Dynamical models of market impact and algorithms for order execution

Jim Gatheral*, Alexander Schied‡§

First version: December 5, 2011
This version: January 24, 2013

Abstract

In this review article, we present recent work on the regularity of dynamical market impact models and their associated optimal order execution strategies. In particular, we address the question of the stability and existence of optimal strategies, showing that in a large class of models, there is price manipulation and no well-behaved optimal order execution strategy. We also address issues arising from the use of dark pools and predatory trading.

1 Introduction

Market impact refers to the fact that the execution of a large order influences the price of the underlying asset. Usually, this influence results in an adverse effect creating additional execution costs for the investor who is executing the trade. In some cases, however, generating market impact can also be the primary goal, e.g., when certain central banks buy government bonds in an attempt to lower the corresponding interest rates.

Understanding market impact and optimizing trading strategies to minimize market impact has long been an important goal for large investors. There is typically insufficient liquidity to permit immediate execution of large orders without eating into the limit order book. Thus, to minimize the cost of trading, large trades are split into a sequence of smaller trades, which are then spread out over a certain time interval.

The particular way in which the execution of an order is scheduled can be critical, as is illustrated by the “Flash Crash” of May 6, 2010. According to CFTC-SEC (2010), an important contribution in triggering this event was the extremely rapid execution of a larger order of certain futures contracts. Quoting from CFTC-SEC (2010):

*Baruch College, CUNY, jim.gatheral@baruch.cuny.edu
‡University of Mannheim, 68131 Mannheim, Germany. schied@uni-mannheim.de
§Support by Deutsche Forschungsgemeinschaft is gratefully acknowledged.
... a large Fundamental Seller [...] initiated a program to sell a total of 75,000 E-Mini contracts (valued at approximately $4.1 billion). [...] On another occasion it took more than 5 hours for this large trader to execute the first 75,000 contracts of a large sell program. However, on May 6, when markets were already under stress, the Sell Algorithm chosen by the large Fundamental Seller to only target trading volume, and not price nor time, executed the sell program extremely rapidly in just 20 minutes.

To generate order execution algorithms, one usually starts by setting up a stochastic market impact model that describes both the volatile price evolution of assets and how trades impact the market price as they are executed. One then specifies a cost criterion that can incorporate both the liquidity costs arising from market impact and the price risk resulting from late execution. Optimal trading trajectories, which are the basis for trading algorithms, are then obtained as minimizers of the cost criterion among all trading strategies that liquidate a given asset position within a given time frame. Some such models admit an optimal order execution strategy. In others, an optimal strategy does not exist or shows unstable behavior.

In this review, we describe some market impact models that appear in the literature and discuss recent work on their regularity. The particular notions of regularity are introduced in the subsequent Section 2. In Section 3, we discuss models with temporary and permanent price impact components such as the Almgren–Chriss or Bertsimas–Lo models. In Section 4, we introduce several recent models with transient price impact. Extended settings with dark pools or several informed agents are briefly discussed in Section 5.

2 Price impact and price manipulation

The phenomenon of price impact becomes relevant for orders that are large in comparison to the instantaneously available liquidity in markets. Such orders cannot be executed at once but need to be unwound over a certain time interval $[0, T]$ by means of a dynamic order execution strategy. Such a strategy can be described by the asset position $X_t$ held at time $t \in [0, T]$. The initial position $X_0$ is positive for a sell strategy and negative for a buy strategy. The condition $X_T + = 0$ assures that the initial position has been unwound by time $T$. The path $X = (X_t)_{t \in [0, T]}$ will be nonincreasing for a pure sell strategy and nondecreasing for a pure buy strategy. A general strategy may consist of both buy and sell trades and hence can be described as the sum of a nonincreasing and a nondecreasing strategy. That is, $X$ is a path of finite variation. See Lehalle (2012) for aspects of the actual order placement algorithm that will be based on such a strategy.

A market impact model basically describes the quantitative feedback of such an order execution strategy on asset prices. It usually starts by assuming exogenously given asset price dynamics $S^0 = (S^0_t)_{t \geq 0}$ for the case when the agent is not active, i.e., when $X_t = 0$ for all $t$. It is reasonable to assume that this unaffected price process $S^0$ is a semimartingale on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ and that all order execution strategies must be predictable with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. When the strategy $X$ is used, the price is changed from $S^0_t$ to $S^X_t$, and each market impact model has a particular way of describing this change.

Typically, a pure buy strategy $X$ will lead to an increase of prices, and hence to $S^X_t \geq S^0_t$ for $t \in [0, T]$, while a pure sell strategy will decrease prices. This effect is responsible for
the liquidation costs that are usually associated with an order execution strategy under price impact. These costs can be regarded as the difference of the face value $X_0 S_0^0$ of the initial asset position and the actually realized revenues. To define these revenues heuristically, let us assume that $X_t$ is continuous in time and that $S_t^X$ depends continuously on the part of $X$ that has been executed by time $t$. Then, at each time $t$, the infinitesimal amount of $-dX_t$ shares is sold at price $S_t^X$. Thus, the total revenues obtained from the strategy $X$ are

$$\mathcal{R}_T(X) = -\int_0^T S_t^X \, dX_t,$$

and the liquidation costs are

$$\mathcal{C}_T(X) = X_0 S_0^0 - \mathcal{R}_T(X).$$

When $X$ is not continuous in time it may be necessary to add correction terms to these formulas.

The problem of optimal order execution is to maximize revenues—or, equivalently, to minimize costs—in the class of all strategies that liquidate a given initial position of $X_0$ shares during a predetermined time interval $[0, T]$. Optimality is usually understood in the sense that a certain risk functional is optimized. Commonly used risk functionals involve expected value as in Bertsimas & Lo (1998), Gatheral (2010) and others, mean-variance criteria as in Almgren & Chriss (1999, 2000), expected utility as in Schied & Schöneborn (2009) and Schöneborn (2011), or alternative risk criteria as in Forsyth, Kennedy, Tse & Windclif (2012) and Gatheral & Schied (2011).

