Forthcoming Review of Finance

Large Orders in Small Markets: Execution with Endogenous Liquidity Supply*

Agostino Capponi† Albert J. Menkveld‡ Hongzhong Zhang§

May 20, 2024

Abstract

We model the execution of a large uninformed sell order in the presence of strategic competitive market makers. We solve for the unique symmetric equilibrium of the model in closed form. Analysis of this equilibrium reveals that large orders unequivocally benefit market makers, while smaller investors stand to benefit only if the order trades with a sufficiently high intensity. The equilibrium results further provide a rationale for the empirically observed patterns of (i) shorter orders trading at higher intensities, and (ii) price pressures potentially subsiding before large orders stop executing.

*We thank Humoud Alsabah, Jonathan Brogaard, Thierry Foucault, Terrence Hendershott, Burton Hollifield, Joshua Mollner, Elmar Nijkamp, Matthew Pritsker, Lucas Saru, Rudi Schadt, Duane Seppi, Ekaterina Serikova, Chester Spatt, and Shihao Yu as well as seminar/conference participants at Bank of America, Carnegie Mellon University, Cass Business School, Memphis University, Florida State University, University of Cyprus, Cambridge University, Humboldt University, NYU Shanghai, EPFL, SUFE Shanghai, Sydney University, UC Berkeley, the Market Microstructure Confronting Many Viewpoints Conference, the Microstructure Exchange online seminar series, the American Finance Association Conference, the 3rd Eastern Conference in Mathematical Finance, the Australian Banking and Finance Conference, and the Utah Winter Finance Conference for helpful comments. Menkveld gratefully acknowledges Netherlands Organisation for Scientific Research (NWO) for a Vici grant. The research of Capponi and Zhang was supported by a grant from the US National Science Foundation (DMS-1716145).

†Columbia University, New York, ac3827@columbia.edu.

‡Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, Netherlands, albertjmenkveld@gmail.com, and Tinbergen Institute.

§Columbia University, New York, hongzhong.z1728@gmail.com.

1. Introduction

The past few decades have seen a gradual rise in intermediated investment in securities markets. For instance, French (2008, Table 1) highlights a notable trend wherein direct holdings of U.S. equity decreased significantly from 47.9% in 1980 to 21.5% in 2007. The Economist reports an even more dramatic trend with individual-investor ownership of U.S. equity declining from 90% in 1950 to 30% in 2014.[1] All this has given rise to large institutional investors for whom asset re-allocations trigger liquidity demands that are large relative to the markets they trade in. In response to concerns regarding liquidity risk management, regulators have undertaken several initiatives of late.[2] This paper focuses on studying liquidity within this particular context.

We investigate the execution of large uninformed sell orders in the presence of strategic competitive market makers (à la Cournot). Other (small) investors are not strategic: they arrive, potentially trade, and immediately leave forever. In the model, we assume the large seller trades at constant intensity. This mirrors the so-called percentage of volume (POV) trading algorithm, one of the most common trade execution strategies.[3]

The model yields a tractable unique symmetric equilibrium with closed-form equilibrium expressions that allow us to answer the following questions:

1. Should regulators worry about large orders? In other words, does the presence of large orders benefit or hurt market makers and/or other (small) investors?

2. Can we rationalize empirically observed patterns of order execution? Data suggest that shorter orders trade at higher intensities (see Brough 2010, Table 2.6). Moreover, price pressure may vanish even before order execution ends (Zarinelli et al. 2015, Figure 8).

[1] See "Reinventing the Deal", The Economist, Oct 24, 2015.

[2] In 2016 the Securities and Exchange Commission (SEC) proposed Rule 22e-4 that requires open-end funds to disclose the liquidity risk associated with their holdings (SEC 2016). It triggered significant pushback from the industry causing delays (see "SEC Tackles Fund Liquidity Complexity with Rule Proposal, Delay, New Guidance," Reuters, March 27, 2018). The European Securities and Markets Authority (ESMA) stopped short of any regulatory steps but instead published guidelines on liquidity stress testing (ESMA 2019).

[3] See, for instance, `https://medium.datadriveninvestor.com/the-types-of-automated-trading-algorithms-228d537254a8`, which lists POV among the top three execution algorithms, and (BIS–Market Committee 2020) which lists POV among the six most common execution algorithms in the FX market. In Remark 1 we show that our setting mirrors POV.

Our analysis contributes to a large and well established theoretical literature on market liquidity in scenarios where information is symmetric across agents. Examples of articles surveying the vast literature on liquidity, including asymmetric-information models, are Madhavan (2000), Vayanos and Wang (2013), and Amihud, Mendelson, and Pedersen (2005). This literature may be structured as follows. One set of models studies best execution in the presence of exogenous liquidity supply, and endogenizes liquidity demand.[4]. Another set of models does the opposite. It assumes liquidity demand to be exogenous and endogenizes liquidity supply.[5] A third set of models considers demand and supply jointly.[6]

Our contribution relative to this literature is to focus on dynamic liquidity supply to a sliced-and-diced large order. The model yields tractable expressions for the endogenous price response to such an order, including a bid-ask spread, depth at the best quotes, and price pressure (i.e., the wedge between the midquote and fundamental value). The tractability of our model enables us to address the two questions mentioned above that, to the best of our knowledge, have not been explored thus far. Particularly novel is our welfare analysis on how the presence of a large seller affects the welfare of the various types of market participants.

Our endogenous price-impact results contribute to the understanding of liquidity by industry professionals, regulators, and retail investors. Institutional investors, in particular, can benefit from our insights. For instance, we demonstrate that each order size necessitates a minimum duration for the market to absorb the order efficiently. If an order of the same size is executed within a shorter period, selling *fewer* shares would result in higher *total* proceeds.[7] Moreover, for any duration beyond the minimum duration, the equilibrium price impact can be computed analytically. The availability of closed-form expressions allows us to perform sensitivity analysis with respect to changing market conditions (e.g., fewer market makers or higher funding cost). This is particularly relevant for complying with liquidity

[4]See, e.g., Bertsimas and Lo (1998), Almgren and Chriss (2001), Huberman and Stanzl (2005), Obizhaeva and Wang (2013), Gatheral and Scheid (2011), Boulatov, Bernhardt, and Larionov (2016), or van Kervel, Kwan, and Westerholm (2018)

[5]See, e.g., Amihud and Mendelson (1980), Ho and Stoll (1981), Aït-Sahalia and Saglam (2024), Grossman and Miller (1988), Weill (2007), Hendershott and Menkveld (2014), Liu and Wang (2016), Gayduk and Nadtochiy (2018), or Bank, Ekren, and Muhle-Karbe (2021).

[6]See, e.g., Vayanos (2001), Goettler, Parlour, and Rajan (2005), Pritsker (2009), Rostek and Weretka (2015), Gabaix et al. (2006), Choi, Larsen, and Seppi (2018), Choi, Larsen, and Seppi (2019), or Fardeau (2019).

[7]This is equivalent to the standard monopolist case where rationing supply can generate larger total profit.

management regulation alluded to earlier (footnote 2). Relatedly, our findings are likely to be of interest to those who offer optimal-execution services such as Goldman Sachs (GSET), BlackRock (Aladdin), or ITG (ACE).[8]

Results. An important baseline result is that the model yields a unique and tractable symmetric equilibrium. Examining this equilibrium provides us with conclusive responses to the key questions formulated in the introduction.

First, we compare the market with and without a large seller to assess whether his presence benefits or hurts others. Market makers unambiguously benefit from the presence of a large seller, but other investors benefit only if he sells at a high enough intensity. The trade-off for these investors is that, on the one hand, they benefit from temporarily depressed prices because they become net buyers in this period. On the other hand, the bid-ask spread is wider while the large seller executes and this hurts them. We show that the former channel dominates the latter, if the large seller sells at a high enough intensity. This would interest regulators who tend to worry about large temporary price "distortions." Our analysis shows that such distortions could be the healthy outcome of a participant paying the price for immediacy that others happily provide (including participants other than the market makers). The analysis, therefore, provide regulators with a comprehensive analysis of market quality, including *all* market participants.

Second, the equilibrium price paths are such that price pressure can subside *before* execution ends (i.e., prices start reverting back to fundamental value before execution ends). This is somewhat surprising since the large seller sells at constant intensity (by assumption, mirroring POV execution). The model can thus endogenously generate this price pattern which has been observed in the data by Zarinelli et al. (2015, Fig. 8). Moreover, it can generate a negative correlation between order duration and execution intensity, for which we provide empirical evidence in Section 5.

Our theoretical results offer a perspective on a large and growing literature on institutional trading. For example, Hu et al. (2018) review the literature based on Abel Noser data

[8]As a matter of fact, ACE is based on an (undisclosed) structural model. Investment Technology Group (ITG) (2007) explicitly advertises its advantage by stating: "The ITG ACE model is not a purely econometric model calibrated based on transaction cost data. Rather, it is a structural model that uses parameters estimated econometrically." Grinold and Kahn (1999, Ch. 15) emphasize the importance of a structural dynamic model for transactions cost in their comprehensive analysis of active portfolio management.

on large order executions of institutional investors. They observe that the size of individual institutional trades declined dramatically over time, suggesting that, indeed, institutional orders are sliced and diced. Puckett and Yan (2011) and Anand et al. (2012) use this data to show that trading skill matters. van Kervel and Menkveld (2019) use more recent data to show that, indeed, for the vast majority of orders high-frequency market makers take the other side when a large institutional investor executes his sliced-and-diced order.

The structure of the paper is as follows. Section 2 presents and motivates the model. Section 3 derives the unique symmetric equilibrium. Section 4 utilizes the tractable equilibrium expressions to address the two main questions. Section 5 considers robustness of these results relative to the model perturbation of making large-order duration completely unknown to market makers. Section 6 calibrates the model, and then shows how it can be used to generate a "liquidity surface." Section 7 concludes. Proofs are delegated to Appendix C.

2. Model

The model primitives capture a setting where a large investor sells at constant intensity for a particular period of time. (The case of a large buyer is completely symmetric.[9]) Other investors arrive continuously and are "small" in the sense that they arrive, trade, and leave forever. The amount they trade depends on the bid and ask prices prevailing in the market. The net order flow is absorbed by a set of homogeneous market makers who compete à la Cournot. They start off with zero inventory. Bid and ask price trajectories are such that the market clears at each point in time. The large investor is large in the sense that he, unlike others, requires an *interval* to trade (i.e., the large investor slices and dices his order whereas others trade in one go).

Before describing the model primitives in detail, it is useful to illustrate the model graphically. Figure 1 provides such illustration (informed by the model calibration presented in Appendix 6). The top graph depicts the net flows by investors. The large investor sells his order at constant intensity until time d (d for duration, red line). Others either buy or sell at a particular intensity. This intensity depends on the relative bid and ask prices (i.e., bid and

[9]We assume that the cost of short sales for market makers is negligible in the case of a large buyer. If this is not true, then, as a first pass, one might simply want to raise inventory costs to reflect costly short sales. However, a precise analysis of how costly short sales affect the equilibrium is out of scope here.

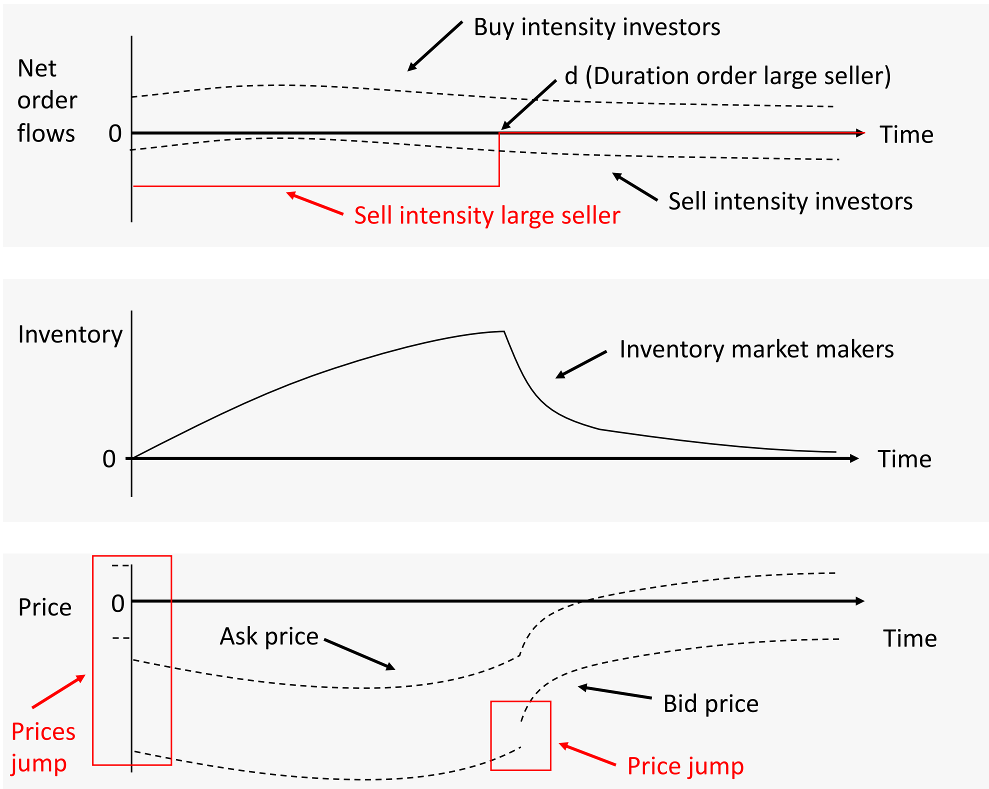


Figure 1: **Model visualization.** This figure illustrates the model by sketching the equilibrium trajectories for key variables. The top graph plots the (continuous) net flow of the large seller as well as the flow of the continuously arriving buy and sell investors. The middle graph plots the inventory of market makers, which results from the net flows in the top graph. The bottom graph plots the bid and ask prices relative to fundamental value.

ask prices minus fundamental value). Buyers, for example, buy more if relative bid prices are more negative. We will refer to these relative prices simply as prices in the remainder of the manuscript for ease of discussion. We will only add "relative" if emphasis is required.

The middle graph illustrates how aggregate inventory of market makers evolves with time. As markets need to clear, their inventory equals the cumulative net flow from the top graph, summed across all investors. Notice how the inventory of market makers increases while the large seller is executing, but mean-reverts quickly to zero after the large seller leaves the market.

The bottom graph illustrates the (relative) bid and ask price trajectories. The bid and ask straddle zero symmetrically just before the large seller starts executing. They jump down the moment the large seller enters which, as we will describe below, is the point at which market makers learn about the presence of the large seller. They observe both his sell intensity and the duration of his order, and respond strategically. Notice further that the bid-ask spread widens while the large seller is executing, seemingly driven by bids adjusting more to his presence than asks.

The remainder of this section describes the model primitives in full detail. To keep track of all notation used, we provide a summary in Appendix A. The model is set in continuous time, features a single security, and three types of agents: a large seller, N market makers, and continuously arriving other investors who are either buyers or sellers.

Security. The security's fundamental (common) value follows a Brownian motion with volatility $\sigma > 0$:

$$\mathrm{d}m_t = \sigma \mathrm{d}Z_t. \tag{1}$$

The fundamental value is common knowledge to all agents in the model.

