$$

in this case. For the volatility correction,

$$\begin{aligned}\mathcal{Z}^\lambda + \widehat{\gamma} Z_t^{1,\lambda} &\approx 0 - \widehat{\gamma}\lambda^{4/7}\left(\frac{1}{2^{1/2}\frac{3}{2}}\right)^{\frac{4}{7}}(\gamma\alpha^2)^{3/7}\left(\frac{\alpha}{\xi}\right)^{2/7}g_{3/2}'\left(-\frac{\xi}{\alpha}\right)\\ &= \left(\frac{2}{9}\right)^{2/7}g_{3/2}'\widehat{\gamma}\lambda^{4/7}\gamma^{3/7}\xi^{5/7}\alpha^{-1/7},\\ &\leq -1.153\widehat{\gamma}\gamma^{3/7}\xi^{5/7}\alpha^{-1/7}\lambda^{4/7},\end{aligned}$$

Compared to the corresponding formulas (5.40) and (5.41), we see that the respective constant and power to which the input parameters are raised changed. In contrast, the comparative statics of the formulas remain almost unchanged. In particular, illiquidity discounts and liquidity premia necessarily correspond to a positive relationship between liquidity and volatility also for non-quadratic costs here.

5.6 Asymptotics for Proportional Costs

Proportional costs $\lambda\Lambda(X_t)$ can be studied as a singular limiting case of the general power costs as $q \downarrow 1$. However, we can also deal with them directly, which, again, inspired by the long-run proportional equilibrium in Chapter 2. More specifically, notice that as $q \downarrow 1$, the solution $\tilde{g}_q$ to the canonical ODE (2.58) converges to

$$\tilde{g}_1(x) = \left(\frac{1}{3}x^3 - \left(\frac{3}{2}\right)^{2/3} x\right) \mathbb{1}_{\{|x| \leq \left(\frac{3}{2}\right)^{1/3}\}} + \mathbb{1}_{\{x < -\left(\frac{3}{2}\right)^{1/3}\}} - \mathbb{1}_{\{x > \left(\frac{3}{2}\right)^{1/3}\}}. \tag{5.42}$$

As a consequence,

$$\tilde{g}_1'(x) = \left(x^2 - \left(\frac{3}{2}\right)^{2/3}\right) \mathbb{1}_{\{|x| \leq \left(\frac{3}{2}\right)^{1/3}\}}, \qquad \tilde{g}_1''(x) = 2x\mathbb{1}_{\{|x| \leq \left(\frac{3}{2}\right)^{1/3}\}}.$$

Define

$$l(t,x) := \left(\frac{3\Lambda\bar{a}\bar{a}^\top}{2\gamma\bar{\sigma}\bar{\sigma}^\top}\right)^{1/3}. \tag{5.43}$$

For small λ, we then use the following ansatz for the approximations of the FBSDE (5.11)-(5.13):

$$\widehat{\mathcal{Y}}_t^\lambda = \lambda^{2/3} A(t, X_t), \tag{5.44}$$

$$\widehat{Y}_t^{1,\lambda} = \lambda\Lambda_t\tilde{g}_1\left(\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t\bar{a}_t\bar{a}_t^\top}\right)^{1/3}\widehat{\Delta}_t^\lambda\right) = \lambda\Lambda_t\tilde{g}_1\left(\left(\frac{3}{2}\right)^{1/3}\frac{\widehat{\Delta}_t^\lambda}{\lambda^{1/3}l_t}\right). \tag{5.45}$$

Note that this coincides with the corresponding approximations for general power costs (5.27) and (5.26), respectively, for $q = 1$. Recall our conclusion from Section 3.6 that trading only occurs when $|\widehat{Y}_t^{1,\lambda}| = \lambda_t$. By the definition of $\tilde{g}_1$, this translates to the condition that $\widehat{\Delta}_t^\lambda = \lambda^{1/3}l_t$. As a result,

$$d\widehat{\Delta}_t^\lambda = dL_t^\lambda - dU_t^\lambda - \bar{a}_t dW_t, \tag{5.46}$$

where L^λ and U^λ are the minimal increasing processes to keep $\widehat{\Delta}_t^\lambda$ inside the interval $[-\lambda^{1/3}l_t, \lambda^{1/3}l_t]$. Now, apply Itô's formula to (5.45), obtaining