This brings us to the issue of regularity of a market impact model. A minimal regularity condition is the requirement that the model does admit optimal order execution strategies for reasonable risk criteria. Moreover, the resulting strategies should be well-behaved. For instance, one would expect that an optimal execution strategy for a sell order $X_0 > 0$ should not involve intermediate buy orders and thus be a nonincreasing function of time (at least as long as market conditions stay within a certain range). To make such regularity conditions independent of particular investors preferences, it is reasonable to formulate them in a risk-neutral manner, i.e., in terms of expected revenues or costs. In addition, we should distinguish the effects of price impact from profitable investment strategies that can arise via trend following. Therefore, we will assume from now on that

$$S^0$$ is a martingale

(1)

when considering the regularity or irregularity of a market impact model. Condition (1) is anyway a standard assumption in the market impact literature, because drift effects can often be ignored due to short trading horizons. We refer to Almgren (2003), Schied (2011), and Lorenz & Schied (2012) for a discussion of the effects that can occur when a drift is added.

The first regularity condition was introduced by Huberman & Stanzi (2004). It concerns the absence of price manipulation strategies, which are defined as follows.

**Definition 1** (Price manipulation). A round trip is an order execution strategy $X$ with $X_0 = X_T = 0$. A price manipulation strategy is a round trip $X$ with strictly positive expected revenues,

$$\mathbb{E}[\mathcal{R}_T(X)] > 0.$$  

(2)
A price manipulation strategy allows price impact to be exploited in a favorable manner. Thus, models that admit price manipulation provide an incentive to implement such strategies, perhaps not even consciously on part of the agent but in hidden and embedded form within a more complex trading algorithm. Moreover, the existence of price manipulation can often preclude the existence of optimal execution strategies for risk-neutral investors, due to the possibility of generating arbitrarily large expected revenues by adding price manipulation strategies to a given order execution strategy. In many cases, this argument also applies to risk-averse investors, at least when risk aversion is small enough.

The concept of price manipulation is clearly related to the concept of arbitrage in derivatives pricing models. In fact, Huberman & Stanzl (2004) showed that, in some models, rescaling and repeating price manipulation can lead to a weak form of arbitrage, called quasi-arbitrage. But there is also a difference between the notions of price manipulation and arbitrage, namely price manipulation is defined as the possibility of average profits, while classical arbitrage is defined in an almost-sure sense. The reason for this difference is the following. In a derivatives pricing model, one is interested in constructing strategies that almost surely replicate a given contingent claim. On the other hand, in a market impact model, one is interested in constructing order execution strategies that are defined not in terms of an almost-sure criterion but as minimizers of a cost functional of a risk averse investor. This fact needs to be reflected in any concept of regularity or irregularity of a market impact model. Moreover, any such concept should be independent of the risk aversion of a particular investor. It is therefore completely natural to formulate regularity conditions for market impact models in terms of expected revenues or costs.

It was observed by Alfonsi, Schied & Slynko (2012) that the absence of price manipulation may not be sufficient to guarantee the stability of a market impact model. There are models that do not admit price manipulation but for which optimal order execution strategies may oscillate strongly between buy and sell trades. This effect looks similar to usual price manipulation, but occurs only when triggered by a given transaction. Alfonsi et al. (2012) therefore introduced the following notion:

**Definition 2** (Transaction-triggered price manipulation). A market impact model admits transaction-triggered price manipulation if the expected revenues of a sell (buy) program can be increased by intermediate buy (sell) trades. That is, there exists $X_0$, $T > 0$, and a corresponding order execution strategy $\tilde{X}$ for which

$$E[R_T(\tilde{X})] > \sup \left\{ E[R_T(X)] \mid X \text{ is a monotone order execution strategy for } X_0 \text{ and } T \right\}.$$ 

Yet another class of irregularities was introduced independently by Klöck, Schied & Sun (2011) and Roch & Soner (2011):

**Definition 3** (Negative expected liquidation costs). A market impact model admits negative expected liquidation costs if there exists $T > 0$ and a corresponding order execution strategy $X$ for which

$$E[C_T(X)] < 0,$$ 

(3)
or, equivalently,
\[ \mathbb{E}[R_T(X)] > X_0 S_0. \]

For round trips, conditions (2) and (3) are clearly equivalent. Nevertheless, there are market impact models that do not admit price manipulation but do admit negative expected liquidation costs. The following proposition, which is taken from Klöck et al. (2011), explains the general relations between the various notions of irregularity we have introduced so far.

**Proposition 1.**  
(a) Any market impact model that does not admit negative expected liquidation costs does also not admit price manipulation.

(b) Suppose that asset prices are decreased by sell orders and increased by buy orders. Then the absence of transaction-triggered price manipulation implies that the model does not admit negative expected liquidation costs. In particular, the absence of transaction-triggered price manipulation implies the absence of price manipulation in the usual sense.

3 Temporary and permanent price impact

In one of the earliest market impact model classes that has so far been proposed, and which has also been widely used in the financial industry, one distinguishes between the following two impact components. The first component is temporary and only affects the individual trade that has also triggered it. The second component is permanent and affects all current and future trades equally.

3.1 The Almgren–Chriss model

In the Almgren–Chriss model, order execution strategies \((X_t)_{t \in [0,T]}\) are assumed to be absolutely continuous functions of time. Price impact of such strategies acts in an additive manner on unaffected asset prices. That is, for two nondecreasing functions \(g, h : \mathbb{R} \to \mathbb{R}\) with \(g(0) = 0 = h(0)\),

\[
S^X_t = S^0_t + \int_0^t g(\dot{X}_s) \, ds + h(\dot{X}_t). \tag{4}
\]

Here, the term \(h(\dot{X}_t)\) corresponds to temporary price impact, while the term \(\int_0^t g(\dot{X}_s) \, ds\) describes permanent price impact. This model is often named after the seminal papers Almgren & Chriss (1999, 2000) and Almgren (2003), although versions of this model appeared earlier; see, e.g., Bertsimas & Lo (1998) and Madhavan (2000).

