ORIGINAL ARTICLE

Liquidity effects of trading frequency

Roman Gayduk | Sergey Nadtochiy

Mathematics Department, University of
Michigan, Michigan

Correspondence
Sergey Nadtochiy, Mathematics Department,
University of Michigan, 530 Church
Street, Ann Arbor, MI 48109, USA.
Email: sergeyn@umich.edu

Abstract

In this paper, we present a discrete-time modeling framework, in which the shape and dynamics of a Limit Order Book (LOB) arise endogenously from an equilibrium between multiple market participants (agents). We use the proposed modeling framework to analyze the effects of trading frequency on market liquidity in a very general setting. In particular, we demonstrate the dual effect of high trading frequency. On the one hand, the higher frequency increases market efficiency, if the agents choose to provide liquidity in equilibrium. On the other hand, it also makes markets more fragile, in the sense that the agents choose to provide liquidity in equilibrium only if they are market neutral (i.e., their beliefs satisfy certain martingale property). Even a very small deviation from market neutrality may cause the agents to stop providing liquidity, if the trading frequency is sufficiently high, which represents an endogenous liquidity crisis (also known as flash crash) in the market. This framework enables us to provide more insight into how such a liquidity crisis unfolds, connecting it to the so-called adverse selection effect.

KEYWORDS

conditional tails of Itô processes, continuum-player games, Limit Order Book, liquidity, trading frequency

1 | INTRODUCTION

This paper is concerned with liquidity effects of trading frequency on an auction-style exchange, in which the participating agents can post limit or market orders. On the one hand, higher trading frequency provides more opportunities for the market participants to trade, hence improving the liquidity of the market and increasing its efficiency. On the other hand, higher trading frequency also provides

more opportunities for some participants to manipulate the price and disrupt the market liquidity. Such a manipulation creates a new type of risk, which reveals itself in unusually high price deviations, which cannot be explained by changes in the fundamental value of the asset. The most famous example of this phenomenon is the "flash crash" of 2010. This example motivates the need for a comprehensive study of the trade-off between the liquidity providing role of strategic players and the liquidity risk they generate, and its relation to trading frequency. The collective liquidity of the market is captured by the *Limit Order Book (LOB)*, which contains all the limit buy and sell orders.

The goal of this paper is twofold. First, we develop a new framework for modeling market microstructure, in which the shape of the LOB, and its dynamics, arise *endogenously* from the interactions between the agents. Among the many advantages of such an approach is the possibility of modeling the market reaction to changes in the rules of the exchange: e.g., limited trading frequency, transaction tax, etc. The second, and most important, goal of the present work is to investigate the *liquidity effects* of *trading frequency*, using the proposed modeling framework. In particular, the main results of this paper (cf. the discussion in Section 3, as well as Theorems 4.8, 4.10, and Corollary 4.9 in Section 4) describe the dual effect of high trading frequency. On the one hand, if the agents choose to provide liquidity in equilibrium, higher trading frequency decreases the bid–ask spread and makes the expected profits of all market participants converge to the same (fundamental) value, thus improving the market efficiency. On the other hand, higher trading frequency also makes the LOB more sensitive to the deviations of the agents' attitudes from market neutrality. It is, of course, clear that a strong bullish or bearish signal induces market participants to trade at a higher or lower price. However, the novelty of our observation is in the role that the trading frequency plays in amplifying this effect. Namely, we show that if the trading frequency is high, even if agents have plenty of inventory, a very small deviation from market neutrality may cause them to stop providing liquidity, by either withdrawing from the market completely, or by posting limit orders far away from the fundamental price. Such actions cause disproportional deviations in the LOB, which cannot be explained by any fundamental reasons; they are much higher than the trading signal (i.e., the expected change in the fundamental price), and they occur without any shortage of supply or demand for the asset. We refer to such a deviation as an *endogenous* liquidity crisis, because it is due to the trading mechanism (i.e., the rules by which the market participants interact), rather than any fundamental reasons (note the similarity with the flash crash). Our framework provides insights into how such a liquidity crisis unfolds, connecting it to the so-called *adverse selection* effect. In particular, Section 3 constructs an equilibrium in which an endogenous liquidity crisis does not occur because of an abnormally large market order, wiping out the liquidity on one side of the LOB, but because the optimal strategies of the agents require them to stop providing liquidity on one side of the LOB. On the mathematical side, our analysis uses the properties of conditional tails of the increments of a general Itô process. The main result in that regard, in Lemma 5.2, provides a uniform exponential bound on the conditional tails of the increments of a general Itô process. We believe that this result is useful in its own right, and to the best of our knowledge, it is not available in the existing literature.

In recent years, we observed an explosion in the amount of literature devoted to the study of market microstructure. In addition to various empirical studies, a large part of the existing theoretical work focuses on the problem of optimal execution; see, among others, Obizhaeva and Wang (2013), Almgren (2003), Schied, Schoneborn, and Tehranchi (2010), Gatheral and Schied (2011), Cont, Stoikov, and Talreja (2010), Bayraktar and Ludkovski (2011), Avellaneda and Stoikov (2008), Cvitanić and Kirilenko (2010), Predoiu, Shaikhet, and Shreve (2011), Guéant and Lehalle (2015), Cartea and Jaimungal (2013), Guilbaud and Pham (2013), Stoikov and Waeber (2012), and references therein. In these articles, the dynamics and shape of the LOB are modeled exogenously, or equivalently, the arrival processes of the limit and market orders are specified exogenously. In particular, none of these articles

attempts to explain the shape and dynamics of the LOB, arising directly from the interaction between the market participants. A different approach to the analysis of market microstructure has its roots in the economic literature. For example, Parlour (1998), Foucault (1999), Goettler, Parlour, and Rajan (2005), Carmona and Webster (n.d.), Lachapelle, Lasry, Lehalle, and Lions (2016), Rosu (2009), Du and Zhu (2014), Bressan and Facchi (2013, 2014), and Bressan and Wei (2016) consider equilibrium models of market microstructure, and they are more closely related to the present work. However, the models proposed in the aforementioned papers do not aim to represent the mechanics of an auction-style exchange with sufficient precision, and in particular, they are not well suited for analyzing the liquidity effects of trading frequency, which is the main focus of this paper. A somewhat related strand of literature focuses on the endogenous formation of LOB in markets with a designated market maker; see, e.g., Glosten and Milgrom (1985), Kyle (1985), Easley and O'Hara (1992), Carmona and Webster (2012), and Adrian, Capponi, Vogt, and Zhang (2016). In these papers, the LOB is not an outcome of a multiagent equilibrium; instead, it is controlled by a single agent, the market maker. In this paper, we model the entire LOB as an output of an equilibrium between a large number of agents, each of whom is allowed to both consume and provide liquidity (in particular, we have no designated market maker). Our setting is related to the literature on *double auctions* (cf. Du & Zhu, 2014; Vayanos, 1999), with the crucial difference that the participants of each auction are allowed to choose two "asymmetric" types of strategies: market or limit orders. In addition, the present framework assumes that, ex ante, all agents have access to the same information, and in this sense, it is similar to Parlour (1998), Goettler et al. (2005), and Rosu (2009). In particular, the *adverse selection* effect, herein, does not arise from any a priori information asymmetry between agents; instead, it is caused by the *mechanics* of the exchange. We formulate the problem as a *continuum-player game*—this abstraction allows us to obtain computationally tractable results (cf. Aumann, 1964; Carmona, 2013; Schmeidler, 1973, for the concept of a continuum-player game, and Cardaliaguet, 2010; Carmona & Delarue, 2013; Lacker, 2016; Lasry & Lions, 2007, for the subclass of mean field games). The paper is organized as follows. Subsection 2.1 describes the probabilistic setting, along with the execution rules of the exchange and the resulting state processes of the agents. Subsection 2.2 defines the equilibrium and introduces the notion of *degeneracy* of the market (which represents an endogenous liquidity crisis). In Section 3, we construct an equilibrium in a simple model, illustrating how an endogenous liquidity crisis unfolds and how it is connected to the adverse selection effect. Theorems 4.8, 4.10, and Corollary 4.9, in Section 4, are the main results of the paper; they formalize and generalize the conclusions of Section 3. In Section 5, we prove the key technical results on the (conditional) tails of marginal distributions of Itô processes. Sections 6 and 7 contain the proofs of the main results. We conclude in Section 8.

2 | MODELING FRAMEWORK FOR A FINITE-FREQUENCY AUCTION-STYLE EXCHANGE

2.1 | Mechanics of the exchange

We consider an exchange in which trading can only occur at discrete times $n = 0, 1, \dots, N$. We assume that the market participants are split into two groups: the *external investors*, who are "impatient," in the sense that they only submit market orders, and the *strategic players*, who can submit both market and limit orders, and who are willing to optimize their actions over a given (short) time horizon, in order to get a better execution price.[1] In our study, we focus on the strategic players, who are referred to as *agents*, and we model the behavior of the external investors exogenously, via an *exogenous demand*. The interpretation of the external investors is clear: These are the investors who either have a longer

term view on the market or who simply need to buy or sell the asset for reasons other than short-term profits. The strategic players (i.e., agents), on the contrary, are short-term traders, who attempt to maximize their objective at a shorter time horizon N. During every time period $[n, n + 1)$, all the orders coming to the exchange are split into *limit* and *market* orders. The limit orders are collected in the so-called *Limit Order Book (LOB)*, and the market orders form the *demand curve*. At time $n + 1$, the market orders in the demand curve are executed against the limit orders in the LOB. Then, this process is repeated in the next time interval. In particular, during a time period $[n, n + 1)$ (for simplicity, we say "at time n"), an agent is allowed to submit a market order, post a limit buy or sell order, or wait (i.e., do nothing). If a limit order is not executed in a given time period, it costs nothing to cancel or reposition it for the next time period. Notice that our framework does not model the time priority of limit orders. However, introducing a time priority would not change the agents' maximum objective value, as the "tick size" is assumed to be zero (i.e., the set of possible price levels is $\mathbb{R}$), and hence, an agent can always achieve a priority by posting her order "infinitesimally" above or below a given competing order. Further details on modeling the formation of an LOB and the execution rules are presented below.

The demand curves are modeled exogenously by a random field $D = (D_n(p))_{p\in\mathbb{R}, n=1,\dots,N}$ on a filtered probability space $(\Omega, \mathbb{F} = (\mathcal{F}_n)_{n=0}^N, \mathbb{P})$, such that $\mathcal{F}_0$ is a trivial sigma-algebra, completed with respect to $\mathbb{P}$. The random variable $D_n^+(p) = \max(D_n(p), 0)$ denotes the number of shares of the asset that the external investors and the agents submitting market orders are willing to purchase at or below the price p, accumulated over the time period $[n - 1, n)$, and $D_n^-(p) = -\min(D_n(p), 0)$ denotes the number of shares of the asset that the external investors and the agents submitting market orders are willing to sell at or above the price p, in the same time period. We assume that $D_n(\cdot)$ is a.s. nonincreasing and measurable with respect to $\mathcal{F}_n \otimes \mathcal{B}(\mathbb{R})$. We denote by $\mathbb{A}$ a Borel space of *beliefs*, and for each $\alpha \in \mathbb{A}$, there exists a *subjective probability measure* $\mathbb{P}^\alpha$ on $(\Omega, \mathcal{F}_N)$, which is absolutely continuous with respect to $\mathbb{P}$. We assume that, for any $n = 0, \dots, N$ and any $\alpha \in \mathbb{A}$, there exists a regular version of the conditional probability $\mathbb{P}^\alpha$ given $\mathcal{F}_n$, denoted $\mathbb{P}_n^\alpha$.[2] We denote the associated conditional expectations by $\mathbb{E}_n^\alpha$. We also need to assume that, for any $\alpha \in \mathbb{A}$, there exists a modification of the family $\{\mathbb{P}_n^\alpha\}_{n=0}^N$, which satisfies the *tower property with respect to* $\mathbb{P}$, in the following sense: for any $n \le m$ and any r.v. ξ, such that $\mathbb{E}^\alpha \xi^+ < \infty$, we have

$$\mathbb{E}_n^\alpha \mathbb{E}_m^\alpha \xi = \mathbb{E}_n^\alpha \xi, \qquad \mathbb{P}\text{-a.s.}$$

There exists such a modification, e.g., if $\mathbb{P}^\alpha \sim \mathbb{P}$. In any market model, for every α, we fix such a modification of conditional probabilities (up to a set of $\mathbb{P}$-measure zero) and assume that all conditional expectations $\{\mathbb{E}_n^\alpha\}$ are taken under this family of measures. The LOB is given by a pair of adapted processes $\nu = (\nu_n^+, \nu_n^-)_{n=0}^N$, such that every ν_n^+ and ν_n^- is a finite sigma-additive random measure on $\mathbb{R}$ (with respect to $\mathcal{F}_n \otimes \mathcal{B}(\mathbb{R})$). Herein, ν_n^+ corresponds to the cumulative limit sell orders, and ν_n^- corresponds to the cumulative limit buy orders, posted at time n. The bid and ask prices at any time $n = 0, \dots, N$ are given by the random variables

$$p_n^b = \sup \operatorname{supp}\left(\nu_n^-\right), \qquad p_n^a = \inf \operatorname{supp}\left(\nu_n^+\right),$$

respectively. Notice that these extended random variables are always well defined but may take infinite values.

We define the *state space* of an agent as $\mathbb{S} = \mathbb{R} \times \mathbb{A}$, where the first component denotes the *inventory* of an agent, and the second component denotes her *beliefs*. Every agent in state (s, α) models the future outcomes using the subjective probability measure $\mathbb{P}^\alpha$. There are infinitely many agents, and their distribution over the state space is given by the *empirical distribution* process $\mu = (\mu_n)_{n=0}^N$, such that

every μ is a finite sigma-additive random measure on $\mathbb{S}$ (with respect to $\mathcal{F}_n \otimes \mathcal{B}(\mathbb{S})$). In particular, the total mass of agents in the set $S \subset \mathbb{S}$ at time n is given by $\mu_n(S)$. The inventory level s represents the *number of shares per agent*, held in state (s, α). In particular, the total number of shares held by all agents in the set $S \subset \mathbb{S}$ is given by $\int_S s\mu_n(ds, d\alpha)$. The interpretation of this definition in a finite-player game is discussed in Remark 2.1. We refer the reader to Carmona (2013) for the general concept of a continuum-player game.

Remark 2.1. *The continuum-player game defined in this section can be related to a finite-player game as follows. Denote by μ_0 the empirical distribution of the agents' states at a given time. Recall that μ_0 is a measure on $\mathbb{S} = \mathbb{R} \times \mathbb{A}$, and assume that it is a finite linear combination of Dirac measures: $\mu_0 = \frac{1}{M} \sum_{i=1}^{M} \delta_{(s^i, \alpha^i)}$. In this case, we interpret s^i as the number of shares per agent held by the agents in the ith group. Let us explain how this notion is related to the actual inventory levels (i.e., the actual numbers of shares held by the agents) in the associated finite-player game. To this end, consider a collection of M agents, whose states are given by their (actual) inventories and beliefs, (s, α), with the current states being $\{(\tilde{s}^i = s^i/M, \alpha^i)\}$. Define the "unit mass" of agents to be M. In this finite-player collection, the mass of agents (measured relative to the unit mass, M) at any state (Ms, α) is precisely $\mu_0(\{(s, \alpha)\})$, and their total inventory is $Ms\mu_0(\{(s, \alpha)\})$. The number of shares per agent is, then, defined as the total inventory held by these agents divided by their mass, and it is equal to Ms. Choosing $s = \tilde{s}^i$, we conclude that, in the finite-player collection, the number of shares per agent held by the agents at state $(\tilde{s}^i, \alpha^i)$ is given by $M\tilde{s}^i = s^i$, which coincides with our interpretation of s^i in the continuum-player game. It is also easy to show that an equilibrium in the proposed continuum-player game (defined in the next subsection) produces an approximate equilibrium in the associated finite-player game, when the inventory levels $\{\tilde{s}^i\}$ are small (cf. subsection 2.3 in the extended version of this paper, Gayduk & Nadtochiy, 2015).*

As the parameter α does not change over time, the state process of an agent, denoted (S_n), is an adapted $\mathbb{R}$-valued process, representing her inventory.[3] The control of every agent is given by a triplet of adapted processes $(p, q, r) = (p_n, q_n, r_n)_{n=0}^{N-1}$ on $(\Omega, \mathbb{F})$, with values in $\mathbb{R}^2 \times \{0, 1\}$. The first coordinate, p_n, indicates the location of a limit order placed at time n, and q_n indicates the size of the order (measured in shares per agent, and with negative values corresponding to buy orders).[4] The last coordinate r_n shows whether the agent submits a market order (if $r_n = 1$) or a limit order (if $r_n = 0$). Assume that an agent posts a limit sell order at a price level p_n. If the demand to buy the asset at this price level, $D_{n+1}^+(p_n)$, exceeds the amount of all limit sell orders posted below p_n at time n, then (and only then) the limit sell order of the agent is executed. Market orders of the agents are always executed at the bid or ask prices available at the time when the order is submitted. We interpret an internal market order (i.e., the one submitted by an agent) as the decision of an agent to join the external investors, in the given time period. Summing up the above, we obtain the following dynamics for the state process of an agent, starting with initial inventory $s \in \mathbb{R}$ at time $m = 0, \dots, N - 1$:

$$S_m^{(p,q,r)}(m, s, \nu) = s, \quad \Delta S_{n+1}^{(p,q,r)}(m, s, \nu) = S_{n+1}^{(p,q,r)}(m, s, \nu) - S_n^{(p,q,r)}(m, s, \nu) = -q_n \mathbf{1}_{\{r_n=1\}}$$

$$-\mathbf{1}_{\{r_n=0\}} \left(q_n^+ \mathbf{1}_{\{D_{n+1}^+(p_n) > \nu_n^+((-\infty, p_n))\}} - q_n^- \mathbf{1}_{\{D_{n+1}^-(p_n) > \nu_n^-((p_n, \infty))\}} \right), \quad n = m, \dots, N - 1. \quad (2.1)$$

The above dynamics represent an optimistic view on the execution by the agents. In particular, they imply that all limit orders at the same price level are executed in full, once the demand reaches them; i.e., each agent believes that her limit order will be executed first among all orders at a given price level. In addition, all agents' market orders are executed at the bid and ask prices; i.e., each agent believes that her market order will be executed first, when the demand curve is cleared against the LOB, at the

end of a given time period. These assumptions can be partially justified by the fact that the agents' orders are infinitesimal: q_n is measured in shares per agent, and an individual agent has zero mass. However, if a nonzero mass of agents submit limit orders at the same price level, or execute market orders, at the same time, then the above state dynamics may violate the market clearing condition: The total size of executed market orders (both in shares and in dollars) may not coincide with the total size of executed limit orders (at least as viewed by the agents). Nevertheless, this issue is resolved if, at any time, the mass of agents posting limit orders at the same price level or posting market orders is zero. In other words, (v, p, q, r) satisfy, $\mathbb{P}$-a.s.: v_n is continuous, as a measure on $\mathbb{R}$ (i.e., it has no atoms), and $r_n = 0$. Such an equilibrium is constructed in section 8 of the extended version of this paper (Gayduk & Nadtochiy, 2015). The general definition of a continuum-player game and its connection to a finite-player game can be found, e.g., in Carmona (2013) and in the references therein (see also subsection 2.3 in the extended version of this paper, Gayduk & Nadtochiy, 2015).

The modeling framework proposed herein has a close connection to the models of *double auctions*, in the economic literature (cf. Du & Zhu, 2014; Vayanos, 1999). The main difference is in the nonstandard design of the auction. Namely, in the proposed setting, the auction participants may choose different styles of trading, i.e., market or limit orders, which generates an ex post information asymmetry between participants: The limit orders have to be submitted before the demand curve is observed, while the market orders are submitted using complete information about the LOB. This difference is not coincidental—it is, in fact, crucial for a realistic representation of the risks associated with each order type, and it is at the core of the results established herein. A more detailed discussion of the information structure is provided in the next subsection.

