Optimal Trading Execution with Nonlinear Market Impact: An Alternative Solution Method

Massimiliano Marzo
Daniele Ritelli
Paolo Zagaglia

Quaderni - Working Paper DSE N° 797
Optimal Trading Execution with Nonlinear Market Impact: An Alternative Solution Method

Massimiliano Marzo, Daniele Ritelli and Paolo Zagaglia*

This version: November 25, 2011

Abstract

We consider the optimal trade execution strategies for a large portfolio of single stocks proposed by Almgren (2003). This framework accounts for a nonlinear impact of trades on average market prices. The execution strategy of Almgren (2003) is based on the assumption that no shares per unit of time are trade at the beginning of the period. We use a general solution method that accomodates the case of positive initial trades. Our results are twofold. First of all, we show that the problem admits a solution with no trading in the opening period only if additional parametric restrictions are imposed. Second, with positive initial trading, the optimal execution time depends on trading activity in the initial period.

Keywords: optimal execution, market impact, trading strategy.
JEL classification: G11, G12.

*Marzo: Department of Economics, Università di Bologna; massimiliano.marzo@unibo.it. Ritelli: Department of Mathematics, Università di Bologna; daniele.ritelli@unibo.it. Zagaglia: Department of Economics, Università di Bologna; paolo.zagaglia@unibo.it.
This is a work in progress. We welcome any comments, suggestions and remarks.
1 Introduction

The execution of large trades in financial markets requires the balance between risks and costs. The main risk concerns the lack of availability of a counterparty, which can lead to a delay in the execution of a transaction. In order to guarantee a fast trade execution, a trader may incur additional costs. As clarified by Hasbruock and Schwartz (1988), a trader faces a choice between a ‘passive’ and an ‘active’ execution strategy.

Given this background, the available models of optimal execution assume that the trading activity of individual investors has an impact on the average price prevailing in the market. The transaction costs are characterized by parametric forms that replicate stylized facts documented in the market microstructure literature (e.g. see Kraus and Stoll, 1972).

Almgren and Chriss (1999, 2000) and Konishi and Makimoto (2001) provide examples of optimal strategies for the execution problem in the stock market. Their models assume that the transaction cost per share is a linear function of the number of shares traded. The only source of uncertainty consists in the volatility of the stock price.

Almgren (2003) suggests that the linearity assumption is largely at odds with reality. First, the average liquidity premium on stocks tends to be either a convex or a concave function of the traded size. This depends on the counterparty’s perception about the reason for the trade, namely on whether it is driven by liquidity or information needs (see Huang and Stoll, 1997). Moreover, the liquidity premium is related to the risk of finding a counterparty. In other words, the lower the probability of finding a counterparty in the market, the higher the liquidity premium.

In this paper, we review the optimal transaction strategy proposed by Almgren (2003). We show that the solution method used by Almgren (2003) is ill-posed. The reason is that it is based on the assumption that no shares per unit of time are traded at the beginning of the period. We use an approach based on the Gaussian hypergeometric function to solve for the case of positive initial trades. Our results differ strongly from those of Almgren (2003). First of all, the problem admits a solution with no trading in the opening period only if additional parametric restrictions are imposed. Second, with positive initial trading, the optimal execution time depends on trading activity in the initial period.

This note is organized as follows. Section 2 provides a selected discussion of the literature on optimal trade execution. Section 3 outlines the structure of the problem. Section 4 proposes a general solution method for positive initial values of the velocity. Section 5 concludes.
2 Overview of the literature

Rebalancing portfolios of assets requires executing trades in the marketplace. With the advent of algorithmic trading and access availability to many alternative trading venues, investors have dedicated increasing resources to the scheduling of trades. The practical setup of the problem is rather intuitive. An investor has a target number of, say shares that it intends to sell or buy within a given time frame. The decision problem requires to compute how many shares to place or demand in the market at each point in time within the trade horizon. The aim of the investor is to minimize the execution costs. These are typically measured as the difference between the price obtained from the market and a benchmark price for the transaction.

There are multiple relevant dimensions to the execution problem. Several contributions have showed that the liquidity premium is time-varying. The reason is that it is determined by the availability of traders willing to act as counterparties, namely traders willing to buy or sell a given quantity of an asset at a desired price. However, as the presence of traders willing to ‘take the other side’ of a trade is uncertain, any trading is characterized by execution risk.

Another relevant aspect is related to the fact that market illiquidity generates transaction costs. This typically takes the form of a large spread between bid and ask prices (Huang and Stoll, 1997). Therefore, as noticed by Wagner and Banks (1992), the minimization of transaction costs is a key aspect of the portfolio optimization problem. As documented in various studies including Chakravarty (2001), Holthausen, Leftwich and Mayers (1990) and Kraus and Stoll (1972), large trades do impact market prices and, thus affecting the bid-ask spreads.