In this model, the unaffected stock price is often taken as a Bachelier model,

\[
S^0_t = S_0 + \sigma W_t, \tag{5}
\]

where \(W\) is a standard Brownian motion and \(\sigma\) is a nonzero volatility parameter. This choice may lead to negative prices of the unaffected price process. In addition, negative prices may occur from the additive price impact components in (4), e.g., when a large asset position is
sold in a very short time interval. With realistic parameter values, however, negative prices normally occur only with negligible probability.

The revenues of an order execution strategy are given by

\[
R_T(X) = -\int_0^T S^X_t dX_t = -\int_0^T S^0_t dX_t - \int_0^T \dot{X}_t \int_0^t g(\dot{X}_s) ds dt - \int_0^T \dot{X}_t h(\dot{X}_t) dt
\]

where

\[
f(x) = xh(x). \tag{6}
\]

For the particular case \( h = 0 \), the next proposition was proved first by Huberman & Stanzl (2004) in a discrete-time version of the Almgren–Chriss model and by Gatheral (2010) in continuous time.

**Proposition 2.** If an Almgren–Chriss model does not admit price manipulation for all \( T > 0 \), then \( g \) must be linear, i.e., \( g(x) = \gamma x \) with a constant \( \gamma \geq 0 \).

**Proof.** For the case in which \( g \) is nonlinear and \( h \) vanishes, Gatheral (2010) constructed a deterministic round trip \((X^1_t)_{0 \leq t \leq T} \) such that \( \int_0^T \dot{X}^1_t \int_0^t g(\dot{X}^1_s) ds dt < 0 \) and such that \( \dot{X}^1_t \) takes only two values. For \( \varepsilon > 0 \), we now define

\[
X^\varepsilon_t = \frac{1}{\varepsilon} X^1_{\varepsilon t}, \quad 0 \leq t \leq T_\varepsilon := \frac{1}{\varepsilon} T.
\]

Then \((X^\varepsilon_t)_{0 \leq t \leq T_\varepsilon} \) is again a round trip with \( \dot{X}^\varepsilon_t = \dot{X}^1_{\varepsilon t} \). Since this round trip is bounded, the expectation of the stochastic integral \( \int_0^{T_\varepsilon} X^\varepsilon_t dS^0_t \) vanishes due to the martingale assumption on \( S^0 \). It follows that

\[
\mathbb{E}[R_{T_\varepsilon}(X^\varepsilon)] = -\int_0^{T_\varepsilon} \dot{X}^\varepsilon_t \int_0^t g(\dot{X}^\varepsilon_s) ds dt - \int_0^{T_\varepsilon} f(\dot{X}^\varepsilon_t) dt
\]

\[
= -\int_0^{T/\varepsilon} \dot{X}^1_{\varepsilon t} \int_0^t g(\dot{X}^1_{\varepsilon s}) ds dt - \int_0^{T/\varepsilon} f(\dot{X}^1_{\varepsilon t}) dt
\]

\[
= \frac{1}{\varepsilon^2} \left( -\int_0^{T} \dot{X}^1_t \int_0^t g(\dot{X}^1_s) ds dt - \varepsilon \int_0^{T} f(\dot{X}^1_t) dt \right).
\]

When \( \varepsilon \) is small enough, the term in parentheses will be strictly positive, and consequently \( X^\varepsilon \) will be a price manipulation strategy.

When \( g(x) = \gamma x \) for some \( \gamma \geq 0 \), the revenues of an order execution strategy \( X \) simplify and are given by

\[
R_T(X) = X_0 S_0^0 + \int_0^T X_t dS^0_t - \frac{\gamma}{2} X_0^2 - \int_0^T f(\dot{X}_t) dt.
\]
**Proposition 3.** Suppose that $g(x) = \gamma x$ for some $\gamma \geq 0$ and the function $f$ in (6) is convex. Then for every $X_0 \in \mathbb{R}$ and each $T > 0$ the strategy

$$X_t^* := \frac{X_0}{t}, \quad 0 \leq t \leq T,$$

maximizes the expected revenues $\mathbb{E}[\mathcal{R}_T(X)]$ in the class of all adaptive and bounded order execution strategies $(X_t)_{0 \leq t \leq T}$. 

**Proof.** When $X$ is bounded, the term $\int_0^T X_t dS_t^0$ has zero expectation. Hence, maximizing the expected revenues reduces to minimizing the expectation $\mathbb{E}[\int_0^T f(\dot{X}_t) dt]$ over the class of order execution strategies for $X_0$ and $T$. By Jensen’s inequality, this expectation has $X^*$ as its minimizer.

By means of Proposition 1, the next result follows.

**Corollary 1.** Suppose that $g(x) = \gamma x$ for some $\gamma \geq 0$ and the function $f$ in (6) is convex. Then the Almgren–Chriss model is free of transaction-triggered price manipulation, negative expected liquidation costs, and price manipulation.

**Remark 1.**

(a) The strategy $X^*$ in (7) can be regarded as a VWAP strategy, where VWAP stands for *volume-weighted average price*, when the time parameter $t$ does not measure physical time but *volume time*, which is a standard assumption in the literature on order execution and market impact.

(b) The assumptions that $g$ is linear and that $f$ is convex are consistent with empirical observation; see Almgren, Thum, Hauptmann & Li (2005), where it was argued that $f(x)$ is well approximated by a multiple of the power law function $|x|^{1+\beta}$ with $\beta \approx 0.6$.

The Almgren–Chriss model is highly tractable and can easily be generalized to multi-asset situations; see, for example, Konishi & Makimoto (2001) or Schöneborn (2011). Accordingly, it has often been the basis for practical applications as well as for the investigation of optimal order execution with respect to various risk criteria. We now discuss some examples of such studies.