Investors (other than the large seller). Investors other than the large seller are small. They arrive, trade, and leave forever. The sense in which they are small is that they can trade all they want instantly. The amount they trade, however, does depend on the bid and ask prices they encounter in the market. Together these investors are modeled as a flow of price-sensitive liquidity demand. The flow of buy demand is:

$$q_t^a \mathrm{d}t = \delta \left((m_t + \omega) - (m_t + p_t^a) \right) \mathrm{d}t = \delta \left(\omega - p_t^a \right) \mathrm{d}t, \tag{2}$$

where p_t^a is the (relative) ask quote. The parameter δ captures the level of trade activity for the security, and ω represents its (relative) reservation value to buyers.

The flow of sell demand is modeled analogously:

$$q_t^b \mathrm{d}t = \delta \left(p_t^b + \omega \right) \mathrm{d}t, \tag{3}$$

where p_t^b is the (relative) bid quote, and $-\omega$ is the (relative) reservation value of sellers.

Large seller. The larger seller arrives at time 0 and leaves at time d. He is not strategic. He sells at constant intensity f and his total order size therefore is: $S = f \times d$. Similar to other investors who sell, his (relative) reservation value for the security is $-\omega$.[10] His surplus from selling discounted to time 0 is:

$$P_d(f) = \int_0^d e^{-\beta t} f \left(\omega + p_t^b\right) \mathrm{d}t, \tag{4}$$

where $\beta > 0$ is the intertemporal discount rate. We do not endow the large seller with an objective function. Instead, we map out all his options and what they mean in terms of price impact (see, e.g., the liquidity surface in Section 6.2).

Market makers. Market makers are Cournot competitors who are in the market continuously. At each instant of time, they decide how much liquidity to supply: how many securities to sell at the ask price and how many to buy at the bid. Denote by $q_{jt}^a = q_{jt}^a(p_t^a)$ the sell intensity of market maker j at the ask price p_t^a at time t. The total sell intensity by market makers at that price therefore equals $\sum_{j=1}^N q_{jt}^a$. Market clearing requires that the intensity of buy investors in Eq. (2) equals the sell intensity of market makers at the ask price:

$$q_t^a = \sum_{j=1}^N q_{jt}^a. \tag{5}$$

Combining this market-clearing condition with Eq. (2) yields an expression for the ask price in terms of the sell intensities of market makers:

$$p_t^a = \omega - \frac{1}{\delta} \sum_{j=1}^N q_{jt}^a. \tag{6}$$

Following the same steps for the other side of the market yields the following expression for the bid price:

$$p_t^b = \frac{1}{\delta} \left(\sum_{j=1}^N q_{jt}^b - f 1_{\{t<d\}} \right) - \omega, \tag{7}$$

[10] Other than maintaining some homogeneity across investors, this assumption avoids the potential trivial solution of market makers setting the bid at the large-seller's reservation value in case it is strictly lower than $-\omega$, thus benefiting from his demand inelasticity.

where 1_A is the indicator function which equals one if A holds and zero otherwise. Observe that the difference between Eq. (6) and Eq. (7) is the (inelastic) sell intensity of the large seller for the period when he is executing.[11]

Note that the pricing equations, Eq. (6) and Eq. (7), relate market-clearing prices to the buy and sell intensities set by market makers. The inventory state of market maker j evolves as

$$\mathrm{d}i_{jt} = -q^a_{jt}\mathrm{d}t + q^b_{jt}\mathrm{d}t. \tag{8}$$

We can now define his wealth dynamics as follows:

$$\mathrm{d}W_{jt} = \left(\omega - \frac{1}{\delta}\sum_{n=1}^{N} q^a_{nt}\right) q^a_{jt}\mathrm{d}t - \left(-\omega + \frac{1}{\delta}\left(\sum_{n=1}^{N} q^b_{nt} - f1_{\{t<d\}}\right)\right) q^b_{jt}\mathrm{d}t. \tag{9}$$

The disutility of non-zero inventory to a market maker enters his objective function as a pecuniary cost that scales with inventory squared (similar to, for example, Madhavan and Smidt 1993, Hendershott and Menkveld 2014, Duffie and Antill 2021).[12] With all these ingredients, the value function of market maker j can now be defined explicitly as:

$$v_j^M\left(i_{jt}, f, d, t\right) = \sup_{q^j|q^{-j}} \int_t^\infty e^{-\beta(s-t)}\left(\mathrm{d}W_{js} - \eta i^2_{js}\mathrm{d}s\right), \tag{10}$$

where q^j_t is shorthand for $\left(q^a_{jt}, q^b_{jt}\right)$ which contains the sell intensity (at the ask) and the buy intensity (at the bid), q^{-j} is a vector with all market maker buy and sell intensities *except* market maker j, and η parameterizes the flow cost associated with non-zero inventory. This cost is increasing in a security's fundamental volatility (σ), in market-maker risk-aversion, or in funding cost.

We can now define the Cournot equilibrium formally as:

[11] Note that the equivalent of Eq. (5) here is

$$f1_{\{t<d\}} + q^b_t = \sum_{j=1}^{N} q^b_{jt}.$$

[12] Almgren and Chriss (2001) develop a model of inventory costs for financial firms that is widely used in the industry and in academic studies. They assume that the rate of inventory costs for trader j at time t is ηi^2_{jt} for a fixed coefficient η.

Definition 1. *A Cournot equilibrium is a strategy profile* $q := (q^1, \ldots, q^N)$ *where each* q^j *solves* (10) *taking* q^{-j} *as given.*

We focus on symmetric strategy profiles only (following standard practice, e.g., Brunnermeier and Pedersen 2005). Note that once equilibrium buy and sell intensities are solved for, equilibrium prices follow immediately by plugging the intensities into Eq. (6) and Eq. (7).

3. Model Equilibrium

This section derives the equilibrium for the baseline case without a large seller and for the case with a large seller. In our equilibrium analysis, we impose as a constraint that traders have the same quadratic value function and employ the same strategy. This yields equal inventory trajectories at all times for all market makers.

Our definition of equilibrium can be rationalized by the assumption that if a market maker considers a different strategy, he expects the other market makers to do the same. Therefore, when calculating the potential benefits of deviating, he does so believing others will deviate in the same way. We limit our attention to this type of symmetric equilibria. Note that our definition is nonstandard, because we do not consider unilateral deviations from the equilibrium strategy in the sense that other traders stick to the equilibrium strategy. An important benefit of our approach is that that it yields clean analytic results.[13]

3.1 Baseline Case: No Large Seller

Proposition 1 (Baseline case: No large seller)**.** *The baseline case without a large seller has a unique symmetric equilibrium. The equilibrium buy and sell intensities are time invariant and depend only on a market maker's inventory state* i*. Note that total inventory summed across all market makers is* Ni *due to symmetry. The (relative) bid and ask prices are:*

$$p^b(i) = -A_\theta - B_\theta N i \,, \qquad \textit{(Bid price)} \qquad (11)$$

$$p^a(i) = A_\theta - B_\theta N i \,. \qquad \textit{(Ask price)} \qquad (12)$$

[13]In Appendix D we explore unilateral deviations in the case of two players. The equilibrium expressions become more involved but show qualitatively similar results with somewhat less aggressive strategies. The intuition is that unilateral deviations are less costly to players, because they do not get followed by others.

The implied bid-ask spread therefore is:

$$s(i) = p^a(i) - p^b(i) = 2A_\theta \qquad \textit{(Bid-ask spread)} \tag{13}$$

and the implied conditional price pressure B_θ (defined as midquote skew per unit of inventory, Hendershott and Menkveld 2014) is:

$$\frac{\left(\frac{p^a(i)+p^b(i)}{2}\right)}{Ni} = B_\theta. \qquad \textit{(Conditional price pressure)} \tag{14}$$

Each market maker chooses the following buy and sell intensity:

$$q^b(i) = C_\theta - D_\theta i, \qquad \textit{(Bid intensity)} \tag{15}$$

$$q^a(i) = C_\theta + D_\theta i. \qquad \textit{(Ask intensity)} \tag{16}$$

The value function for a market maker is:

$$v^M(i) = E_\theta - F_\theta i^2. \qquad \textit{(Value to market maker)} \tag{17}$$

If market makers start with a zero inventory (i.e., $i_0 = 0$), then the inventory path is:

$$i_t = 0 \quad \forall t > 0. \qquad \textit{(Market maker inventory)} \tag{18}$$

The constants $A_\theta, \ldots, F_\theta > 0$ are closed-form expressions in the model parameters, whose expressions are specified in Eq. (A14)*.*

We discuss the economic insights that follow from Proposition 1 in two stages. We first consider the special case of perfect competition for which the equilibrium expressions become particularly friendly. We then present comparative statics by signing the partial derivatives for various equilibrium variables, such as the bid-ask spread and conditional price pressure.

Perfect competition. The case of perfect competition is obtained by taking the number of market makers to infinity as is done in the following corollary:

Corollary 1 (Perfect competition)**.** *The liquidity supplied by market makers in the case of* *perfect competition (i.e., $N \to \infty$) can be characterized as follows. The bid-ask spread is:*

$$\lim_{N\to\infty} p^a(i) - p^b(i) = \lim_{N\to\infty} 2A_\theta = 0, \qquad \textit{(Bid-ask spread)} \tag{19}$$

the conditional price pressure is:

$$\lim_{N\to\infty} B_\theta = 2\frac{\eta}{\beta}, \qquad \textit{(Conditional price pressure)} \tag{20}$$

and the aggregate trading intensities the market makers supply at the bid and the ask price, respectively, are:

$$\lim_{N\to\infty} NC_\theta + ND_\theta i = \delta\omega - 2\delta\frac{\eta}{\beta}i, \qquad \textit{(Bid depth)} \tag{21}$$

$$\lim_{N\to\infty} NC_\theta + ND_\theta i = \delta\omega + 2\delta\frac{\eta}{\beta}i. \qquad \textit{(Ask depth)} \tag{22}$$

The value function for a market maker is:

$$\lim_{N\to\infty} v^M(0) = 0. \qquad \textit{(Value single market maker)} \tag{23}$$

Corollary 1 yields several insights. First, the bid-ask spread becomes zero when infinitely many market makers enter. Second, the sum of bid and ask depth reaches its maximum value: $2NC_\theta = 2\delta\omega$.[14] Third, conditional price pressure remains strictly positive. Fourth, investors only pay or earn such price pressure if market maker inventory is non-zero. This, however, never happens if market makers start off with zero inventory.

The expression for conditional price pressure in Eq. (20) yields some additional insight. First, if holding inventory becomes more costly for market makers (i.e., higher η), then price pressure increases. This result is intuitive. Second, if market makers care less about the future (i.e., higher β) then price pressure declines. This finding is less intuitive. The reason is that β affects a fundamental trade-off underlying equilibrium price pressure. The trade-off

[14] This result follows from the following two observations. First, from Eq. (2) and Eq. (3), it follows that total depth is:

$$q_t^a + q_t^b = \delta\left(2\omega - \left(p_t^a - p_t^b\right)\right). \tag{24}$$

Second, the bid-ask spread is constrained to be non-negative. It then immediately follows that the maximum value of the right-hand side of Eq. (24) is $2\delta\omega$.

Table 1: Comparative statics equilibrium liquidity supply

This table summarizes how equilibrium liquidity supply in the baseline model changes when one of the parameters is changed. The proof for these results is provided in Appendix C. Liquidity supply is characterized by the bid-ask spread and conditional price pressure (i.e., price change per unit of aggregate inventory change). The results are obtained by signing partial derivatives.

Parameter	Bid-ask spread	Conditional price pressure
Inventory-holding cost (η)	Unchanged	+
Number of market makers (N)	−	−
Discount rate (β)	Unchanged	−
Level of trade activity (δ)	Unchanged	−
Reservation value investors (ω)	+	Unchanged

for market makers is spending more money now to mean-revert out of non-zero inventory (i.e., more price pressure) or saving this money but then staying longer on the inventory and suffer the pecuniary costs associated with it (captured by η). If the future is discounted more, then this trade-off shifts in favor of spending less now and therefore reducing price pressure. Note that this effect cushions liquidity deterioration at times of crisis when market makers much prefer a dollar now over a dollar in the future (i.e., β becomes higher).

Comparative statics. Table 1 provides comparative statics on equilibrium liquidity supply. Such supply is fully characterized by two complementary measures: the bid-ask spread and conditional price pressure (i.e., the price change on a unit of aggregate inventory change). The comparative statics serve mostly as a sanity check. The table, for example, shows that price pressure increases when the cost of holding the security is higher, liquidity demand is lower, or fewer market makers supply liquidity.

Some results in Table 1 are less trivial. First, conditional price pressure decreases when the discount rate increases. This result reflects an increased desire of market makers to save a dollar now, less price pressure, and, in return, accept a lower speed of inventory mean-reversion (see the above discussion of Eq. (20)).

Second, if investors have higher reservation values (i.e., higher ω), then the bid-ask spread increases and conditional price pressure remains unchanged. The spread result can be explained by the Cournot competition: Market makers squeeze out additional rent if demand surplus is increased by raising ω. The unchanged price pressure, on the other hand, can be understood by observing that changing ω does *not* change the sensitivity of liquidity demand to price changes. If the speed of inventory mean-reversion is unchanged, then market makers have no reason to change conditional price pressure.

3.2 Large-Seller Case

Proposition 2 (Large-seller case)**.** *Consider the baseline case and add a large seller who sells at constant rate f during the interval $[0, d]$. The symmetric equilibrium remains unique but equilibrium trading intensities are no longer time invariant. They now depend on both (individual) market-maker inventory i and time t. The relative bid and ask price are:*

$$p^b(i, f, d, t) = p^b(i) - 1_{\{t \leq d\}} f \left(G_\theta \left(d - t\right) + H_\theta\right), \qquad \text{(Bid price)} \tag{25}$$

$$p^a(i, f, d, t) = p^a(i) - 1_{\{t \leq d\}} f \left(G_\theta \left(d - t\right)\right), \qquad \text{(Ask price)} \tag{26}$$

where $p^b(i)$ and $p^a(i)$ are the bid and ask price in the baseline case of Proposition 1. $G_\theta(\tau)$ is zero for $\tau = 0$, strictly concave and strictly increasing in τ, and H_θ is a positive constant. Therefore, the bid-ask spread is:

$$s(i, f, d, t) = s(i) + 1_{\{t \leq d\}} f H_\theta, \qquad \text{(Bid-ask spread)} \tag{27}$$

where $s(i)$ is the bid-ask spread in the baseline case. The conditional price pressure is the same as in the baseline case:

$$B_\theta = \frac{1}{Ni} \frac{p^a(i) + p^b(i)}{2}. \qquad \text{(Conditional price pressure)} \tag{28}$$

Each market maker chooses the following buy and sell intensity:

$$q^b(i, f, d, t) = q^b(i) - 1_{\{t \leq d\}} f \frac{\delta}{N} \left(G_\theta \left(d - t\right) - N H_\theta\right), \qquad \text{(Bid depth)} \tag{29}$$

$$q^a(i, f, d, t) = q^a(i) + 1_{\{t \leq d\}} f \frac{\delta}{N} \left(G_\theta \left(d - t\right)\right). \qquad \text{(Ask depth)} \tag{30}$$

The value function for a market maker is:

$$v^M(i,f,d,t) = v^M(i) + 1_{\{t\leq d\}} \left(I_\theta \left(d-t, f \right) + f J_\theta \left(d-t \right) i \right), \qquad \textit{(Value)} \qquad (31)$$

where $I_\theta(\tau, f) = 0$ *for* $\tau = 0$ *or* $f = 0$ *and convex in* f*.* $J_\theta(\tau) = 0$ *for* $\tau = 0$ *and strictly decreases in* τ*.*

Proposition 2 describes how the equilibrium changes when, relative to the baseline case, a large seller arrives at $t = 0$ and sells at constant intensity f until $t = d$ (d for duration). Market makers learn all this information at the time the large seller arrives, and respond strategically. It is useful to compare the large-seller case to the baseline case. Such comparison yields the following insights.