2.2 | Equilibrium

The objective function of an agent, starting at the initial state $(s, \alpha) \in \mathbb{S}$, at any time $m = 0, \dots, N$, and using the control (p, q, r), is given by the $\mathcal{F}_m$-measurable random variable

$$
J^{(p,q,r)}(m, s, \alpha, v) = \mathbb{E}_m^\alpha \left[\left(S_N^{(p,q,r)}(m, s, v) \right)^+ p_N^b - \left(S_N^{(p,q,r)}(m, s, v) \right)^- p_N^a \right.
$$

$$
\left. - \sum_{n=m}^{N-1} \left(p_n \mathbf{1}_{\{r_n=0\}} + p_n^a \mathbf{1}_{\{r_n=1, q_n<0\}} + p_n^b \mathbf{1}_{\{r_n=1, q_n>0\}} \right) \Delta S_{n+1}^{(p,q,r)}(m, s, v) \right], \tag{2.2}
$$

where we assume that $0 \cdot \infty = 0$. In the above expression, we assume that, at the final time $n = N$, each agent is forced to liquidate her position at the bid or ask prices available at that time. Alternatively, one can think of it as *marking to market* the residual inventory, right after the last external market order is executed.

Definition 2.2. *For a given LOB v, integer $m = 0, \dots, N - 1$, and state $(s, \alpha) \in \mathbb{S}$, the triplet of adapted processes (p, q, r) is an admissible control if the positive part of the expression inside the expectation in (2.2) has a finite expectation under $\mathbb{P}^\alpha$.*

For a given LOB v, an initial condition (m, s, α), and a triplet of $\mathbb{F} \times \mathcal{B}(\mathbb{S})$-adapted random fields (p, q, r), we identify the latter (whenever it causes no confusion) with stochastic processes (p, q, r) via:

$$
p_n = p_n \left(S_n^{(p,q,r)}(m, s, v), \alpha \right), \quad q_n = q_n \left(S_n^{(p,q,r)}(m, s, v), \alpha \right), \quad r_n = r_n \left(S_n^{(p,q,r)}(m, s, v), \alpha \right),
$$

and the state dynamics (2.1), for $n = m, \dots, N$. This system determines (p, q, r) and $S^{(p,q,r)}$ recursively.

Definition 2.3. *For a given LOB v, we call the triplet of progressively measurable random fields (p, q, r) an optimal control if, for any $m = 0, \ldots, N$ and any $(s, \alpha) \in \mathbb{S}$, we have:*

- *(p, q, r) is admissible,*
- *$J^{(p,q,r)}(m, s, \alpha, v) \geq J^{(p',q',r')}(m, s, \alpha, v)$, $\mathbb{P}$-a.s., for any admissible control (p', q', r').*

In the above, we make the standard simplifying assumptions of continuum-player games: Each agent is too small to affect the empirical distribution of cumulative controls (reflected in v) when she changes her control (cf. Carmona, 2013). Note also that our definition of the optimal control implies that it is time consistent: Reevaluation of the optimality at any future step, using the same terminal criteria, must lead to the same optimal strategy. Next, we discuss the notion of equilibrium in the proposed game. First, we notice that, if p_N^b or p_N^a becomes infinite, the agents with positive or negative inventory may face the objective value of "$-\infty$", for any control they use. In such a case, their optimal controls may be chosen in an arbitrary way, resulting in unrealistic equilibria. To avoid this, we impose the additional regularity condition on v.

Definition 2.4. *A given LOB v is admissible if, for any $m = 0, \ldots, N - 1$ and any $\alpha \in \mathbb{A}$, we have, $\mathbb{P}$-a.s.:*

$$\mathbb{E}_m^\alpha |p_N^a| \vee |p_N^b| < \infty.$$

Let us consider the (stochastic) value function of an agent for a fixed (m, s, α, v):

$$V_m^v(s, \alpha) = \mathrm{esssup}_{p,q,r} J^{(p,q,r)}(m, s, \alpha, v), \tag{2.3}$$

where the essential supremum is taken under $\mathbb{P}$, over all admissible controls (p, q, r), and $J^{(p,q,r)}$ is given by (2.2). Appendix A shows that for any admissible v, $V_m^v(\cdot, \alpha)$ has a continuous modification under $\mathbb{P}$, which we refer to as the value function of an agent with beliefs α. Using the Dynamic Programming Principle, Appendix A provides an explicit system of recursive equations that characterize optimal strategies and the value function. In particular, the results of Appendix A (cf. Corollary A.2) yield the following proposition.

Proposition 2.5. *Assume that, for an admissible LOB v, there exists an optimal control $(\hat{p}, \hat{q}, \hat{r})$. Then, for any $(s, \alpha) \in \mathbb{S}$, the following holds $\mathbb{P}$-a.s., for all $n = 0, \ldots, N - 1$:*

$$V_n^v(s, \alpha) = s^+ \lambda_n^a(\alpha) - s^- \lambda_n^b(\alpha),$$

with some adapted processes $\lambda^a(\alpha)$ and $\lambda^b(\alpha)$, such that $\lambda_N^a(\alpha) = p_N^b$ and $\lambda_N^b(\alpha) = p_N^a$.

The values of $\lambda^a(\alpha)$ and $\lambda^b(\alpha)$ can be interpreted as the *expected execution prices* of the agents with beliefs α, who are long and short the asset, respectively.

Definition 2.6. *Consider an empirical distribution process $\mu = (\mu_n)_{n=0}^N$ and a market model, as described in Subsection 2.1. We say that a given LOB process v and a control (p, q, r) form an equilibrium, if there exists a Borel set $\tilde{\mathbb{A}} \subset \mathbb{A}$, called the* support *of the equilibrium, such that:*

(1) *$\mu_n(\mathbb{R} \times (\mathbb{A} \setminus \tilde{\mathbb{A}})) = 0$, $\mathbb{P}$-a.s., for all n,*

(2) *v is admissible, and (p, q, r) is an optimal control for v, on the state space $\tilde{\mathbb{S}} = \mathbb{R} \times \tilde{\mathbb{A}}$,*

(3) *and, for any $n = 0, \ldots, N - 1$, we have, $\mathbb{P}$-a.s.,*

$$v_n^+((-\infty, x]) = \int_{\tilde{\mathbb{S}}} \mathbf{1}_{\{p_n(s,\alpha) \leq x, r_n(s,\alpha) = 0\}} \, q_n^+(s, \alpha) \mu_n(ds, d\alpha), \quad \forall x \in \mathbb{R}, \tag{2.4}$$

$$v_n^-((-\infty, x]) = \int_{\tilde{\mathbb{S}}} \mathbf{1}_{\{p_n(s,\alpha) \leq x, r_n(s,\alpha) = 0\}} \, q_n^-(s, \alpha) \mu_n(ds, d\alpha), \quad \forall x \in \mathbb{R}. \tag{2.5}$$

Remark 2.7. *It follows from Proposition 2.5 that, in equilibrium, it is optimal for an agent with zero initial inventory to do nothing. Hence, in equilibrium, roundtrip strategies are impossible. To allow for roundtrip strategies in equilibrium, one can, e.g., introduce an upper bound on $|q|$ or on the total inventory of an agent (as it is done, e.g., in Brunnermeier & Pedersen, 2005). However, we do not believe that such a modification would change the qualitative behavior of market liquidity as a function of trading frequency, which is the main focus of this paper.*

Notice that, because the optimal controls are required to be time consistent under $\mathbb{P}$, the above definition, in fact, defines a *subgame perfect equilibrium*. It is also worth mentioning that Definition 2.6 defines a *partial equilibrium*, as the empirical distribution process μ is given exogenously. A more traditional version of Nash equilibrium would require μ to be determined by the initial distribution and the values of the state processes:

$$\mu_n = \mu_0 \circ \left((s, \alpha) \mapsto \left(S_n^{(p,q,r)}(0, s, \nu), \alpha \right) \right)^{-1}, \tag{2.6}$$

which must hold $\mathbb{P}$-a.s., for all $n = 0, \ldots, N$, with $S_n^{(p,q,r)}(0, s, \nu)$ defined via (2.1), in addition to the other conditions in Definition 2.6. Nevertheless, we choose not to enforce the condition (2.6) in the definition of equilibrium, in order to allow new agents to enter the game, which, in effect, amounts to modeling μ exogenously. If one assumes that no new agent arrives to the market, then the fixed-point condition (2.6) has to be enforced. Note also that our interpretation of the demand curve $D_n(\cdot)$ implies that it consists of both the external (i.e., due external investors) and internal (i.e., due to the agents) market orders. Therefore, it may be reasonable to consider an additional consistency condition for an equilibrium. A part of this condition is to ensure that a nonzero mass of agents submit market buy orders only if the fundamental price rises above the ask price (i.e., only if a market buy order is actually executed), and similarly, a nonzero mass of agents submit market sell orders only if the fundamental price falls below the bid price. We assume that the agents' market orders enter into the demand curve with the highest level of priority; e.g., their market buy orders enter the demand curve at the price level infinitesimally close to, but below, the fundamental price, in order to guarantee that they are the first ones to be executed. Thus, another part of the aforementioned consistency condition is to ensure that the absolute value of the demand curve to the left or to the right of the fundamental price is sufficiently large to account for all internal market orders. Mathematically, such consistency condition can be formulated as follows:

$$d_n^b := \mu_n(\{(s, \alpha) \; : \; q_n(s, \alpha) < 0, \, r_n(s, \alpha) = 1\}) > 0 \; \Rightarrow \; p_{n+1}^0 > p_n^a, \; \lim_{p \uparrow p_{n+1}^0} D_{n+1}^+(p) \geq d_n^b, \tag{2.7}$$

$$d_n^a := \mu_n(\{(s, \alpha) \; : \; q_n(s, \alpha) > 0, \, r_n(s, \alpha) = 1\}) > 0 \; \Rightarrow \; p_{n+1}^0 < p_n^b, \; \lim_{p \downarrow p_{n+1}^0} D_{n+1}^-(p) \geq d_n^a. \tag{2.8}$$

The above conditions become redundant if the agents never submit market orders in equilibrium. Section 8 of the extended version of this paper (Gayduk & Nadtochiy, 2015), shows how to construct an equilibrium that satisfies condition (2.6), and in which the agents never submit market orders (hence, (2.7) and (2.8) are also satisfied). However, it is important to emphasize that the main results of the present work (cf. Section 4) provide necessary conditions for *all* equilibria: for those satisfying the conditions (2.6), (2.7), (2.8) and for the ones that do not.

Remark 2.8. *Let us comment on the information structure of the game. In the present setting, all agents observe the same information, given by the filtration $\mathbb{F}$. We consider an open-loop Nash equilibrium, in which the agent's strategy is viewed as an adapted stochastic process (rather than a function of the states and controls of other players), and the definition of optimality is chosen accordingly. In addition, as μ is adapted to $\mathbb{F}$, each agent has complete information about the present and past states of other agents, and their beliefs. However, as the agents use different (subjective) measures $\{\mathbb{P}^\alpha\}$, their views on the future values of μ may be different. Of course, it would be more realistic to assume that the agents do not have complete information about each other's current states, but this would make the problem significantly more complicated. In the present setting, the agents also have complete information about the current location of the fundamental price. In our follow-up paper, Gayduk and Nadtochiy (2018), we relax this assumption, which allows us to develop a more realistic model for the "local" behavior of an individual agent. However, such a relaxation does not seem necessary for the questions analyzed herein.*

As all agents use the same information, this paper belongs to the strand of literature that attempts to explain microstructure phenomena without information asymmetry (cf. Foucault, 1999; Goettler et al., 2005; Parlour, 1998; Rosu, 2009). Nevertheless, it is important to mention that information asymmetry arises ex post, between the market participants submitting market and limit orders. This asymmetry is not due to superior information a priori available to any of the agents. Instead, it stems from the very nature of limit orders, which are "passive" by design (cf. the discussion on the last paragraph of Subsection 2.1). Similar observation is made in Goettler et al. (2005).

Next, we need to add another condition to the notion of equilibrium. Notice that equations (2.4) and (2.5) should serve as the fixed-point constraints that enable one to obtain the optimal controls (p, q, r), along with the LOB ν. However, these equations only hold for $n = 0, \dots, N - 1$; indeed, the agents do not need to choose their controls at time $n = N$, as the game is over and their residual inventory is marked to the bid and ask prices. However, the terminal bid and ask prices are determined by the LOB ν_N, which, in turn, can be chosen arbitrarily. To avoid such ambiguity, we impose an additional constraint on the equilibria studied herein. First, we introduce the notion of a *fundamental price*.

Definition 2.9. *Assume that $\mathbb{P}$-a.s., for any $n = 1, \dots, N$, there exists a unique p_n^0 satisfying $D_n(p_n^0) = 0$. Then, the adapted process $(p_n^0)_{n=1}^N$ is called the fundamental price process.*

Whenever the notion of a fundamental price is invoked, we assume that it is well defined. The intuition behind p^0 is clear: It is a price level at which the immediate demand is balanced. However, it is important to stress that we do not assume that the asset can be traded at the fundamental price level. Rather, p^0 is a feature of the abstract current demand curve, whereas all actual trading happens on the exchange, against the current LOB. This aspect of our setting differs from many other approaches in the literature.

Definition 2.10. *Assume that the fundamental price is well defined and denote $\xi_N = p_N^0 - p_{N-1}^0$. Then, an equilibrium with LOB ν is linear at terminal crossing (LTC) if*

$$\nu_N = \nu_{N-1} \circ (x \mapsto x + \xi_N)^{-1}, \qquad \mathbb{P}\text{-a.s.} \tag{2.9}$$

The above definition assumes that the terminal LOB v_N is obtained from v_{N-1} by a simple shift, with the size of the shift equal to the increment in the fundamental price. This definition connects the LOB at the terminal time with the demand process, ruling out many unnatural equilibria. In particular, the question of existence of an equilibrium becomes nontrivial. However, the mere existence of an equilibrium is not the main focus of the present work: The existence results, established herein, are limited to Section 3, which constructs an LTC equilibrium in a specific Gaussian random walk model (a slightly more general existence result is given in section 8 of the extended version of this paper, Gayduk & Nadtochiy, 2015). What is central to the present investigation is the observation that the agents may reach an equilibrium in which one side of the LOB becomes empty (as demonstrated by the example of Section 3). We call such LOB, and the associated equilibrium, *degenerate*.

Definition 2.11. *We say that an equilibrium with LOB v is nondegenerate if $v_n^+(\mathbb{R}) > 0$ and $v_n^-(\mathbb{R}) > 0$, for all $n = 0, \ldots, N-1$, $\mathbb{P}$-a.s.*

Intuitively, the degeneracy of the LOB refers to a situation where, with positive probability, one side of the LOB disappears from the market; i.e., $v_n^+(\mathbb{R})$ or $v_n^-(\mathbb{R})$ becomes zero. Clearly, this happens when the agents who are supposed to provide liquidity choose to post market orders (i.e., consume liquidity) or wait (neither provide nor consume liquidity). Such a degeneracy can be interpreted as the *endogenous liquidity crisis*—the one that arises purely from the interaction between the agents, and cannot be justified by any fundamental economic reasons (e.g., the external demand for the asset may still be high, on both sides). Taking an optimistic point of view, we assume that the agents choose a nondegenerate equilibrium, whenever one is available. However, if a nondegenerate equilibrium does not exist, an endogenous liquidity crisis may occur with positive probability. One of the main goals of this paper is to provide insights into the occurrence of an endogenous liquidity crisis and its relation to trading frequency.

3 | EXAMPLE: A GAUSSIAN RANDOM WALK MODEL

In this section, we consider a specific market model for the external demand D, to construct a non-degenerate LTC equilibrium. More importantly, using this model, we illustrate the liquidity effects of trading frequency. The present example, albeit very simplistic, enables us to identify the important changes in the optimal strategies of the agents (and, hence, to the LOB) as the trading frequency increases. In particular, we demonstrate how the *adverse selection* effect may be amplified disproportionally by the high trading frequency and may cause a liquidity crisis. Note that the adverse selection phenomenon, in the present setting, is not a consequence of any ex ante information asymmetry but is due to the mechanics of the exchange (i.e., the nature of limit orders), which is similar to the phenomena documented in Goettler et al. (2005) and Foucault (1999). In the rest of the paper, we show that the conclusions of this section are not due to the particular choice of a model made in the present section and, in fact, persist in a much more general setting.

On a complete stochastic basis $(\Omega, \tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{t \in [0,T]}, \mathbb{P})$, we consider a continuous-time process $\tilde{p}_0$:

$$\tilde{p}_t^0 = p_0^0 + \alpha t + \sigma W_t, \quad p_0^0 \in \mathbb{R}, \quad t \in [0, T], \tag{3.1}$$

where $\alpha \in \mathbb{R}$ and $\sigma > 0$ are constants, and W is a Brownian motion. We also consider an arbitrary progressively measurable random field $(\tilde{D}_t(p))$, s.t., $\mathbb{P}$-a.s., the function $\tilde{D}_t(\cdot) - \tilde{D}_s(\cdot)$ is strictly decreasing and vanishing at zero, for any $0 \le s < t \le T$. Finally, we introduce the empirical distribution process $(\tilde{\mu}_t)$, with values in the space of finite sigma-additive measures on $\mathbb{S}$. We partition the time interval

$[0, T]$ into N subintervals of size $\Delta t = T/N$. A discrete-time model is obtained by discretizing the continuous-time one[5]

$$\mathcal{F}_n = \tilde{\mathcal{F}}_{n\Delta t}, \quad p_n^0 = \tilde{p}_{n\Delta t}^0, \quad D_n(p) = (\tilde{D}_{n\Delta t} - \tilde{D}_{(n-1)\Delta t})(p - p_n^0), \quad \mu_n = \tilde{\mu}_{n\Delta t}.$$

In this section, for simplicity, we assume that the set of agents' beliefs is a singleton: $\mathbb{A} = \{\alpha\}$ and $\mathbb{P}^\alpha = \mathbb{P}$. We also assume that (at least from the agents' point of view) there are always some long and short agents present in the market: $\mu_n((0, \infty) \times \mathbb{A}), \mu_n((-\infty, 0) \times \mathbb{A}) > 0$, $\mathbb{P}$-a.s., for all n. Clearly, N represents the trading frequency, and the continuous-time model represents the "limiting model," which the agents use as a benchmark, in order to make consistent predictions in the markets with different trading frequencies. We assume that the benchmark model is fixed, and N is allowed to vary. In the remainder of this section, we propose a method for constructing a nondegenerate LTC equilibrium in the above discrete-time model. We show that the method succeeds for any (N, σ) if $\alpha = 0$. However, for $\alpha \neq 0$, we demonstrate numerically that the method fails as N becomes large enough. We show why, precisely, the proposed construction fails, providing an economic interpretation of this phenomenon. Moreover, we analyze the market close to the moment when a nondegenerate equilibrium fails to exist and demonstrate that the agents' behavior at this time follows the pattern typical for an endogenous liquidity crisis.