Asset price volatility is a source for execution risk. The reason is that it affects the probability of finding a suitable counterparty. Hence, it affects the sucessfulness of a trading strategy execution. The recent literature has focused on the specific aspect of volatility, namely the increased uncertainty in execution price incurred by rapid execution of large share blocks. In fact, Hasbrouck and Seppi (2001) show that liquidity fluctuates due to intrinsic variations in market activity independently of trade size.

Based on the considerations outlined earlier, what are the properties of an ‘optimal’ execution strategy? What defines an ‘optimal execution price’? Bertsimas and Lo (1998) argue that ‘best’ execution can be thought of as a dynamic strategy that minimizes ‘liquidation’ costs. They show that dynamic programming techniques can be used fruitfully to characterize these strategies.

Almgren and Chriss (2000) compute optimal trajectories for trading prices that are obtained by balancing market impact costs. The optimal profiles provide a motivation for low execution speed.\footnote{The literature has proposed two main alternative benchmarks. These consist of average prices that materialize within the trading horizon, and are characterized as a time-weighted and volume-weighted average price.} This results arises from the balance between the need to reduce the
expected value of execution, and the need to minimize the adverse effects of market volatility. While the first factor provides a reason for slower execution, the second factor lays the ground for rapid execution. That would, in fact, reduce execution risk in the form of the variance of execution cost with respect to the benchmark price. In short, early execution reduces execution risk, whereas a delayed execution is more geared towards minimizing execution costs. Evidently, the degree of investor risk aversion determines how early within the trading horizon the execution starts. The shape of the schedule instead depends on the form of the assumed market impact model.

Konishi and Makimoto (2001) makes the assumption that the market impact of trading is a linear function of trade size. In this case, the optimality frontier representing the combinations of minimized costs and market volatility has an analytical solution. Value-at-risk utility functions are then used to select the first-best solution. This choice of utility function provides a natural testing ground for the concept of liquidity-adjusted VaR, which explicitly considers the best trade-off between volatility risk and liquidation cost. Almgren (2003) generalizes the results of Konishi and Makimoto (2001) to the case of nonlinear functions for the market impact of trades. In this framework, the assumption is that the market impact cost per share follows a power law function of the trading rate.

3 The optimal execution problem

We follow the general framework of Almgren and Chriss (2000). At time \( t = 0 \), an investor holds \( X \) shares of an asset. The problem is to sell these shares by by time \( t = T \). We should stress that this is the statement of a general framework. In fact, the initial size \( X \) can either be positive or negative. In the first case, the investor needs to schedule a selling program. In the second case, the investor looks at a buying program. In this paper, for simplicity, we focus on the case \( X > 0 \).

The model proposed by Almgren (2003) delivers the following optimization problem:

\[
\begin{aligned}
& \min_{x(t)} \int_0^T F(x, \dot{x}) \, dt \\
& x(0) = X, \quad x(T) = 0
\end{aligned}
\]  

(3.1)

where \( F(x, y) \) is

\[
F(x, y) = -\gamma xy + \eta(-y)^{k+1} + \lambda \sigma^2 x^2, \quad \gamma, \eta, \lambda, k > 0.
\]  

(3.2)

The problem is to determine the optimal function \( x(e) \) so as to minimize a chosen cost

\[ F(x, \dot{x}) - \dot{x} \frac{\partial F}{\partial y}(x, \dot{x}) = \text{constant} \]  

(3.3)
evaluating the constant of integration at the end time \( T \) we are lead to the differential equation

\[ \lambda \sigma^2 - k \eta (-\dot{x})^{k+1} = -k \eta (-\dot{x}(T))^{k+1} \]  

(3.4)

Almgren (2003) proposes a solution based only on the “elementary” case \( v_0 = 0 \). In this note, we show how to compute the constant \( v_0^{k+1} := (-\dot{x}(T))^{k+1} \) using the initial condition \( x(0) = X \).

3.1 The case \( k = 1 \)
The case \( k = 1 \) is straightforward since it gives rise to a linear ordinary differential equation, whose solution is better found starting from the classical Euler Lagrange equation:

\[ \frac{d}{dt} \frac{\partial L}{\partial \dot{y}}(x, \dot{x}) = \frac{\partial L}{\partial x}(x, \dot{x}) \iff 2\lambda \sigma^2 x - 2\eta \dot{x} = 0 \]  

(3.5)

imposing the boundary conditions \( x(0) = X, x(T) = 0 \) we find:

\[ x(t) = X \frac{\sinh \left( \frac{\sqrt{\lambda}}{\sqrt{\eta}} \sigma(T-t) \right)}{\sinh \left( \frac{\sqrt{\lambda}}{\sqrt{\eta}} \sigma T \right)} \]  

(3.6)

It is worth noting that for \( k = 1 \) (3.6) gives a minimizer of (3.1) since in this case Legendre condition reads:

\[ \frac{\partial^2 F}{\partial y^2}(x, \dot{x}) = 2\eta > 0 \]  