First, the bid and ask prices are skewed downwards by an amount that is larger the further t is from duration d. This skew is captured by the term $G_\theta(t-d)$, which is an analytic function of the model parameters collected in the vector θ. While the large seller is trading, the ask price smoothly converges to the baseline ask price that prevails just after he finished trading at $t = d$ (i.e., mathematically, $G_\theta(0) = G'_\theta(0) = 0$).

The bid price, on the other hand, jumps up at $t = d$ by $H_\theta > 0$. Note that, relative to the ask price, the bid price is shaded by this additional amount H_θ while the large seller is trading. This is the result of market makers being strategic and understanding that part of the liquidity demand of sellers is price *inelastic* in this period. In other words, they extract additional rent from the fact that the large seller is trading at constant intensity f (i.e., f is price insensitive).[15]

Second, because the bid is skewed more than the ask while the large seller is executing, the bid-ask spread is wider in this period. This finding contributes to explain why Comerton-Forde et al. (2010) find that spreads were wider when the NYSE specialist was carrying larger inventories. The baseline model cannot explain this finding since, in that model, the spread does not depend on inventory (similar to Hendershott and Menkveld 2014, Section 3.2.2.). It is the presence of a price-insensitive large investor combined with imperfectly competitive market making that drives this result.

[15]Observe that it is never optimal for market makers to let the bid price drop to $-\infty$ if the large seller liquidates at a finite rate. If this were the case, the demand of the price-sensitive end investors would cause huge inventory imbalances and the resulting costs exceed trading gains.

Remark 1. POV determines the distribution of order trading intensity over the day using a target percent of traded volume. Combining equations (2) and (3), we deduce that the (gross) volume traded by the price-sensitive investors in each infinitesimal interval is $\delta\left(2\omega + p_t^b - p_t^a\right) \mathrm{d}t$. This volume is constant because the bid-ask spread $p_t^a - p_t^b$ is constant during the execution period $[0, d]$, as seen from Eq. (27). This implies that, by selling at a constant intensity f, the larger seller in our model is effectively implementing a POV strategy.

4. Results

We use the equilibrium developed in the previous section to provide answers to the two main questions of our study. Section 4.1 discusses how welfare of other market participants is affected by the presence of a large seller. We henceforth assume that market makers start with zero inventory. Section 4.2 establishes results on the equilibrium pattern of order execution.

4.1 Value of the Presence of a Large Seller to Others

In this final section, we analyze how the presence of a large seller affects the value of others: market makers and other (small) investors. The benchmark in this analysis is the baseline case where there is no large seller (i.e., we compare the equilibrium in Proposition 2 with that in Proposition 1).

Market makers Market makers unequivocally benefit from the presence of a large seller. Proposition 3 states this result formally. This finding should not be too surprising given that the presence of a large seller adds to total liquidity demand.

Proposition 3 (The value of the presence of a large seller to market makers)**.** *Market makers with zero initial inventory are better off in case there is a large seller when compared to the baseline case (without such a seller):*

$$v^M(i_0 = 0, f, d, t = 0) > v^M(i_0 = 0) \quad \textit{for} \quad f, d > 0. \tag{32}$$

Other investors. While the presence of a large seller benefits market makers unambiguously, it benefits other investors only if the seller sells at a high enough intensity. The threshold level decreases with duration. In a sense, it seems that *directionality* in liquidity demand needs to be painful enough for market makers to make them share part of the gains from trade with other investors. Directionality becomes painful to them either if intensity is extremely high or duration is very long.

To state this result formally, we first need to define the value of trading to investors other than the large seller. A natural definition is to set it equal to the expected realized surplus of all investors, discounted appropriately, assuming that market makers start with zero inventory. Mathematically, this definition is:

$$v^I = \int_0^\infty e^{-\beta s} \left(\frac{1}{2} \delta \left(\omega - p_s^a \right)^2 + \frac{1}{2} \delta \left(\omega + p_s^b \right)^2 \right) \mathrm{d}s, \tag{33}$$

with $i_0 = 0$.

Proposition 4 (Value of large-seller presence to other investors)**.** *If market makers start with zero inventory, then the realized surplus of investors other than the large seller, is convex in large-seller trade intensity f. There exists a threshold $\bar{f}(d)$ such that investors are*

- *strictly worse for $f < \bar{f}(d)$,*
- *equally well off when $f = \bar{f}(d)$, and*
- *strictly better off when $f > f(\bar{d})$.*

The threshold level $\bar{f}(d)$ strictly decreases in duration d.

Proposition 4 states formally that the presence of a large seller can hurt or benefit other investors. Fixing d, the proposition says it only benefits investors if f is high enough. This result reflects the interplay of two opposing forces. First, the presence of a large seller temporarily increases the bid-ask spread which hurts other investors, because they pay the spread. The bid-ask spread increases linearly in f (see Proposition 2).

Second, the presence of a large seller causes prices to be pressured downward which benefits other investors because they are net buyers in the period when market makers carry positive inventory. Note that *conditional* price pressure (i.e., price pressure per unit of inventory) is unaffected by the presence of large seller. It is equal to B_θ in both the baseline model and the large-seller model (see Proposition 1 and 2, respectively). Yet, more intense selling implies that market makers carry larger inventories which, in turn, causes prices to become more pressured. This effect is super-linear given that increasing sell intensity adds to price pressure contemporaneously, but also in the future given that inventory needs to mean-revert.

Taken together, these two observations on how the presence of a large seller affects other investors explains the threshold result on sell intensity. The cost of a higher spread is linear in sell intensity, while the benefit of stronger price pressure is super-linear in sell intensity. It is therefore not surprising that there is a threshold intensity level above which the benefit dominates the cost.

4.2 Equilibrium Pattern of Order Execution

The model equilibrium implies two key execution patterns that match documented empirical patterns. These are developed in the next two subsections.

4.2.1 Order Duration and Execution Intensity

The following proposition shows how our model can reproduce the negative relation between order duration and execution intensity documented above.

Proposition 5 (Optimal liquidation intensity at a given duration)**.** *For every duration $d > 0$, there exists an optimal intensity $f^\star > 0$ such that for any $f \neq f^\star$, we have $P_d(f^\star) > P_d(f)$. This maximum intensity $f^\star$ is decreasing in d.*

Proposition 5 essentially says that there is a saturation point for each duration. That is, if one trades at an intensity higher than this saturation point, then this reduces *total* proceeds are reduced. In other words, for each duration, there is a maximum size for large orders (which is equal to the expression $f^\star \times d$ in the proposition).

4.2.2 Price Pressure Can Subside Before Execution Ends

The equilibrium result developed in Section 3.2 yields tractable expressions for the path of market-maker inventory and, relatedly, the path of bid and ask prices. These are characterized in the following lemmas.

Lemma 1 (Market-maker inventory path)**.** *If market makers start with zero inventory, then their inventory*

- *strictly increases in time while the large seller is executing,*
- *is homogeneous of degree one in the large seller's trade intensity, and*
- *exhibits exponential decay towards zero after the seller completed the execution of his order.*

Lemma 1 establishes that, if market makers start with zero inventory, they accumulate inventory while the large seller is executing his order. When he is done, inventory decays back to zero exponentially. Interestingly, if the large seller increases his trade intensity then inventory increases proportionally (i.e., homogeneity of degree one). These results for inventory paths can now be used to characterize price paths since these are linear in inventory. This leads to the following lemma.

Lemma 2 (Equilibrium bid and ask price path)**.** *If the market makers start with zero inventory, then bid and ask price for* $t \leq d$*:*

- *jump down at* $t = 0$*,*
- *are convex in* t*,*
- *strictly decline initially,*
- *but potentially increase in* t *provided* d *is large enough.*

For $t > d$ *the price paths*

- *are concave in* t *and*
- *strictly increase in* t*.*

At time $t = d$*, only the bid price jumps up.*

The characterization of equilibrium price paths in Lemma 2 yields the following insights. First, if market makers start with zero inventory (i.e., the steady state in the baseline case), then (relative) prices straddle symmetrically around zero, but jump down when the large seller starts executing. The reason is that market makers observe entry of the large seller and understand that this implies directional liquidity demand until $t = d$.

Second, prices keep declining after the initial jump, but might show reversal *before* the seller is done executing. This result is somewhat surprising given that the large seller does *not* release his sell pressure in the final stages of his order execution. This result demonstrates the power of our model being a dynamic one. The core trade-off market makers face at any point in time is whether to skew quotes more (by adjusting trading intensities) and thus pay investors a larger subsidy to revert out of non-zero inventory, *or* stay longer on costly non-zero inventory. This trade-off explains why in the final stages of the sell-order execution, market makers might become less willing to pay a subsidy given that they will soon be able to mean-revert paying a lower subsidy (i.e., after the large seller has left the market). This pattern of prices reverting before an execution ends, has been documented based on Abel Noser data (Zarinelli et al. 2015, Fig. 8). Our equilibrium result can rationalize this hitherto somewhat puzzling empirical finding.

5. Robustness to Unpredictable Duration

In the main model presented in Section 2, we assume information symmetry in the sense that order flow duration is perfectly known to both the large seller and the market makers. More precisely, market makers perfectly learn both sell intensity f and duration d of a large sell order as soon as the large seller starts selling.

To show that our main results are robust, we now consider the other polar case of the order flow being completely unpredictable. Information on duration becomes asymmetric in the sense that market makers continue to observe sell intensity immediately when the large seller starts selling, but they do *not* learn duration (which now is private information to the large seller). To ensure that they remain uninformed while the large order executes, we assume an exponential distribution with mean d for duration. The memoryless property of the exponential distribution ensures that the time elapsed since the start has no value for predicting how long the large order continues to execute.

Similar to the predictable-duration case, there is a unique symmetric equilibrium in the unpredictable-duration case. The formal statement of this result along with related findings mirroring those developed so far for the predictable-duration case are not repeated here, but are relegated to Appendix B. The remainder of this section discusses how the results for the unpredictable-duration case differ, in particular those results that speak to the two main objectives.

Characterization of the unpredictable-duration equilibrium. The results for the unpredictable-duration case are illustrated by Figure 2, which is the equivalent of Figure 1 for the predictable-duration case. The two cases differ in the following ways. First, in the case duration is unknown, both the bid and the ask price jump at $t = d$. This is not surprising because, at that time, market makers observe that the order intensity drops from f to zero, and therefore learn that the large sell order stopped executing. Second, the bid and ask price never revert *before* the large seller is done executing. After their initial jump down, they continue to decline while the large order is executing. Third, the initial price jump is strictly smaller for the case when the order duration is unknown to market makers. Lemma 3 states these results formally.

Lemma 3. *Relative to the large-seller model where market makers know the duration, the model with unknown duration features a strictly higher bid and ask price at* $t = 0$ *(after the initial jump down):*

$$\tilde{p}_0^b - p_0^b > 0 \quad \textit{and} \quad \tilde{p}_0^a - p_0^a > 0, \tag{34}$$

where the tilde corresponds to the model with unknown duration.

The intuition for the smaller initial price jump in the case of unknown duration is as follows. Define the *shadow cost* of inventory as the component of relative prices that is not explained by the current inventory level, but rather by expected incoming inventory flow. If duration is known to market makers, then they can compute the shadow cost of initial inventory with certainty given that future liquidity demand is deterministic. If, however, duration is unknown to them, then they necessarily have to resort to an expectation of this cost. Since the shadow cost of a unit of inventory is concave in duration due to the time

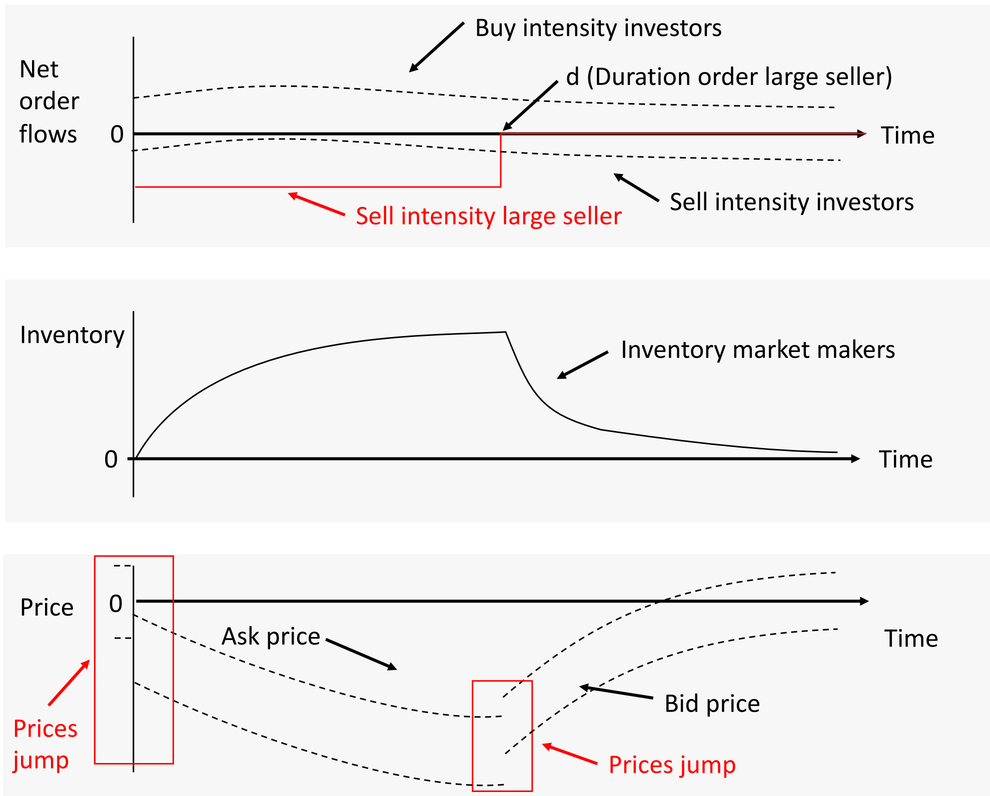


Figure 2: **Model visualization for unknown order duration.** This figure illustrates the model by plotting the equilibrium trajectories for key variables, if market makers start with zero inventory. It mirrors Figure 1 with the exception that the duration of the large sell order is not known to market makers. The top graph plots the (continuous) net flow of the large seller as well as the flow of the continuously arriving buy and sell investors. The middle graph plots the inventory of market makers, which results from the net flows in the top graph. The bottom graph plots the bid and ask prices relative to fundamental value.

discounting, the expectation of this shadow cost is lower than the shadow cost at the mean duration (i.e., at d).

The impact of large seller on market maker and other investors. In spite of the somewhat different equilibrium patterns discussed thus far, the results that speak to the two main objectives are the same, with one exception. In both cases, market makers benefit from the presence of a large seller, and other investors do so only if sell intensity is high enough.