In view of Proposition 2.5, in order to construct a nondegenerate LTC equilibrium, we need to find a control $(\hat{p}, \hat{q}, \hat{r})$, and the expected execution prices $(\hat{\lambda}^a, \hat{\lambda}^b)$, s.t. the value function of an agent with inventory s is given by $V_n(s) = s^+ \hat{\lambda}_n^a - s^- \hat{\lambda}_n^b$, and it is attained by the strategy $(\hat{p}, \hat{q}, \hat{r})$. In addition, we need to find a nondegenerate LOB ν, s.t. (2.4), (2.5), and (2.9) hold. Our ansatz is as follows:

$$\nu_n = \left(h_n^a \delta_{p_n^a}, h_n^b \delta_{p_n^b} \right), \quad p_n^a = \hat{p}_n^a + p_n^0, \quad p_n^b = \hat{p}_n^b + p_n^0, \quad -\infty < \hat{p}_n^b, \hat{p}_n^a < \infty,$$

$$\hat{p}_n(s) = p_n^a \mathbf{1}_{\{s > 0\}} + p_n^b \mathbf{1}_{\{s < 0\}}, \quad \hat{q}_n(s) = s, \quad \hat{r}_n(s) = 0, \quad \lambda_n^a = \hat{\lambda}_n^a + p_n^0, \quad \lambda_n^b = \hat{\lambda}_n^b + p_n^0,$$

where δ is the Dirac measure, $(\hat{p}^a, \hat{p}^b, \hat{\lambda}^a, \hat{\lambda}^b)$ are deterministic processes, and $h_n^a = \int_0^\infty s \mu_n(ds) > 0$, $h_n^b = \int_{-\infty}^0 |s| \mu_n(ds) > 0$. With such an ansatz, conditions (2.4) and (2.5) are satisfied automatically. Thus, we only need to choose finite deterministic processes $(\hat{p}^a, \hat{p}^b, \hat{\lambda}^a, \hat{\lambda}^b)$ s.t.: $\hat{p}_N^a = \hat{p}_{N-1}^a, \hat{p}_N^b = \hat{p}_{N-1}^b$ (so that the equilibrium is LTC) and the associated $(\hat{p}, \hat{q}, 0)$ form an optimal control, producing the value function $V_n(s) = s^+ \lambda_n^a - s^- \lambda_n^b$. Appendix A contains necessary and sufficient conditions for characterizing such families $(p^a, p^b, \lambda^a, \lambda^b)$. In particular, we deduce from Corollaries A.2 and A.3 that $(\hat{p}_{N-1}^a, \hat{p}_{N-1}^b, \hat{\lambda}_{N-1}^a, \hat{\lambda}_{N-1}^b)$ form a suitable family in a single-period case, $[N-1, N]$, if they solve the following system:

$$\begin{cases} \hat{p}_{N-1}^a \in \arg\max_{p \in \mathbb{R}} \mathbb{E}\left(\left(p - \hat{p}_{N-1}^b - \xi \right) \mathbf{1}_{\{\xi > p\}} \right), \quad \hat{p}_{N-1}^b < 0, \\[2mm] \hat{p}_{N-1}^b \in \arg\max_{p \in \mathbb{R}} \mathbb{E}\left(\left(\hat{p}_{N-1}^a - p + \xi \right) \mathbf{1}_{\{\xi < p\}} \right), \quad \hat{p}_{N-1}^a > 0, \\[2mm] \hat{\lambda}_{N-1}^a = \hat{p}_{N-1}^b + \alpha \Delta t + \mathbb{E}\left(\left(\hat{p}_{N-1}^a - \hat{p}_{N-1}^b - \xi \right) \mathbf{1}_{\{\xi > \hat{p}_{N-1}^a\}} \right), \\[2mm] \hat{\lambda}_{N-1}^b = \hat{p}_{N-1}^a + \alpha \Delta t - \mathbb{E}\left(\left(\hat{p}_{N-1}^a - \hat{p}_{N-1}^b + \xi \right) \mathbf{1}_{\{\xi < \hat{p}_{N-1}^b\}} \right), \\[2mm] \hat{p}_{N-1}^b \leq \hat{\lambda}_{N-1}^a, \quad \hat{\lambda}_{N-1}^b \leq \hat{p}_{N-1}^a, \quad \hat{p}_{N-1}^a \geq \hat{p}_{N-1}^b + |\alpha| \Delta t, \end{cases} \quad (3.2)$$

where $\xi = \Delta p_N^0 \sim \mathcal{N}(\alpha \Delta t, \sigma^2 \Delta t)$. Let us comment on the economic meaning of the equations in (3.2). The expectations in the first two lines represent the *relative expected profit* from executing a limit order
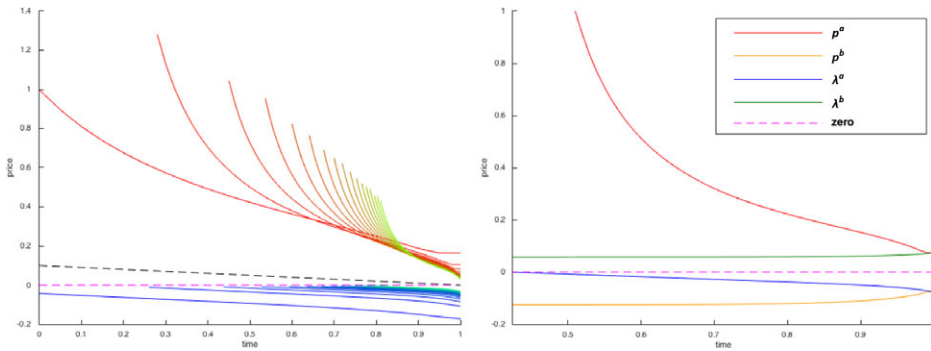
FIGURE 3.1 (Left) Ask price $\hat{p}^a$ (in red) and the associated expected execution prices $\hat{\lambda}^a$ (in blue); different curves correspond to different trading frequencies ($N = 20, \ldots, 500$); black dashed line is the expected change in the fundamental price $\alpha(T - t)$. (Right) Ask price $\hat{p}^a$ (in red) and the associated expected execution price $\hat{\lambda}^a$ (in blue), bid price $\hat{p}^b$ (in orange), and the associated expected execution price $\hat{\lambda}^b$ (in green), for $N = 100$. Nondegenerate equilibrium exists only on a time interval where $\hat{\lambda}^a < 0$. All prices are measured relative to the fundamental price and are plotted as functions of time. Positive drift: $\alpha = 0.1$, $\sigma = 1$, $T = 1$ [Color figure can be viewed at wileyonlinelibrary.com]

at time N, at the chosen price level $p + p^0_{N-1}$, versus marking the inventory to market at time N, at the best price available on the other side of the book, i.e., $p^b_N = \hat{p}^b_{N-1} + \xi + p^0_{N-1}$ or $p^a_N = \hat{p}^a_{N-1} + \xi + p^0_{N-1}$. Notice that a limit order is executed if and only if the fundamental price at time N is above or below the chosen limit order, i.e., if $p^0_{N-1} + \xi > p + p^0_{N-1}$ or $p^0_{N-1} + \xi < p + p^0_{N-1}$.[6] Clearly, it is only optimal for an agent to post a limit order if the relative expected profit is nonnegative, which is the case if and only if $\hat{p}^b_{N-1} < 0 < \hat{p}^a_{N-1}$. The third and fourth lines in (3.2) represent the expected execution prices of the agents at time $N - 1$, assuming they use the controls given by $(\hat{p}^a_{N-1}, \hat{p}^b_{N-1})$. Each of the right-hand sides is a sum of two components: the relative expected profit from posting a limit order and the expected value of marking to market at time N, measured relative to p^0_{N-1}. Let us analyze the inequalities in the last line of (3.2). If the bid price at time $N - 1$ exceeds the expected execution price of a long agent, i.e., $\hat{p}^b_{N-1} + p^0_{N-1} > \hat{\lambda}^a_{N-1} + p^0_{N-1}$, then every agent with positive inventory prefers to submit a market order, rather than a limit order, at time $N - 1$, which causes the ask side of the LOB to degenerate. Similarly, we establish $\hat{\lambda}^b_{N-1} \leq \hat{p}^a_{N-1}$. Finally, if $\alpha > 0$ and $\hat{p}^a_{N-1} < \hat{p}^b_{N-1} + \alpha\Delta t$, an agent may buy the asset using a market order at time $N - 1$, at the price $\hat{p}^a_{N-1} + p^0_{N-1}$, and sell it at time N, at the expected price $\hat{p}^b_{N-1} + p^0_{N-1} + \alpha\Delta t > \hat{p}^a_{N-1} + p^0_{N-1}$ (a reverse strategy works for $\alpha < 0$). This strategy can be scaled to generate infinite expected profit and, hence, is excluded by the last inequality in the last line of (3.2).

We construct a solution to (3.2) by solving a fixed-point problem given by the first two lines of (3.2) and verifying that the desired inequalities hold.[7] We implement this computation in MatLab, and the results can be seen as the right-most points on the graphs in Figure 3.1. From the numerical solution, we see that, whenever Δt is small enough, the conditions $\hat{p}^b_{N-1} \leq \hat{\lambda}^a_{N-1}$ and $\hat{\lambda}^b_{N-1} \leq \hat{p}^a_{N-1}$ are satisfied (cf. the right part of Figure 3.1).[8] In addition, for $\alpha \geq 0$, we have

$$0 < \mathbb{E}\left(\hat{p}^a_{N-1} - \hat{p}^b_{N-1} - \xi \mid \xi > \hat{p}^a_{N-1}\right) = \hat{p}^a_{N-1} - \hat{p}^b_{N-1} - \mathbb{E}\left(\xi \mid \xi > \hat{p}^a_{N-1}\right) \leq \hat{p}^a_{N-1} - \hat{p}^b_{N-1} - \alpha\Delta t,$$

which yields the last inequality in (3.2). The case of $\alpha < 0$ is treated similarly. Notice that $\hat{\lambda}^a_N = \hat{p}^b_N = \hat{p}^b_{N-1}$ and $\hat{p}^a_{N-1} = \hat{p}^a_N = \hat{\lambda}^b_N$. Thus, the single-period equilibrium we have constructed

satisfies

$$\hat{p}_n^b \le \hat{\lambda}_n^a, \quad \hat{\lambda}_n^b \le \hat{p}_n^a, \quad \hat{\lambda}_{n+1}^a < 0, \quad \hat{\lambda}_{n+1}^b > 0 \tag{3.3}$$

for $n = N - 1$. If one of the first two inequalities in (3.3) fails, the agents choose to submit market orders, as opposed to limit orders, which leads to *degeneracy* of the LOB—one side of it disappears. If one of the last two inequalities fails, the execution of a limit order, at any price level, yields a negative relative expected profit for the agents on one side of the book (given by the expectation in the first or second line of (3.2)). As a result, it becomes optimal for all such agents to stop posting any limit orders, and the LOB degenerates. The latter is interpreted as the *adverse selection* effect. For example, if the third inequality in (3.3) fails, then every long agent believes that, no matter the price at which her limit order is posted, if it is executed in the next time period, her expected execution price at the next time step will be higher than the price at which the limit order is executed. Hence, it suboptimal to post a limit order at all.

In a single period $[N - 1, N]$, by choosing small enough Δt, we can ensure that the inequalities in (3.3) are satisfied. However, it turns out that as we progress recursively, constructing an equilibrium, we may encounter a time step at which one of the inequalities in (3.3) fails, implying that a nondegenerate LTC equilibrium cannot be constructed for the given time period (at least using the proposed method). To see this, consider the recursive equations for $(\hat{p}^a, \hat{\lambda}^a)$ (which are chosen to satisfy the conditions of Corollary A.2, in Appendix A, given our ansatz):

$$\begin{cases} \hat{p}_n^a \in \arg\max_{p \in \mathbb{R}} \mathbb{E}\left(\left(p - \hat{\lambda}_{n+1}^a - \xi \right) \mathbf{1}_{\{\xi > p\}} \right), \\ \hat{\lambda}_n^a = \hat{\lambda}_{n+1}^a + \alpha \Delta t + \mathbb{E}\left(\left(\hat{p}_n^a - \hat{\lambda}_{n+1}^a - \xi \right) \mathbf{1}_{\{\xi > \hat{p}_n^a\}} \right) < 0, \end{cases} \tag{3.4}$$

and similarly for $(\hat{p}^b, \hat{\lambda}^b)$. Using the properties of the Gaussian distribution, it is easy to see that if $\hat{\lambda}_{n+1}^a < 0$, we have $\hat{p}_n^a > 0$. A similar conclusion holds for $(\hat{\lambda}^b, \hat{p}^b)$. Thus, if $\hat{\lambda}_k^a < 0 < \hat{\lambda}_k^b$, for $k = n + 1, \dots, N$, our method allows us to construct a nondegenerate LTC equilibrium on the time interval $[n, N]$, with $\hat{p}^b < 0 < \hat{p}^a$. Such a construction always succeeds if the agents are market neutral, i.e., $\alpha = 0$. Indeed, in this case, assuming $\hat{\lambda}_{n+1}^a < 0 < \hat{\lambda}_{n+1}^b$, we have $\hat{p}_n^b < 0 < \hat{p}_n^a$ and

$$\hat{\lambda}_{n+1}^a + \left(\mathbb{E}\left(\left(\hat{p}_n^a - \hat{\lambda}_{n+1}^a - \xi \right) \mathbf{1}_{\{\xi > \hat{p}_n^a\}} \right) \right)^+ = \mathbb{E}\left(\hat{\lambda}_{n+1}^a \mathbf{1}_{\{\xi > \hat{p}_n^a\}} \right) + \mathbb{E}\left(\left(\hat{p}_n^a - \xi \right) \mathbf{1}_{\{\xi > \hat{p}_n^a\}} \right) < 0.$$

Hence, $\hat{\lambda}_n^a < 0$, and, similarly, we deduce that $\hat{\lambda}_n^b > 0$. By induction, we obtain a nondegenerate LTC equilibrium on $[0, N]$, for any (N, σ), as long as $\alpha = 0$. Corollary 4.9 shows that, as $N \to \infty$, the processes $(\hat{\lambda}^a, \hat{\lambda}^b)$ converge to zero, which means that the expected execution prices converge to the fundamental price. The latter is interpreted as *market efficiency* in the high-frequency trading regime: Any market participant expects to buy or sell the asset at the fundamental price. The left-hand side of Figure 3.2 shows that the bid and ask prices also converge to the fundamental price if $\alpha = 0$. This can be interpreted as a *positive liquidity effect* of increasing the trading frequency.

However, the situation is quite different if $\alpha \ne 0$. Assume, e.g., that $\alpha > 0$. Then, the second line of (3.4) implies that $\hat{\lambda}^a$ increases by, at least, $\alpha \Delta t$ at each step of the (backward) recursion. Recall that the number of steps is $N = T / \Delta t$, hence, $\hat{\lambda}_0^a \ge \hat{\lambda}_N^a + \alpha T$. If $|\hat{\lambda}_N^a|$ is small (which is typically the case if N is large), then we may obtain $\hat{\lambda}_{n+1}^a \ge 0$, at some time n, which violates the third inequality in (3.3), or, equivalently, implies that the objective in the first line of (3.4) is strictly negative for all p. The latter implies that it is suboptimal for the agents with positive inventory to post limit orders, and the proposed method fails to produce a nondegenerate LTC equilibrium in the interval $[n, N]$.
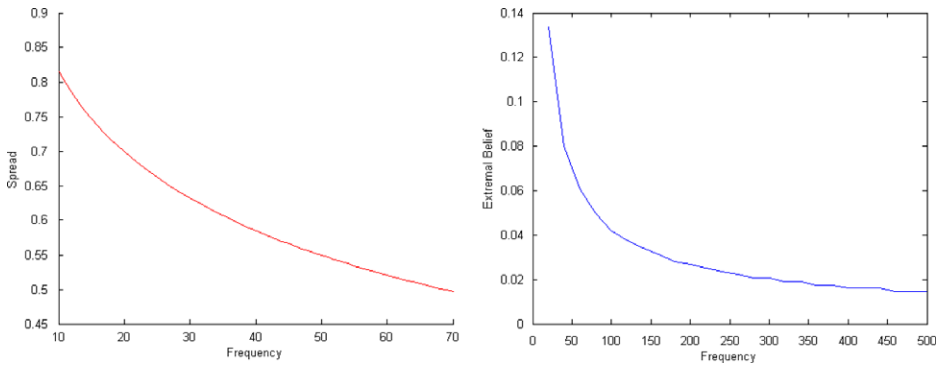
FIGURE 3.2   The horizontal axis represents trading frequency, measured in the number of steps N. (Left) Time zero bid–ask spread in the zero-drift case ($\alpha = 0$). (Right) The maximum value of drift α for which a nondegenerate equilibrium exists on the entire time interval. Parameters: $\sigma = 1$, $T = 1$ [Color figure can be viewed at wileyonlinelibrary.com]

Figure 3.1 shows that this does, indeed, occur. Figures 3.1 and 3.2 also show that for a given (finite) frequency N, if $|\alpha|$ is small enough, a nondegenerate equilibrium may still be constructed. Nevertheless, for any $|\alpha| \neq 0$, however small it is, there exists a large enough N, s.t. the nondegenerate LTC equilibrium fails to exist (at least within the class defined by the proposed method). This is illustrated in Figure 3.2.

It is important to provide an economic interpretation of why such degeneracy occurs. A careful examination of Figure 3.1 reveals that around the time when $\hat{\lambda}^a$ becomes nonnegative, the ask price $\hat{p}^a$ explodes. This means that the agents who want to sell the asset are only willing to sell it at a very high price. Notice also that this price is several magnitudes larger than the expected change in the fundamental price (represented by the black dashed line in the left-hand side of Figure 3.1). Hence, such a behavior cannot be justified by the behavior of the fundamental. Indeed, this is precisely what is called an *endogenous liquidity crisis*. So, what causes such a liquidity crisis? Recall that there are two potential reasons for the market to degenerate: Agents may choose to submit market orders (if $\hat{p}^b_n > \hat{\lambda}^a_n$ or $\hat{p}^a_n < \hat{\lambda}^b_n$), or they may choose to wait and do nothing (if $\hat{\lambda}^a_{n+1} \geq 0$ or $\hat{\lambda}^b_{n+1} \leq 0$). The right-hand side of Figure 3.1 shows that the degeneracy is caused by the second scenario. This means that the naive explanation of an endogenous liquidity crisis, based on the claim that in a bullish market, those who need to buy the asset will submit market orders wiping out liquidity on the sell side of the book, is wrong. Instead, if the agents on the sell side of the book have the same beliefs, they will increase the ask price so that it is no longer profitable for the agents who want to buy the asset to submit market buy orders. In fact, the ask price may increase disproportionally to the expected change in the fundamental price (i.e., the signal), and this is what causes an endogenous liquidity crisis. The size of the resulting change in the bid or ask price depends not only on the signal, but also on the trading frequency, which demonstrates the *negative liquidity effect* of increasing the trading frequency: It fragilizes the market with respect to deviations of the agents from market neutrality. The latter, in turn, is explained by the fact that higher trading frequency exacerbates the *adverse selection* effect. To see this, consider, e.g., an agent who is trying to sell one share of the asset. Increasing the trading frequency increases the expected execution value of this agent, bringing it closer to the fundamental price: This corresponds to $\hat{\lambda}^a$ approaching zero (from below). Assume that the agent posts a limit sell order at a price level p. If this order is executed in the next period, then the agent receives p, but for this to happen, the fundamental price value at the next time step, p^0_{n+1}, has to be above p. On the other hand, the expected execution price

of the agent at the next time step is $p_{n+1}^0 + \hat{\lambda}_{n+1}^a$. Thus, the expected relative profit, given the execution of her limit order, is $\mathbb{E}_n(p - p_{n+1}^0 - \hat{\lambda}_{n+1}^a \mid p_{n+1}^0 > p)$. The latter expression cannot be positive, unless $\hat{\lambda}_{n+1}^a < 0$ and $|\hat{\lambda}_{n+1}^a|$ is sufficiently large. Therefore, if $|\hat{\lambda}_{n+1}^a|$ is small relative to $\mathbb{E}_n(p_{n+1}^0 - p \mid p_{n+1}^0 > p)$, the agent is reluctant to post a limit order at the price level p. Hence, p needs to be sufficiently large, to ensure that $\mathbb{E}_n(p_{n+1}^0 - p \mid p_{n+1}^0 > p)$ is smaller than $|\hat{\lambda}_{n+1}^a|$ (in the Gaussian model of this section, the latter expectation vanishes as $p \to \infty$)—and the extent to which p needs to increase determines the effect of adverse selection. It turns out that, if the agents are market neutral (i.e., $\alpha = 0$), as the frequency N increases, the quantity $\mathbb{E}_n(p_{n+1}^0 - p \mid p_{n+1}^0 > p)$, for any fixed p, converges to zero at the same rate as $|\hat{\lambda}_{n+1}^a|$; hence, the above adverse selection effect does not get amplified. On the contrary, if the agents are not market neutral, $\hat{\lambda}_{n+1}^a$ reaches zero at some high enough (but finite) frequency, while $\mathbb{E}_n(p_{n+1}^0 - p \mid p_{n+1}^0 > p)$ remains strictly positive, for any finite p, which produces an "infinite" adverse selection effect and causes the market to degenerate. Of course, so far, these conclusions are based on a very specific example and on a particular method of constructing an equilibrium. The next section shows that they remain valid in any model (with, possibly, heterogeneous beliefs) in which the fundamental price is given by an Itô process.

It is worth mentioning that a similar adverse selection effect arises in Goettler et al. (2005), and it is referred to as the "winner's curse" in Foucault (1999). However, the latter papers do not investigate the nature of this phenomenon and focus on other questions instead. In the literature on double auctions (cf. Du & Zhu, 2014; Vayanos, 1999), a similar effect arises when the auction participants choose to decrease their trading activity in a given auction, because they expect many more opportunities to trade in the future. The latter is similar to the agents choosing to forgo limit orders and wait, in the present example.