- **Mean-variance optimization** corresponds to maximization of a mean-variance functional of the form

$$\mathbb{E}[\mathcal{R}_T(X)] - \lambda \text{var}(\mathcal{R}_T(X)), \quad (8)$$

where var $(Y)$ denotes the variance with respect to $\mathbb{P}$ of a random variable $Y$ and $\lambda \geq 0$ is a risk aversion parameter. This problem was studied by Almgren & Chriss (1999, 2000), Almgren (2003), and Lorenz & Almgren (2011). The first three papers solve the problem for deterministic order execution strategies, while the latter one gives results on mean-variance optimization over adaptive strategies. This latter problem is much more difficult than the former, mainly due to the time inconsistency of the mean-variance functional. Konishi & Makimoto (2001) study the closely related problem of maximizing the functional for which variance is replaced by standard deviation, i.e., by the square root of the variance.
Expected-utility maximization corresponds to the maximization of
\[ \mathbb{E}[u(\mathcal{R}_T(X))] \]
where \( u : \mathbb{R} \to \mathbb{R} \) is a concave and increasing utility function. In contrast to the mean-variance functional, expected utility is time consistent, which facilitates the use of stochastic control techniques. For the case in which \( S^0 \) is a Bachelier model, this problem was studied in Schied & Schöneborn (2009), Schied, Schöneborn & Tehranchi (2010), and Schöneborn (2011); see also Schöneborn (2008). In these papers, it is shown in particular that the maximization of expected exponential utility over adaptive strategies is equivalent to mean-variance optimization over deterministic strategies.

A time-averaged risk measure was introduced by Gatheral & Schied (2011). Optimal sell order execution strategies for this risk criterion minimize a functional of the form
\[ \mathbb{E}\left[ C_T(X) + \int_0^T (\lambda X_t S^0_t + \nu X^2_t) \, dt \right], \]
where \( \lambda \) and \( \nu \) are two nonnegative constants. Here, optimal strategies may become negative, but this effect occurs only in extreme market scenarios or for values of \( \lambda \) that are too large. On the other hand, when \( h \) is linear, optimal strategies can be computed in closed form, they react on changes in asset prices in a reasonable way, and, as shown in Schied (2011), they are robust with respect to misspecifications of the probabilistic dynamics of \( S^0 \).

### 3.2 The Bertsimas–Lo model

The Bertsimas–Lo model was introduced in Bertsimas & Lo (1998) to remedy the possible occurrence of negative prices in the Almgren–Chriss model. In the following continuous-time variant of the Bertsimas–Lo model, the price impact of an absolutely continuous order execution strategy \( X \) acts in a multiplicative manner on unaffected asset prices:
\[ S^X_t = S^0_t \exp \left( \int_0^t g(\dot{X}_s) \, ds + h(\dot{X}_t) \right), \]
for two nondecreasing functions \( g, h : \mathbb{R} \to \mathbb{R} \) with \( g(0) = 0 = h(0) \) that describe the respective permanent and temporary impact components. The unaffected price process \( S^0 \) is often taken as (risk-neutral) geometric Brownian motion:
\[ S^0_t = \exp \left( \sigma W_t - \frac{\sigma^2}{2} t \right), \]
where \( W \) is a standard Brownian motion and \( \sigma \) is a nonzero volatility parameter. The following result was proved by Forsyth et al. (2012).

**Proposition 4.** When \( g(x) = \gamma x \) for some \( \gamma \geq 0 \), the Bertsimas–Lo model does not admit price manipulation in the class of bounded order execution strategies.
The computation of optimal order execution strategies is more complicated in this model than in the Almgren–Chriss model. We refer to Bertsimas & Lo (1998) for a dynamic programming approach to the maximization of the expected revenues in the discrete-time version of the model. Forsyth et al. (2012) use Hamilton–Jacobi–Bellman equations to analyze order execution strategies that optimize a risk functional consisting of the expected revenues and the expected quadratic variation of the portfolio value process. Kato (2011) studies optimal execution in a related model with nonlinear price impact under the constraint of pure sell or buy strategies.

### 3.3 Further models with permanent or temporary price impact

An early market impact model described in the academic literature is the one by Frey & Stremme (1997). In this model, price impact is obtained through a microeconomic equilibrium analysis. As a result of this analysis, permanent price impact of the following form is obtained:

\[ S^X_t = F(t, X_t, W_t) \]

for a function \( F \) and a standard Brownian motion \( W \). This form of permanent price impact has been further generalized by Baum (2001) and Bank & Baum (2004) by assuming a smooth family \( (S_t(x))_{x \in \mathbb{R}} \) of continuous semimartingales. The process \( (S_t(x))_{t \geq 0} \) is interpreted as the asset price when the investor holds the constant amount of \( x \) shares. The price of a strategy \( (X_t)_{0 \leq t \leq T} \) is then given as

\[ S^X_t = S_t(X_t). \]

The dynamics of such an asset price can be computed via the Itô-Wentzell formula. This analysis reveals that continuous order execution strategies of bounded variation do not create any liquidation costs (Bank & Baum 2004, Lemma 3.2). Since any reasonable trading strategy can be approximated by such strategies (Bank & Baum 2004, Theorem 4.4), it follows that, at least asymptotically, the effects of price impact can always be avoided in this model.

A related model for temporary price impact was introduced by Çetin, Jarrow & Protter (2004). Here, a similar class \( (S_t(x))_{x \in \mathbb{R}} \) of processes is used, but the interpretation of \( x \mapsto S_t(x) \) is now that of a supply curve for shares available at time \( t \). Informally, the infinitesimal order \( dX_t \) is then executed at price \( S_t(dX_t) \). Also in this model, continuous order execution strategies of bounded variation do not create any liquidation costs (Çetin et al. 2004, Lemma 2.1). The model has been extended by Roch (2011) so as to allow for additional price impact components. We also refer to the survey paper Gökay, Roch & Soner (2011) for an overview for further developments and applications of this model class and for other, related models.