Both predictable and unpredictable duration cases imply that shorter orders likely trade at higher intensities. More precisely, the saturation point, beyond which total proceeds would decline, decreases in (expected) duration. In other words, if a large seller picks a shorter duration, then he is able to sell at a higher intensity. The only difference between the two cases is that there no longer is an equilibrium pattern of price pressures declining before execution ends in the unpredictable-duration case. The intuition is that market makers are unable to predict when execution ends and, therefore, it is no longer optimal for them to relieve price pressure close to the end of the large order's execution.

In sum, the results for the unpredictable-duration case show that our conclusions are robust to this model perturbation. The empirical finding in Zarinelli et al. (2015, Figure 8) of price pressure subsiding before executions end suggests that market makers likely had some ability to predict the end times.[16] But, the welfare results are robust across the two cases, as is the implied relationship between order duration and execution intensity.

6. Model Calibration

We explain how the proposed model can be calibrated. We then use the calibrated model to estimate a liquidity surface, and then quantify various aspects of liquidity as implied by the surface.

6.1 Calibration Procedure

We pick the seven model parameters to match equally many real-world variables for the baseline case where time is measured in days:

- $\beta = -\log(0.9998)$: In their discrete-time model, Hendershott and Menkveld (2014) set the daily discount rate equal to 0.9998. The continuous-time analogue is the negative of the log.
- $N = 10$: We believe a natural real-world equivalent of the model's market makers are high-frequency traders. Therefore the number of market makers was set to ten based

[16] This is an important reason for why we feature the predictable-duration case in the main model specification.

on Scandinavian equity trading on Nasdaq-OMX.[17]

- The remaining three parameters are based on Menkveld (2013) who analyzed trading by one large high-frequency market maker for a 2007-2008 sample of Dutch blue chip equities. Its particular appeal is that it contains estimates of the conditional price pressure.[18] The calibration focuses on afternoon trading in large-cap stocks.

  – $\eta = 0.2434$ bps/(€1000), $\delta =$ €18×10^6/bps, and $\omega = 17.6$ bps where "bps" is basis points: These parameters are set at values to match the following three variables: Bid-ask spread (3.2 bps), conditional price pressure (0.026 bps/€1000), and the average traded volume of (€$31,400 \times 17,000+$ €$2500 \times 16,800 =$ €$575,800,000$ from incumbent and entrant markets altogether) (Menkveld 2013, Table 1). The value of these parameters are recovered as the unique solution of a system of three equations in three unknowns:

$$\begin{aligned} \frac{\omega}{N+1} &= 1.6, \\ 2 \times \delta \left(\omega - \frac{\omega}{N+1} \right) &= 575800, \\ \frac{2A^\star}{N+1} = \frac{2}{N+1} \frac{2\eta}{\beta + \sqrt{\beta^2 + \frac{32\delta\eta}{(N+1)^2}}} &= 0.0026. \end{aligned}$$

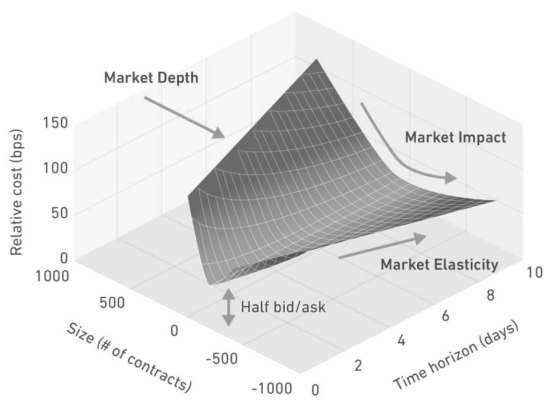
  To derive the equations above, we have applied Proposition 1 in the case the market makers' initial inventory is zero. In particular, the daily trading volume is given by the trading flow with the end investors when the relative ask/bid price is given by $\pm\frac{\omega}{N+1}$. We have also used Eq. (A16) in Appendix C to deduce the third equality.

[17] "Europe's Top 10 High-Frequency Kingmakers (in Scandinavia, at Least)," Financial News, 2014.

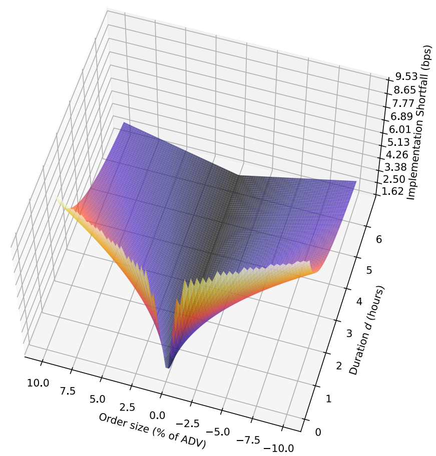
[18] In this period, we believe it is reasonable to assume that there were ten market makers in the model (see earlier point). Unfortunately, Menkveld (2013) presents detailed trade results for only one of them who happened to be (temporarily) large in the sample. We interpret this HFT as being under strong competitive threat. We caution that the calibration only serves to find "reasonable" parameters. Estimation of the model is beyond reach as it requires detailed HFT data that is not available to us. We leave that for future research.

6.2 Model-Implied Liquidity Surface

Regulators want investors to manage and report their liquidity risk (see footnote 2). A critical part of such liquidity management is a reliable methodology to predict the price impact that investors incur when they need to suddenly liquidate part of their portfolio. Such a need may arise when, for example, investors seek to redeem their shares in an open-end fund (e.g., Capponi, Glasserman, and Weber 2020).



(a) MSCI liquidity surface



(b) Model-implied liquidity surface

Order size (% of ADV)	Order duration (hours) 1	2	3	4	5	6
0	1.60	1.60	1.60	1.60	1.60	1.60
1	2.68	2.15	1.97	1.88	1.82	1.79
2	3.76	2.71	2.34	2.16	2.05	1.97
3	4.84	3.26	2.72	2.44	2.27	2.16
4	5.91	3.81	3.09	2.72	2.50	2.35
5	6.99	4.37	3.46	3.00	2.72	2.54
6	8.07	4.92	3.83	3.28	2.95	2.72
7	9.15	5.47	4.20	3.56	3.17	2.91

(c) Model-implied liquidity surface, panel **(b)** in numbers

Figure 3: **Liquidity surface.** This figure plots the liquidity surface: the price impact of a large order as a function of its size and duration. The top-left graph depicts the MSCI version of it (MSCI 2019, p. 2). The top-right graph plots the model-implied liquidity surface. The bottom panel presents the numbers on which this plot is based.

The price impact of orders is often thought of as a "liquidity surface." It is a 3D plot with impact as a funtion of order size *and* duration. This is not merely an academic construct, because it is known as such in industry as well. The plot in Panel (a) of Figure 3, for example, is taken from an MSCI white paper (MSCI 2019).

We construct a model-implied liquidity surface for a typical blue-chip stock using model calibration of Section 6.1. One might argue that a model is not needed if one has historical data on price impact, size, and duration. This, however, *assumes* that market conditions now are comparable to historical conditions. For example, there is no entry or exit in the market-making sector or investor participation remains unchanged. These assumptions are strong ones in particular for crisis scenarios that are of interest when assessing liquidity risk. A model-based liquidity surface can handle these cases in a logically consistent manner. This is especially helpful for a fund to manage its liquidity and meet its regulatory duties.

The new SEC regulation for open-end funds referred to in the introduction serves as a good example (SEC 2016). A model-based approach to liquidity will allow such a fund to gauge which securities qualify as "highly liquid," depending on market conditions. SEC (2016) requires that at least 85% of the assets to be highly liquid.[19]

Panel (b) in Figure 3 plots the liquidity surface based on the calibrated model.[20] It plots the price impact of orders up to 10% of ADV and lasting for up to a day.[21] Let us visit the four aspects of liquidity depicted in the MSCI liquidity surface in panel (a), and discuss the model-implied liquidity surface accordingly. First, (infinitely) small orders pay half the *bid-ask* spread which is 1.62 basis points.

Second, large orders pay more than the half spread to compensate market makers for carrying costly inventory through time. If one executes an order that is seven percent of ADV over a one-hour period, then the model-implied *price impact* is 9.15 basis points. Coincidentally, this price impact of 1.62 and 9.15 basis points for small and large orders is comparable to the $0.5 \times 0.0345 = 1.7$ and 11.9 basis points, respectively, reported for US institutional orders in 2007 (Anand et al. 2013, Table 1 and 2).[22]

[19]See also "Get Used to It: New Regulations on Liquidity Risk," Haaretz, December 1, 2018.

[20]Given an order of size S and duration d, we first compute the implied sell intensity $f = S/d$. We then estimate the price impact as

$$\frac{1}{d}\int_0^d p^b\left(i_s, f, d, s\right) ds \quad \text{assuming} \quad i_0 = 0.$$

We further assume symmetry across buy and sell orders to create the plot.

[21]It is common practice to express the size of large order size in ADV units (see, e.g., Anand et al. 2013). One reason is that this makes order size more comparable across stocks.

[22]The average order size in Anand et al. (2013) is 2.2% of ADV (Table 1). They do not report durations

Third, *market elasticity* is such that if for the 7%-ADV order, one takes two hours instead of one, the price impact shrinks by about 40%, from 9.15 to 5.4 basis points. This dimension of liquidity is more commonly referred to as market resiliency in the academic literature (e.g., Obizhaeva and Wang 2013).

Fourth, (maximum) *market depth* is implicitly part of the plot, because this market depth determines the edges of the liquidity surface. That is, for each order size, there is a minimum duration. If one were to execute the order over a period slightly shorter than this minimum, then the total proceeds would be strictly *higher* if a slightly smaller order were executed. Figure 4 plots makes this minimum duration explicit by plotting it as a function of order size. Not surprisingly, larger orders require longer durations. This result is closely related to the saturation point result of Proposition 5. It is stated formally in the next proposition.

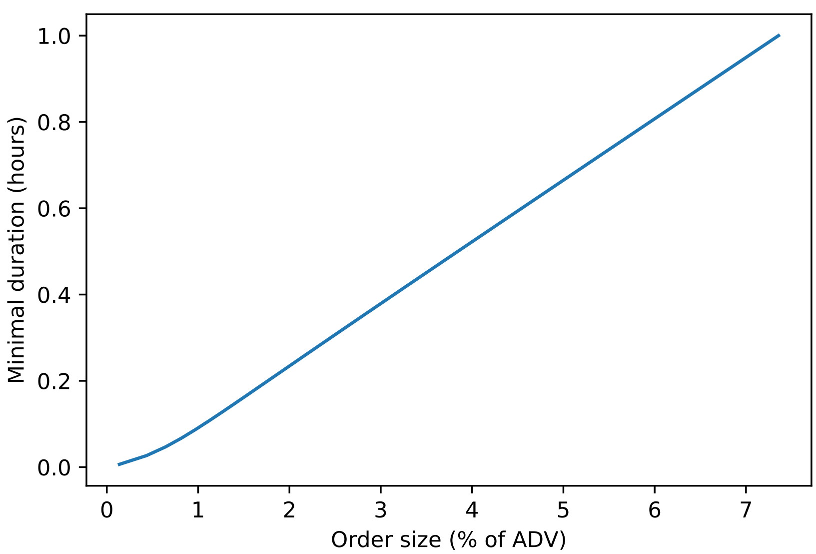


Figure 4: **Minimum duration as a function of order size.** This plot illustrates how long a large seller needs to take to sell an order of a particular size (expressed as a percentage of average daily volume, ADV). It is a minimum in the sense that if he were to execute over a slightly shorter period, then his total proceeds would be strictly *higher* if a slightly smaller order were traded. The plot is based on the calibration of our model (see Appendix 6).

but if we take our Abel Noser duration mean of 1.23 hours as a guideline, then their order executes at an average intensity of $2.2/(1.23/6.50) = 11.6\%$ (note that they also use Abel Noser data).

Proposition 6 (Minimum duration)**.** *For every order size $S > 0$, there exists a minimum duration $\underline{d} > 0$ such that for any $0 < d' < \underline{d}$ there exists an $S' < S$ such that $P_{d'}(\frac{S'}{d'}) > P_{d'}(\frac{S}{d'})$. This minimum duration $\underline{d}$ is increasing in S.*

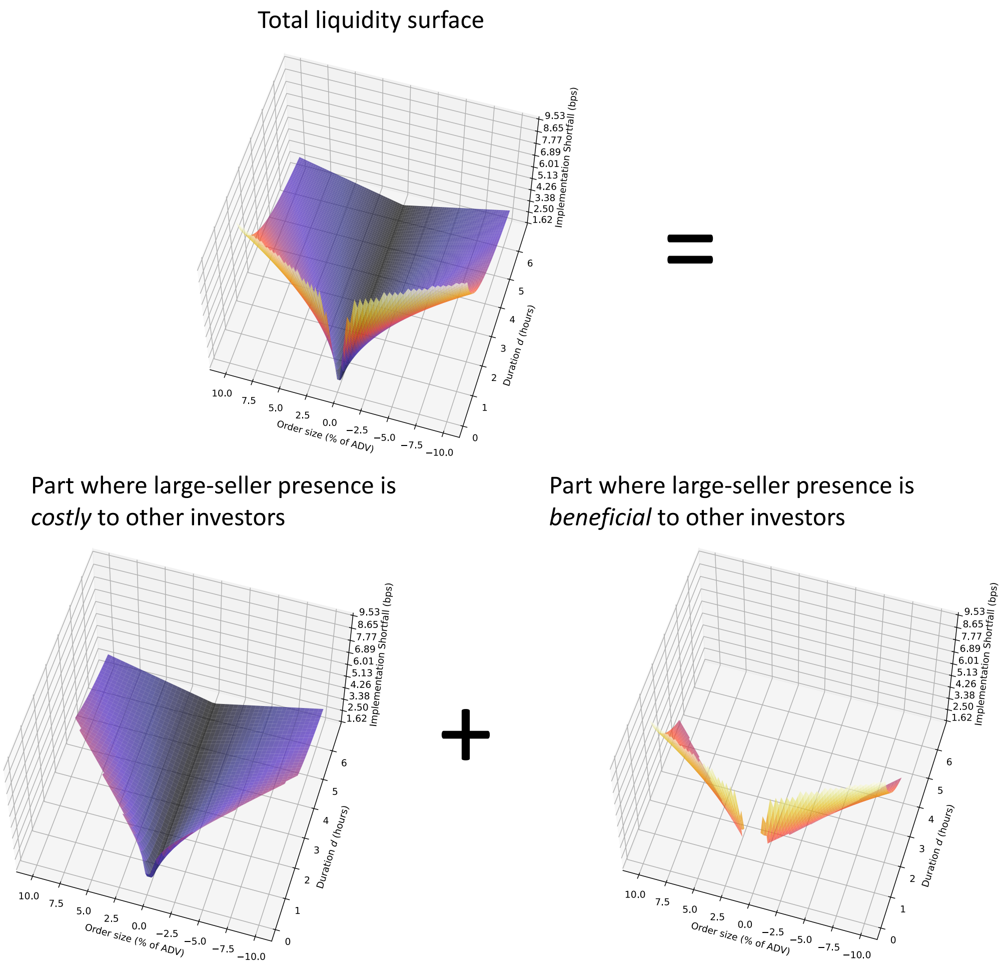


Figure 5: **When do other investors benefit from the presence of a large seller?** This figure re-plots the model-implied liquidity surface of panel (b) in Figure 3 to illustrate for which part of the liquidity surface, the presence of a large seller hurts other investors (bottom-left panel) and for which part it benefits them (bottom-right part).