4 | MAIN RESULTS

In this section, we generalize the previous conclusions so that they hold in a general model and for any equilibrium. As before, we begin with the "limiting" continuous-time model. Consider a terminal time horizon $T > 0$ and a complete stochastic basis $(\Omega, \tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_t)_{t \in [0,T]}, \mathbb{P})$, with a Brownian motion W on it.[9] We define the adapted process $\tilde{p}^0$ as a continuous modification of

$$\tilde{p}_t^0 = p_0^0 + \int_0^t \sigma_s dW_s, \qquad p_0^0 \in \mathbb{R}, \tag{4.1}$$

where σ is a progressively measurable locally square-integrable process.

Assumption 4.1. *There exists a constant $C > 1$, such that, $1/C \leq \sigma_t \leq C$, for all $t \in [0,T]$, $\mathbb{P}$-a.s.*

Consider a Borel set of beliefs $\mathbb{A}$ and the associated family of measures $\{\mathbb{P}^\alpha\}_{\alpha \in \mathbb{A}}$ on $(\Omega, \tilde{\mathcal{F}}_T)$, absolutely continuous with respect to $\mathbb{P}$. Then, for any $\alpha \in \mathbb{A}$, we have

$$\tilde{p}_t^0 = p_0^0 + A_t^\alpha + \int_0^t \sigma_s dW_s^\alpha, \quad p_0^0 \in \mathbb{R}, \quad \mathbb{P}^\alpha\text{-a.s.}, \ \forall t \in [0,T],$$

where W^α is a Brownian motion under $\mathbb{P}^\alpha$, and A^α is a process of finite variation. We assume that A^α is absolutely continuous; i.e., for any $\alpha \in \mathbb{A}$, there exists a locally integrable process μ^α, such that

$$A_t^\alpha = \int_0^t \mu_s^\alpha ds, \quad \mathbb{P}^\alpha\text{-a.s.}, \ \forall t \in [0,T].$$

Assumption 4.2. *For any $\alpha \in \mathbb{A}$, the process μ^α is $\mathbb{P}$-a.s. right-continuous, and there exists a constant $C > 0$, such that $|\mu^\alpha_t| \leq C$, for all $t \in [0, T]$, $\mathbb{P}$-a.s.*

Thus, we can rewrite the dynamics of $\tilde{p}^0$, under each $\mathbb{P}^\alpha$, as follows: $\mathbb{P}^\alpha$-a.s., the following holds for all $t \in [0, T]$

$$\tilde{p}^0_t = p^0_0 + \int_0^t \mu^\alpha_s \, ds + \int_0^t \sigma_s \, dW^\alpha_s, \quad p^0_0 \in \mathbb{R}. \tag{4.2}$$

In addition, we modify the above stochastic integral on a set of $\mathbb{P}^\alpha$-measure zero, so that (4.2) holds for *all* (t, ω). In what follows, we often need to analyze the future dynamics of $\tilde{p}^0$ under $\mathbb{P}^\alpha$, conditional on $\tilde{\mathcal{F}}_t$, for various (t, α) simultaneously. This is why we need the following joint regularity assumption.

Assumption 4.3. *There exists a modification of regular conditional probabilities*

$$\left\{ \tilde{\mathbb{P}}^\alpha_t = \mathbb{P}^\alpha \left(\cdot \mid \tilde{\mathcal{F}}_t \right) \right\}_{t \in [0,T], \, \alpha \in \mathbb{A}},$$

such that it satisfies the tower property with respect to $\mathbb{P}$ (as described in Section 2.1).

Assumption 4.3 is satisfied, e.g., if $\mathbb{P}^\alpha \sim \mathbb{P}$, for all $\alpha \in \mathbb{A}$, or if the set $\mathbb{A}$ is countable. Throughout the rest of the paper, $\tilde{\mathbb{P}}^\alpha_t$ refers to a member of the family appearing in Assumption 4.3. All conditional expectations $\tilde{\mathbb{E}}^\alpha_t$ are taken under such $\tilde{\mathbb{P}}^\alpha_t$.

The main results of this section require additional continuity assumptions on σ and μ^α. The following assumption can be viewed as a stronger version of $\mathbb{L}^2$-continuity of σ.

Assumption 4.4. *There exists a function $\varepsilon(\cdot) \geq 0$, such that $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, $\mathbb{P}$-a.s.,*

$$\tilde{\mathbb{P}}^\alpha_t \left(\mathbb{E}^\alpha \left(\left(\sigma_{s \vee \tau} - \sigma_\tau \right)^2 \mid \mathcal{F}_\tau \right) \leq \varepsilon(\Delta t) \right) = 1$$

holds for all $t \in [0, T - \Delta t]$, all $s \in [t, t + \Delta t]$, all stopping times $t \leq \tau \leq s$, and all $\alpha \in \mathbb{A}$.

The above assumption is satisfied, e.g., if σ is an Itô process with bounded drift and diffusion coefficients. Next, we state a continuity assumption on the drift, which can be interpreted as a uniform right-continuity in probability of the martingale $\tilde{\mathbb{E}}^\alpha_t \mu^\alpha_s$.

Assumption 4.5. *For any $\alpha \in \mathbb{A}$ and any $t \in [0, T)$, there exists a deterministic function $\varepsilon(\cdot) \geq 0$, such that $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, $\mathbb{P}^\alpha$-a.s.,*

$$\tilde{\mathbb{P}}^\alpha_{t'} \left(\left| \int_t^T \left(\tilde{\mathbb{E}}^\alpha_{t''} \mu^\alpha_s - \tilde{\mathbb{E}}^\alpha_{t'} \mu^\alpha_s \right) ds \right| \geq \varepsilon(\Delta t) \right) \leq \varepsilon(\Delta t)$$

holds for all $t \leq t' \leq t'' \leq t + \Delta t \leq T$.

Notice that Assumptions 4.3, 4.4, and 4.5 are not quite standard. Therefore, below, we describe a more specific (although still rather general) diffusion-based framework, in which Assumptions 4.1–4.5 reduce to standard regularity conditions on the diffusion coefficients, and are easily verified. To this end, consider a model in which $\mu^\alpha_t = \bar{\mu}^\alpha(t, Y_t)$, $\sigma_t = \bar{\sigma}(t, Y_t)$, and under $\mathbb{P}$, the process Y is a diffusion taking values in $\mathbb{R}^d$

$$dY_t = \Gamma(t, Y_t) dt + \Sigma(t, Y_t) d\bar{B}_t,$$

where $\Gamma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\Sigma = (\Sigma^{i,j})$ is a mapping on $[0, T] \times \mathbb{R}^d$ with values in the space of $d \times m$ matrices, and $\bar{B}$ is m-dimensional Brownian motion under $\mathbb{P}$ (on the original stochastic basis). We

assume that Γ and Σ possess enough regularity to conclude that Y is a strongly Markov process. Notice that Assumptions 4.1 and 4.2 reduce to the upper and lower bounds on the functions $\bar{\mu}^\alpha$ and $\bar{\sigma}$. Assumption 4.3 is satisfied if we assume that $\mathbb{P}^\alpha \sim \mathbb{P}$, for all $\alpha \in \mathbb{A}$. Let us further assume that the Radon–Nikodym derivative of each measure is in Girsanov form:

$$\frac{d\mathbb{P}^\alpha}{d\mathbb{P}} = \exp\left(-\frac{1}{2}\int_0^t \|\gamma^\alpha(s, Y_s)\|^2 ds + \int_0^t \gamma^\alpha(s, Y_s) d\bar{B}_s\right),$$

with an $\mathbb{R}^d$-valued function γ^α, for each $\alpha \in \mathbb{A}$. Let us assume that all entries of Γ, γ^α, and Σ are absolutely bounded by a constant (uniformly over $\alpha \in \mathbb{A}$). Assuming, in addition, that $\bar{\sigma}$ is globally Lipschitz, we easily verify Assumption 4.4. In order to verify Assumption 4.5, we assume that the quadratic form generated by $A(t, y) := \Sigma(t, y)\Sigma^T(t, y)$ is bounded away from zero, uniformly over all (t, y), and that the entries of Γ, γ^α, and Σ are continuously differentiable with absolutely bounded derivatives (uniformly over $\alpha \in \mathbb{A}$). Then, the Feynman–Kac formula implies that, for any $t \leq s$,

$$\tilde{\mathbb{E}}_t^\alpha \mu_s^\alpha = u^{s,\alpha}(t, Y_t),$$

where $u^{s,\alpha}$ is the unique solution to the associated partial differential equation (PDE)

$$\partial_t u^{s,\alpha} + \sum_{i=1}^d \Gamma^{\alpha,i} \partial_{y_i} u^{s,\alpha} + \frac{1}{2}\sum_{i,j=1}^d A^{i,j}\partial_{y_i y_j}^2 u^{s,\alpha} = 0, \quad (t, y) \in (0, s) \times \mathbb{R}^d, \quad u^{s,\alpha}(s, y) = \bar{\mu}^\alpha(s, y),$$

and $\Gamma^\alpha = \Gamma + \Sigma\gamma^\alpha$. Assume that, for each $s \in [0, T]$, the function $\bar{\mu}^\alpha(s, \cdot)$ is continuously differentiable with absolutely bounded derivatives, uniformly over all (s, α). Then, the standard Gaussian estimates for derivatives of the fundamental solution to the above PDE (cf. theorem 9.4.2 in Friedman, 1964) imply that every $\partial_{y_i} u^{s,\alpha}$ is absolutely bounded, uniformly over all (s, α). Then, Itô's formula and Itô's isometry yield, for all $t' \leq t''$ and $s \geq t'$:

$$\tilde{\mathbb{E}}_{t'}^\alpha \left(\tilde{\mathbb{E}}_{t''}^\alpha \mu_s^\alpha - \tilde{\mathbb{E}}_{t'}^\alpha \mu_s^\alpha\right)^2 = \sum_{j=1}^m \int_{t'}^{t'' \wedge s} \tilde{\mathbb{E}}_{t'}^\alpha \left(\sum_{i=1}^d \partial_{y_i} u^{s,\alpha}(v, Y_v)\Sigma^{i,j}(v, Y_v)\right)^2 dv \leq C_1(t'' \wedge s - t'),$$

with some constant $C_1 > 0$. The above estimate and Jensen's inequality imply the statement of Assumption 4.5 and complete the description of the diffusion-based setting.

As in Section 3, we also consider a progressively measurable random field $\tilde{D}$, s.t. $\mathbb{P}$-a.s. the function $\tilde{D}_t(\cdot) - \tilde{D}_s(\cdot)$ is strictly decreasing and vanishing at zero, for any $0 \leq s < t \leq T$. We assume that the demand curve, $\tilde{D}_t(\cdot) - \tilde{D}_s(\cdot)$, cannot be "too flat."

Assumption 4.6. *There exists $\varepsilon > 0$, s.t., for any $0 \leq t - \varepsilon \leq s < t \leq T$, there exists a $\tilde{\mathcal{F}}_s \otimes \mathcal{B}(\mathbb{R})$-measurable random function $\kappa_s(\cdot)$, s.t., $\mathbb{P}$-a.s., $\kappa_s(\cdot)$ is strictly decreasing and $|\tilde{D}_t(p) - \tilde{D}_s(p)| \geq |\kappa_s(p)|$, for all $p \in \mathbb{R}$.*

Finally, we introduce the empirical distribution process $(\tilde{\mu}_t)$, with values in the space of finite sigma-additive measures on $\mathbb{S}$. The next assumption states that every $\tilde{\mu}_t$ is dominated by a deterministic measure.

Assumption 4.7. *For any $t \in [0, T]$, there exists a finite sigma-additive measure μ_t^0 on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$, s.t., $\mathbb{P}$-a.s., $\tilde{\mu}_t$ is absolutely continuous with respect to μ_t^0.*

We partition the time interval $[0, T]$ into N subintervals of size $\Delta t = T/N$. A discrete-time model is obtained by discretizing the continuous-time one

$$\mathcal{F}_n = \tilde{\mathcal{F}}_{n\Delta t}, \quad p_n^0 = \tilde{p}_{n\Delta t}^0, \quad D_n(p) = (\tilde{D}_{n\Delta t} - \tilde{D}_{(n-1)\Delta t})(p - p_n^0), \quad \mu_n = \tilde{\mu}_{n\Delta t}.$$

Before we present the main results, let us comment on the above assumptions. These assumptions are important from a technical point of view; however, some of them have an economic interpretation that may provide (partial) intuitive explanations of the results that follow. In particular, Assumption 4.1 ensures that the fundamental price remains "noisy," which implies that an agent can execute a limit order very quickly by posting it close to the present value of p^0, if there are no other orders posted there. In combination with Assumption 4.6, the latter implies that when the frequency, N, is high, an agent has a lot of opportunities to execute her limit order at a price close to the fundamental price (at least if no other orders are posted too close to the fundamental price). Intuitively, this means that the agent's execution value should improve as the frequency increases. Assumption 4.5 ensures that if an agent has a signal about the direction of the fundamental price, this signal is persistent—i.e., it is continuous in the appropriate sense. When the trading frequency N is large, such persistency means that an agent has a large number of opportunities to exploit the signal, implying that she is in no rush to have her order executed immediately. The main results of this work, presented below, along with their proofs, confirm that these heuristic conclusions are, indeed, correct.

As mentioned in the preceding sections, our main goal is to analyze the liquidity effects of increasing the trading frequency. Therefore, we fix a limiting continuous-time model, and consider a sequence of discrete-time models, obtained from the limiting one as described above, for $N \to \infty$. This can be interpreted as observing the same population of agents, each of whom has a fixed continuous-time model for future demand, in various exchanges that allow for different trading frequencies. We begin with the following theorem, which shows that if every market model in a given sequence admits a nondegenerate equilibrium, then the terminal bid and ask prices converge to the fundamental price, as the trading frequency goes to infinity.

Theorem 4.8. *Let Assumptions 4.1–4.4, 4.6, and 4.7 hold. Consider a family of uniform partitions of a given time interval $[0, T]$, with diameters $\{\Delta t = T/N > 0\}$ and with the associated family of discrete-time models, and denote the associated fundamental price process by $p^{0,\Delta t}$. Assume that every such model admits a nondegenerate LTC equilibrium, and denote the associated bid and ask prices by $p^{b,\Delta t}$ and $p^{a,\Delta t}$, respectively. Then, there exists a deterministic function $\varepsilon(\cdot)$, s.t. $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, for all small enough $\Delta t > 0$, the following holds $\mathbb{P}$-a.s.:*

$$\left| p_N^{a,\Delta t} - p_N^{0,\Delta t} \right| + \left| p_N^{b,\Delta t} - p_N^{0,\Delta t} \right| \le \varepsilon(\Delta t).$$

The above theorem has a useful corollary, which can be interpreted as follows: *if the market does not degenerate as the frequency increases, then such an increase improves market efficiency.* Here, we understand the "improving efficiency" in the sense that the expected execution price (i.e., the price per share that an agent expects to receive or pay by the end of the game) of every agent converges to the fundamental price.

Corollary 4.9. *Under the assumptions of Theorem 4.8, denote the support of every equilibrium by $\tilde{\mathbb{A}}^{\Delta t}$ and the associated expected execution prices by $\lambda^{a,\Delta t}$ and $\lambda^{b,\Delta t}$. Then, there exists a deterministic function $\varepsilon(\cdot)$, such that $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, $\mathbb{P}$-a.s.,*

$$\sup_{n=0,\dots,N,\, \alpha \in \tilde{\mathbb{A}}^{\Delta t}} \left(\left| \lambda_n^{a,\Delta t}(\alpha) - p_n^{0,\Delta t} \right| + \left| \lambda_n^{b,\Delta t}(\alpha) - p_n^{0,\Delta t} \right| \right) \le \varepsilon(\Delta t)$$

for all small enough $\Delta t > 0$.

Proof. Denote $\mathbb{E}_n^\alpha = \tilde{\mathbb{E}}_{n\Delta t}^\alpha$. It follows from Corollary A.2, in Appendix A, and the definition of LTC equilibrium that $\lambda_N^{a,\Delta t}(\alpha) = p_N^{b,\Delta t}$ and $\lambda_N^{b,\Delta t}(\alpha) = p_N^{a,\Delta t}$. It also follows from Corollary A.2 (or, more generally, from the definition of a value function) that $\lambda^{a,\Delta t}(\alpha)$ is a supermartingale, and $\lambda^{b,\Delta t}(\alpha)$ is a submartingale, under $\mathbb{P}^\alpha$. Thus, we have: $\lambda_n^{a,\Delta t}(\alpha) \geq \mathbb{E}_n^\alpha p_N^{b,\Delta t}$ and $\lambda_n^{b,\Delta t}(\alpha) \leq \mathbb{E}_n^\alpha p_N^{a,\Delta t}$. On the other hand, notice that we must have $\lambda_n^{a,\Delta t}(\alpha) \leq \mathbb{E}_n^\alpha p_N^{a,\Delta t}$ and $\lambda_n^{b,\Delta t}(\alpha) \geq \mathbb{E}_n^\alpha p_N^{b,\Delta t}$. Assume, e.g., that $\lambda_n^{a,\Delta t}(\alpha) > \mathbb{E}_n^\alpha p_N^{a,\Delta t}$ on the event Ω' of positive $\mathbb{P}^\alpha$-probability. Consider an agent at state $(0, \alpha)$, who follows the optimal strategy of an agent at state $(1, \alpha)$, starting from time n and onward, on the event Ω' (otherwise, she does not do anything). It is easy to see that the objective value of this strategy is

$$\mathbb{E}^\alpha \left(\mathbf{1}_{\Omega'} \left(\lambda_n^{a,\Delta t}(\alpha) - \mathbb{E}_n^\alpha p_N^{a,\Delta t} \right) \right) > 0,$$

which contradicts Corollary A.2. The second inequality is shown similarly. Thus, we conclude that for any $n = 0, \dots, N-1$, both $\lambda_n^{a,\Delta}(\alpha)$ and $\lambda_n^{b,\Delta}(\alpha)$ belong to the interval

$$\left[\mathbb{E}_n^\alpha p_N^{b,\Delta t}, \mathbb{E}_n^\alpha p_N^{a,\Delta t} \right],$$

which, in turn, converges to zero, as $\Delta t \to 0$, due to the deterministic bounds obtained in the proof of Proposition 4.8. □

The results of Theorem 4.8 and Corollary 4.9 can be viewed as a specific case of a more general observation: Markets become more efficient as the frictions become smaller. In the present setting, the limited trading frequency is viewed as friction, and the market efficiency is measured by the difference between the bid and ask prices, or between the expected execution prices. Many more instances of analogous results can be found in the literature, depending on the choice of a friction type. For example, the markets become efficient in Glosten and Milgrom (1985) and Kyle (1985) as the number of insiders vanishes. Similarly, the markets become efficient in Du and Zhu (2014) as the trading frequency increases and the size of private signals vanishes. It is also mentioned in Brunnermeier and Pedersen (2005) that the market would become efficient if there was no restriction on the size of agents' inventories therein.

The above results demonstrate the positive role of high trading frequency. However, they are based on the assumption that the market does not degenerate as frequency increases. In the context of Section 3, we saw that the markets do not degenerate only if the agents are market-neutral (i.e., $\alpha = 0$). If this condition is violated and the frequency N is sufficiently high, the market does not admit any nondegenerate equilibrium (i.e., there exists no safe regime, in which the liquidity crisis would never occur). It turns out that this conclusion still holds in the general setting considered herein.