### 4 Transient price impact

**Transience** of price impact means that this price impact will decay over time, an empirically well-established feature of price impact as well-described in Moro, Vicente, Moyano, Gerig, Farmer, Vaglica, Lillo & Mantegna (2009) for example.
4.1 Linear transient price impact

One of the first models for linear transient price impact was proposed by Obizhaeva & Wang (2013) for the case of exponential decay of price impact. Within the class of linear price impact models, this model was later extended by Alfonsi et al. (2012) and Gatheral, Schied & Slynko (2012). In this extended model, an order for \(dX_t\) shares placed at time \(t\) is interpreted as market order to be placed in a limit order book, in which \(qd\sigma\) limit orders are available in the infinitesimal price interval from \(s\) to \(s + d\sigma\). In other words, limit orders have a continuous and constant distribution. We also neglect the bid-ask spread (see Alfonsi, Fruth & Schied (2008) and Section 2.6 in Alfonsi & Schied (2010) on how to incorporate a bid-ask spread into this model). If the increment \(dX_t\) of an order execution strategy has negative sign, the order \(dX_t\) will be interpreted as a sell market order, otherwise as a buy market order. This market order will be matched against all limit orders that are located in the price range between \(S_t^X\) and \(S_{t+}^X\), i.e.,

\[
dX_t = \frac{1}{q}(S_{t+}^X - S_t^X);
\]

see Figure 4.1. Thus, the price impact of the order \(dX_t\) is \(S_{t+}^X - S_t^X = q dX_t\). The decay of price impact is modeled by means of a (typically nonincreasing) function \(G: \mathbb{R}_+ \to \mathbb{R}_+\), the decay kernel or resilience function. We assume for the moment that \(q = G(0) < \infty\). Then the price impact created at time \(t\) by the order \(dX_t\) is equal to \(G(0) dX_t\). By some later time \(u > t\), this price impact will have decayed to \(G(u - t) dX_t\). Thus, the price process resulting from an order execution strategy \((X_t)\) is modeled as

\[
dS_t^X = S_0^t + \int_{[0,t]} G(t - s) dX_s. \tag{13}
\]

Figure 1: For a supply curve with a constant density \(q\) of limit buy orders, the price is shifted from \(S_t^X\) to \(S_{t+}^X = S_t^X + q dX_t\) when a market sell order of size \(dX_t < 0\) is executed.
One shows that the expected costs of an order execution strategy are
\[ \mathbb{E}[\mathcal{C}_T(X)] = \frac{1}{2} \mathbb{E} \left[ \int_{[0,T]} \int_{[0,T]} G(|t-s|) \, dX_s \, dX_t \right]; \]
(Gatheral et al. 2012, Lemma 2.3). The next result follows from Bochner’s theorem, which was first formulated in Bochner (1932).

**Proposition 5.** Suppose that \( G \) is continuous and finite. Then the following are equivalent.

(a) The model does not admit negative expected liquidation costs.

(b) \( G \) is positive definite in the sense of Bochner (1932).

(c) \( G(| \cdot |) \) is the Fourier transform of a nonnegative finite Borel measure \( \mu \) on \( \mathbb{R} \).

In particular, the model does not admit price manipulation when these equivalent conditions are satisfied.

It follows from classical results by Carathéodory (1907), Toeplitz (1911), and Young (1913) that \( G(| \cdot |) \) is positive definite in the sense of Bochner if \( G : \mathbb{R}_+ \to \mathbb{R}_+ \) is convex and nondecreasing (see Proposition 2 in Alfonsi et al. (2012) for a short proof). This fact is sometimes also called “Pólya criterion” after Pólya (1949).

When \( G(| \cdot |) \) is positive definite, a deterministic order execution strategy \( X^* \) for which the measure \( dX^*_t \) is supported in a given compact set \( \mathbb{T} \subset \mathbb{R}_+ \) minimizes the expected costs in the class of all bounded order execution strategies supported on \( \mathbb{T} \) if and only if there exists \( \lambda \in \mathbb{R} \) such that \( X^* \) is a measure-valued solution to the following Fredholm integral equation of the first type,
\[ \int_{\mathbb{T}} G(|t-s|) \, dX^*_s = \lambda \quad \text{for all } t \in \mathbb{T}; \]
(14)
see Theorem 2.11 in Gatheral et al. (2012). This observation can be used to compute order execution strategies for various decay kernels. One can also take \( \mathbb{T} \) as a discrete set of time points. In this case, (14) is a simple matrix equation that can be solved by standard techniques. For instance, when taking \( \mathbb{T} = \{ \frac{k}{N}T \mid k = 0, \ldots, N \} \) for various \( N \) and comparing the corresponding optimal strategies for the two decay kernels
\[ G(t) = \frac{1}{(1+t)^2} \quad \text{and} \quad G(t) = \frac{1}{1+t^2} \]
one gets the optimal strategies in Figure 2. In the case of the first decay kernel, which is convex and decreasing, strategies are well behaved. In the case of the second decay kernel, however, strategies oscillate more and more strongly between alternating buy and sell trades. These oscillations become stronger and stronger as the time grid of trading dates becomes finer. That is, there is transaction-triggered price manipulation. But since the function \( G(t) = 1/(1+t^2) \) is positive definite as the Fourier transform of the measure \( \mu(dx) = \frac{1}{2}e^{-|x|}dx \), the corresponding model admits neither negative expected liquidation costs nor price manipulation.
Figure 2: Trade sizes $dX_t^*$ for optimal strategies for the decay kernels $G(t) = 1/(1 + t)^2$ (left column) and $G(t) = 1/(1 + t^2)$ (right column), with equidistant trading dates $t = \frac{k}{N} T$, $k = 0, \ldots, N$. Horizontal axes correspond to time, vertical axes to trade size. We chose $X_0 = -10$, $T = 10$, and $N = 10, 30, 50, 100$. 


So the condition that \( G \) is positive definite does not yet guarantee the regularity of the model.

The following result was first obtained as Theorem 1 in Alfonsi et al. (2012) in discrete time. By approximating continuous-time strategies with discrete-time strategies, this result can be carried over to continuous time, as was observed in Theorem 2.20 of Gatheral et al. (2012).

**Theorem 1.** Let \( G \) be a nonconstant nonincreasing convex decay kernel. Then there exists a unique optimal strategy \( X^* \) for each \( X_0 \) and \( T \). Moreover, \( X^*_t \) is a monotone function of \( t \). That is, there is no transaction-triggered price manipulation.