$$\begin{aligned}
d\widehat{Y}_t^{1,\lambda} &= \lambda\Lambda_t\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t\bar{a}_t\bar{a}_t^\top}\right)^{1/3}\left(\tilde{g}_1' d\widehat{\Delta}_t^\lambda + \frac{1}{2}\bar{a}_t\bar{a}_t^\top\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\lambda\Lambda_t\bar{a}_t\bar{a}_t^\top}\right)^{1/3}\tilde{g}_1' dt\right) + O(\lambda) \\
&= \gamma\bar{\sigma}_t\bar{\sigma}_t^\top\widehat{\Delta}_t^\lambda dt + \lambda^{2/3}\Lambda_t\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\Lambda_t\bar{a}_t\bar{a}_t^\top}\right)^{1/3}\tilde{g}_1'\left(dL_t^\lambda - dU_t^\lambda - \bar{a}_t dW_t\right) + O(\lambda).
\end{aligned}$$

Comparing the diffusion term to its counterpart in (5.13), we find that

$$\widehat{Z}_t^{1,\lambda} = -\lambda^{2/3}\Lambda_t\left(\frac{\gamma\bar{\sigma}_t\bar{\sigma}_t^\top}{\Lambda_t\bar{a}_t\bar{a}_t^\top}\right)^{1/3}\tilde{g}_1'\bar{a}_t + O(\lambda). \tag{5.47}$$

Similarly, Itô's formula applied to (5.44) yields

$$d\widehat{\mathfrak{Y}}_t^\lambda = \lambda^{2/3}\left[\left(\partial_t A(t,X_t) + \mathcal{L}^X A(t,X_t)\right)dt + \nabla_x A(t,X_t)^\top a_t dW_t\right] + O(\lambda). \tag{5.48}$$

Whence, comparison with the diffusion part in (5.12) gives

$$\widehat{\mathcal{Z}}_t^\lambda = \lambda^{2/3}\nabla_x A(t,X_t)^\top a_t + O(\lambda). \tag{5.49}$$

Now, compare the leading-order drift rate of order $O(\lambda^{1/q+2})$ in (5.48) to its counterpart in (5.12), also taking into account (5.47) and (5.49). This shows that the function A needs to satisfy a linear PDE with source term:

$$\begin{aligned}
&\partial_t A + \mathcal{L}^X A - \bar{\gamma}(2s\bar{\sigma} + \xi^1 + \xi^2)a^\top \nabla_x A \\
&= -\widehat{\gamma}\bar{\gamma}\left(\frac{\gamma\bar{\sigma}\bar{\sigma}^\top\Lambda^2}{\bar{a}\bar{a}^\top}\right)^{\frac{1}{3}}(2s\bar{\sigma} + \xi^1 + \xi^2)\bar{a}^\top\tilde{g}_q'\left(\left(\frac{\gamma\bar{\sigma}\bar{\sigma}^\top}{\lambda\Lambda\bar{a}\bar{a}^\top}\right)^{\frac{1}{3}}\widehat{\Delta}^\lambda\right), \\
&A(T,x) = 0.
\end{aligned} \tag{5.50}$$

Now by (Cayé et al., 2020, Theorem 1.2), as $\lambda \downarrow 0$, we can approximate the illiquidity discount $\widehat{\mathfrak{Y}}^\lambda = \lambda^{2/q+2}A(t,x)$ where $A(t,x)$ is approximated as:

$$\begin{aligned}
&A(t,x) \\
&= \widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\left(\frac{\gamma\bar{\sigma}_u\bar{\sigma}_u^\top\Lambda_u^2}{\bar{a}_u\bar{a}_u^\top}\right)^{\frac{1}{3}}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\tilde{g}_q'\left(\left(\frac{\gamma\bar{\sigma}_u\bar{\sigma}_u^\top}{\lambda\Lambda_u\bar{a}_u\bar{a}_u^\top}\right)^{\frac{1}{3}}\widehat{\Delta}_u^\lambda\right)du\right] \\
&\approx \widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\frac{3\Lambda_u}{2l_u}\left(\frac{1}{3}-1\right)\mathbb{1}_{[|\widehat{\Delta}_u|\le\lambda^{1/3}l_u]}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top du\right] \\
&= -\widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\frac{\Lambda_u}{l_u}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top du\right] \\
&= -\widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\left(\frac{2\gamma\Lambda_u^2\bar{\sigma}_u\bar{\sigma}_u^\top}{3\bar{a}_u\bar{a}_u^\top}\right)^{1/3}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top du\right].
\end{aligned} \tag{5.51}$$

Up to a change of powers and constant, this again closely parallels the formulas we have for superlinear costs above.