Theorem 4.10. *Let Assumptions 4.1–4.7 hold. Consider a family of uniform partitions of a given time interval $[0, T]$, with diameters $\{\Delta t = T/N > 0\}$, containing arbitrarily small Δt, and with the associated family of discrete-time models. Assume that every such model admits a nondegenerate LTC equilibrium, with the same support $\tilde{\mathbb{A}}$. Then, for all $\alpha \in \tilde{\mathbb{A}}$, we have: $\tilde{p}^0$ is a martingale under $\mathbb{P}^\alpha$.*

The above theorem shows that the market degenerates even if the signal μ^α is very small (but nonzero), provided the trading frequency N is large enough. Therefore, as discussed at the end of Section 3, such degeneracy cannot be attributed to any fundamental reasons, and we refer to it as the *endogenous liquidity crisis*. Let us provide an intuitive (heuristic) argument for why the statement of Theorem 4.10 holds. Assume, first, that all long agents (i.e., those having positive inventory) are bullish about the asset (i.e., have a positive drift μ^α). Then, similar to Section 3, the higher trading frequency amplifies the *adverse selection effect*, forcing the long agents to withdraw liquidity from the market

(i.e., they prefer to do nothing and wait for a higher fundamental price level). Note that in the present setting, the agents may have different beliefs, the LOB may have a complicated shape and dynamics, and the expected execution prices are no longer deterministic. All this makes it difficult to provide a simple description of how the high frequency amplifies the adverse selection. Nevertheless, the general analysis of this case is still based on the idea discussed at the end of Section 3: It has to do with how fast $\tilde{\mathbb{E}}^\alpha_{n\Delta t}(p^0_{n+1} - p \mid p^0_{n+1} > p)$ vanishes (as the frequency increases), relative to the rate at which the expected execution prices approach the fundamental price. Thus, to avoid market degeneracy, there must be a nonzero mass of long agents who are market neutral or bearish. As the trading frequency grows, these agents will post their limit orders at lower levels. Next, assume that there exists a bullish agent (long, short, or with zero inventory). Then, at a sufficiently high trading frequency, the agent's expected value of a long position in a single share of the asset will exceed the ask prices posted by the market-neutral and bearish long agents. In this case, the bullish agent prefers to buy more shares at the posted ask price, in order to sell them later. As the agents are small and their objectives are linear, the bullish agent can scale up her strategy to generate infinite expected profits. This contradicts the definition of optimality and implies that an equilibrium fails to exist. Thus, all agents have to be either market neutral or bearish. Applying a symmetric argument, we conclude that all agents must be market neutral.[10] A rigorous formulation of the above arguments, which constitutes the proof of Theorem 4.10, is given in Section 7.

It is worth mentioning that the possible degeneracy of the LOB is also documented in Glosten and Milgrom (1985) and is referred to as a "market shut down." The setting used in the latter paper is very different: It analyzes a quote-driven exchange (i.e., the one with a designated market maker) and assumes the existence of insiders with superior information. Nevertheless, it is possible to draw a parallel with the LOB degeneracy in the present setting. Namely, the degeneracy in Glosten and Milgrom (1985) occurs when the number of insiders increases, which implies that the signal, generated by the insiders' trading, becomes sufficiently large. The latter is similar to the deviation from martingality of the fundamental price in the present setting. However, an increase in the number of insiders in Glosten and Milgrom (1985) also implies an increase in frictions (since the insiders can be interpreted as friction in Glosten & Milgrom, 1985). Theorem 4.10, on the other hand, states that a market degeneracy will occur when the frictions are sufficiently small. Perhaps, this dual role of the number of insiders did not allow for a detailed analysis of market shut downs in Glosten and Milgrom (1985). Many other models of market microstructure (cf. Du & Zhu, 2014; Foucault, 1999; Goettler et al., 2005; Parlour, 1998; Rosu, 2009) are not well suited for the analysis of market degeneracy, either because the agents in these models pursue "one-shot" strategies (i.e., they cannot choose to wait and post a limit order later) or because the fundamental price (or its analog) is restricted to be a martingale.

5 | CONDITIONAL TAILS OF THE MARGINAL DISTRIBUTIONS OF ITÔ PROCESSES

As follows from the discussion in the preceding sections, in order to prove the main results of the paper, we need to investigate the properties of marginal distributions of the fundamental price $\tilde{p}^0$ (more precisely, the distributions of its increments). In order to prove Theorem 4.8, we need to show that the difference between the fundamental price and the bid or ask prices converges to zero, as the frequency N increases to infinity. It turns out that, for this purpose, it suffices to show that the distribution of a normalized increment of $\tilde{p}^0$ converges to the standard normal distribution. The following lemma summarizes these results. It is rather simple, but technical; hence, its proof is postponed to Appendix B. In order to formulate the result (and to facilitate the derivations in subsequent sections),

we introduce additional notation. For notational convenience, we drop the superscript Δt for some variables (we only emphasize this dependence when it is important). For any market model on the time interval $[0, T]$, associated with a uniform partition with diameter $\Delta t = T/N > 0$, and having a fundamental price process p^0, we define

$$\xi_n = p_n^0 - p_{n-1}^0 = \tilde{p}_{t_n}^0 - \tilde{p}_{t_{n-1}}^0, \quad \mathbb{E}_n^\alpha = \tilde{\mathbb{E}}_{t_n}^\alpha, \quad \mathbb{P}_n^\alpha = \tilde{\mathbb{P}}_{t_n}^\alpha, \quad t_n = n\Delta t, \quad n = 1, \dots, NT/\Delta t. \quad (5.1)$$

We denote by η_0 a standard normal random variable (on a, possibly, extended probability space), which is independent of $\mathcal{F}_N$ under every $\mathbb{P}^\alpha$.

Lemma 5.1. *Let Assumptions 4.1–4.4 hold. Then, there exists a function $\varepsilon(\cdot) \geq 0$, s.t. $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and the following holds $\mathbb{P}$-a.s., for all $p \in \mathbb{R}$, all $\alpha \in \mathbb{A}$, and all $n = 1, \dots, N$,*

(i) $(|p| \vee 1)|\mathbb{P}_{n-1}^\alpha(\frac{\xi_n}{\sqrt{\Delta t}} > p) - \mathbb{P}_{n-1}^\alpha(\sigma_{t_{n-1}}\eta_0 > p)| \leq \varepsilon(\Delta t)$,

(ii) $|\mathbb{E}_{n-1}^\alpha(\frac{\xi_n}{\sqrt{\Delta t}}\mathbf{1}_{\{\xi_n/\sqrt{\Delta t}>p\}}) - \mathbb{E}_{n-1}^\alpha(\sigma_{t_{n-1}}\eta_0\mathbf{1}_{\{\sigma_{t_{n-1}}\eta_0>p\}})| \leq \varepsilon(\Delta t)$.

In addition, the above estimates hold if we replace (ξ_n, η_0, p) by $(-\xi_n, -\eta_0, -p)$.

In order to prove Theorem 4.10, we need to compare the rates at which the conditional expectations $\mathbb{E}_n^\alpha(p_{n+1}^0 - p \mid p_{n+1}^0 > p)$ vanish (as the frequency N goes to infinity) to the rate at which the expected execution prices converge to the fundamental price. This requires a more delicate analysis—in particular, the mere proximity of the distribution of a (normalized) fundamental price increment to the Gaussian distribution is no longer sufficient. In fact, what we need is a precise uniform estimate of the conditional tail of the distribution of a fundamental price increment. The desired property is formulated in the following lemma, which, we believe, is valuable in its own right. This result enables us to estimate the tails of the conditional marginal distribution of an Itô process X uniformly by an exponential. To the best of our knowledge, this result is new. The main difficulties in establishing the desired estimates are: (a) the fact that we estimate the *conditional*, as opposed to the regular, tail, and (b) the fact that the estimates need to be uniform over the values of the argument. Note that even in the case of a diffusion process X, the classical Gaussian-type bounds for the tails of the marginal distributions of X are not sufficient to establish the desired estimates. The reason is that, in general, the Gaussian estimates of the regular tails from above and from below have different orders of decay, for the large values of the argument, which makes them useless for estimating the conditional tail (which is a ratio of two regular tails).

Lemma 5.2. *Consider the following continuous semimartingale on a stochastic basis $(\hat{\Omega}, (\hat{\mathcal{F}}_t)_{t \in [0,1]}, \hat{\mathbb{P}})$:*

$$X_t = \int_0^t \hat{\mu}_u du + \int_0^t \hat{\sigma}_u dB_u, \quad t \in [0, 1],$$

where B is a Brownian motion (with respect to the given stochastic basis), $\hat{\mu}$ and $\hat{\sigma}$ are progressively measurable processes, such that the above integrals are well defined. Assume that for any stopping time τ with values in $[0, 1]$, $c \leq |\hat{\sigma}_\tau| \leq C$ holds a.s. with some constants $c, C > 0$. Then, there exists $\varepsilon > 0$, depending only on (c, C), s.t., if

$$\hat{\mu}_\tau^2 \leq \varepsilon, \quad \hat{\mathbb{E}}\left((\hat{\sigma}_{s \vee \tau} - \hat{\sigma}_\tau)^2 \mid \hat{\mathcal{F}}_\tau\right) \leq \varepsilon \quad a.s.$$

for all $s \in [0, 1]$ *and all stopping time τ, with values in* $[0, 1]$*, then for any $c_1 > 0$, there exists $C_1 > 0$, depending only on* (c, C, ε, c_1)*, s.t. the following holds:*

$$\hat{\mathbb{P}}(X_1 > x + z \mid X_1 > x) \le C_1 e^{-c_1 z}, \quad \forall x, z \ge 0.$$

Proof. In the course of this proof, we will use the shorthand notation, $\hat{\mathbb{E}}_\tau$ and $\hat{\mathbb{P}}_\tau$, to denote the conditional expectation and the conditional probability with respect to $\hat{\mathcal{F}}_\tau$. We also denote

$$A_t = \int_0^t \hat{\mu}_u du, \qquad G_t = \int_0^t \hat{\sigma}_u d B_u.$$

For any $x \ge 0$, let us introduce $\tau_x = 1 \wedge \inf\{t \in [0, 1] : X_t = x\}$. Then,

$$\hat{\mathbb{P}}(X_1 > x + z) \le \hat{\mathbb{P}}\left(\sup_{t \in [0,1]} X_t > x + z\right) = \hat{\mathbb{E}}\left(\mathbf{1}_{\{\tau_x < 1\}} \hat{\mathbb{P}}_{\tau_x}\left(\sup_{s \in [\tau_x, 1]} (X_s - x) > z\right)\right).$$

Notice that, on $\{\tau_x \le s\}$, we have $X_s - x = A_{s \vee \tau_x} - A_{\tau_x} + G_{s \vee \tau_x} - G_{\tau_x}$. In addition, the process $(Y)_{s \in [0,1]}$, with $Y_s = A_{s \vee \tau_x} - A_{\tau_x}$, is adapted to the filtration $(\hat{\mathcal{F}}_{\tau_x \vee s})$, while the process $(Z)_{s \in [0,1]}$, with $Z_s = G_{s \vee \tau_x} - G_{\tau_x}$, is a martingale with respect to it. Next, on $\{\tau_x < 1\}$, we have

$$\hat{\mathbb{P}}_{\tau_x}\left(\sup_{s \in [\tau_x, 1]} (X_s - x) > z\right) = \hat{\mathbb{P}}_{\tau_x}\left(\sup_{s \in [0,1]} (Y_s + Z_s) > z\right)$$

$$\le \hat{\mathbb{P}}_{\tau_x}\left(\sup_{s \in [0,1]} \exp\left(c_1 Z_s - \frac{1}{2} c_1^2 \langle Z \rangle_s\right) > \exp\left(c_1 z - c_1 \sqrt{\varepsilon} - \frac{1}{2} c_1^2 C^2\right)\right),$$

where we used the fact that $\langle Z \rangle_s \le \langle X \rangle_1 \le C^2$, for all $s \in [0, 1]$. Using the Novikov's condition, it is easy to check that

$$M_s = \exp\left(c_1 Z_s - \frac{1}{2} c_1^2 \langle Z \rangle_s\right), \qquad s \in [0, 1]$$

is a true martingale, and hence, we can apply the Doob's martingale inequality to obtain, on $\{\tau_x < 1\}$:

$$\hat{\mathbb{P}}_{\tau_x}\left(\sup_{s \in [0,1]} \exp\left(c_1 Z_s - \frac{1}{2} c_1^2 \langle Z \rangle_s\right) > \exp\left(c_1 z - c_1 \sqrt{\varepsilon} - \frac{1}{2} c_1^2 C^2\right)\right)$$

$$\le \exp\left(-c_1 z + c_1 \sqrt{\varepsilon} + \frac{1}{2} c_1^2 C^2\right).$$

Collecting the above inequalities, we obtain

$$\hat{\mathbb{P}}(X_1 > x + z) \le \hat{\mathbb{P}}\left(\sup_{t \in [0,1]} X_t > x + z\right) \le C_2(\varepsilon) e^{-c_1 z} \hat{\mathbb{P}}(\tau_x < 1) = C_2(\varepsilon) e^{-c_1 z} \hat{\mathbb{P}}\left(\sup_{t \in [0,1]} X_t > x\right). \quad (5.2)$$

The next step is to estimate the distribution tails of a running maximum via the tails of the distribution of X_1. To do this, we proceed as before:

$$\hat{\mathbb{P}}(X_1 > x) = \hat{\mathbb{E}}\left(\mathbf{1}_{\{\tau_x < 1\}} \hat{\mathbb{P}}_{\tau_x}\left(Y_1 + Z_1 > 0\right)\right) \quad (5.3)$$

with Y and Z defined above. Notice that, on $\{\tau_x < 1\}$,

$$\hat{\mathbb{P}}_{\tau_x}(Y_1 + Z_1 > 0) = \hat{\mathbb{P}}_{\tau_x}\left(\hat{\sigma}_{\tau_x}\frac{B_1 - B_{\tau_x}}{\sqrt{1 - \tau_x}} + \frac{1}{\sqrt{1 - \tau_x}}\int_{\tau_x}^1 \hat{\mu}_u du\right.$$

$$\left. + \frac{1}{\sqrt{1 - \tau_x}}\int_0^1 (\hat{\sigma}_{u \vee \tau_x} - \hat{\sigma}_{\tau_x})dB_u^x > 0\right),$$

where $B_s^x = B_{s \vee \tau_x}$ is a continuous square-integrable martingale with respect to $(\hat{\mathcal{F}}_{s \vee \tau_x})$. Denote

$$R_s = \int_0^s (\hat{\sigma}_{u \vee \tau_x} - \hat{\sigma}_{\tau_x})dB_u^x, \quad s \in [0, 1],$$

and notice that it is a square-integrable martingale with respect to $(\hat{\mathcal{F}}_{s \vee \tau_x})$. Then, on $\{\tau_x < 1\}$ (possibly, without a set of measure zero), we have

$$\hat{\mathbb{E}}_{\tau_x}\left(\frac{1}{\sqrt{1 - \tau_x}}R_1\right)^2 = \frac{1}{1 - \tau_x}\hat{\mathbb{E}}_{\tau_x}R_1^2 \leq \frac{1}{1 - \tau_x}\int_{\tau_x}^1 \hat{\mathbb{E}}_{\tau_x}(\hat{\sigma}_{u \vee \tau_x} - \hat{\sigma}_{\tau_x})^2 du \leq \varepsilon.$$

In addition,

$$\hat{\mathbb{E}}_{\tau_x}\left(\frac{1}{\sqrt{1 - \tau_x}}\int_{\tau_x}^1 \hat{\mu}_u du\right)^2 \leq \varepsilon.$$

Collecting the above and using Chebyshev's inequality, we obtain, on $\{\tau_x < 1\}$:

$$\left|\hat{\mathbb{P}}_{\tau_x}(Y_1 + Z_1 > 0) - \hat{\mathbb{P}}_{\tau_x}\left(\hat{\sigma}_{\tau_x}\frac{B_1 - B_{\tau_x}}{\sqrt{1 - \tau_x}} \leq -\varepsilon^{1/3}\right)\right| \leq 2\varepsilon^{1/6}.$$

On the other hand, due to the strong Markov property of Brownian motion, on $\{\tau_x < 1\}$, we have, a.s.

$$\hat{\mathbb{P}}_{\tau_x}\left(\hat{\sigma}_{\tau_x}\frac{B_1 - B_{\tau_x}}{\sqrt{1 - \tau_x}} \leq -\varepsilon^{1/3}\right) = \hat{\mathbb{P}}\left(\xi \leq -\frac{\varepsilon^{1/3}}{\sigma}\right)\Big|_{\sigma = \hat{\sigma}_{\tau_x}},$$

where ξ is a standard normal. As $\hat{\sigma}_{\tau_x} \in [c, C]$, we conclude that the right-hand side of the above converges to $1/2$, as $\varepsilon \to 0$, uniformly over almost all random outcomes in $\{\tau_x < 1\}$. In particular, for all small enough $\varepsilon > 0$, we have

$$\mathbf{1}_{\{\tau_x < 1\}}\left|\hat{\mathbb{P}}_{\tau_x}(Y_1 + Z_1 \leq 0) - \hat{\mathbb{P}}_{\tau_x}(Y_1 + Z_1 > 0)\right| \leq \mathbf{1}_{\{\tau_x < 1\}}\delta(\varepsilon) < 1,$$

and, in view of (5.3),

$$\hat{\mathbb{P}}(X_1 > x) \geq \hat{\mathbb{E}}\left(\mathbf{1}_{\{\tau_x < 1\}}\hat{\mathbb{P}}_{\tau_x}(Y_1 + Z_1 \leq 0)\right) - \delta(\varepsilon)\hat{\mathbb{P}}(\tau_x < 1).$$

Summing up the above inequality and (5.3), we obtain

$$2\hat{\mathbb{P}}(X_1 > x) \geq (1 - \delta(\varepsilon))\hat{\mathbb{P}}(\tau_x < 1) = (1 - \delta(\varepsilon))\hat{\mathbb{P}}\left(\sup_{t \in [0,1]} X_t > x\right),$$

which, along with (5.2), yields the statement of the lemma. □

6 | PROOF OF THEOREM 4.8

Within the scope of this proof, we adopt the notation introduced in (5.1) and use the following convention:

Notational Convention 6.1. *The LOB, the bid and ask prices, the expected execution prices, and the demand, are all measured relative to p^0. Namely, we use ν_n to denote $\nu_n \circ (x \mapsto x + p_n^0)^{-1}$, p_n^a to denote $p_n^a - p_n^0$, p_n^b to denote $p_n^b - p_n^0$, λ_n^a to denote $\lambda_n^a - p_n^0$, λ_n^b to denote $\lambda_n^b - p_n^0$, and $D_n(p)$ to denote $D_n(p_n^0 + p)$.*

Herein we are only concerned with what happens in the last trading period—at time $(N-1)$, where $N = T/\Delta t$. Hence, we omit the subscript $N-1$ whenever it is clear from the context. In particular, we write p^a and p^b for p_{N-1}^a and p_{N-1}^b, ν for ν_{N-1}, and ξ for ξ_N. Note also that in an LTC equilibrium, we have $p^a = p_N^a = p_{N-1}^a$, with similar equalities for p^b and ν. For convenience, we also drop the superscript Δt in the LOB and the associated bid and ask prices. Finally, we denote by $\tilde{\mathbb{A}}$ the support of a given equilibrium. As the roles of p^a and p^b in our model are symmetric, we will only prove the statement of the proposition for p^b. We are going to show that, under the assumptions of the theorem, there exists a constant $C_0 > 0$, depending only on the constant C in Assumptions 4.1 and 4.2, such that, for all small enough Δt, we have, $\mathbb{P}$-a.s.

$$-C_0 \leq p^b / \sqrt{\Delta t} < 0. \tag{6.1}$$

First, we introduce $\hat{A}^\alpha(p; x)$, which we refer to as the simplified objective:

$$\hat{A}^\alpha(p; x) = \mathbb{E}_{N-1}^\alpha \left((p - x - \xi) \mathbf{1}_{\{\xi > p\}} \right). \tag{6.2}$$

Recall that the expected relative profit from posting a limit sell order at price level p, in the last time period,[11] is given by $A^\alpha(p; p_N^b)$, where

$$A^\alpha(p; x) = \mathbb{E}_{N-1}^\alpha \left((p - x - \xi) \mathbf{1}_{\{D_N^+(p-\xi) > \nu^+((-\infty, p))\}} \right). \tag{6.3}$$

The simplified objective is similar to A^α, but it assumes that there are no orders posted at better prices than the one posted by the agent. In particular, $\hat{A}^\alpha(p; x) = A^\alpha(p; x)$ for $p \leq p^a$. Corollary A.2, in Appendix A, states that in equilibrium, $\mathbb{P}$-a.s., if the agents in the state (s, α) post limit sell orders, then they post them at a price level p that maximizes the true objective $A^\alpha(p; p^b)$. The following lemma shows that the value of the modified objective becomes close to the value of the true objective, for the agents posting limit sell orders close to the ask price.

Lemma 6.2. $\mathbb{P}$-*a.s., either $\nu^+(\{p^a\}) > 0$ or we have*

$$\left| A^\alpha(p; p^b) - \hat{A}^\alpha(p^a; p^b) \right| \to 0,$$

as $p \downarrow p^a$, uniformly over all $\alpha \in \tilde{\mathbb{A}}$.