Finally, the calibrated model allows to determine if the presence of large orders benefits or hurts other investors. To this end, Figure 5 decomposes the liquidity surface into an area where investors benefit from large orders and an area where they are hurt by them. The plot suggests that for the majority of size-duration combinations, investors are hurt by the presence of large orders.

7. Conclusion

We analyze a model in which a large seller trades continuously at a constant intensity for a particular period of time. This mirrors the slicing and dicing of a large uninformed order in real-world markets targeting a POV execution strategy. Liquidity is supplied dynamically by a set of Cournot-competitive market makers.

A particularly attractive feature of our setting is that there is a unique symmetric equilibrium that can be solved analytically. This allows us to generate two main findings: (i) the presence of a large seller benefits markets makers always, and other investors only if he sells at a high enough intensity, and (ii) the model-implied patterns of order execution are able to match empirically observed patterns of order flow execution by institutional investors.

The tractability of our model setting opens up possibilities for further investigation. How do the results change if one expands the strategy space of the large seller? For instance, can execution patterns other than POV reduce total price impact? How do they change if multiple large orders from different investors are executed in potentially overlapping periods, including both buy and sell orders? Or, how do they change when market makers can only detect a large order with a delay? How do they change if market makers are heterogeneous? These are all interesting areas for future research, in particular given the trend of large orders becoming more prevalent.

7. Data Availability Statement

The data underlying this article are available in the article and in its online supplementary material.

A. Notation

The following schema summarizes the notation used throughout the paper.

β	Intertemporal discount rate used by all agents in the market.
δ	The level of trade activity for the security.
η	The flow cost of carrying inventory for a market maker is ηi^2, where i denotes his inventory.
ω	The relative reservation value of the security to investors (i.e., the reservation value minus common value m_t). It is ω for buy investors and it is $-\omega$ for sell investors and the large seller.
σ	The standard deviation of the security's fundamental value m_t.
θ	Vector that includes all parameters: $\theta = (\beta, N, \eta, \delta, \omega)$.
d	The duration of the sell order that the large investor is executing.
i	A single market maker's inventory.
p_t^a	The relative ask price at time t (i.e., the ask price minus m_t).
p_t^b	The relative bid price at time t (i.e., the bid price minus m_t).
m_t	The security's common value at time t (i.e., fundamental value).[23]
N	The number of market makers.
q_t^a	Trade intensity of buy investors at the ask price at time t.
q_t^b	Trade intensity of sell investors (excluding the large seller) at the bid price at time t.

B. Unpredictable-Duration Case

This appendix states various propositions and lemmas for the case in which a large order executes and its duration is unpredictable for market makers. All results mirror the results in the main text. For each result, we refer to the result in the main text that it mirrors. Proofs are in Appendix C.

The following proposition mirrors Proposition 2:

[23] "m" stands for martingale.

Proposition B.1 (Price trajectories, market makers' traded quantities and value functions if order flow duration is unknown)**.** *Assume a large seller sells at constant intensity f from $t = 0$ until $t = d$, where market makers assume that duration follows an exponential distribution with mean d. Then the results developed in Proposition 1 change in the following ways. The equilibrium remains unique with relative bid and ask prices for each market maker equal to*

$$\tilde{p}^a(i, f, d, t) = p^a(i) - 1_{\{t \leq d\}} f \left(\tilde{G}_\theta\right), \qquad \text{(Relative Ask price)} \qquad \text{(A1)}$$

$$\tilde{p}^b(i, f, d, t) = p^b(i) - 1_{\{t \leq d\}} f \left(\tilde{G}_\theta + H_\theta\right), \qquad \text{(Relative Bid price)} \qquad \text{(A2)}$$

where $p^a(i)$ and $p^b(i)$ are the ask and bid quotes in the baseline case (Proposition 1) and H_θ is a positive constant (featured also in Proposition 2). The associated quantities are

$$\tilde{q}^a(i, f, d, t) = q^a(i) + 1_{\{t \leq d\}} f \frac{\delta}{N} \left(\tilde{G}_\theta\right), \qquad \text{(Relative Ask quantity)} \qquad \text{(A3)}$$

$$\tilde{q}^b(i, f, d, t) = q^b(i) - 1_{\{t \leq d\}} f \frac{\delta}{N} \left(\tilde{G}_\theta - N H_\theta\right), \qquad \text{(Relative Bid quantity)} \qquad \text{(A4)}$$

and the bid-ask spread therefore is

$$\tilde{s}(i, f, d, t) = s(i) + 1_{\{t \leq d\}} f\, H_\theta = s(i, f, t). \qquad \text{(Bid-ask spread)} \qquad \text{(A5)}$$

The value function for a market maker is

$$\tilde{v}(i, f, d, t) = v^M(i) + 1_{\{t \leq d\}} \left(\tilde{I}_\theta(f) + f \tilde{J}_\theta i\right), \qquad \text{(Value market makers)} \qquad \text{(A6)}$$

where $\tilde{I}_\theta(f) = 0$ for $f = 0$ and convex in f, and $\tilde{J}_\theta < 0$.

B.1 Value of the Presence of Large Seller to Others

The following proposition mirrors Proposition 3:

Proposition B.2 (The value of the presence of a large seller to market makers, when order duration is unpredictable)**.** *Market makers with zero initial inventory are better off in case there is a large seller when compared to the baseline case (without such a seller).*

The following proposition mirrors Proposition 4:

Proposition B.3 (Value of large-seller presence to other investors, when order duration is unpredictable)**.** *If market makers start with zero inventory, then the realized surplus of investors other than the large seller, is convex in large-seller trade intensity f. There exists a threshold $\bar{f}(\nu)$ such that investors are*

- *strictly worse for $f < \bar{f}(\nu)$,*
- *equally well off when $f = \bar{f}(\nu)$, and*
- *strictly better off when $f > \bar{f}(\nu)$.*

B.2 Equilibrium Pattern of Order Execution

The following proposition mirrors Proposition 5:

Proposition B.4 (Optimal liquidation intensity at a given duration mean, when order duration is unpredictable)**.** *For every duration $d > 0$, there exists an optimal intensity $f^\star > 0$ such that for any $f \neq f^\star$, we have $P_d(f^\star) > P_d(f)$. This maximum intensity $f^\star$ is decreasing in d.*

The following lemma mirrors Lemma 1:

Lemma B.1 (Market-maker inventory path, when order duration is unpredictable)**.** *The inventory of a market maker*

- *strictly increases and is concave in time while the large seller is executing,*
- *is homogeneous of degree one in the large seller's trade intensity,*
- *exhibits exponential decay towards zero after the seller disappears.*

The following lemma mirrors Lemma 2:

Lemma B.2 (Equilibrium bid and ask price path, when order duration is unpredictable)**.** *If market makers start with zero inventory, then the equilibrium bid and ask price for $t \leq d$:*

- *jump down at $t = 0$,*
- *are convex in t, and*

- *strictly decline in t.*

For $t > d$ the price paths

- *are concave in t and*
- *strictly increase in t.*

At time $t = d$, both the bid and the ask price jump up. The bid-ask spread is unchanged relative to the case where the order flow duration is known to market makers.

C. Proofs of Lemmas, Corollaries, and Propositions

Proof of Proposition 1. In the absence of the institutional investor, the trading intensities by each market maker are time homogenous and only depend on the size of that market maker's inventory:

$$q^a_{jt} = q^a(i_{jt}), \quad q^b_{jt} = q^b(i_{jt}),$$

where $q^a(i)$ and $q^b(i)$ are deterministic functions of the inventory. We can then rewrite the equation that governs the dynamics of i_{jt}, given in (8), as a time-homogenous autonomous equation:

$$\mathrm{d}i_{jt} = -q^a(i_{jt})\mathrm{d}t + q^b(i_{jt})\mathrm{d}t, \quad \forall t > 0.$$

Similarly, the trading revenue of each market maker follows the dynamics

$$\mathrm{d}W_{jt} = \left(\omega - \frac{1}{\delta}\sum_{n=1}^{N} q^a(i_{nt})\right) q^a(i_{jt})\mathrm{d}t - \left(\frac{1}{\delta}\sum_{n=1}^{N} q^b(i_{nt}) - \omega\right) q^b(i_{jt})\mathrm{d}t, \quad \forall t > 0.$$

By virtue of the dynamic programming principle, the value function $v^M_j(i)$ of the control problem solved by the j-th market maker is the solution to the Bellman equation:

$$\begin{aligned}\eta i^2 + \beta v^M_j(i) = \sup_{q^a_j, q^b_j} \Bigg[&\left(\omega - \frac{1}{\delta}\sum_{n=1}^{N} q^a_n\right) q^a_j - \frac{\partial}{\partial i} v^M_j(i) q^a_j \\ &- \left(\frac{1}{\delta}\sum_{n=1}^{N} q^b_n - \omega\right) q^b_j + \frac{\partial}{\partial i} v^M_j(i) q^b_l \Bigg]\Bigg|_{q^a_n = q^{a,\star}_j, q^b_n = q^{b,\star}_j \text{ for all } n \neq j},\end{aligned} \tag{A7}$$

where $(q_j^{a,\star}, q_j^{b,\star})$ is the optimizer of the Hamiltonian above, in which we set $q_n^a = q_j^{a,\star}, q_n^b = q_j^{b,\star}$ for all $n \neq j$ because we are considering a symmetric Markov perfect equilibrium. We make the ansatz that the value function $v_j^M(i)$ is quadratic and concave in i. Moreover, since all market makers are identical, value function and strategies are the same for all market makers, i.e.,

$$v_j^M(i) = v^M(i) = -Ai^2 + Bi + C, \tag{A8}$$

for some constant $A > 0$. It then follows that the optimal control strategy for each market maker is given by

$$q^a = \frac{\omega + 2Ai - B}{N+1}\delta, \qquad q^b = \frac{\omega - 2Ai + B}{N+1}\delta \tag{A9}$$

Using the expressions above, we can then rewrite (A7) as

$$(\eta - \beta A)i^2 + \beta Bi + \beta C = \frac{8A^2\delta}{(N+1)^2}i^2 + \frac{8\delta A}{(N+1)^2}Bi + \frac{2\delta(B^2 + \omega^2)}{(N+1)^2}.$$

By matching the coefficients of i^2, i and the constant term on both sides, we obtain that (A, B, C) must satisfy the following equations

$$\eta - \beta A = \frac{8\delta A^2}{(N+1)^2}, \tag{A10}$$

$$\beta B = \frac{8\delta A}{(N+1)^2}B, \tag{A11}$$

$$\beta C = \frac{2\delta(B^2 + \omega^2)}{(N+1)^2}. \tag{A12}$$

Hence, $A^\star$ must be the unique positive solution to quadratic equation (A10), and $B^\star = 0$ otherwise (A11) would not hold. Define

$$C^\star = \frac{1}{\beta}\frac{2\delta\omega^2}{(N+1)^2}. \tag{A13}$$

Then we obtain that a solution to (A10)-(A12) is given by $(A^\star, 0, C^\star)$. Hence, the value function is given by

$$v(i) = -A^\star i^2 + C^\star.$$

The ask and bid price policy functions follow from the defining expression (7), and the expressions for the traded intensities given by (A9) in which A and C are replaced, respectively, by the coefficients $A^\star$ and $C^\star$ of the optimal value function $v^M(i)$. Overall, for parameter $\theta = (\beta, \nu, N, \eta, \delta, \omega)$, the results in Proposition 1 holds with positive constants:

$$\begin{aligned} A_\theta &= \frac{\omega}{N+1}, \quad B_\theta = \frac{2A^\star}{N+1}, \quad C_\theta = \frac{\delta\omega}{N+1}, \\ D_\theta &= \frac{2\delta A^\star}{N+1}, \quad E_\theta = C^\star, \qquad F_\theta = A^\star. \end{aligned} \tag{A14}$$

□

Proof of Corollary 1. We notice that the right hand side of equation (A10) is strictly decreasing in N and strictly increasing in A, whereas the left hand side is strictly decreasing in A. Hence, the unique positive solution to (A10), $A^\star$, is strictly increasing in N, and

$$\lim_{N\to\infty} A^\star = \frac{\eta}{\beta}. \tag{A15}$$

It follows that

$$\lim_{N\to\infty} A_\theta = \lim_{N\to\infty} B_\theta = \lim_{N\to\infty} C_\theta = \lim_{N\to\infty} D_\theta = \lim_{N\to\infty} E_\theta = 0, \quad \lim_{N\to\infty} F_\theta = \frac{\eta}{\beta}.$$

Moreover, we also have

$$\lim_{N\to\infty} NC_\theta = \delta\omega, \quad \lim_{N\to\infty} ND_\theta = 2\delta\frac{\eta}{\beta}.$$

This completes the proof. □

Proof of Results in Table 1. Recall that $A^\star$ is given by

$$A^\star = \frac{-\beta + \sqrt{\beta^2 + \frac{32\delta\eta}{(N+1)^2}}}{\frac{16\delta}{(N+1)^2}} = \frac{2\eta}{\beta + \sqrt{\beta^2 + \frac{32\delta\eta}{(N+1)^2}}} \tag{A16}$$

Clearly, $A^\star$ is strictly increasing in η, and is strictly increasing in N, and is strictly decreasing in β, and is strictly decreasing in δ, and does not depend on ω.

Because the spread in the baseline is equal to $\frac{2\omega}{N+1}$, we immediately know that the claims in the column for the bid-ask spread hold.