Remark 5.9. Observe that $\tilde{g}_1'(0) = -\left(\frac{3}{2}\right)^{2/3}$ and $\tilde{v}_1 = \frac{1}{3}\left(\frac{3}{2}\right)^{2/3}$. Plugging this into the expansion (5.37) for general power costs and sending $q \to 1$ then yields the same approximation as (5.51):

$$\begin{aligned}
&\widehat{\gamma}\bar{\gamma}\left(\frac{\gamma}{4}\right)^{\frac{1}{1+2}}(\tilde{g}_1'(0) + \tilde{v}_1)\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\left(\frac{2\Lambda_u\left(\bar{\sigma}_u\bar{\sigma}_u^\top\right)^{\frac{1}{2}}}{\left(\bar{a}_u\bar{a}_u^\top\right)^{\frac{2-1}{2}}}\right)^{\frac{1}{1+2}}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\,dr\right] \\
&= -\widehat{\gamma}\bar{\gamma}\left(\frac{\gamma}{4}\right)^{\frac{1}{1+2}}\frac{2}{3}\left(\frac{3}{2}\right)^{\frac{2}{3}}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\left(\frac{4\Lambda_u^2\bar{\sigma}_u\bar{\sigma}_u^\top}{(\bar{a}_u\bar{a}_u^\top)}\right)^{\frac{1}{3}}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top\,dr\right] \\
&= -\widehat{\gamma}\bar{\gamma}\widehat{\mathbb{E}}_{t,x}\left[\int_t^T\left(\frac{2\gamma\Lambda_u^2\bar{\sigma}_u\bar{\sigma}_u^\top}{3\bar{a}_u\bar{a}_u^\top}\right)^{1/3}(2s\bar{\sigma}_u + \xi_u^1 + \xi_u^2)\bar{a}_u^\top du\right].
\end{aligned}$$

Bibliography

V. V. Acharya and L. H. Pedersen. Asset pricing with liquidity risk. *J. Fin. Econ.*, 77(2):375–410, 2005.

K. Adam, J. Beutel, A. Marcet, and S. Merkel. Can a financial transaction tax prevent stock price booms? *J. Mon. Econ.*, 76:S90–S109, 2015.

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.

Z. Allen-Zhu, Y. Li, and Z. Song. A convergence theory for deep learning via over-parameterization. Preprint, 2018.

R. F. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Appl. Math. Finance*, 10(1):1–18, 2003.

R. F. Almgren and N. Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3:5–40, 2001.

R. F. Almgren and T. M. Li. Option hedging with smooth market impact. *Market Microstucture Liq.*, 2(1), 2016.

R. F. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *RISK*, July, 2005.

Y. Amihud and H. Mendelson. Asset pricing and the bid-ask spread. *J. Fin. Econ.*, 17(2):223–249, 1986.

Y. Amihud, H. Mendelson, and L. H. Pedersen. Liquidity and asset prices. *Foundations and Trends in Finance*, 1(4):269–364, 2006.

S. Ankirchner and T. Kruse. Optimal position targeting with stochastic linear–quadratic costs. *Banach Center Publ.*, 104(1):9–24, 2015.

P. Bank, H. M. Soner, and M. Voß. Hedging with temporary price impact. *Math. Fin. Econ.*, 11(2):215–239, 2017.

E. Bayraktar, T. Cayé, and I. Ekren. Asymptotics for small nonlinear price impact: a PDE homogenization approach to the multidimensional case. Preprint, 2018.

B. Bouchard, M. Fukasawa, M. Herdegen, and J. Muhle-Karbe. Equilibrium returns with transaction costs. *Finance Stoch.*, 22(3):569–601, 2018.

M. J. Brennan and A. Subrahmanyam. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. *J. Fin. Econ.*, 41 (3):441–464, 1996.

F. Bucci, M. Benzaquen, F. Lillo, and J.-P. Bouchaud. Crossover from linear to square-root market impact. *Phys. Rev. Lett.*, 122(10):108302, 2019.

A. Buss and B. Dumas. The dynamic properties of financial-market equilibrium with trading fees. *J. Finance*, 74(2):795–844, 2019.

A. Buss, B. Dumas, R. Uppal, and G. Vilkov. The intended and unintended consequences of financial-market regulations: A general-equilibrium analysis. *J. Mon. Econ.*, 81:25–43, 2016.

J. Cai, M. Rosenbaum, and P. Tankov. Asymptotic lower bounds for optimal tracking: a linear programming approach. *Ann. Appl. Probab.*, 27(4):2455–2514, 2017.