Proof. If $\nu^+(\{p^a\}) = 0$, then ν^+ is continuous at p^a, and $\nu^+((-\infty, p]) \to 0$, as $p \downarrow p^a$. Then, we have

$$\left| A^\alpha(p; p^b) - \hat{A}^\alpha(p^a; p^b) \right|$$

$$= \left| \mathbb{E}_{N-1}^\alpha \left((p - p^b - \xi) \mathbf{1}_{\{D_N^+(p-\xi) > \nu^+((-\infty, p))\}} \right) - \mathbb{E}_{N-1}^\alpha \left((p^a - p^b - \xi) \mathbf{1}_{\{\xi > p^a\}} \right) \right|$$

$$\leq |p - p^a| + \|p^a - p^b - \xi\|_{\mathbb{L}^2\left(\mathbb{P}_{N-1}^\alpha\right)} \mathbb{P}_{N-1}^\alpha \left(\xi > p^a, \, D_N^+(p - \xi) \leq \nu^+((-\infty, p)) \right).$$

Thus, it suffices to show that: (i) $\|p^a - p^b - \xi\|_{\mathbb{L}^2(\mathbb{P}^\alpha_{N-1})}$ is bounded by a finite random variable independent of α, and (ii)

$$\mathbb{P}^\alpha_{N-1}\left(\xi_N > p^a, \; D^+_N(p - \xi) \le v^+((-\infty, p))\right) \to 0, \quad \mathbb{P}\text{-a.s.,}$$

as $p \downarrow p^a$, uniformly over α. For (i), we have

$$\|p^a - p^b - \xi\|_{\mathbb{L}^2\left(\mathbb{P}^\alpha_{N-1}\right)} \le |p^a - p^b| + \|\xi\|_{\mathbb{L}^2\left(\mathbb{P}^\alpha_{N-1}\right)} \le |p^a - p^b| + 2C\sqrt{\Delta t},$$

where the constant C appears in Assumptions 4.1 and 4.2. For (ii), we note that

$$\{\xi_N > p^a, \; D^+_N(p - \xi) \le v^+((-\infty, p))\} = \{\xi_N > p^a, \; \xi \le p - D^{-1}_N\left(v^+((-\infty, p))\right)\},$$

as $D_N(\cdot)$ is strictly decreasing, with $D_N(0) = 0$. Assumption 4.6 implies that

$$\kappa^{-1}(v^+((-\infty, p))) \le D^{-1}_N(v^+((-\infty, p))) < 0,$$

where κ is known at time $N - 1$. Therefore,

$$\mathbb{P}^\alpha_{N-1}\left(\xi > p^a, \; D^+_N(p - \xi) \le v^+((-\infty, p))\right) \le \mathbb{P}^\alpha_{N-1}\left(\xi \in \left(p^a, p - \kappa^{-1}(v^+((-\infty, p)))\right]\right).$$

It remains to show that, $\mathbb{P}$-a.s., the right-hand side of the above converges to zero, uniformly over all α. Assume that it does not hold. Then, with positive probability $\mathbb{P}$, there exists $\varepsilon > 0$ and a sequence of (p_k, α_k), such that $p_k \downarrow p^a$ and

$$\mathbb{P}^{\alpha_k}_{N-1}\left(\xi \in (p^a, p_k - \kappa^{-1}(v^+((-\infty, p_k)))]\right) \ge \varepsilon.$$

Notice that, $\mathbb{P}$-a.s., the family of measures $\left\{\hat{\mu}_k = \mathbb{P}^{\alpha_k}_{N-1} \circ \xi^{-1}\right\}_k$ is tight. The latter follows, e.g., from the fact that, $\mathbb{P}$-a.s., the conditional second moments of ξ are bounded uniformly over all α (which, in turn, is a standard exercise in stochastic calculus). Prokhorov's theorem, then, implies that there is a subsequence of these measures that converges weakly to some measure $\hat{\mu}$ on $\mathbb{R}$. Next, notice that for any fixed k in the chosen subsequence, there exists a large enough k', such that

$$\left|\hat{\mu}\left(\left(p^a, p_k - \kappa^{-1}(v^+((-\infty, p_k)))\right]\right) - \mu_{k'}\left(\left(p^a, p_k - \kappa^{-1}(v^+((-\infty, p_k)))\right]\right)\right| \le \varepsilon/2.$$

Thus, for any k in the subsequence, we have

$$\hat{\mu}\left(\left(p^a, p_k - \kappa^{-1}(v^+((-\infty, p_k)))\right]\right) \ge \varepsilon/2.$$

The above is a contradiction, as the intersection of the corresponding intervals, $(p^a, p_k - \kappa^{-1}(v^+((-\infty, p_k)))]$, over all k is empty. □

Now we are ready to prove the upper bound in (6.1).

Lemma 6.3. *In any nondegenerate LTC equilibrium, $p^b < 0 < p^a$, $\mathbb{P}$-a.s.*

Proof. We only show that $p^b < 0$ holds, the other inequality being very similar. Assume that $p^b \ge 0$ on some positive $\mathbb{P}$-probability set $\Omega' \in \mathcal{F}_{N-1}$. We are going to show that this results in a contradiction. First, Corollary A.2, in Appendix A, implies that, $\mathbb{P}$-a.s., if the agents in state (s, α) post a limit sell order, then we must have $\sup_{p \in \mathbb{R}} A^\alpha(p; p^b) \ge 0$. In addition, on Ω', we have $\hat{A}^\alpha(p^a; p^b) < 0$ for all $\alpha \in \tilde{\mathbb{A}}$, as ξ has full support in $\mathbb{R}$ under every $\mathbb{P}^\alpha_{N-1}$ (which, in turn, follows from the fact that σ is bounded

uniformly away from zero). Then, Lemma 6.2 implies that there exists a $\mathcal{F}_{N-1}$-measurable $\bar{p} \geq p^a$, such that, on Ω', the following holds a.s.: if $v^+(\{p^a\}) = 0$ then $\bar{p} > p^a$, and, in all cases,

$$A^\alpha(p; p^b) < 0, \qquad \forall p \in [p^a, \bar{p}], \quad \forall \alpha \in \tilde{\mathbb{A}}. \tag{6.4}$$

Clearly, it is suboptimal for an agent to post a limit sell order below $\bar{p}$. However, an agent's strategy only needs to be optimal up to a set of $\mathbb{P}$-measure zero, and these sets can be different for different (s, α). Therefore, a little more work is required to obtain the desired contradiction. Consider the set $B \subset \Omega' \times \mathbb{R} \times \tilde{\mathbb{A}}$:

$$B = \{(\omega, s, \alpha) \mid \hat{q}(s, \alpha) > 0, \ \hat{p}(s, \alpha) \leq \bar{p}\}.$$

This set is measurable with respect to $\mathcal{F}_{N-1} \otimes \mathcal{B}(\mathbb{R} \times \tilde{\mathbb{A}})$, due to the measurability properties of $\hat{q}$ and $\hat{p}$. Notice that, due to the above discussion and the optimality of agents' actions (cf. Corollary A.2 in Appendix A), for any $(s, \alpha) \in \mathbb{R} \times \tilde{\mathbb{A}}$, we have:

$$\mathbb{P}(\{\omega \mid (\omega, s, \alpha) \in B\}) = 0,$$

and hence

$$\mathbb{E}_{N-1} \int_{\mathbb{R} \times \tilde{\mathbb{A}}} \mathbf{1}_B(\omega, s, \alpha) \mu_{N-1}(ds, d\alpha) = \int_{\mathbb{R} \times \tilde{\mathbb{A}}} \mathbb{E}_{N-1} \left(\mathbf{1}_B(\omega, s, \alpha) \rho_{N-1}(\omega, s, \alpha)\right) \mu_{N-1}^0(ds, d\alpha) = 0,$$

where ρ_{N-1} is the Radon–Nikodym density of μ_{N-1} with respect to the deterministic measure μ_{N-1}^0 (cf. Assumption 4.7). The above implies that, $\mathbb{P}_{N-1}$-a.s., $\mathbf{1}_B(\omega, s, \alpha) \rho_{N-1}(\omega, s, \alpha) = 0$, for μ_{N-1}^0-a.e. (s, α). Notice also that, for all $(\omega, s, \alpha) \in \Omega' \times \mathbb{R} \times \tilde{\mathbb{A}}$,

$$\mathbf{1}_{\{\hat{p}(s,\alpha) \leq \bar{p}\}} \hat{q}^+(s, \alpha) \mathbf{1}_{B^c} = 0.$$

From the above observations and the condition (2.4) in the definition of equilibrium (cf. Definition 2.6), we conclude that, on Ω', the following holds a.s.:

$$v^+([p^a, \bar{p}]) = 0,$$

where $\bar{p} \geq p^a$, and, if $v^+(\{p^a\}) = 0$, then $\bar{p} > p^a$. This contradicts the definition of p^a (recall that p^a is $\mathbb{P}$-a.s. finite, due to nondegeneracy of the LOB). □

It only remains to prove the lower bound on p^b in (6.1). Assume that it does not hold. That is, assume that there exists a family of equilibria, with arbitrary small Δt, and positive $\mathbb{P}$-probability $\mathcal{F}_{N-1}$-measurable sets $\Omega^{\Delta t}$, such that $p^b < -C_0 \sqrt{\Delta t}$ on $\Omega^{\Delta t}$. We are going to show that this leads to a contradiction with $p^a > 0$. To this end, assume that the agents maximize the simplified objective function, $\hat{A}^\alpha$, instead of the true one, A^α. Then, if p^b is negative enough, the optimal price levels become negative for all α. The precise formulation of this is given by the following lemma.

Lemma 6.4. *There exists a constant $C_0 > 0$, s.t., for any small enough Δt, there exist constants $\epsilon, \delta > 0$, s.t., $\mathbb{P}$-a.s., we have*

$$\hat{A}^\alpha(-\delta; x) \geq \epsilon + \sup_{y \geq 0} \hat{A}^\alpha(y; x)$$

for all $\alpha \in \tilde{\mathbb{A}}$ and all $x \leq -C_0 \sqrt{\Delta t}$.

Proof. Denote $\bar{\xi} = \xi/\sqrt{\Delta t}$ and consider the random function

$$\bar{A}^\alpha(p; x) = \mathbb{E}^\alpha_{N-1}\left((p - x - \bar{\xi})\mathbf{1}_{\{\bar{\xi}>p\}}\right).$$

Notice that

$$\hat{A}^\alpha(p; x) = \sqrt{\Delta t}\,\bar{A}^\alpha\left(p/\sqrt{\Delta t}; x/\sqrt{\Delta t}\right),$$

and, hence, we can reformulate the statement of the lemma as follows: There exists a constant $C_0 > 0$, s.t., for any small enough Δt, there exist constants $\epsilon, \delta > 0$, s.t., $\mathbb{P}$-a.s., we have

$$\bar{A}^\alpha(-\delta; x) \geq \epsilon + \sup_{y \geq 0} \bar{A}^\alpha(y; x)$$

for all $\alpha \in \tilde{\mathbb{A}}$ and all $x \leq -C_0$. Notice that

$$\bar{A}^\alpha(-\delta; x) - \bar{A}^\alpha(y; x) = -x\mathbb{E}^\alpha_{N-1}\left(\mathbf{1}_{\{-\delta<\bar{\xi}\leq y\}}\right) - \mathbb{E}^\alpha_{N-1}\left(\xi\mathbf{1}_{\{-\delta<\bar{\xi}\leq y\}}\right) - \delta\mathbb{E}^\alpha_{N-1}\left(\mathbf{1}_{\{\bar{\xi}>-\delta\}}\right) - y\mathbb{E}^\alpha_{N-1}\left(\mathbf{1}_{\{\bar{\xi}>y\}}\right)$$

is nonincreasing in x, and, hence, such is $\bar{A}^\alpha(-\delta; x) - \sup_{y\geq 0} \bar{A}^\alpha(y; x)$. Hence, it suffices to prove the above statement for $x = -C_0$. Next, consider the deterministic function $A_\sigma(p; x)$, defined via

$$A_\sigma(p; x) = \hat{\mathbb{E}}\left((p - x - \sigma\eta_0)\mathbf{1}_{\{\sigma\eta_0>p\}}\right), \tag{6.5}$$

where η_0 is a standard normal random variable on some auxiliary probability space $(\hat{\Omega}, \hat{\mathbb{P}})$. It follows from Lemma 5.1 that there exists a function $\varepsilon_2(\cdot) \geq 0$, s.t. $\varepsilon_2(\Delta t) \to 0$, as $\Delta t \to 0$, and, $\mathbb{P}$-a.s., we have

$$\left|\bar{A}^\alpha(p; -C_0) - A_{\sigma_{t_{N-1}}}(p; -C_0)\right| \leq \varepsilon_2(\Delta t)$$

for all $\alpha \in \tilde{\mathbb{A}}$ and all $p \in \mathbb{R}$. Then, as we can always choose Δt small enough, so that $\varepsilon_2(\Delta t) < \epsilon$, the statements of the lemma would follow if we can show that there exist constants $\epsilon, \delta, C_0 > 0$, s.t., $\mathbb{P}$-a.s.,

$$A_{\sigma_{t_{N-1}}}(-\delta; -C_0) \geq 3\epsilon + \sup_{y\geq 0} A_{\sigma_{t_{N-1}}}(y; -C_0).$$

As $\sigma_{t_{N-1}}(\omega) \in [1/C, C]$, $\mathbb{P}$-a.s., it suffices to find $\epsilon, \delta, C_0 > 0$, s.t.

$$A_\sigma(-\delta; -C_0) \geq 3\epsilon + \sup_{y\geq 0} A_\sigma(y; -C_0), \quad \forall \sigma \in [1/C, C].$$

Note that the above inequality does not involve ω or ξ, and it is simply a property of a deterministic function. Notice also that $A_\sigma(p; x) = \sigma A_1(p/\sigma; x/\sigma)$, with A_1 given in (6.5). Then, if we denote by $F(x)$ and $f(x)$, respectively, the cdf and pdf of a standard normal, we obtain

$$A_1(p; x) = (p - x)(1 - F(p)) - \int_p^\infty t f(t)dt.$$

A straightforward calculation gives us the following useful properties of A_1 and A_σ:

(i) For any $\sigma > 0$ and any $x < 0$, the function $p \mapsto A_\sigma(p; x)$ has a unique maximizer $p_\sigma(x)$, in particular, it is increasing in $p \leq p_\sigma(x)$ and decreasing in $p \geq p_\sigma(x)$.

(ii) The function

$$x \mapsto p_\sigma(x) = \sigma p_1(x/\sigma) = \sigma \left((1 - F)/f\right)^{-1} (-x/\sigma)$$

is increasing in $x < 0$ and converges to $-\infty$, as $x \to -\infty$.

Then, choosing C_0 large enough, so that $p_1(-C_0/C) < 0$, ensures $p_\sigma(-C_0) < 0$, for all $\sigma \in [1/C, C]$. Setting $\delta = -p_1(-C_0/C)/C$ guarantees that $p_\sigma(-C_0) \leq -\delta$, for all $\sigma \in [1/C, C]$. Then, by property (i) above, we have, for all $\sigma \in [1/C, C]$

$$A_\sigma(-\delta; -C_0) > A_\sigma(0; -C_0) = \sup_{y \geq 0} A_\sigma(y; -C_0).$$

Finally, as $A_\sigma(-\delta; -C_0) - A_\sigma(0; -C_0)$ is a continuous function of $\sigma \in [1/C, C]$, we can find ϵ, such that

$$A_\sigma(-\delta; -C_0) \geq 3\epsilon + \sup_{y \geq 0} A_\sigma(y; -C_0), \quad \forall \sigma \in [1/C, C]. \qquad \square$$

Recall that our assumption is that $p^b < -C_0\sqrt{\Delta t}$ holds on a set $\Omega^{\Delta t}$ of positive $\mathbb{P}$-measure. Recall also that $p^a > 0$, $\mathbb{P}$-a.s., due to Lemma 6.3. Then, Lemmas 6.2 and 6.4 imply that there exists $\mathcal{F}_{N-1}$-measurable $\bar{p} \geq p^a$, s.t., on $\Omega^{\Delta t}$, we have a.s.: if $\nu^+(\{p^a\}) = 0$ then $\bar{p} > p^a$, and, in all cases,

$$A^\alpha(p; p^b) < \sup_{p' \in \mathbb{R}} A^\alpha(p'; p^b), \quad \forall p \in [p^a, \bar{p}], \quad \forall \alpha \in \tilde{\mathbb{A}}.$$

It is intuitively clear that posting limit sell orders at the above price levels p must be suboptimal for the agents. However, the above inequality, on its own, does not yield a contradiction, as the agents' strategies are only optimal up to a set of $\mathbb{P}$-probability zero, and these sets may be different for different states (s, α). To obtain a contradiction with the definition of p^a, we simply repeat the last part of the proof of Lemma 6.3 (following equation (6.4)). This ensures that (6.1) holds and completes the proof of the theorem.

7 | PROOF OF THEOREM 4.10

Within the scope of this proof, we adopt the notation introduced in (5.1) and use Notational Convention 6.1 (i.e., we measure the LOB, the expected execution prices, and the demand, relative to p^0, but keep the same variables' names). Assume that the statement of the theorem does not hold; i.e., there exists $\alpha_0 \in \tilde{\mathbb{A}}$, such that $\tilde{p}^0$ is not a martingale under $\mathbb{P}^{\alpha_0}$. Then, there exists $s \in [0, T)$, s.t., with positive probability $\mathbb{P}^{\alpha_0}$, we have

$$\tilde{\mathbb{E}}_s^{\alpha_0} \tilde{p}_T^0 \neq \tilde{p}_s^0.$$

Without loss of generality, we assume that there exists a constant $\delta > 0$ and a set $\Omega' \in \mathcal{F}_s$, having positive probability $\mathbb{P}^{\alpha_0}$ (and hence $\mathbb{P}$), s.t. for all random outcomes in Ω', we have

$$\tilde{\mathbb{E}}_s^{\alpha_0} \left(\tilde{p}_T^0 - \tilde{p}_s^0\right) \geq \delta \tag{7.1}$$

(the case of negative values is analogous). Next, we fix an arbitrary Δt from a given family and consider the associated nondegenerate LTC equilibrium.