Moreover, recall that the conditional price pressure $B_\theta = \frac{2A^\star}{N+1}$, so we know all sensitivity results with respect to $\eta, \beta, \delta, \omega$ hold. Finally, from

$$\frac{A^\star}{N+1} = \frac{2\eta}{\beta(N+1) + \sqrt{\beta^2(N+1)^2 + 32\delta\eta}}, \tag{A17}$$

we deduce that the conditional price pressure $B_\theta = \frac{2A^\star}{N+1}$ is strictly decreasing in N. This completes the proof. $\square$

Proof of Proposition 2. The market is completely the same as in case of no large seller if $t > d$. Hence, the results are those given in Proposition 1. For $t \leq d$, by the dynamic programming principle the value function $v^M(i, f, d, t)$ of each market makers' control problem is the solution to the Bellman equation:

$$\begin{aligned} 0 = -\eta i^2 - \beta v^M + \frac{\partial}{\partial t} v^M + \sup_{q_j^a, q_j^b} \Bigg[\left(\omega - \frac{1}{\delta} \sum_{n=1}^N q_n^a \right) q_j^a - \frac{\partial}{\partial i} v^M \, q_j^a \\ - \left(-\omega + \frac{1}{\delta} \left(\sum_{n=1}^N q_n^b - f \right) \right) q_j^b + \frac{\partial}{\partial i} v^M \, q_j^b \Bigg] \Bigg|_{q_n^a = q_j^{a,\star},\, q_n^b = q_j^{b,\star} \text{ for all } n \neq j}, \end{aligned} \tag{A18}$$

where $(q_j^{a,\star}, q_j^{b,\star})$ is the optimizer of Hamiltonian above in which we set $q_n^a = q_j^{a,\star}, q_n^b = q_j^{b,\star}$ for all $n \neq j$ because we are focusing symmetric Markov perfect equilibria. We make the following quadratic ansatz for the value function

$$v^M(i, f, d, t) = -A(t, f, d) i^2 + B(t, f, d) i + C(t, f, d), \quad \forall t \geq 0, \tag{A19}$$

where for notational simplicity we have suppressed f in the arguments of the right hand side above, and noted that the value function is the same for all market makers given that they have the same inventory costs and equally split the incoming trading flow. It then follows from the HJB equation (A18) that the optimal control $(q_j^{a,\star}, q_j^{b,\star})$ is the same for each market marker, and given by

$$q^a(i,t,f,d) = \frac{\omega + 2A(t,f,d)i - B(t,f,d)}{N+1}\delta$$
$$q^b(i,t,f,d) = \frac{\omega + \frac{f}{\delta} - 2A(t,f,d)i + B(t,f,d)}{N+1}\delta,$$

as long as $A(t,f,d) > 0$. By using (A19), and matching the coefficient of i^2 in (A18), we obtain an ordinary differential equation in time

$$\eta - \beta A(t) + A'(t) = \frac{8\delta(A(t))^2}{(N+1)^2}, \quad A(d) = A^\star, \tag{A20}$$

where we only highlight the dependence on the time variable above. It follows from the definition of $A^\star$ is the unique positive solution to (A10) that $A'(d) = 0$. By the uniqueness of the solution to an ODE, we obtain that $A(t) \equiv A^\star$. Similarly, by matching the coefficient of i in (A18), we obtain an ODE in the time variable for $B(t,f,d)$, where we only highlight the dependence on the time variable:

$$\begin{aligned}
0 &= -\beta B(t) + B'(t) - \frac{4A^\star(f + 2\delta B(t))}{(N+1)^2} \\
&= -\left(\beta + \frac{8\delta A^\star}{(N+1)^2}\right)B(t) + B'(t) - \frac{4A^\star}{(N+1)^2}f \\
&= -\frac{\eta}{A^\star}B(t) + B'(t) - \frac{4A^\star}{(N+1)^2}f.
\end{aligned}$$

subject to the terminal condition $B(d) = 0$. The solution of the above ODE is

$$B(t,f,d) = -f\frac{4A^\star}{(N+1)^2}\frac{1 - e^{-\frac{\eta}{A^\star}(d-t)}}{\eta/A^\star} =: fb(d-t), \quad \forall t \in [0,d]. \tag{A21}$$

The function $C(t,f,d)$ can be determined similarly: for $t \in [0,d]$, we have the following ODE in the time variable t, where again we only emphasize the dependence on t:

$$0 = C'(t) - \beta C(t) + \frac{2\delta^2(B^2(t) + \omega^2) + 2\delta(B(t)+\omega)f + f^2}{(N+1)^2\delta}.$$

Letting $C_1(t) := C(t) - E_\theta$, then $C_1(d) = 0$ and

$$0 = C_1'(t) - \beta C_1(t) + \frac{2\delta^2 B^2(t) + 2\delta(B(t)+\omega)f + f^2}{(N+1)^2\delta}.$$

It follows that the solution is

$$\begin{aligned} C(t,f,d) = E_\theta + 1_{\{t\leq d\}} \frac{1}{\delta(N+1)^2} \\ \left[2\delta\omega f \frac{1-e^{-\beta(d-t)}}{\beta} f + f^2 \left(\int_0^{d-t} e^{-\beta(d-t-s)}(1+2\delta^2 b^2(s) + 2\delta b(s))\mathrm{d}s\right)\right], \end{aligned}$$

where E_θ is given in the proof of Proposition 1.

Finally, using the expressions for the ask and bid price policy functions in the absence of liquidation, the expression for the optimal trading intensities, and the expressions for the value function given in the proof of Proposition 1, we obtain that the ask and bid relative price policy functions in the presence of liquidation are given by

$$\begin{aligned} p^a(i,f,d,t) &= p^a(i) - 1_{\{t\leq d\}} f\left(G_\theta(d-t)\right), \\ p^b(i,f,d,t) &= p^b(i) - 1_{\{t\leq d\}} f\left(G_\theta(d-t) + H_\theta\right), \end{aligned}$$

where $G_\theta(t) = -Nb(t)/(N+1)$ and $H_\theta = 1/(\delta(N+1))$. The optimal control is given by

$$\begin{aligned} q^a(i,f,d,t) &= q^a(i) + 1_{\{t\leq d\}} f \frac{\delta}{N}\left(G_\theta(d-t)\right), \\ q^b(i,f,d,t) &= q^b(i) - 1_{\{t\leq d\}} f \frac{\delta}{N}\left(G_\theta(d-t) - NH_\theta\right), \end{aligned}$$

Finally, the value function takes the form

$$v^M(i,f,d,t) = v^M(i) + 1_{\{t\leq d\}}\left(I_\theta\left(d-t,f\right) + fJ_\theta\left(d-t\right)i\right),$$

where $I_\theta(d-t,f) = C(t) - E_\theta$, which is strictly increasing in variable "$d-t$" for $t \leq d$ and strictly convex in f, and $J_\theta(d-t) = B(t)$, which is negative and increasing in t. This completes the proof. $\square$

Proof of Lemma 1. We first derive an explicit expression for the inventory. We use (8), (29) and (30) to obtain the dynamics of a market maker's inventory:

$$\begin{aligned}\mathrm{d}i_{jt} =& -\left(C_\theta + D_\theta i_{jt} + 1_{\{t\le d\}}\delta\frac{f}{N}G_\theta(d-t)\right)\mathrm{d}t \\ &+ \left(C_\theta - D_\theta i_{jt} - 1_{\{t\le d\}}\delta\frac{f}{N}\left(G_\theta(d-t) - NH_\theta\right)\right)\mathrm{d}t.\end{aligned}$$

It follows that i_{jt} satisfies the ODE:

$$\frac{\mathrm{d}}{\mathrm{d}t}i_{jt} = -2D_\theta i_{jt} + 1_{\{t\le d\}}\delta\frac{f}{N}\left(NH_\theta - 2G_\theta(d-t)\right), \tag{A22}$$

subject to the initial condition $i_{j0}=0$. Solving the ODE we obtain that $i_{jt} = f\times g(t)$, where

$$g(t) = \begin{cases}\dfrac{M}{4\delta A^\star}\left[\dfrac{\beta A^\star}{\eta}\dfrac{1-e^{-Mt}}{M} + \left(1-\dfrac{\beta A^\star}{\eta}\right)\dfrac{e^{\frac{\eta}{A^\star}t} - e^{-Mt}}{\frac{\eta}{A^\star}+M}\right], & t\le d,\\ g(d)e^{-M(t-d)}, & t>d,\end{cases}$$

where $M = \frac{4\delta A^\star}{N+1}$. A straightforward calculation yields $\frac{\partial g(t)}{\partial t} > 0$ for $t \le d$. This completes the proof. □

Proof of Lemma 2. By combining (25) with the explicit expression for i_{jt} in the proof of Lemma 1, we obtain that the relative bid price is given as follows: for $t\le d$,

$$\begin{aligned}p^b(i_{jt},f,d,t) =& -\frac{\omega}{N+1} - \frac{f}{2\delta(N+1)}\times\\ &\left[2 + N\left(1 - \frac{\frac{\eta}{A^\star}-\beta}{M+\frac{\eta}{A^\star}}e^{-\frac{\eta}{A^\star}(d-t)} - \frac{\beta A^\star}{\eta}e^{-Mt} - \frac{\frac{\eta}{A^\star}-\beta}{\frac{\eta}{A^\star}}\frac{M}{M+\frac{\eta}{A^\star}}e^{-\frac{\eta}{A^\star}d - Mt}\right)\right],\end{aligned} \tag{A23}$$

for $t>d$,

$$\begin{aligned}p^b(i_{jt},f,d,t) =& -\frac{\omega}{N+1} - \frac{fN}{2\delta(N+1)}\times\\ &\left(\frac{\beta A^\star}{\eta}\left[1-e^{-Md}\right] + \frac{(\eta-\beta A^\star)}{\eta}\frac{M}{M+\frac{\eta}{A^\star}}\left[1-e^{-(M+\frac{\eta}{A^\star})d}\right]\right)e^{-M(t-d)},\end{aligned}$$

where $M = \frac{4\delta A^\star}{N+1}$. Straightforward calculation yields that $p^b(i_{jt}, t)$ is strictly convex in t for $t < d$. Moreover, because[24]

$$\frac{\partial}{\partial t}\, p^b(i_{jt}, f, d, t)\Big|_{t=0} \propto \frac{\eta - \beta A^\star}{M A^\star + \eta}\frac{\eta}{A^\star} e^{-\frac{\eta}{A^\star} d} - \left(\frac{\beta A^\star}{\eta} + \frac{\eta - \beta A^\star}{\eta}\frac{M}{M + \frac{\eta}{A^\star}} e^{-\frac{\eta}{A^\star} d}\right) M.$$

Clearly, the right hand side of the above expression is monotone in d, and is negative if $d > 0$ is sufficiently large. As $d \to 0$, this expression converges to

$$\frac{-2\beta + \frac{\eta}{A^\star} - M}{M + \frac{\eta}{A^\star}} = \frac{1}{M + \frac{\eta}{A^\star}}\left(- \beta - M\frac{N-1}{N+1}\right) < 0.$$

Thus, we know that for all $d > 0$ it holds that $\frac{\partial}{\partial t} p^b(i_{jt}, f, d, t)|_{t=0} < 0$, and the relative bid price is decreasing in t for small t, as claimed. Moreover,

$$\frac{\partial}{\partial t}\, p^b(i_{jt}, f, d, t)\Big|_{t=d} \propto \frac{\eta - \beta A^\star}{M A^\star + \eta}\frac{\eta}{A^\star} - \left(\frac{\beta A^\star}{\eta} + \frac{\eta - \beta A^\star}{\eta}\frac{M}{M + \frac{\eta}{A^\star}} e^{-\frac{\eta}{A^\star} d}\right) M e^{-Md}.$$

Thus, if $d > 0$ is sufficiently large, we have $\frac{\partial}{\partial t} p^b(i_{jt}, f, d, t)|_{t=d} > 0$, so the relative bid price is increasing in t if t and d are both sufficiently large. The properties for relative bid price after d hold obviously. For the relative ask price, we obtain the results via relationship

$$p^a(i_{jt}, f, d, t) = p^b(i_{jt}, f, d, t) + 2A_\theta + f\, H_\theta 1_{\{t \le d\}}.$$

This completes the proof. □

Proof of Lemma 5. Using (4) and (A23), we can express the objective of the large seller in (4) as

$$\sup_{f \ge 0} \Big(P(d)\, f - Q(d)\, f^2 \Big)\,, \tag{A24}$$

[24]The symbol $\propto$ means proportional to, i.e., the left hand side is a positive constant multiple of the right hand side.

where $P(d)$ and $Q(d)$ are positive functions given by

$$
\begin{aligned}
P(d) =& \frac{N\omega}{N+1} h_\beta(d), && \text{(A25)}\\
Q(d) =& \frac{h_\beta(d)/(2\delta)}{N+1}\left\{(N+2) + N\left[\frac{\beta}{M+\frac{\eta}{A^\star}} - \frac{\beta A^\star}{\eta}\frac{h_{M+\beta}(d)}{h_\beta(d)}\right.\right.\\
&\left.\left. - \frac{\beta}{(M+\beta)}\frac{h_{\frac{\eta}{A^\star}}(d)}{h_\beta(d)} - \frac{(\eta-\beta A^\star)}{\eta}\frac{M}{M+\frac{\eta}{A^\star}}\frac{M+\beta+\frac{\eta}{A^\star}}{M+\frac{\eta}{A^\star}}\frac{h_{M+\beta+\frac{\eta}{A^\star}}(d)}{h_\beta(d)}\right]\right\}, && \text{(A26)}
\end{aligned}
$$

with $M = \frac{4\delta A^\star}{N+1}$ and $h_\gamma(u) = \frac{1-e^{-\gamma u}}{\gamma}$ for any positive parameter $\gamma > 0$. It follows that the optimal liquidation intensity is given by

$$
f^*(d) = \frac{P(d)}{2Q(d)}.
$$

Next, we show the monotonicity of $f^*(d)$ in d. First, we notice that

$$
\begin{aligned}
\frac{1}{f^*(d)} =& \frac{2Q(d)}{P(d)}\\
=& \frac{1}{\delta N\omega}\left\{(N+2) + N\left[\frac{(\frac{\eta}{A^\star}-\beta)}{\frac{\eta}{A^\star}+M}\frac{\beta A^\star}{\eta-\beta A^\star} - \frac{\beta A^\star}{\eta}\frac{h_{M+\beta}(d)}{h_\beta(d)}\right.\right.\\
&\left.\left. - \frac{\beta}{(M+\beta)}\frac{h_{\frac{\eta}{A^\star}}(d)}{h_\beta(d)} - \frac{(\eta-\beta A^\star)}{\eta}\frac{M}{M+\frac{\eta}{A^\star}}\frac{M+\beta+\frac{\eta}{A^\star}}{M+\beta}\frac{h_{M+\beta+\frac{\eta}{A^\star}}(d)}{h_\beta(d)}\right]\right\} && \text{(A27)}
\end{aligned}
$$

To finalize the proof we need the following additional technical result.

Lemma C.3. *Let $h_\gamma(u) = \frac{1-e^{-\gamma u}}{\gamma}$ for $\gamma > 0$. If $\beta_1 > \beta_2$, the function $h_{\beta_1}(d)/h_{\beta_2}(d)$ is strictly decreasing in d, and strictly convex in d.*

Proof. Let $h(d) = \log(h_{\beta_1}(d)/h_{\beta_2}(d))$, we first prove that $h(d)$ is strictly decreasing in d, and strictly convex in d. To that end, taking derivative

$$
\begin{aligned}
h'(d) &= \frac{\beta_1 e^{-\beta_1 d}}{1-e^{-\beta_1 d}} - \frac{\beta_2 e^{-\beta_2 d}}{1-e^{-\beta_2 d}}\\
&= \beta_2 - \beta_1 + \frac{\beta_1}{1-e^{-\beta_1 d}} - \frac{\beta_2}{1-e^{-\beta_2 d}},\\
h''(d) &= -\frac{\beta_1^2 e^{-\beta_1 d}}{(1-e^{-\beta_1 d})^2} + \frac{\beta_2^2 e^{-\beta_2 d}}{(1-e^{-\beta_2 d})^2}.
\end{aligned}
$$

Note that $h'(+\infty) = 0$. To proceed, let us define $g(y, c) = \frac{c^2}{1-y^c}$, then to prove $h''(d) > 0$, it suffices to demonstrate that

$$g\left(e^{-\beta_2 x}, \frac{\beta_1}{\beta_2}\right) - g\left(e^{-\beta_2 x}, 1\right) < 0.$$

In other words, with $y = e^{-\beta_2 x} \in (0, 1)$, and $c = \frac{\beta_1}{\beta_2} > 1$, we need to prove that

$$g(y, c) - g(y, 1) < 0,$$

for all $0 < y < 1$ and $c > 1$. Because $g(y, c = 1) - g(y, 1) = 0$, it suffices to prove that $g(y, c)$, or equivalently, $\log g(y, c)$ is strictly decreasing in c for $c > 1$. The latter monotone property of $\log g(y, c)$ can be verified with standard analysis. Therefore, we know that $h(d)$ is also strictly decreasing in d, and strictly convex in d.