In (Alfonsi et al. 2012, Proposition 2) it is shown that transaction-triggered price manipulation exists as soon as \( G \) violates the convexity condition in a neighborhood of zero, i.e.,

\[
\text{there are } s, t > 0, s \neq t, \text{ such that } G(0) - G(s) < G(t) - G(t + s).
\] (15)

The oscillations in the right-hand part of Figure 2 suggest that there is no convergence of optimal strategies as the time grid becomes finer. One would expect as a consequence that optimal strategies do not exist for continuous trading throughout an interval \([0, T]\). In fact, it is shown in (Gatheral et al. 2012, Theorem 2.15) that there do not exist order execution strategies minimizing the expected cost among all strategies on \([0, T]\) when \( G(|\cdot|) \) is the Fourier transform of a measure \( \mu \) that has an exponential moment: \[ \int e^{\epsilon x} \mu(dx) < \infty \text{ for some } \epsilon \neq 0. \] Moreover, mean-variance optimization can lead to sign switches in optimal strategies even when \( G \) is convex and decreasing (Alfonsi et al. 2012, Section 7). Being nonincreasing and convex is a monotonicity condition on the first two derivatives of \( G \). When an alternating monotonicity condition is imposed on all derivatives of a smooth decay kernel \( G \), then \( G \) is called completely monotone. Alfonsi & Schied (2012) show how optimal execution strategies for such decay kernels can be computed by means of singular control techniques.

Theorem 1 extends also to the case of a decay kernel \( G \) that is weakly singular in the following sense:

\[
G : (0, \infty) \to [0, \infty) \text{ is nonconstant, nonincreasing, convex, and } \int_0^1 G(t) \, dt < \infty; \quad (16)
\]

see Theorem 2.24 in Gatheral et al. (2012). This includes power-law decay kernels such as \( G(t) = t^{-\gamma} \) with \( 0 < \gamma < 1 \).

**Remark 2.** Dramatic oscillatory effects such as those on the right-hand side of Figure 2 will of course never appear in the practical implementation of an order execution strategy of a single trader. But the following quote from CFTC-SEC (2010) indicates that similar effects can appear in reality through the interaction of the trading algorithms of several high-frequency traders (HFT), and it seems possible that the transience of price impact has a certain role in this.

\[ \ldots \text{HFTs began to quickly buy and then resell contracts to each other—generating a “hot-potato” volume effect as the same positions were rapidly passed back and forth. Between 2:45:13 and 2:45:27, HFTs traded over 27,000 contracts, which accounted for about 49 percent of the total trading volume, while buying only about 200 additional contracts net.} \]
Some results can still be obtained when model parameters are made time dependent or even stochastic. For instance, Alfonsi et al. (2008) consider exponential decay of price impact with a deterministic but time-dependent rate ($\rho_t$): the price impact $q dX_t$ generated at time $t$ will decay to $qe^{-\int_t^u \rho_s \, ds} dX_t$ by time $u > t$. Also this model does not admit transaction-triggered price manipulation (Alfonsi et al. 2008, Theorem 3.1). Fruth, Schöneborn & Urusov (2011), further extend this model by allowing the parameter $q$ to become time-dependent. In this case, the price process $S^X_t$ associated with an order execution strategy $X$ is given by

$$S^X_t = S^0_t + \int_{(0,t)} q_s e^{-\int_s^t \rho_r \, dr} \, dX_s.$$  

Proposition 8.3 and Corollary 8.5 in Fruth et al. (2011) give conditions under which (transaction-triggered) price manipulation does or does not exist. Moreover, it is argued in (Fruth et al. 2011, Proposition 3.4) that ordinary and transaction-triggered price manipulation can be excluded by considering a two-sided limit order book in which buy orders affect mainly the ask side and sell orders affect mainly the bid side, and which has a nonzero bid-ask spread.

4.2 Limit order book models with general shape

The assumption of a constant density of limit orders in the preceding section was relaxed in Alfonsi, Fruth & Schied (2010) by allowing the density of limit orders to vary as a function of the price. Thus, $f(s) \, ds$ limit orders are available in the infinitesimal price interval from $s$ to $s + ds$, where $f : \mathbb{R} \to (0, \infty)$ is called the shape function of the limit order book model. Such a varying shape fits better to empirical observations than a constant shape; see, e.g., Weber & Rosenow (2005). The volume of limit orders that are offered in a price interval $[s, s']$ is then given by $F(s') - F(s)$, where

$$F(x) = \int_0^x f(y) \, dy$$  

is the antiderivative of the shape function $f$. Thus, volume impact and price impact of an order are related in a nonlinear manner. We define the volume impact process $E^X_t$ with time-dependent exponential resilience rate $\rho_t$ as

$$E^X_t = \int_{(0,t)} e^{-\int_s^t \rho_r \, dr} \, dX_s.$$  

The corresponding price impact process $D^X_t$ is defined as

$$D^X_t = F^{-1}(E^X_t),$$  

and the price process associated with the order execution strategy $X$ is

$$S^X_t = S^0_t + D^X_t.$$  

see Figure 4.2.

The following result is taken from Corollary 2.12 in Alfonsi & Schied (2010).
Theorem 2. Suppose that $F(x) \to \pm \infty$ as $x \to \pm \infty$ and that $f$ is nondecreasing on $\mathbb{R}_-$ and nonincreasing on $\mathbb{R}_+$ or that $f(x) = \lambda |x|^\alpha$ for constants $\lambda, \alpha > 0$. Suppose moreover trading is only possible at a discrete time grid $T = \{t_0, t_1, \ldots, t_N\}$. Then the model admits neither standard nor transaction-triggered price manipulation.

Instead of assuming volume impact reversion as in (18), one can also consider a variant of the preceding model, defined via price impact reversion. In this model, we retain the relation (19) between volume impact $E^X_t$ and price impact $D^X_t$, but now price impact decays exponentially:

$$dD^X_t = -\rho_t D^X_t \, dt \quad \text{when } dX_t = 0.$$  

In this setting, a version of Theorem 2 remains true (Alfonsi & Schied 2010, Corollary 2.18).

We also refer to Alfonsi & Schied (2010) for formulas of optimal order execution strategies in discrete time and for their continuous-time limits. A continuous-time generalization of the volume impact version of the model has been introduced by Predoiu, Shaikhet & Shreve (2011). In this model, $F$ may be the distribution function of a general nonnegative measure, which, in view of the discrete nature of real-world limit order books, is more realistic than the requirement (17) of absolute continuity of $F$. Moreover, the resilience rate $\rho_t$ may be a function of $E^X_t$; we refer to Weiss (2009) for a discussion of this assumption. Predoiu et al. (2011) obtain optimal order execution strategies in their setting, but they restrict trading to buy-only or sell-only strategies. So price manipulation is excluded by definition.