Lemma 7.1. *There exists a deterministic function $\varepsilon(\cdot) \geq 0$, s.t. $\varepsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, for any small enough $\Delta t > 0$, there exists $n = 0, \dots, N - 3$ and $\Omega'' \in \mathcal{F}_n$, s.t. $\mathbb{P}_n^{\alpha_0}(\Omega'') > 0$ and the following holds on Ω''*

$$\mathbb{P}_{n+2}^{\alpha_0} \left(\mathbb{E}_{n+3}^{\alpha_0} \left(p_N^0 - p_{n+3}^0 \right) \leq \delta/2 \right) \leq \varepsilon(\Delta t).$$

Proof. The proof follows from Assumption 4.5. Consider $t = t' = s$ and $t'' = t_{n+2}$. Then, Assumption 4.5 implies

$$\tilde{\mathbb{P}}_s^{\alpha_0} \left(\left| \tilde{\mathbb{E}}_{t_{n+2}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du - \tilde{\mathbb{E}}_s^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du \right| \geq \varepsilon(\Delta t) \right) \leq \varepsilon(\Delta t)$$

on Ω', a.s. Notice also that

$$\tilde{\mathbb{E}}_s^{\alpha_0} \left(\tilde{p}_T^0 - \tilde{p}_s^0 \right) = \tilde{\mathbb{E}}_s^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du.$$

Then, assuming that $\varepsilon(\Delta t)$ is small enough and recalling (7.1), we obtain

$$\tilde{\mathbb{P}}_s^{\alpha_0} \left(\tilde{\mathbb{E}}_{t_{n+2}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du \leq 3\delta/4 \right) \leq \varepsilon(\Delta t),$$

on Ω'. Therefore, there exists a set $\Omega'' \in \mathcal{F}_s \subset \mathcal{F}_{t_n}$, s.t. $\tilde{\mathbb{P}}_{t_n}^{\alpha_0}(\Omega'') > 0$ and

$$\tilde{\mathbb{E}}_{t_{n+2}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du \geq 3\delta/4,$$

on Ω''. Next, we choose $t = s$, $t' = t_{n+2}$, $t'' = t_{n+3}$, and use Assumption 4.5, to obtain

$$\tilde{\mathbb{P}}_{t_{n+2}}^{\alpha_0} \left(\left| \tilde{\mathbb{E}}_{t_{n+3}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du - \tilde{\mathbb{E}}_{t_{n+2}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du \right| \geq \varepsilon(\Delta t) \right) \leq \varepsilon(\Delta t),$$

on Ω'', a.s. Assuming that $\varepsilon(\Delta t)$ is small enough and using the last two inequalities, we obtain

$$\tilde{\mathbb{P}}_{t_{n+2}}^{\alpha_0} \left(\tilde{\mathbb{E}}_{t_{n+3}}^{\alpha_0} \int_s^T \mu_u^{\alpha_0} du \leq \delta/2 \right) \leq \varepsilon(\Delta t).$$

Finally, due to Assumption 4.2, and as Δt is small, we can replace $\int_s^T \mu_u^{\alpha_0} du$ by $\int_{t_{n+3}}^T \mu_u^{\alpha_0} du$, and $\delta/2$ by $\delta/4$, in the above equation. This completes the proof of the lemma. $\qquad\square$

Using the strategy at which the agent in state $(1, \alpha_0)$ waits until the last moment $n = N$, we conclude that the process $(\lambda_n^a(\alpha_0) + p_n^0)$ must be a supermartingale under $\mathbb{P}^{\alpha_0}$. More precisely, due to the definition of an optimal strategy, we have, $\mathbb{P}$-a.s.

$$\lambda_{n+2}^a(\alpha_0) \geq \mathbb{E}_{n+2}^{\alpha_0} \lambda_N^a(\alpha_0) + \mathbb{E}_{n+2}^{\alpha_0} \left(\mathbb{E}_{n+3}^{\alpha_0} \left(p_N^0 - p_{n+3}^0 \right) + \xi_{n+3} \right).$$

Recall that $\lambda_N^a(\alpha_0) = p_N^b$ and, due to Theorem 4.8 (more precisely, it follows from the proof of the theorem), there exists a constant $C_0 > 0$, s.t., for all small enough $\Delta t > 0$, the following holds $\mathbb{P}$-a.s.

$$-C_0 \sqrt{\Delta t} \leq p_N^b < 0 < p_N^a \leq C_0 \sqrt{\Delta t}.$$

Thus, we have, $\mathbb{P}$-a.s.

$$\lambda_{n+2}^a(\alpha_0) \geq -C_0\sqrt{\Delta t} + \mathbb{E}_{n+2}^{\alpha_0}\left(\mathbb{E}_{n+3}^{\alpha_0}\left(p_N^0 - p_{n+3}^0\right)\right) + \mathbb{E}_{n+2}^{\alpha_0}\xi_{n+3}. \tag{7.2}$$

Due to Assumption 4.2, we have, $\mathbb{P}$-a.s.

$$\mathbb{E}_{n+2}^{\alpha_0}\xi_{n+3} \leq C\Delta t, \qquad \left|\mathbb{E}_{n+3}^{\alpha_0}\left(p_N^0 - p_{n+3}^0\right)\right| \leq CT,$$

and, hence,

$$\lambda_{n+2}^a(\alpha_0) \geq -C_0\sqrt{\Delta t} + CT + C\Delta t.$$

In addition, making use of Lemma 7.1, we conclude that, for any small enough Δt, there exist $n = 0, \dots, N-2$ and $\Omega'' \in \mathcal{F}_n$, s.t. $\mathbb{P}_n^{\alpha_0}(\Omega'') > 0$ and

$$\mathbb{P}_{n+2}^\alpha\left(\mathbb{E}_{n+3}^\alpha\left(p_N^0 - p_{n+3}^0\right) \leq \delta/2\right) \leq \varepsilon(\Delta t), \quad \text{on } \Omega''.$$

Using (7.2) and assuming that Δt is small enough, we obtain

$$\lambda_{n+2}^a(\alpha_0) \geq \delta/4, \quad \text{on } \Omega''.$$

Next, Corollary A.2, in Appendix A, implies that, $\mathbb{P}$-a.s.,

$$p_{n+1}^b \geq \mathbb{E}_{n+1}^{\alpha_0}\left(\lambda_{n+2}^a(\alpha_0) + \xi_{n+2}\Big|\xi_{n+2} < p_{n+1}^b\right).$$

Thus, on Ω'', we obtain

$$p_{n+1}^b - \mathbb{E}_{n+1}^{\alpha_0}\left(\xi_{n+2}\Big|\xi_{n+2} < p_{n+1}^b\right) \geq \delta/4. \tag{7.3}$$

The following lemma shows that, for any number p, the conditional expectation of the fundamental price increment, $\mathbb{E}_{n+1}^{\alpha_0}(\xi_{n+2}|\xi_{n+2} < p)$, approaches p as the size of the time interval vanishes. This result follows from Lemma 5.2.

Lemma 7.2. *There exists a constant $C_3 > 0$, s.t., for all small enough $\Delta t > 0$, and for any $t \in [0, T - \Delta t]$, the following holds $\mathbb{P}$-a.s.*

$$\sup_{p \leq 0}\left|p - \tilde{\mathbb{E}}_t^{\alpha_0}\left(\tilde{p}_{t+\Delta t}^0 - \tilde{p}_t^0 \Big| \tilde{p}_{t+\Delta t}^0 - \tilde{p}_t^0 < p\right)\right| \leq C_3\sqrt{\Delta t}.$$

Proof. Fix t and $\Delta t > 0$ and consider the evolution of $\tilde{p}_s^0$, for $s \in [t, t + \Delta t]$, under $\mathbb{P}_t^{\alpha_0}$

$$\tilde{p}_s^0 - \tilde{p}_t^0 = \int_t^s \mu_u^{\alpha_0}du + \int_t^s \sigma_u dW_u^{\alpha_0},$$

where W^{α_0} is a Brownian motion under $\mathbb{P}^{\alpha_0}$. Rescaling by $\sqrt{\Delta t}$, we obtain

$$\left(\tilde{p}_s^0 - \tilde{p}_t^0\right)/\sqrt{\Delta t} = X_{(s-t)/\Delta t}, \quad X_s = \int_0^s \hat{\mu}_u du + \int_0^s \hat{\sigma}_u d\hat{W}_u, \quad s \in [0, 1],$$

with

$$\hat{\mu}_s = \sqrt{\Delta t}\,\mu_{t+s\Delta t}^{\alpha_0}, \quad \hat{\sigma}_s = \sigma_{t+s\Delta t}, \quad \hat{W}_s = \frac{1}{\sqrt{\Delta t}}\left(W_{t+s\Delta t}^{\alpha_0} - W_t^{\alpha_0}\right), \quad s \in [0, 1].$$

Notice that the above processes are adapted to the new filtration $\hat{\mathbb{F}}$, with $\hat{\mathcal{F}}_s = \tilde{\mathcal{F}}_{t+s\Delta t}$, and, $\mathbb{P}$-a.s., under $\tilde{\mathbb{P}}_t^{\alpha_0}$, $\hat{W}$ is a Brownian motion with respect to $\hat{\mathbb{F}}$. Next, due to Assumptions 4.1 and 4.4, for any small enough $\Delta t > 0$, $\mathbb{P}$-a.s., the dynamics of $(-X_s)$, under $\tilde{\mathbb{P}}_t^{\alpha_0}$, satisfy all the assumptions of Lemma 5.2. As a result, we obtain

$$\tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < -x - z) \leq C_1 e^{-z} \tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < -x), \qquad \forall x, z \geq 0.$$

Finally, we notice that

$$\sup_{p \leq 0} \left| p - \tilde{\mathbb{E}}_t^{\alpha_0} \left(\tilde{p}_{t+\Delta t}^0 - \tilde{p}_t^0 \middle| \tilde{p}_{t+\Delta t}^0 - \tilde{p}_t^0 < p \right) \right| = \sqrt{\Delta t} \sup_{p \leq 0} \left| p - \tilde{\mathbb{E}}_t^{\alpha_0} \left(X_1 \middle| X_1 < p \right) \right|$$

$$= \sqrt{\Delta t} \sup_{p \leq 0} \left| p - \frac{\int_{-p}^{\infty} x \, d \, \tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < -x)}{\tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < p)} \right| = \sqrt{\Delta t} \sup_{p \leq 0} \left| \frac{\int_0^{\infty} \tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < p - z) dz}{\tilde{\mathbb{P}}_t^{\alpha_0}(X_1 < p)} \right| \leq C_1 \sqrt{\Delta t},$$

which completes the proof of the lemma. □

Using (7.3) and Lemma 7.2, we conclude that, for all small enough Δt, we have: $p_{n+1}^b > 0$ on Ω'', $\mathbb{P}$-a.s. In addition, Corollary A.2, in Appendix A, implies that, for any $\alpha \in \tilde{\mathbb{A}}$, the following holds $\mathbb{P}$-a.s.

$$\lambda_{n+1}^a(\alpha) \geq p_{n+1}^b.$$

Next, with a slight abuse of notation (similar notation was introduced in the proof of Proposition 4.8), we consider the simplified objective of an agent who posts a limit sell order at the ask price p_n^a

$$\hat{A}^\alpha \left(p_n^a; \lambda_{n+1}^a \right) = \mathbb{E}_n^\alpha \left(p_n^a - \lambda_{n+1}^a - \xi_{n+1} \mid \xi_{n+1} > p_n^a \right).$$

The above estimates imply that, on Ω'', we have, $\mathbb{P}$-a.s.

$$\hat{A}^\alpha \left(p_n^a; \lambda_{n+1}^a \right) \leq \mathbb{E}_n^\alpha \left(p_n^a - \xi_{n+1} \mid \xi_{n+1} > p_n^a \right) - \mathbb{E}_n^\alpha \left(p_{n+1}^b \mathbf{1}_{\Omega''} \mid \xi_{n+1} > p_n^a \right) < 0, \qquad \forall \alpha \in \tilde{\mathbb{A}}. \quad (7.4)$$

To obtain the last inequality in the above, we recall that $\Omega'' \in \mathcal{F}_n$ and, $\mathbb{P}$-a.s., $\mathbf{1}_{\Omega''}\mathbb{P}_n(\Omega \setminus \Omega'') = 0$, $p_{n+1}^b > 0$ on Ω'', and $\mathbb{P}_n^\alpha(\xi_{n+1} > p_n^a) > 0$, for all $\alpha \in \tilde{\mathbb{A}}$. Next, repeating the proof of Lemma 6.2 (and using the fact that λ_{n+1}^a is absolutely bounded, as shown in Corollary 4.9), we conclude that, $\mathbb{P}$-a.s., either $v_n^+(\{p_n^a\}) > 0$, or we have

$$\left| A^\alpha \left(p; \lambda_{n+1}^a \right) - \hat{A}^\alpha \left(p_n^a; \lambda_{n+1}^a \right) \right| \to 0,$$

as $p \downarrow p^a$, uniformly over all $\alpha \in \tilde{\mathbb{A}}$, where we introduce the true objective,

$$A^\alpha \left(p; \lambda_{n+1}^a \right) = \mathbb{E}_n^\alpha \left(\left(p - \lambda_{n+1}^a - \xi_{n+1} \right) \mathbf{1}_{\{ D_{n+1}^+(p - \xi_{n+1}) > v_n^+((-\infty, p)) \}} \right).$$

This convergence, along with (7.4), implies that there exists a $\mathcal{F}_n$-measurable $\bar{p} \geq p_n^a$, such that, on Ω'', the following holds $\mathbb{P}$-a.s.: if $v_n^+(\{p_n^a\}) = 0$ then $\bar{p} > p_n^a$, and, in all cases,

$$A^\alpha \left(p; \lambda_{n+1}^a \right) < 0, \qquad \forall p \in \left[p_n^a, \bar{p} \right], \quad \forall \alpha \in \tilde{\mathbb{A}}.$$

Finally, we repeat the last part of the proof of Lemma 6.3 (following equation (6.4)), to obtain a contradiction with the definition of p_n^a, and complete the proof of the theorem. The last argument also shows

that when Δt is small enough, it becomes suboptimal for the agents to post limit sell orders, as the expected relative profit from this action becomes negative, causing the market to degenerate.

8 | SUMMARY AND FUTURE WORK

In this paper, we present a new framework for modeling market microstructure, which does not require the existence of a designate market maker, and in which the LOB arises endogenously, as a result of equilibrium between multiple strategic players (also known as agents). This framework is based on a continuum-player game. It closely approximates the mechanics of an auction-style exchange, so that, in particular, it can be used to analyze the liquidity effects of changes in the rules of the exchange. We use the proposed framework to study the liquidity effects of high trading frequency. In particular, we demonstrate the dual nature of high trading frequency. On the one hand, in the absence of a bullish or bearish signal about the asset, the higher trading frequency improves the efficiency of the market. On the other hand, at a sufficiently high trading frequency, even a very small trading signal may amplify the adverse selection effect, creating a disproportionally large change in the LOB, which is interpreted as an endogenous liquidity crisis.

This paper raises many questions for further research. Notice that our main results are of a qualitative nature: They demonstrate the general behavior of the LOB, as a function of trading frequency, but do not immediately allow for computations. It would also be interesting to establish quantitative results. In particular, we would like to construct an equilibrium in a more realistic, and more concrete, model than the one used in Section 3. Such a model would allow for heterogeneous beliefs, and it would prescribe the specific sources of information (i.e., relevant market factors) used by the agents to form their beliefs. A model of this type could be calibrated to market data and used to study the effects of changes in relevant market parameters on the LOB. Finally, it would be interesting to develop a continuous-time version of the proposed framework, in order to better capture the present state of the markets, where the trading frequency is not restricted. All these questions are the subject of our follow-up paper Gayduk and Nadtochiy (2018).

ENDNOTES

[1] We do not distinguish the "aggressive" limit orders, which are posted at the price level of an opposite limit order, and treat them as market orders. This causes no loss of generality, as the market participants in our setting have a perfect observation of the LOB.

[2] This assumption holds, e.g., if $\mathcal{F}_N$ is generated by a random element with values in a standard Borel space.

[3] Note that although $\mathbb{P}^\alpha$ does not change over time, the conditional distribution of the future demand, as perceived by the agent, changes dynamically, according to the new information received.

[4] Note each agent is only allowed to place her limit order at a single price level, at any given time. However, this entails no loss of optimality. Indeed, using the Dynamic Programming Principle derived in Appendix A, one can show, by induction, that, in equilibrium, an agent does not benefit from posting multiple limit orders at the same time. As shown in Schmeidler (1973), this is typical for a continuum-player game.

[5] In order to ensure the existence of regular conditional probabilities for the discrete-time model, we can, e.g., assume that $\tilde{\mathcal{F}}_T$ is generated by a random element with values in a standard Borel space.

[6] The execution of limit orders simplifies in the chosen ansatz, because the agents on each side of the book (i.e., long or short) post orders at the same prices.

[7] In fact, it is not difficult to prove rigorously that, for any (α, σ), there exists a unique solution to such a system, provided Δt is small enough. We omit this result for the sake of brevity.

[8] This is easy to explain intuitively, as the optimal objective values in the first two lines of (3.2) are of the form $C\sqrt{\Delta t} + \underline{\alpha O(\Delta t)}$.

[9] In order to ensure the existence of regular conditional probabilities for the discrete-time model, we can, e.g., assume that $\tilde{\mathcal{F}}_T$ is generated by a random element with values in a standard Borel space.

[10] This argument, along with the fact that Definition 2.3 requires an optimal control to be optimal for *all* α, explains why the statement of Theorem 4.10 holds for *all*, as opposed to μ_n-a.e., $\alpha \in \tilde{\mathbb{A}}$.

[11] Recall that everything is measured relative to the fundamental price, according to the Notational Convention 6.1.

[12] The admissibility constraint does not cause any difficulties here, as, in the case where (p_n, q_n, r_n) do not attain the supremum, they can be improved, so that (p_n, q_n) increase by no more than a fixed constant.

REFERENCES

Adrian, T., Capponi, A., Vogt, E., & Zhang, H. (2016). Intraday market making with overnight inventory costs. Preprint. Retrieved from https://doi.org/ssrn.com/abstract=2844881

Aliprantis, C., & Border, K. (2006). *Infinite dimensional analysis: A hitchhiker's guide*. Berlin: Springer-Verlag.

Almgren, R. (2003). Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, *10*, 1–18.

Aumann, R. (1964). Markets with a continuum of traders. *Econometrica*, *32*, 39–50.

Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, *8*(3), 217–224.

Bayraktar, E., & Ludkovski, M. (2011). Optimal trade execution in illiquid markets. *Mathematical Finance*, *21*(4), 681–701.

Bressan, A., & Facchi, G. (2013). A bidding game in a continuum limit order book. *SIAM Journal on Control and Optimization*, *51*, 3459–3485.

Bressan, A., & Facchi, G. (2014). Discrete bidding strategies for a random incoming order. *SIAM Journal on Financial Mathematics*, *5*, 50–70.

Bressan, A., & Wei, D. (2016). Dynamic stability of the Nash equilibrium for a bidding game. *Analysis and Applications*, *14*(4), 1–24.

Brunnermeier, M. K., & Pedersen, L. H. (2005). Predatory trading. *Journal of Finance*, *60*(4), 1825–1863.

Cardaliaguet, P. (2010). *Notes on mean field games* (from P.-L. Lions' lectures at Collège de France). (Technical report. Retrieved from https://doi.org/www.ceremade.dauphine.fr/~cardalia/MFG20130420.pdf

Carmona, G. (2013). *Existence and stability of Nash equilibrium*. Singapore: World Scientific Publishing.

Carmona, R., & Delarue, F. (2013). Probabilistic analysis of mean-field games. *SICON*, *51*(4), 2705–2734.

Carmona, R., & Webster, K. (n.d.). *A belief-driven order book model*. Working paper. Retrieved from http://doi.org/www.princeton.edu/~rcarmona/download/fe/CW2.pdf

Carmona, R., & Webster, K. (2012). High frequency market making. Preprint, arXiv:1210.5781.

Cartea, A., & Jaimungal, A. (2013). General intensity shapes in optimal liquidation. *Applied Mathematical Finance*, *20*(6), 512–547.

Cont, R., Stoikov, S., & Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, *58*(3), 549–563.

Cvitanić, J., & Kirilenko, A. A. (2010). *High frequency traders and asset prices*. Retrieved from https://doi.org/10.2139/ssrn.1569067

Du, S., & Zhu, H. (2014). *Welfare and optimal trading frequency in dynamic double auctions*. NBER Working Paper No. 20588.

Easley, D., & O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance*, *47*(2), 577–605.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, *2*, 99–134.

Friedman, A. (1964). *Partial differential equations of parabolic type*. Englewood Cliffs, NJ: Prentice Hall.

Gatheral, J., & Schied, A. (2011). Optimal trade execution under geometric Brownian motion in the Almgren and Chriss framework. *International Journal on Theoretical and Applied Finance*, *14*(3), 353–368.

Gayduk, R., & Nadtochiy, S. (2015). *Liquidity effects of trading frequency (extended version)*. Technical report, arXiv:1508.07914.

Gayduk, R., & Nadtochiy, S. (2018). Endogenous formation of limit order books: Dynamics between trades. *SIAM Journal on Control and Optimization*, *56*(3), 1577–1619. https://doi.org/10.1137/16M1078045

Glosten, L., & Milgrom, P. P. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*, 71–100.

Goettler, R. L., Parlour, C. A., & Rajan, U. (2005). Equilibrium in a dynamic limit order market. *Journal of Finance*, *60*(5), 2149–2192.

Guéant, O., & Lehalle, C.-A. (2015). General intensity shapes in optimal liquidation. *Mathematical Finance*, *25*(3), 457–495.

Guilbaud, F., & Pham, H. (2013). Optimal high-frequency trading with limit and market orders. *Quantitative Finance*, *13*(1), 79–94.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., & Lions, P.-L. (2016). Efficiency of the price formation process in presence of high frequency participants: A mean field game analysis. *Mathematics and Financial Economics*, *10*(3), 223–262. https://doi.org/10.1007/s11579-015-0157-1

Lacker, D. (2016). A general characterization of the mean field limit for stochastic differential games. *Probability Theory and Related Fields*, *165*(3–4), 581–648. https://doi.org/10.1007/s00440-015-0641-9

Lasry, J.-M., & Lions, P.-L. (2007). Mean field games. *Japanese Journal of Mathematics*, *2*, 229–260.

Obizhaeva, A. A., & Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, *16*, 1–32.

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, *11*(4), 789–816.