Finally, note that $e^{h(d)}$ is also strictly decreasing in d, and from

$$(e^{h(d)})'' = e^{h(d)}[h''(d) + (h'(d))^2] > 0,$$

we know that $e^{h(d)} = h_{\beta_1}(d)/h_{\beta_2}(d)$ is strictly convex in d. □

Using Lemma C.3, we can express $1/f^*(d)$ as

$$\frac{1}{f^*(d)} = A - h(d),$$

where $A > 0$ is a constant, and $h(d) > 0$ is a positive function that is strictly decreasing in d, and strictly convex in d. Therefore, we know that $f^*(d)$ is strictly decreasing in d. Moreover, using standard analysis one can verify that $0 < f^*(\infty) < f^*(0) < \frac{N\delta\omega}{2}$, so it holds that $A - h(0) > 0$.

□

Proof of Proposition 6. We continue the proof of Proposition 5. For the optimal intensity $f^\star(d)$ defined in (A27), we have the following expression for the derivative

$$\left(\frac{1}{df^*(d)}\right)' = \frac{-A + h(d) - h'(d)d}{d^2}.$$

Because $h(d)$ is strictly convex in d, we know that $h(d) - h(0) < h'(d)d$, and then

$$\left(\frac{1}{df^*(d)}\right)' < \frac{-A + h(0)}{d^2} < 0.$$

It follows that $df^*(d)$ is strictly increasing. Moreover, because $0 < f^*(\infty) < f^*(0) < \frac{N\delta\omega}{2}$, we know that the range of the liquidation size $df^*(d)$ is $(0, \infty)$.

Hence, for any size $S > 0$, there is a unique $\underline{d} > 0$ defined as $\underline{d} \cdot f^*(\underline{d}) = S$, and $\underline{d}$ is strictly increasing with S. For any $d' < \underline{d}$, from $d' f^*(d') < \underline{d} f^*(\underline{d}) = S$ we know that the implied liquidation rate with size S and duration d', S/d', is larger than the optimal liquidation rate $f^*(d')$. By the optimality of $f^*(d')$, we know that $P_{d'}(S/d') < P_{d'}(f^*(d'))$. In other words, liquidating size S in duration d' (which is shorter than $\underline{d}$) is worse than liquidating a smaller size $S' = d' \times f^*(d') < S$ with the same duration d'. □

Proof of Proposition 3.

$$\begin{aligned}
&v^M(i_0 = 0, f, d, t = 0) - v^M(i_0 = 0) \\
&= I_\theta(d, f) \\
&= \frac{1}{\delta(N+1)^2}\left[2\delta\omega f \frac{1 - e^{-\beta d}}{\beta} f + f^2 \left(\int_0^d e^{-\beta(d-s)}(1 + 2\delta^2 b^2(s) + 2\delta b(s))\mathrm{d}s\right)\right] \\
&> \frac{1}{\delta(N+1)^2}\left[2\delta\omega f \frac{1 - e^{-\beta d}}{\beta} f + f^2 \left(\int_0^d e^{-\beta(d-s)}(1 + \delta b(s))^2\mathrm{d}s\right)\right] \\
&> 0.
\end{aligned}$$

Above, the first inequality came from $1+2\delta^2 b^2(s)+2\delta b(s) > 1+\delta^2 b^2(s)+2\delta b(s) = (1+\delta b(s))^2$ for $b(s) < 0$. This completes the proof. □

Proof of Proposition 4. Let us denote the surplus without the large seller by

$$v^I(i,t) = \int_t^\infty e^{-\beta(s-t)}\left(\frac{1}{2}\delta\left(\omega - p_s^a\right)^2 \mathrm{d}s + \frac{1}{2}\delta\left(\omega + p_s^b\right)^2 \mathrm{d}s\right),$$

where $i_{jt} = i$. If there is no liquidation, the market makers use stationary strategies, so $v^I(t,i)$ will be independent of time t. In this case, $v^I(0,i)$ solves the equation

$$\beta v^I(i,0) = \Bigg(\frac{\delta}{2}(\omega - p_t^a)^2 + \frac{\partial}{\partial i} v^I(i,0) \cdot \\ \Bigg(-\frac{\delta}{N}(\omega - p_t^a)\Bigg) + \frac{\delta}{2}(\omega + p_t^b)^2 + \frac{\partial}{\partial i} v^I(i,0) \cdot \frac{\delta}{N}(p_t^b + \omega)\Bigg),$$

where $p_t^a = p^a(i)$ and $p_t^b = p^b(i)$ are given in (12) and (11), respectively. In this case, following a similar argument to that in the proof of Proposition 1, we obtain that

$$v^I(i) = E_\theta^I + F_\theta^I i^2,$$

where

$$E_\theta^I = \frac{\delta N^2}{\beta(N+1)^2}\omega^2, \quad F_\theta^I = A_\omega,$$

with

$$A_\omega = \frac{4\delta N^2 (A^\star)^2}{\beta(N+1)^2 + 8\delta(N+1)A^\star} > 0, \tag{A28}$$

Notice that this also gives the welfare of the end-investor for $t > d$ if the liquidation duration is $d > 0$.

Similarly, when a larger seller who liquidates at intensity $f > 0$ for a duration $d > 0$ is present, we denote by v^I the surplus of end-user investors. Then $v^I(i,d,f,t)$ solves

$$0 = -\beta v^I + \frac{\partial v^I}{\partial t} + \Bigg(\frac{\delta}{2}(\omega - p_t^a)^2 + \frac{\partial}{\partial i} v^I(i,t) \cdot \Bigg(-\frac{\delta}{N}(\omega - p_t^a) + \frac{f}{N}\Bigg) \\ + \frac{\delta}{2}(\omega + p_t^b)^2 + \frac{\partial}{\partial i} v^I(i,t) \cdot \frac{\delta}{N}(p_t^b + \omega)\Bigg),$$

for $t \in [0,d]$, where $p_t^a = p^a(i,d,f,t)$ and $p_t^b = p^b(i,d,f,t)$ are given in (26) and (25), respectively.

Following the argument used in the proof of Proposition 2, we can show that

$$v^I(i,d,f,t) = v^I(i) + 1_{\{t \le d\}} \left(I_\theta^I(d-t,f) + J_\theta^I(d-t) f\, i \right),$$

where

$$J_\theta^I(u) = \int_0^u e^{-(\beta+M)(u-s)} \cdot \frac{2}{N}\left(A_\omega + \frac{[N^2A^\star - (N+1)A_\omega](1-2\delta N b(s))}{(N+1)^2}\right)\mathrm{d}s, \qquad \text{(A29)}$$

and $b(s)$ is defined in (A21). Because $b(s) < 0$ for all $s > 0$, and

$$N^2A^\star - (N+1)A_\omega = N^2A^\star(N+1)\frac{\beta(N+1) + 4\delta A^\star}{\beta(N+1)^2 + 8\delta(N+1)A^\star} > 0,$$

we know that

$$J_\theta^I(u) > \frac{2}{N}A_\omega \int_0^u e^{-(\beta+M)(u-s)}\mathrm{d}s > 0.$$

Moreover,

$$\begin{aligned} I_\theta^I(u,f) =& -f\frac{\omega N}{(N+1)^2}\frac{1-e^{-\beta u}}{\beta} + f^2\int_0^u e^{-\beta(u-s)}\frac{(1+2\delta b(s))J_\theta^I(s)}{N+1}\mathrm{d}s \\ &+ f^2\int_0^u e^{-\beta(u-s)}\frac{(1-2\delta N b(s) + 2\delta^2N^2b^2(s))}{2\delta(N+1)^2}\mathrm{d}s. \end{aligned} \qquad \text{(A30)}$$

Because for any $s > 0$ we know that $J_\theta^I(s) > 0$, and $0 > 2\delta b(s) > -\frac{\eta-\beta A^\star}{\eta} > -1$, we deduce that the coefficient of f^2 in (A30) is positive, while that of f in (A30) is negative. Hence, there exists $\bar{f} > 0$ such that $I_\theta^I(d,f) > 0$ if and only if $f > \bar{f}$, and $I_\theta^I(d,f) < 0$ if and only if $0 < f < \bar{f}$.

□

Proof of Proposition B.1. Prior to the end of liquidation the market maker faces, at any point in time, the same average amount of time left. This is due to the memoryless property of the exponential distribution of duration. Specifically, the market maker treats the duration of the liquidation as an exponential random variable with mean d. Hence, before d, the value function and the strategy of each market maker, as well as the price policies, can be obtained by taking the expectation of the corresponding quantities in Proposition 2 and assuming the distribution of time remaining till end of liquidation is $\mathrm{Exp}(\nu)$, $\nu = \frac{1}{d}$, in the event $\{t \le d\}$. In particular, one can show that

$$\tilde{G}_\theta = -\frac{\eta - \beta A^\star}{2\delta A^\star}\frac{N}{N+1}\frac{A^\star}{\nu A^\star + \eta} \equiv -\frac{4A^\star}{(N+1)^2}\frac{N}{N+1}\frac{A^\star}{\nu A^\star + \eta}, \tag{A31}$$

$$\tilde{I}_\theta(f) = \int_0^\infty \nu e^{-\nu u} I_\theta(u, f)\mathrm{d}u, \tag{A32}$$

$$\tilde{J}_\theta = -\frac{\eta - \beta A^\star}{2\delta A^\star}\frac{A^\star}{\nu A^\star + \eta} \equiv -\frac{4A^\star}{(N+1)^2}\frac{A^\star}{\nu A^\star + \eta}. \tag{A33}$$

Here, the last equality in (A31) and (A33) are due to (A10). In particular, notice that the bid and ask prices both jump to the corresponding quantities in the baseline case at the moment the large seller stops trading, because the term $\tilde{G}_\theta$ is constant and thus does not converge to zero as time progresses towards d as in the case of known duration. After time d, there is no liquidation. Hence, the value function and the strategy of any market maker, as well as the corresponding price policies, are those given in Proposition 1. The only salient difference is that the time remaining until the large seller leaves is no longer a state variable because a market maker does not know it if the duration of the order flow of the large seller is unknown. For example, the coefficient $\tilde{J}_\theta$ in the value differential across the setting with unknown and known order duration does not depend on $(d-t)$. It however varies with other parameters in the same way as $J_\theta(d-t)$ does. The same observation holds for $\tilde{I}_\theta$. All these differences reflect what the lack of common knowledge about order duration does to the market. □

Proof of Proposition B.2. Proposition B.2 echoes Proposition 3. The proof follows using similar arguments and is omitted here. □

Proof of Proposition B.3 The proof for Proposition B.3 for the unpredictable duration mirrors mirrors the proof of Proposition 4. After standard calculations, one obtains that the end-user investor's surplus, denoted by $\tilde{v}^I(i,t)$, is given by

$$\tilde{v}^I(i,t) = v^I(i) + 1_{\{t\le d\}}\left(\tilde{I}^I_\theta(f) + \tilde{J}^I_\theta f\, i\right),$$

where

$$\tilde{J}^I_\theta = \frac{\nu}{\nu+\beta+M}\int_0^\infty e^{-\nu s}\cdot\frac{2}{N}\left(A_\omega + \frac{(N^2A^\star-(N+1)A_\omega)(1-2\delta\lambda N b(s))}{(N+1+2\delta A^\star)^2}\right)\mathrm{d}s. \tag{A34}$$

The latter quantity can be shown (as above for the proof of Proposition 4) to be a positive constant, and

$$\begin{aligned}\tilde{I}^I_\theta(f) =& -f\frac{\omega(N^2+2\delta A_\omega)}{N(N+1+2\delta A^\star)^2}\frac{\nu}{\nu+\beta} + f^2\frac{\nu}{\nu+\beta}\int_0^\infty e^{-\nu s}\frac{(1+\frac{2\delta A^\star}{N}+2\delta\lambda b(s))J^I_\theta(s)}{N+1+2\delta A^\star}\mathrm{d}s\\ &+ f^2\frac{\nu}{\nu+\beta}\int_0^\infty e^{-\nu s}\frac{(N^2+2\delta A_\omega)(1-2\delta\lambda N b(s)+2\delta^2\lambda^2N^2b^2(s))}{2\delta\lambda N^2(N+1+2\delta A^\star)^2}\mathrm{d}s,\end{aligned} \tag{A35}$$

which is again a convex, quadratic function of f with a negative linear coefficient. Hence, we obtain qualitatively the same result as in the case when the duration is fully known. $\square$

Proof of Lemma B.1. Following the same argument as in the proof of Lemma 1, one can obtain that $i_{jt} = f\times\tilde{g}(t)$, where

$$\tilde{g}(t) = \begin{cases}\frac{N+2\delta A^\star}{N}\frac{1}{4\delta\lambda A^\star}\frac{\nu+\beta}{\nu+\frac{\eta}{A^\star}}(1-e^{-Mt}), & t\le d,\\ g(d)e^{-M(t-d)}, & t>d,\end{cases}$$

and $M = \frac{4\delta\lambda A^\star}{N+1+2\delta A^\star}$. This completes the proof. $\square$

Proof of Lemma B.2 By combining (A2) with (A23), we obtain that, for $t\le d$, the bid price is equal to

$$\frac{-\omega(1+2\delta A^\star)}{N+1+2\delta A^\star} - \frac{f\left[2+(N+2\delta A^\star)\left(\frac{\eta-\beta A^\star}{\eta+\nu A^\star}+\frac{1}{2A^\star}\frac{\beta A^\star+\nu A^\star}{\eta+\nu A^\star}(1-e^{-Mt})\right)\right]}{2\delta\lambda(N+1+2\delta A^\star)}, \tag{A36}$$

for $t>d$, the bid price is equal to

$$\frac{-\omega(1+2\delta A^\star)}{N+1+2\delta A^\star} - \frac{f(N+2\delta A^\star)\left(\frac{1}{2A^\star}\frac{\beta A^\star+\nu A^\star}{\eta+\nu A^\star}(1-e^{-Md})\right)}{2\delta\lambda(N+1+2\delta A^\star)}e^{-M(t-d)},$$

where $M = \frac{4\delta\lambda A^\star}{N+1+2\delta A^\star}$. The result for the ask price follows from the fact that the bid-ask spread is a constant before and after d. $\square$

D. Equilibrium Strategies Relaxing the Symmetric Response Assumption

We develop an extension of the model presented in the main body. Therein, we have imposed as a constraint that the value function of each market maker is quadratic, and that each market maker adopts the same linear strategy. Such a strategy is obtained in Proposition 1 by maximizing the Hamiltonian. Economically, this means that at any time t, each market maker solves for the optimal tradeoff between maximizing gains from trades in the infinitesimal period $[t, t + dt]$ and the continuation value at the post-trading inventory prevailing at $t + dt$, net of the inventory cost incurred in the period $[t, t + dt]$.

In this section, we consider the two market maker case. We maintain the assumption that the value function of each market maker is quadratic, but no longer impose the constraint that the strategy of market makers are identical. Rather, the strategy of each market maker is a linear best response function to the strategy of the other market maker. However, to maintain the analysis tractable we can only solve the problem in a "static" manner. This means that, at the outset (time zero), each market maker determines once and for all an optimal strategy that he will follow from time 0 to ∞. This strategy is designed to best respond, on average, to the strategies of other market makers over the same infinite time period. We illustrate the result in case of two market makers, and in a baseline setting without large seller.