T. Cayé, M. Herdegen, and J. Muhle-Karbe. Trading with small nonlinear price impact. *Ann. Appl. Probab.*, to appear, 2019.

T. Cayé, M. Herdegen, and J. Muhle-Karbe. Scaling limits of processes with fast nonlinear mean reversion. *Stoch. Process. Appl.*, 130(4):1994–2031, 2020.

C Chicone. *Ordinary differential equations with applications*. Springer, New York, 1999.

J. H. Choi and K. Larsen. Taylor approximation of incomplete Radner equilibrium models. *Finance Stoch.*, 19:653–679, 2015.

J. H. Choi, K. Larsen, and D. J. Seppi. Smart TWAP trading in continuous-time equilibria. Preprint, 2018.

P. Collin-Dufresne, K. Daniel, and M. Sağlam. Liquidity regimes and optimal dynamic asset allocation. *J. Fin. Econ.*, 136(2):379–406, 2020.

A. Danilova and C. Julliard. Understanding volatility, liquidity, and the Tobin tax. *Preprint*, 2019.

M. H. A. Davis and A. R. Norman. Portfolio selection with transaction costs. *Math. Oper. Res.*, 15(4):676–713, 1990.

J. De Lataillade, C. Deremble, M. Potters, and J.-P. Bouchaud. Optimal trading with linear costs. *RISK*, 1(3):2047–1246, 2012.

S. S. Du, J. D. Lee, H. Li, L. Wang, and X. Zhai. Gradient descent finds global minima of deep neural networks. Preprint, 2018.

B. Dumas and E. Luciano. An exact solution to a dynamic portfolio choice problem under transactions costs. *J. Finance*, 46(2):577–595, 1991.

I. Ekeland and R. Temam. *Convex analysis and variational problems.* SIAM, Philadelphia, PA, 1999.

N. Garleanu and L. H. Pedersen. Dynamic portfolio choice with frictions. *J. Econ. Theory*, 165:487–516, 2016.

S. Gerhold, P. Guasoni, J. Muhle-Karbe, and W. Schachermayer. Transaction costs, trading volume, and the liquidity premium. *Finance Stoch.*, 18(1): 1–37, 2014.

L. Gonon, J. Muhle-Karbe, and X. Shi. Asset pricing with general transaction costs: Theory and numerics. Preprint, 2019.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, Cambridge, MA, 2016.

S. J. Grossman and J. E. Stiglitz. On the impossibility of informationally efficient markets. *Am. Econ. Rev.*, 70(3):393–408, 1980.

P. Guasoni and M. Rásonyi. Hedging, arbitrage and optimality with superlinear frictions. *Ann. Appl. Probab.*, 25(4):2066–2095, 2015.

P. Guasoni and M. Weber. Dynamic trading volume. *Math. Finance*, 27(2): 313–349, 2017.

P. Guasoni and M. Weber. Nonlinear price impact and portfolio choice. *Math. Finance*, to appear, 2018.

J. Han and J. Long. Convergence of the deep BSDE method for coupled FBSDEs. Preprint, 2018.

J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci.*, 115(34):8505–8510, 2018.

G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, Cambridge, UK, 1952.

H. Hau. The role of transaction costs for financial volatility: Evidence from the Paris bourse. *J. Europ. Econ. Ass.*, 4(4):862–890, 2006.

J. Heaton and D. J. Lucas. Evaluating the effects of incomplete markets on risk sharing and asset pricing. *J. Pol. Econ.*, 104(3):443–487, 1996.

M. Herdegen and J. Muhle-Karbe. Stability of Radner equilibria with respect to small frictions. *Finance Stoch.*, 22(2):443–502, 2018.

M. Herdegen, J. Muhle-Karbe, and D. Possamaï. Equilibrium asset pricing with transaction costs. Preprint, 2019.

S. L. Heston and S. Nandi. Preference-free option pricing with path-dependent volatility: A closed-form approach. Discussion Paper, Federal Reserve Bank of Atlanta, 1998.

K. Janeček and S. E. Shreve. Futures trading with transaction costs. *Illinois J. Math.*, 54(4):1239–1284, 2010.

C. M. Jones and P. J. Seguin. Transaction costs and price volatility: Evidence from commission deregulation. *Am. Econ. Rev.*, 4(87):728–737, 1997.

J. Kallsen and J. Muhle-Karbe. Option pricing and hedging with small transaction costs. *Math. Finance*, 25(4):702–723, 2015.

J. Kallsen and J. Muhle-Karbe. The general structure of optimal investment and consumption with small transaction costs. *Math. Finance*, 27(3):659–703, 2017.