Predoiu, S., Shaikhet, G., & Shreve, S. (2011). Optimal execution of a general one-sided limit-order book. *SIAM Journal on Financial Mathematics*, *2*, 183–212.

Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, *22*, 4601–4641.

Schied, A., Schoneborn, T., & Tehranchi, M. (2010). Optimal basket liquidation for CARA investors is deterministic. *Applied Mathematical Finance*, *17*, 471–489.

Schmeidler, D. (1973). Equilibrium points of nonatomic games. *Journal of Statistical Physics*, *7*(4), 295–300.

Stoikov, S., & Waeber, R. (2012). *Optimal asset liquidation using limit order book information*. Working Paper, https://doi.org/10.2139/ssrn.2113827

Vayanos, D. (1999). Strategic trading and welfare in a dynamic market. *Review of Economic Studies*, *66*, 219–254.

APPENDIX A

This section contains several useful technical results on the representation of the value function of an agent in the proposed game. Notice that (2.1) and (2.2) imply that if v is admissible, then, for any (α, m, p, q, r), we have, $\mathbb{P}$-a.s.

$$|J^{(p,q,r)}(m, s, \alpha, v) - J^{(p,q,r)}(m, s', \alpha, v)| \leq |s - s'| \mathbb{E}_m^\alpha \left| p_N^a \right| \vee \left| p_N^b \right|, \qquad \forall s, s' \in \mathbb{R}.$$

This implies that every $J^{(p,q,r)}(m, \cdot, \alpha, v)$ and $V_m^v(\cdot, \alpha)$ has a continuous modification under $\mathbb{P}$. Thus, whenever v is admissible, we define the value function of an agent as the aforementioned continuous modification of the left-hand side of (2.3).

Lemma A.1. *Assume that an optimal control exists for an admissible LOB v. Assume also that, for any $\alpha \in \mathbb{A}$, the associated value function $V_n^v(\cdot, \alpha)$, defined in (2.3), is measurable with respect to $\mathcal{F}_n \otimes \mathcal{B}(\mathbb{R})$. Then, it satisfies the following Dynamic Programming Principle:*

- *For $n = N$ and all $(s, \alpha) \in \mathbb{S}$, we have, $\mathbb{P}$-a.s.*

$$V_N^v(s, \alpha) = s^+ p_N^b - s^- p_N^a. \tag{A.1}$$

- *For all $n = N - 1, \ldots, 0$ and all $(s, \alpha) \in \mathbb{S}$, we have*

$$V_n^v(s, \alpha) = \operatorname{esssup}_{p,q,r} \left\{ \mathbf{1}_{\{r_n=0\}} \mathbb{E}_n^\alpha \left(V_{n+1}^v(s, \alpha) + \left(q_n p_n + V_{n+1}^v(s - q_n, \alpha) - V_{n+1}^v(s, \alpha) \right) \right. \right.$$

$$\left. \times \left(\mathbf{1}_{\left\{ q_n \geq 0, \, D_{n+1}^+(p_n) > v_n^+((-\infty, p_n)) \right\}} + \mathbf{1}_{\left\{ q_n < 0, \, D_{n+1}^-(p_n) > v_n^-((p_n, \infty)) \right\}} \right) \right)$$

$$\left. + \mathbf{1}_{\{r_n=1\}} \left(q_n^+ p_n^b - q_n^- p_n^a + \mathbb{E}_n^\alpha V_{n+1}^v(s - q_n, \alpha) \right) \right\}, \tag{A.2}$$

where the essential supremum is taken under $\mathbb{P}$, over all admissible controls (p, q, r).

Proof. The most important step is to show that, for all $n = 0, \ldots N - 1$ and $(s, \alpha) \in \mathbb{S}$,

$$V_n^v(s, \alpha) = \operatorname{esssup}_{p,q,r} \mathbb{E}_n^\alpha \left(V_{n+1}^v \left(S_{n+1}^{n,s,(p,q,r)}, \alpha \right) - g_n^v \left(p_n, q_n, r_n, D_{n+1} \right) \right), \tag{A.3}$$

where the essential supremum is taken under $\mathbb{P}$, over all admissible controls (p, q, r), and

$$g_n^v \left(p_n, q_n, r_n, D_{n+1} \right) = \left(p_n \mathbf{1}_{\{r_n=0\}} + p_n^a \mathbf{1}_{\{r_n=1, q_n<0\}} + p_n^b \mathbf{1}_{\{r_n=1, q_n>0\}} \right) \Delta S_{n+1}^{n,s,(p,q,r)}$$

does not depend on s. Assume that $J^{(p,q,r)}(n, \cdot, \alpha, v)$ is a continuous modification of the objective function. Notice that, for all $m \leq k \leq n$, we have, $\mathbb{P}$-a.s.

$$\mathbb{E}_k^\alpha J^{(p,q,r)} \left(n, S_n^{m,s,(p,q,r)}, \alpha, v \right) = J^{(p,q,r)} \left(k, S_k^{m,s,(p,q,r)}, \alpha, v \right) + \mathbb{E}_k^\alpha \sum_{j=k}^{n-1} g_j^v(p_j, q_j, r_j, D_{j+1}).$$

Notice also that for any (p, q, r) we have, $\mathbb{P}$-a.s.: $J^{(p,q,r)}(m, s, \alpha, v) \leq V_m^v(s, \alpha)$, for all $s \in \mathbb{S}$. Let us show that the left-hand side of (A.3) is less than its right-hand side

$$V_m^v(s, \alpha) = \operatorname{essup}_{p,q,r} J^{(p,q,r)} \left(m, S_m^{m,s,(p,q,r)}, \alpha, v \right)$$

$$= \operatorname{essup}_{p,q,r} \mathbb{E}_m^\alpha \left(J^{(p,q,r)} \left(m + 1, S_{m+1}^{m,s,(p,q,r)}, \alpha, v \right) - g_m^v(p_m, q_m, r_m, D_{m+1}) \right)$$

$$\leq \operatorname{essup}_{p,q,r} \mathbb{E}_m^\alpha \left(V_{m+1}^v \left(S_{m+1}^{m,s,(p,q,r)}, \alpha \right) - g_m^v(p_m, q_m, r_m, D_{m+1}) \right).$$

Next, we show that the right-hand side of (A.3) is less than its left-hand side. For any (p, q, r), we have, $\mathbb{P}$-a.s.

$$
\mathbb{E}_m^\alpha \left(V_{m+1}^\nu \left(S_{m+1}^{m,s,(p,q,r)}, \alpha \right) - g_m^\nu(p_m, q_m, r_m, D_{m+1}) \right)
$$

$$
= \mathbb{E}_m^\alpha \left(J^{(\hat{p},\hat{q},\hat{r})} \left(m+1, S_{m+1}^{m,s,(p,q,r)}, \alpha, \nu \right) - g_m^\nu(p_m, q_m, r_m, D_{m+1}) \right)
$$

$$
= J^{(\tilde{p},\tilde{q},\tilde{r})} (m, s, \alpha, \nu) \leq V_m^\nu(s, \alpha),
$$

where $(\tilde{p}_n, \tilde{q}_n, \tilde{r}_n)$ coincide with $(\hat{p}_n, \hat{q}_n, \hat{r}_n)$, for $n \geq m+1$, while they are equal to (p_m, q_m, r_m), for $n = m$. The proof is completed easily by plugging the dynamics of the state process, (2.1), into (A.3). □

The following corollary provides a more explicit recursive formula for the value function and optimal control. In particular, it states that the value function of an agent at any time remains linear in s, in both positive and negative half lines (with possibly different slopes).

Corollary A.2. *Assume that an admissible LOB ν has an optimal control $(\hat{p}, \hat{q}, \hat{r})$. Then, for any $(s, \alpha) \in \mathbb{S}$, the following holds $\mathbb{P}$-a.s., for all $n = 0, \ldots, N-1$:*

(1) *$V_n^\nu(s, \alpha) = s^+ \lambda_n^a(\alpha) - s^- \lambda_n^b(\alpha)$, with some adapted processes $\lambda^a(\alpha)$ and $\lambda^b(\alpha)$, such that $\lambda_N^a(\alpha) = p_N^b$ and $\lambda_N^b(\alpha) = p_N^a$;*

(2) *$p_n^a \geq \mathbb{E}_n^\alpha(\lambda_{n+1}^a(\alpha))$ and $p_n^b \leq \mathbb{E}_n^\alpha(\lambda_{n+1}^b(\alpha))$;*

(3) *if, for some $p \in \mathbb{R}$, $\mathbb{P}_n^\alpha(D_{n+1}^+(p) > \nu_n^+((-\infty, p))) > 0$, then*

$$
p \leq \mathbb{E}_n^\alpha \left(\lambda_{n+1}^b(\alpha) \mid D_{n+1}^+(p) > \nu_n^+((-\infty, p)) \right);
$$

(4) *if, for some $p \in \mathbb{R}$, $\mathbb{P}_n^\alpha(D_{n+1}^-(p) > \nu_n^-((p, \infty))) > 0$, then*

$$
p \geq \mathbb{E}_n^\alpha \left(\lambda_{n+1}^a(\alpha) \mid D_{n+1}^-(p) > \nu_n^-((p, \infty)) \right);
$$

(5) *for all $s > 0$,*

- *$\lambda_n^a(\alpha) = \max\{p_n^b, \mathbb{E}_n^\alpha \lambda_{n+1}^a(\alpha) + (\sup_{p \in \mathbb{R}} \mathbb{E}_n^\alpha((p - \lambda_{n+1}^a(\alpha))\mathbf{1}_{\{D_{n+1}^+(p) > \nu_n^+((-\infty, p))\}}))^+\}$,*
- *if $\hat{q}_n(s, \alpha) \neq 0$ and $\hat{r}_n(s, \alpha) = 0$, then*

$$
\lambda_n^a(\alpha) = \mathbb{E}_n^\alpha \lambda_{n+1}^a(\alpha) + \sup_{p \in \mathbb{R}} \mathbb{E}_n^\alpha \left((p - \lambda_{n+1}^a(\alpha)) \mathbf{1}_{\left\{ D_{n+1}^+(p) > \nu_n^+((-\infty, p)) \right\}} \right),
$$

  and $p = \hat{p}_n(s, \alpha)$ attains the above supremum,
- *if $\hat{q}_n(s, \alpha) = 0$ and $\hat{r}_n(s, \alpha) = 0$, then $\lambda_n^a(\alpha) = \mathbb{E}_n^\alpha \lambda_{n+1}^a(\alpha)$,*
- *if $\hat{r}_n(s, \alpha) = 1$, then $\lambda_n^a(\alpha) = p_n^b$;*

(6) *for all $s < 0$,*

- *$\lambda_n^b(\alpha) = \min\{p_n^a, \mathbb{E}_n^\alpha \lambda_{n+1}^b(\alpha) - (\sup_{p \in \mathbb{R}} \mathbb{E}_n^\alpha((\lambda_{n+1}^b(\alpha) - p)\mathbf{1}_{\{D_n^-(p) > \nu_{n-1}^-((p, \infty))\}}))^+\}$,*

- *if $\hat{q}_n(s, \alpha) \neq 0$ and $\hat{r}_n(s, \alpha) = 0$, then*

$$\lambda_n^b(\alpha) = \mathbb{E}_n^\alpha \lambda_{n+1}^b(\alpha) - \sup_{p \in \mathbb{R}} \mathbb{E}_n^\alpha \left(\left(\lambda_{n+1}^b(\alpha) - p \right) \mathbf{1}_{\left\{ D_n^-(p) > v_{n-1}^-((p, \infty)) \right\}} \right),$$

  and $p = \hat{p}_n(s, \alpha)$ attains the above supremum,
- *if $\hat{q}_n(s, \alpha) = 0$ and $\hat{r}_n(s, \alpha) = 0$, then $\lambda_n^b(\alpha) = \mathbb{E}_n^\alpha \lambda_{n+1}^b(\alpha)$,*
- *if $\hat{r}_n(s, \alpha) = 1$, then $\lambda_n^b(\alpha) = p_n^a$.*

Proof. Let us plug the piecewise-linear form of the value function into (A.2)

$$V_n^v(s, \alpha) = \operatorname{esssup}_{p,q,r} \left\{ \mathbf{1}_{\{r_n=0\}} \left(s^+ \mathbb{E}_n^\alpha \lambda_{n+1}^a(\alpha) - s^- \mathbb{E}_n^\alpha \lambda_{n+1}^b(\alpha) \right. \right.$$

$$+ \mathbb{E}_n^\alpha \left(\left(q_n p_n + (s - q_n)^+ \lambda_{n+1}^a(\alpha) - (s - q_n)^- \lambda_{n+1}^b(\alpha) - s^+ \lambda_{n+1}^a(\alpha) + s^- \lambda_{n+1}^b(\alpha) \right) \right.$$

$$\left. \left. \left. \times \left(\mathbf{1}_{\left\{ q_n \geq 0, D_{n+1}^+(p_n) > v_n^+((-\infty, p_n)) \right\}} + \mathbf{1}_{\left\{ q_n < 0, D_{n+1}^-(p_n) > v_n^-((p_n, \infty)) \right\}} \right) \right) \right)$$

$$+ \mathbf{1}_{\{r=1\}} \left(q_n^+ p_n^b - q_n^- p_n^a + (s - q_n)^+ \mathbb{E}_n^\alpha \lambda_{n+1}^a(\alpha) - (s - q_n)^- \mathbb{E}_n^\alpha \lambda_{n+1}^b(\alpha) \right) \right\}.$$

First, notice that it suffices to consider the essential supremum over all random variables (p_n, q_n, r_n).[12] Moreover, the essential supremum can be replaced by the supremum over all deterministic $(p_n, q_n, r_n) \in \mathbb{R}^2 \times \{0, 1\}$. To see the latter, it suffices to assume that the supremum is not attained by the optimal strategy (with positive probability), and construct a superior strategy via the standard measurable selection argument (cf. corollary 18.27 and theorem 18.26 in Aliprantis & Border, 2006), which results in a contradiction. It is easy to see that for any fixed (p_n, s, r_n), the above function is piecewise linear in q_n, with the slope changing at $q_n = 0$ and $q_n = s$. Hence, for a finite maximum to exists, the slope of this function must be nonnegative, at $q_n \to -\infty$, and nonpositive, at $q_n \to \infty$. This must hold for any (p_n, r_n, s), to ensure that the value function of an agent is finite; otherwise, an agent can scale up her position to increase the value function arbitrarily. Considering $r_n = 1$, we obtain condition 2 of the corollary. The case $r_n = 0$ yields conditions 3 and 4. Notice also that the maximum of the aforementioned function is always attained at $q_n = 0$ or $q_n = s$. Considering all possible cases: $r_n = 0, 1$, $q_n = 0, s$, $s = 0$, $s > 0$, and $s < 0$—we obtain the recursive formulas for λ_n^a and λ_n^b (i.e., conditions 5 and 6 of the corollary). In addition, as the optimal q_n takes values 0 and s, it is easy to see that the piecewise linear structure of the value function in s is propagated backward, and, hence, condition 1 of the corollary holds. □

It is also useful to have a converse statement.

Corollary A.3. *Consider an admissible LOB v and admissible control $(\hat{p}, \hat{q}, \hat{r})$, such that $\hat{q}_n(s, \alpha) \in \{0, s\}$. Assume that, for any $\alpha \in \mathbb{A}$ and any $n = 0, \dots, N$, there exists a progressively measurable random function $V_n^v(\cdot, \alpha)$, such that, for any $s \in \mathbb{R}$, $\mathbb{P}$-a.s., $(\hat{p}, \hat{q}, \hat{r}, V^v)$ satisfy the conditions 1–6 of Corollary A.2. Then, $(\hat{p}, \hat{q}, \hat{r})$ is an optimal control for the LOB v.*

Proof. It suffices to revert the arguments in the proof of Corollary A.2, and recall that $\hat{q}$ can always be chosen to be equal to 0 or s, without compromising the optimality. □

APPENDIX B

Proof of Lemma 5.1. The following lemma shows that the normalized price increments are close to Gaussian in the conditional $\mathbb{L}^2$ norm.

Lemma B.1. *Let Assumptions 4.1–4.4 hold. Then, there exists a deterministic function $\epsilon(\cdot) \geq 0$, such that $\epsilon(\Delta t) \to 0$, as $\Delta t \to 0$, and, $\mathbb{P}$-a.s., for all $\alpha \in \mathbb{A}$ and all $n = 1, \dots, N$, we have*

$$\mathbb{E}_{n-1}^{\alpha} \left(\left(\xi_n/\sqrt{\Delta t} - \sigma_{t_{n-1}}(W_{t_n}^{\alpha} - W_{t_{n-1}}^{\alpha})/\sqrt{\Delta t} \right)^2 \right) \leq \epsilon(\Delta t).$$

Proof. Notice: $\xi_n/\sqrt{\Delta t} - \sigma_{t_{n-1}}(W_{t_n}^{\alpha} - W_{t_{n-1}}^{\alpha})/\sqrt{\Delta t} = \frac{1}{\sqrt{\Delta t}} \int_{t_{n-1}}^{t_n} \mu_s^{\alpha} ds + \frac{1}{\sqrt{\Delta t}} \int_{t_{n-1}}^{t_n} (\sigma_s - \sigma_{t_{n-1}}) dW_s^{\alpha}.$

Then, using Assumptions 4.2, 4.4, and Itô's isometry, we obtain the statement of the lemma. ☐

The next lemma connects the proximity in terms of $\mathbb{L}^2$ norm and the proximity of expectations of certain functions of random variables. This result would follow trivially from the classical theory, but in the present case, we require additional uniformity—hence, a separate lemma is needed (whose proof is, nevertheless, quite simple).

Lemma B.2. *For any constant $C > 1$, there exists a deterministic function $\gamma(\cdot) \geq 0$, s.t. $\gamma(\varepsilon) \to 0$, as $\varepsilon \to 0$, and, for any $\varepsilon > 0$, $\sigma \in [1/C, C]$, and any random variables $\eta \sim \mathcal{N}(0, \sigma^2)$ and ξ (the latter is not necessarily Gaussian), satisfying $\mathbb{E}(\xi - \eta)^2 \leq \varepsilon$, the following holds for all $p \in \mathbb{R}$*

 (i) $(|p| \vee 1) |\mathbb{P}(\xi > p) - \mathbb{P}(\eta > p)| \leq \gamma(\varepsilon)$,
(ii) $|\mathbb{E}(\xi \mathbf{1}_{\{\xi > p\}}) - \mathbb{E}(\eta \mathbf{1}_{\{\eta > p\}})| \leq \gamma(\varepsilon)$.

Proof. (ii) Note that

$$|\mathbb{E}(\xi \mathbf{1}_{\{\xi > p\}}) - \mathbb{E}(\eta \mathbf{1}_{\{\eta > p\}})| \leq |\mathbb{E}((\xi - \eta) \mathbf{1}_{\{\xi > p\}})| + |\mathbb{E}(\eta (\mathbf{1}_{\{\xi > p\}} - \mathbf{1}_{\{\eta > p\}}))|$$

$$\leq \sqrt{\varepsilon} + \|\eta\|_2 \sqrt{\mathbb{P}(\xi > p, \eta \leq p) + \mathbb{P}(\xi \leq p, \eta > p)},$$

and

$$\mathbb{P}(\xi > p, \eta \leq p) \leq \mathbb{P}(p \geq \eta \geq p - \sqrt[3]{\varepsilon}) + \mathbb{P}(|\xi - \eta| > \sqrt[3]{\varepsilon}) \leq M \sqrt[3]{\varepsilon} + \frac{\mathbb{E}(\xi - \eta)^2}{(\sqrt[3]{\varepsilon})^2} \leq (M + 1) \sqrt[3]{\varepsilon},$$

where we used the fact that η has a density bounded by a fixed constant M. We can similarly show that $\mathbb{P}[\xi \leq p, \eta > p] \leq (M + 1) \sqrt[3]{\varepsilon}$. The resulting estimates yield the statement of the lemma. ☐

Taking $\epsilon(\Delta t) = \gamma(\epsilon(\Delta t))$ and applying the above lemmas, we get the statement of Lemma 5.1, with $(W_{t_n}^{\alpha} - W_{t_{n-1}}^{\alpha})/\sqrt{\Delta t}$ in place of η_0. Finally, we note that the laws of the two random variables coincide under $\mathbb{P}_{n-1}^{\alpha}$, and the statement depends only on these laws. The last statement of Lemma 5.1 follows from the fact that Lemma B.2 is stable under analogous substitution.