Denote by i_{1t} and i_{2t} the inventories of market maker 1 and 2 at time t. Each market maker follows a buy and sell strategy which is linear in his own inventory: for positive numbers $a_0, a_1, b_0, b_1 > 0$, the amount marker maker 1 sells at time t is

$$q_{1t}^a = a_0 + a_1 x_t, \tag{A37}$$

and the amount market maker 1 buys at time t is

$$q_{1t}^b = a_0 - a_1 x_t. \tag{A38}$$

Similarly, the amount marker maker 2 sells at time t is

$$q_{2t}^a = b_0 + b_1 x_t, \tag{A39}$$

and the amount market maker 1 buys at time t is

$$q_{2t}^b = b_0 - b_1 x_t. \tag{A40}$$

Above, we have assumed that the supply and demand functions of each market maker are symmetric in the market maker's inventory. This is justified by the fact that each market maker has no preference in taking long or short position of the security, and that each market maker prefers to remain at zero inventory if his current inventory level is zero. Under these strategies, the inventory dynamics of the two market makers are described by the following ODE:

$$di_{1t} = \; -(a_0 + a_1 i_{1t})dt + (a_0 - a_1 i_{1t})dt, \tag{A41}$$

$$di_{2t} = \; -(b_0 + b_1 i_{2t})dt + (b_0 - b_1 i_{2t})dt. \tag{A42}$$

It can be easily seen that the solution to the above ODE is given by

$$i_{1t} \;=\; e^{-2a_1 t} i_{10}, \tag{A43}$$

$$i_{2t} \;=\; e^{-2b_1 t} i_{20} \tag{A44}$$

Fixing a_0, a_1, b_0, b_1, and assume that two market makers begin at same inventory levels $i_{10} = i_{20}$. Then one can explicitly compute the value of these market makers' strategies. In particular, the time zero value of the total inventory cost incurred by market maker 1 is given by

$$\int_0^\infty e^{-\beta t} \eta i_{1t}^2 dt = \eta x_0^2 \int_0^\infty e^{-\beta t - 4a_1 t} dt = \frac{\eta i_{10}^2}{\beta + 4a_1} \tag{A45}$$

The time zero value of market maker 1's proceeds from the sale is

$$\int_0^\infty e^{-\beta t}[\omega - \frac{1}{\delta}(a_0 + b_0 + a_1 i_{1t} + b_1 i_{2t})](a + a_1 i_{1t})dt$$
$$= [\omega - \frac{1}{\delta}(a_0 + b_0)](\frac{a_0}{\beta} + \frac{a_1 i_{10}}{\beta + 2a_1}) - \frac{a_0 a_1 i_{10}}{\delta}\frac{1}{\beta + 2a_1} - \frac{a_0 b_1 i_{10}}{\delta}\frac{1}{\beta + 2b_1} - \frac{a_1^2 i_{10}^2}{\delta}\frac{1}{\beta + 4a_1} \tag{A46}$$
$$- \frac{a_1 b_1 i_{10}^2}{\delta}\frac{1}{\beta + 2a_1 + 2b_1} \tag{A47}$$

and his proceeds from buying are given by

$$\int_0^\infty e^{-\beta t}[\omega - \frac{1}{\delta}(a_0 + b_0 - a_1 x_t - b_1 i_{2t})](a - a_1 i_{1t})dt$$
$$= [\omega - \frac{1}{\delta}(a_0 + b_0)](\frac{a_0}{\beta} - \frac{a_1 i_{10}}{\beta + 2a_1}) + \frac{a_0 a_1 i_{10}}{\delta}\frac{1}{\beta + 2a_1} + \frac{a_0 b_1 i_{10}}{\delta}\frac{1}{\beta + 2b_1} - \frac{a_1^2 i_{10}^2}{\delta}\frac{1}{\beta + 4a_1} \tag{A48}$$
$$- \frac{a_1 b_1 i_{10}^2}{\delta}\frac{1}{\beta + 2a_1 + 2b_1} \tag{A49}$$

Combining (A45), (A47) and (A49)

$$v_1(a_0, a_1; b_0, b_1) = 2[\omega - \frac{1}{\delta}(a_0 + b_0)]\frac{a_0}{\beta} - \left(\frac{2a_1^2 + \delta\eta}{\beta + 4a_1} + \frac{2a_1 b_1}{\beta + 2a_1 + 2b_1}\right)\frac{i_{10}^2}{\delta}. \tag{A50}$$

Fix b_0, b_1. The best response for market maker 1 is a pair of (a_0, a_1) that solves the following system

$$\begin{cases} 2a_0 + b_0 = \delta\omega \\ \dfrac{4\beta a_1 + 8a_1^2 - 4\delta\eta}{(\beta + 4a_1)^2} + \dfrac{2b_1(\beta + 2b_1)}{(\beta + 2a_1 + 2b_1)^2} = 0 \end{cases}. \tag{A51}$$

Similarly, for fixed a_0, a_1 are given, the best response for market maker 2 is a pair (b_0, b_1) that solves the following system

$$\begin{cases} a_0 + 2b_0 = \delta\omega \\ \dfrac{4\beta b_1 + 8b_1^2 - 4\delta\eta}{(\beta + 4b_1)^2} + \dfrac{2a_1(\beta + 2a_1)}{(\beta + 2a_1 + 2b_1)^2} = 0 \end{cases}. \tag{A52}$$

A Nash equilibrium is defined as a system of pairs (a_0, a_1) and (b_0, b_1) that solves (A51) and (A52) simultaneously. First, we observe that the intercepts of the linear strategies need to satisfy

$$a_0 = b_0 = \frac{\delta\omega}{3}. \tag{A53}$$

Hence, the intercept of the market makers' buy and sell strategies is the same as the one in the main body of the paper, and given in Proposition 1. We next look for a symmetric equilibrium, and thus impose that the slopes of the two market makers' strategies are the same, i.e., $a_1 = b_1$. This yields a quadratic equation for a_1:[25]

$$6a_1^2 + 3\beta a_1 - 2\delta\eta = 0. \tag{A54}$$

Solving the above we obtain that $a_1 = \frac{\sqrt{9\beta^2+48\delta\eta}-3\beta}{12} = \frac{4\delta\eta}{\sqrt{9\beta^2+48\delta\eta}+3\beta}$.

We next compare the equilibrium strategies obtained above with the equilibrium strategies in Eq. (A3) in the main body. A direct comparison between the expressions shows that the slope of the buy and sell strategies is higher than the one in Eq. (A3).

Recall from Proposition 1 in the main body and its corresponding proof that the slope D_θ of the equilibrium strategy is $\frac{2\delta}{3}\frac{\sqrt{\beta^2+\frac{32}{9}\delta\eta}-\beta}{16\delta/9} = \frac{4\delta\eta}{\sqrt{9\beta^2+32\delta\eta}+3\beta}$, which is smaller than the slope $\frac{4\delta\eta}{\sqrt{9\beta^2+48\delta\eta}+3\beta}$ obtained above.

The reason for this discrepancy can be understood as follows. In the equilibrium setup considered in the main body, the inventories of all market makers are identically at all time, because we impose as a constraint that they follow the same linear strategy. This inventory co-movement implies that they all have the same sensitivity of inventory holdings. By contrast, in the formulation considered in this appendix, inventories do not need to stay the same at all times: if two market makers with the same initial inventories choose different linear strategies at time zero and then always trade according to them, their inventories after time 0 will deviate from each other. As a result, each market maker would worry more of being in a disadvantageous position relative to the other in terms of inventory holdings. Hence, the buy and sell strategy of each market maker is more sensitive to his current

[25]Notice that this is different from the setting in the main body of the manuscript. Therein, we have constrained the market makers' strategies to be the same. Here, instead, we let each market maker to best respond to the linear strategy of the others (with possibly different intercept and slope), and then we look for the symmetric equilibria.

inventory level, compared to the setting in the main body.

The higher sensitivity to the inventory of each market maker in this setting is also reflected in the speed at which his inventory revert to the zero level. Compare the exponential decay in Eq. (A44) with that of the inventory in the equilibrium setting considered in the main body of the paper (see Eq. (A16) therein and set $f = 0$). Because $a_1 > D_\theta$, it follows immediately that the inventory reverts faster when one consider the expanded strategy space in this section.

D. References

Aït-Sahalia, Yacine and Mehmet Saglam. 2024. "High Frequency Market Making: The Role of Speed." *Journal of Econometrics* 239 (2):105421.

Almgren, Robert and Neil Chriss. 2001. "Optimal Execution of Portfolio Transactions." *Journal of Risk* 3:5–40.

Amihud, Yakov and Haim Mendelson. 1980. "Dealership Market: Market-Making with Inventory." *Journal of Financial Economics* 8:31–53.

Amihud, Yakov, Haim Mendelson, and Lasse H. Pedersen. 2005. "Liquidity and Asset Prices." *Foundations and Trends in Finance* 1:269–364.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman. 2012. "Performance of Institutional Trading Desks: An Analysis of Persistence in Trading Costs." *Review of Financial Studies* 25:557–598.

———. 2013. "Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis." *Journal of Financial Economics* 108:773–797.

Bank, Peter, Ibrahim Ekren, and Johannes Muhle-Karbe. 2021. "Liquidity in Competitive Dealer Markets." *Mathematical Finance* 31:825–1095.

Bertsimas, Dimitris and Andrew W. Lo. 1998. "Optimal Control of Execution Costs." *Journal of Financial Markets* 1:1–50.

BIS–Market Committee. 2020. "FX execution algorithms and market functioning." Manuscript, Bank for International Settlements.

Boulatov, Alex, Dan Bernhardt, and Ivan Larionov. 2016. "Predatory and Defensive Trading in a Dynamic Model of Optimal Execution by Multiple Traders." Manuscript, Higher School of Economics (Moscow).

Brough, Tyler Jon. 2010. "An Empirical Analysis of Institutional Liquidity Trading." Phd thesis, University of Arizona.

Brunnermeier, Markus K. and Lasse H. Pedersen. 2005. "Predatory Trading." *Journal of Finance* 60:1825–1863.

Capponi, Agostino, Paul Glasserman, and Marko Weber. 2020. "Swing Pricing for Mutual Funds: Breaking the Feedback Loop between Fire Sales and Fund Redemptions." *Management Science* 66 (8):3581–3602.

Choi, Jin Hyuk, Kasper Larsen, and Duane J. Seppi. 2018. "Smart TWAP Trading in Continuous-Time Equilibria." Manuscript, Carnegie Mellon University.

———. 2019. "Information and Trading Targets in a Dynamic Market Equilibrium." *Journal of Financial Economics* 132:22–49.

Comerton-Forde, Carole, Terrence Hendershott, Charles M. Jones, Mark S. Seasholes, and Pamela C. Moulton. 2010. "Time Variation in Liquidity: The Role of Market Maker Inventories and Revenues." *Journal of Finance* :295–331.

Duffie, D. and S. Antill. 2021. "Augmenting Markets with Mechanisms." *Review of Economic Studies* 88 (4):1665–1719.

ESMA. 2019. "Guidelines on liquidity stress testing in UCITS and AIFs." Manuscript, European Securities and Markets Authority.

Fardeau, Vincent. 2019. "Strategic Trading around Anticipated Supply/Demand Shocks." Manuscript, National Research University Higher School of Economics.

French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing." *Journal of Finance* 63:1537–1573.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vlasiliki Plerou, and H.Eugene Stanley. 2006. "Institutional Investors and Stock Market Volatility." *Quarterly Journal of Economics* 121:461–504.

Gatheral, Jim and Alexander Scheid. 2011. "Optimal Trade Execution under Geometric Brownian Motion in the Almgren and Chriss Framework." *International Journal of Theoretical and Applied Finance* 14:353–368.

Gayduk, R. and Sergey Nadtochiy. 2018. "Liquidity effects of trading frequency." *Mathematical Finance* 28(3):839–876.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan. 2005. "Equilibrium in a Dynamic Limit Order Market." *Journal of Finance* 60(5):2149–2192.

Grinold, Richard C. and Ronald N. Kahn. 1999. *Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling.* McGraw-Hill Education.

Grossman, Sanford J. and Merton H. Miller. 1988. "Liquidity and Market Structure." *Journal of Finance* 43:617–633.

Hendershott, Terrence and Albert J. Menkveld. 2014. "Price Pressures." *Journal of Financial Economics* 114:405–423.

Ho, Thomas and Hans R. Stoll. 1981. "Optimal Dealer Pricing under Transaction Cost and Return Uncertainty." *Journal of Financial Economics* 9:47–73.

Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie. 2018. "Institional Trading and Abel Noser Data." *Journal of Corporate Finance* 52:143–167.

Huberman, Gur and Werner Stanzl. 2005. "Optimal Liquidity Trading." *Review of Finance* 9:165–200.

Investment Technology Group (ITG). 2007. "ITG ACE - Agency Cost Estimator: A Model Description." Manuscript, ITG.

Liu, Hong and Yajun Wang. 2016. "Market Making with Asymmetric Information and Inventory Risk." *Journal of Economic Theory* 163:73–109.

Madhavan, Ananth and Seymour Smidt. 1993. "An Analysis of Changes in Specialist Inventories and Quotations." *Journal of Finance* 48:1595–1628.

Madhavan, Ananth N. 2000. "Market Microstructure: A Survey." *Journal of Financial Markets* 3(3):205–258.

van Kervel, Vincent, Amy Kwan, and Joakim Westerholm. 2018. "Order Splitting and Searching for a Counterparty." Manuscript, Pontificia Universidad Católica de Chile.

Menkveld, Albert J. 2013. "High Frequency Trading and the New-Market Makers." *Journal of Financial Markets* 16:712–740.

MSCI. 2019. "LiquidityMetrics." Report, MSCI.

Obizhaeva, Anna A. and Jiang Wang. 2013. "Optimal Trading Strategy and Supply/Demand Dynamics." *Journal of Financial Markets* 16:1–32.

Pritsker, Matthew. 2009. "Large Investors: Implications for Equilibrium Asset Returns, Shock Absorption, and Liquidity." Manuscript, Federal Reserve Board.

Puckett, Andy and Xuemin (Sterling) Yan. 2011. "The Interim Trading Skills of Institutional Investors." *Journal of Finance* 66:601–633.

Rostek, Marzena and Marek Weretka. 2015. "Dynamic Thin Markets." *Review of Financial Studies* 28:2946–2992.

SEC. 2016. "Investment Company Liquidity Risk Management Programs." Release No. 33–10233.

van Kervel, Vincent and Albert J. Menkveld. 2019. "High-Frequency Trading around Large Institutional Orders." *Journal of Finance* 74:1091–1137.

Vayanos, Dimitri. 2001. "Strategic Trading in a Dynamic Noisy Market." *Journal of Finance* 56:131–171.

Vayanos, Dimitri and Jiang Wang. 2013. "Market Liquidity: Theory and Empirical Evidence." In *Handbook of the Economics of Finance*, edited by George Constantinides, Milton Harris, and René Stulz. Amsterdam: Elsevier Press, 1289–1361.

Weill, Pierre-Olivier. 2007. "Leaning Against the Wind." *Review of Economic Studies* 74:1329–1354.

Zarinelli, Elia, Michele Treccani, J. Doyne Farmer, and Fabrizio Lillo. 2015. "Beyond the Square Root: Evidence for Logarithmic Dependence of Market Impact on Size and Participation Rate." *Market Microstructure and Liquidity* 1:1–32.