There are numerous other approaches to modeling limit order books and to discuss optimal order execution in these models. We refer to Avellaneda & Stoikov (2008), Bayraktar & Ludkovski (2011), Bouchard, Dang & Lehalle (2011), Cont & de Larrard (2010), Cont & de Larrard (2011), Cont, Kukanov & Stoikov (2010), Cont, Stoikov & Talreja (2010), Guéant, Lehalle & Tapia (2012), Kharrouri & Pham (2010), Lehalle, Guéant & Razafinimanana (2011), and Pham, Ly Vath & Zhou (2009).
4.3 The JG model

In the model introduced by Gatheral (2010), an absolutely continuous order execution strategy $X$ results in a price process of the form

$$
S_t^X = S_0 + \int_0^t h(X_s)G(t-s) \, ds.
$$

(21)

Here, $h$ is a nondecreasing impact function and $G : (0, \infty) \to \mathbb{R}_+$ is a decay kernel as in Section 4.1. When $h$ is linear, we recover the model dynamics (13) from Section 4.1. We refer to Gatheral, Schied & Slynko (2011) for discussion of the relations between the model (21) and the limit order book models in Section 4.2. An empirical analysis of this model is given in Lehalle & Dang (2010). The next result is taken from Section 5.2.2 in Gatheral (2010).

**Theorem 3.** Suppose that $G(t) = t^{-\gamma}$ for some $\gamma \in (0, 1)$ and that $h(x) = c|x|^{\delta}\text{sign } x$ for some $c, \delta > 0$. Then price manipulation exists when $\gamma + \delta < 1$.

That it is necessary to consider decay kernels that are weakly singular in the sense of (16), such as power-law decay $G(t) = t^{-\gamma}$, follows from the next result, which is taken from Gatheral et al. (2011).

**Proposition 6.** Suppose that $G(t)$ is finite and continuous at $t = 0$ and that $h : \mathbb{R} \to \mathbb{R}$ is not linear. Then the model admits price manipulation.

The preceding proposition immediately excludes exponential decay of price impact, $G(t) = e^{-\rho t}$ (Gatheral 2010, Section 4.2). It also excludes discrete-time versions of the model (21), because $G(0)$ must necessarily be finite in a discrete-time version of the model. An example is the following version that was introduced by Bouchaud, Gefen, Potters & Wyart (2004); see also Bouchaud (2010):

$$
S_{t_n}^X = S_0^{t_n} + \sum_{k=0}^{n-1} \varepsilon_k G(t_n - t_k)|\xi_{t_k}|^{\delta}\text{sign } \xi_{t_k}
$$

(22)

Here, trading is possible at times $t_0 < t_1 < \cdots$ with discrete trade sizes $\xi_{t_k}$ at time $t_k$, and the $\varepsilon_k$ are positive random variables. The parameter $\delta$ satisfies $0 < \delta < 1$, and $G(t) = c(1+t)^{-\gamma}$. That this model admits price manipulation can either be shown by using discrete-time variants of the arguments in the proof of Proposition 6, or by using (22) as a discrete-time approximation of the model (6).

**Remark 3.** The model of Theorem 3 with $\delta \approx 0.5$ and $\gamma \approx 0.5$ is consistent with the empirical rule-of-thumb that market impact is roughly proportional to the square-root of the trade size and not very dependent on the trading rate. Tóth, Lemprière, Deremble, de Lataillade, Kockelkoren & Bouchaud (2011) verify the empirical success of this simple rule over a very large range of trade sizes and suggest a possible mechanism: The ultimate submitters of large orders are insensitive to changes in price of the order of the daily volatility or less during execution of their orders.

These observations are also not completely inconsistent with the estimate $\beta \approx 0.6$ of Almgren et al. (2005) noted previously in Remark 1.
5 Further extensions

5.1 Adding a dark pool

Recent years have seen a mushrooming of alternative trading platforms called dark pools. Orders placed in a dark pool are not visible to other market participants and thus do not influence the publicly quoted price of the asset. Thus, when dark-pool orders are executed against a matching order, no direct price impact is generated, although there may be certain indirect effects. Dark pools therefore promise a reduction of market impact and of the resulting liquidation costs. They are hence a popular platform for the execution of large orders.

A number of dark-pool models have been proposed in the literature. We mention in particular Laruelle, Lehalle & Pagès (2010), Kratz & Schöneborn (2010), and Klöck et al. (2011). Kratz & Schöneborn (2010) use a discrete-time model and discuss existence and absence of price manipulation in their Section 7. Here, however, we will focus on the model and results of Klöck et al. (2011), because these fit well into our discussion of the Almgren–Chriss model in Section 3.1.

In the extended dark pool model, the investor will first place an order of $\hat{X} \in \mathbb{R}$ shares in the dark pool. Then the investor will choose an absolutely continuous order execution strategy for the execution of the remaining assets at the exchange. The derivative of this latter strategy will be described by a process $(\xi)$. Moreover, until fully executed, the remaining part of the order $\hat{X}$ can be cancelled at a (possibly random) time $\rho < T$. Let

$$Z_t = \sum_{i=1}^{N_t} Y_i$$

denote the total quantity executed in the dark pool up to time $t$, $Y_i$ denoting the size of the $i$th trade and $N_t$ the number of trades up to time $t$. Then the number of shares held by the investor at time $t$ is

$$X_t := X_0 + \int_0^t \xi_s ds + Z^\rho_{t-},$$

where $Z^\rho_{t-}$ denotes the left-hand limit of $Z^\rho_t = Z_{\rho \land t}$. In addition, the liquidation constraint

$$X_0 + \int_0^T \xi_t dt + Z_{\rho} = 0$$

must be $\mathbb{P}$-a.s. satisfied. As in (4), the price at which assets can be traded at the exchange is defined as

$$S_t = S^0_t + \gamma \left( \int_0^t \xi_s ds + \alpha Z^\rho_{t-} \right) + h(\xi_t).$$