I. Karatzas and S. E. Shreve. Brownian motion. In *Brownian Motion and Stochastic Calculus*. Springer, New York, 1998.

C. Kardaras, H. Xing, and G. Žitković. Incomplete stochastic equilibria with exponential utilities: close to Pareto optimality. Preprint, 2015.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. Preprint, 2014.

M. Kohlmann and S. Tang. Global adapted solution of one-dimensional backward stochastic Riccati equations, with application to the mean–variance hedging. *Stoch. Process. Appl.*, 97(2):255–288, 2002.

D. Kramkov and S. Pulido. A system of quadratic BSDEs arising in a price impact model. *Ann. Appl. Probab.*, 26(2):794–817, 2016a.

D. Kramkov and S. Pulido. Stability and analytic expansions of local solutions of systems of quadratic BSDEs with applications to a price impact model. *SIAM J. Fin. Math.*, 7(1):567–587, 2016b.

L. Kruk, J. Lehoczky, K. Ramanan, and S. E. Shreve. An explicit formula for the Skorokhod map on $[0, a]$. *Ann. Probab.*, 35(5):1740–1768, 2007.

A. S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6): 1315–1335, 1985.

F. Lillo, J. D. Farmer, and R. N. Mantegna. Master curve for price-impact function. *Nature*, 421:129–130, 2003.

A. W. Lo and J. Wang. Trading volume: definitions, data analysis, and implications of portfolio theory. *Rev. Fin. Stud.*, 13(2):257–300, 2000.

A. W. Lo, H. Mamaysky, and J. Wang. Asset prices and trading volume under fixed transaction costs. *J. Pol. Econ.*, 112(5):1054–1090, 2004.

A. W. Lynch and S. Tan. Explaining the magnitude of liquidity premia: The roles of return predictability, wealth shocks, and state-dependent transaction costs. *J. Finance*, 66(4):1329–1368, 2011.

R. Martin. Optimal trading under proportional transaction costs. *RISK*, 27 (8):54–59, 2014.

R. Martin and T. Schöneborn. Mean reversion pays, but costs. *RISK*, 24(2): 84–89, 2011.

L. Moreau, J. Muhle-Karbe, and H. M. Soner. Trading with small price impact. *Math. Finance*, 27(2):350–400, 2017.

J. Muhle-Karbe, M. Reppen, and H. M. Soner. A primer on portfolio choice with small transaction costs. *Ann. Rev. Fin. Econ.*, 9:301–331, 2017.

J. Muhle-Karbe, M. Nutz, and X. Tan. Asset pricing with heterogenous beliefs and illiquidity. *Math. Finance*, to appear, 2020.

R. Novy-Marx and M. Velikov. A taxonomy of anomalies and their trading costs. *Rev. Fin. Stud.*, 29(1):104–147, 2016.

L. Pástor and R. F. Stambaugh. Liquidity risk and expected stock returns. *J. Pol. Econ.*, 111(3):642–685, 2003.

Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl Dickstein. On the expressive power of deep neural networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, pages 2847–2854, 2017.

Y. Sannikov and A. Skrzypacz. Dynamic trading: price inertia and tront-running. Preprint, 2016.

N. Z. Shor. *Minimization methods for non-differentiable functions.* Springer, Berlin, 2012.

S. E. Shreve and H. M. Soner. Optimal investment and consumption with transaction costs. *Ann. Appl. Probab.*, 4(3):609–692, 1994.

H. M. Soner and N. Touzi. Homogenization and asymptotics for small transaction costs. *SIAM J. Control Optim.*, 51(4):2893–2921, 2013.

M. Taksar, M. J. Klass, and D. Assaf. A diffusion model for optimal portfolio selection in the presence of brokerage fees. *Math. Oper. Res.*, 13(2):277–294, 1988.

S. R. Umlauf. Transaction taxes and the behavior of the Swedish stock market. *J. Fin. Econ.*, 2(33):227–240, 1993.

D. Vayanos. Transaction costs and asset prices: a dynamic equilibrium model. *Rev. Fin. Stud.*, 11(1):1–58, 1998.

A. Y. Veretennikov. On polynomial mixing bounds for stochastic differential equations. *Stoch. Process. Appl.*, 70(1):115–127, 1997.

H. Xing and G. Žitković. A class of globally solvable Markovian quadratic BSDE systems and applications. *Ann. Probab.*, 46(1):491–550, 2018.

G. Žitković. An example of a stochastic equilibrium with incomplete markets. *Finance Stoch.*, 16(2):177–206, 2012.