Here $\alpha \in [0, 1]$ describes the possible permanent impact of an execution in the dark pool on the price quoted at the exchange. This price impact can be understood in terms of a deficiency in opposite price impact. The price at which the $i$th incoming order is executed in the dark pool will be

$$S^0_{\tau_i} + \gamma \left( \int_0^{\tau_i} \xi_s ds + \alpha Z_{\tau_i-} + \beta Y_i \right) + g(\xi_{\tau_i}) \quad \text{for } \tau_i = \inf \{ t \geq 0 \mid N_t = i \}.$$
In this price, orders executed at the exchange have full permanent impact, but their possible temporary impact is described by a function \( g : \mathbb{R} \rightarrow \mathbb{R} \). The parameter \( \beta \geq 0 \) in (26) describes additional “slippage” related to the dark-pool execution, which will result in transaction costs of the size \( \beta \gamma Y_i^2 \). We assume that \( \alpha \in [0, 1] \), \( \beta \geq 0 \), that \( h \) is increasing, and that \( f(x) := xh(x) \) is convex. We assume moreover that \( g \) either vanishes identically or satisfies the same conditions as \( h \). See Theorem 4.1 in Klück et al. (2011) for the following result, which holds under fairly mild conditions on the joint laws of the sizes and arrival times of incoming matching orders in the dark pool (see Klück et al. (2011) for details).

**Theorem 4.** For given dark-pool parameters, the following conditions are equivalent.

(a) For any Almgren–Chriss model, the dark-pool extension has positive expected liquidation costs.

(b) For any Almgren–Chriss model, the dark-pool extension does not admit price manipulation for every time horizon \( T > 0 \).

(c) The parameters \( \alpha, \beta, \) and \( g \) satisfy \( \alpha = 1, \beta \geq \frac{1}{2} \) and \( g = 0 \).

The most interesting condition in the preceding theorem is the requirement \( \beta \geq \frac{1}{2} \). It means that the execution of a dark-pool order of size \( Y_i \) needs to generate transaction costs of at least \( \frac{1}{2} \gamma Y_i^2 \), which is equal to the costs from permanent impact one would have incurred by executing the order at the exchange. It seems that typical dark pools do not charge transaction costs or taxes of this magnitude. Nevertheless, Theorem 4 requires this amount of transaction costs to exclude price manipulation.

In Theorem 4, it is crucial that we may vary the underlying Almgren-Chriss model. When the Almgren–Chriss model is fixed, the situation becomes more subtle. We refer to Klück et al. (2011) for details.

### 5.2 Multi-agent models

If a financial agent is liquidating a large asset position, other informed agents could try to exploit the resulting price impact. To analyze this situation mathematically, we assume that there are \( n + 1 \) agents active in the market who all are informed about each other’s asset position at each time. The asset position of agent \( i \) will be given as an absolutely continuous order execution strategy \( X_i^t, i = 0, 1, \ldots, n \). Agent 0 ("the seller") has an initial asset position of \( X_0^0 > 0 \) shares that need to be liquidated by time \( T_0 \). All other agents ("the competitors") have initial asset positions \( X_i^0 = 0 \). They may acquire arbitrary positions afterwards but need to liquidate these positions by time \( T_1 \). Assuming a linear Almgren–Chriss model, the asset price associated with these trading strategies is

\[
S_t^X = S_t^0 + \gamma \sum_{i=0}^{n} (X_t^i - X_0^i) + \eta \sum_{i=0}^{n} \dot{X}_t^i. \tag{27}
\]

Consider a competitor who is aware of the fact that the seller is unloading a large asset position by time \( T_0 \). Probably the first guess is that the seller will start shortening the asset
in the beginning of the trading period $[0, T_0]$ and then close the short position by buying back toward the end of the trading period when prices have been lowered by the seller’s pressure on prices. Since such a strategy decreases the revenues of the seller it is called a predatory trading strategy. When such a strategy uses advance knowledge and anticipates trades of the seller, it can be regarded as a market manipulation strategy and classified as illegal front running.

Predatory trading is indeed found to be the optimal strategy by Carlin, Lobo & Viswanathan (2007) when $T_0 = T_1$; see also Brunnermeier & Pedersen (2005). The underlying analysis is carried out by establishing a Nash equilibrium between all agents active in the market. This Nash equilibrium can in fact be given in explicit form. Building on Carlin et al. (2007), Schöneborn & Schied (2009) showed that the picture can change significantly, when the competitors are given more time to close their positions than the seller, i.e., when $T_1 > T_0$. In this case, the behavior of the competitors in equilibrium is determined in a subtle way by the relations of the permanent impact parameter $\gamma$, the temporary impact parameter $\eta$, and the number $n$ of competitors. For instance, it can happen that it is optimal for the competitors to build up long positions rather than short positions during $[0, T_0]$ and to liquidate these during $[T_0, T_1]$. This happens in markets that are elastic in the sense that the magnitude of temporary price impact dominates permanent price impact. That is, the competitors engage in liquidity provision rather than in predatory trading and their presence increases the revenues of the seller. When, on the other hand, permanent price impact dominates, markets have a plastic behavior. In such markets, predatory trading prevails. Nevertheless, it is shown in Schöneborn & Schied (2009) that, for large $n$, the return of the seller is always increased by additional competitors, regardless of the values of $\gamma$ and $\eta$.

References


URL: http://ssrn.com/abstract=1983943


URL: http://ssrn.com/abstract=1712822


URL: http://dx.doi.org/10.1287/opre.1090.0780


URL: http://ssrn.com/paper=1925808


21
URL: http://ssrn.com/paper=1785409


URL: http://ssrn.com/abstract=1344583

URL: http://arxiv.org/abs/0910.1166


URL: http://ssrn.com/abstract=1993103


URL: http://ssrn.com/paper=1923840

URL: http://ssrn.com/paper=1991097


URL: http://ssrn.com/paper=1343985


URL: http://arxiv.org/abs/0904.4131v2