(3.7)

Turning back to equation (3.4) for general \( k \), we write it solving with respect to \( \dot{x} \):

\[
\begin{cases}
\dot{x} = -\left( v_0^{k+1} + \frac{\lambda \sigma^2}{k\eta} x^2 \right)^{\frac{1}{k+1}} \\
x(0) = X
\end{cases}
\]  

(3.8)

In this general case observe that, since:

\[ \frac{\partial^2 F}{\partial y^2}(x, \dot{x}) = \eta k(k + 1)(-\dot{x})^{k-1} \]  

(3.9)
since solution to (3.8) is decreasing we infer, for the Legendre condition, the minimality of the extremal \( x(t) \).
4 A general solution method for the case of positive initial trades

In this section, we propose a solution method that holds when there are positive stock trades in the initial period, namely for $v_0 > 0$. This solution $x$ to (3.8) is implicitly defined by

$$
\int_x^X \left( v_0^{k+1} + \frac{\lambda \sigma^2}{k \eta} z^2 \right)^{-\frac{1}{k+1}} dz = t. \tag{4.1}
$$

Integral in the right hand side of (4.1) can be evaluated by means of the Gauss hypergeometric function ${}_2F_1$ whose definition and basic properties are given in the appendix. After some changes of variables which allows to rewrite (4.1) as:

$$
\frac{1}{2v_0} \left\{ X \int_0^1 \frac{s^{-\frac{1}{2}}}{\left( 1 + \frac{\lambda \sigma^2 X^2}{k \eta v_0^{k+1}} s^2 \right)^{\frac{k+1}{2}}} ds - x \int_0^1 \frac{s^{-\frac{1}{2}}}{\left( 1 + \frac{\lambda \sigma^2 x^2}{k \eta v_0^{k+1}} s^2 \right)^{\frac{k+1}{2}}} ds \right\} = t \tag{4.2}
$$

we can finally use the integral representation Theorem, see equation (A.4) in the appendix, for the hypergeometric function to obtain

$$
\frac{1}{v_0} \left\{ X {}_2F_1 \left( \frac{1}{2}, \frac{1}{k+1} | - \frac{\lambda \sigma^2 X^2}{k \eta v_0^{k+1}} \right) - x {}_2F_1 \left( \frac{1}{2}, \frac{1}{k+1} | - \frac{\lambda \sigma^2 x^2}{k \eta v_0^{k+1}} \right) \right\} = t \tag{4.3}
$$

To obtain $x(t)$ from equation (4.3) observe that the function

$$
x \mapsto x {}_2F_1 \left( \frac{1}{2}, \frac{1}{k+1} | - \frac{\lambda \sigma^2 x^2}{k \eta v_0^{k+1}} \right) \tag{4.4}
$$

is strictly decreasing function for values of the independent variable $> 0$ and so is possible to revert and obtain $x(t)$ from (4.3), if needed, numerically or, better, using the Lagrange power series reversion when:

$$
\left| \frac{\lambda \sigma^2 x^2}{k \eta v_0^{k+1}} \right| < 1 \tag{4.5}
$$

But in (4.3) there is no determination of $v_0$ which is essential for the full solution of the problem. If we limit to assign some specific values for $v_0$ as in the “easy case” $v_0 = 0$ we lose control on the initial value $x(0) = X$.

The way to obtain the final velocity in order to fit with the initial value $x(0) = X$ is explained below. We use the so called “shooting method” as presented for instance in Stoer and Burlish (1993) section 7.3.1 pages 502-507. The starting point is the Euler Lagrange
equation with initial values in $T$

$$\begin{align*}
\dot{x} &= -\frac{2\lambda\sigma^2}{\eta k(k + 1)} x(-\dot{x})^{1-k} \\
x(T) &= 0, \quad \dot{x}(T) = -v_0
\end{align*}$$  \hspace{1cm} (4.6)

To solve (4.6) we use the change of variables $u := x$, $y := \dot{x}$ following Murphy (1960) section 2.3 pages 160-161 and, since

$$\ddot{x} = y \frac{dy}{du}$$  \hspace{1cm} (4.7)

equation (4.6) is transformed in

$$\begin{align*}
\frac{dy}{du} &= -\frac{2\lambda\sigma^2}{\eta k(k + 1)} (-y)^{-k} u \\
y(0) &= -v_0
\end{align*}$$  \hspace{1cm} (4.8)

Since (4.8) is separable we can integrate it, so that, returning back to the original variables we find:

$$\begin{align*}
\dot{x} &= -\left(\frac{v_0^{k+1} + \frac{\lambda\sigma^2}{\eta k} x^2}{v_0^{k+1}}\right)^{\frac{1}{k+1}} \\
x(T) &= 0
\end{align*}$$  \hspace{1cm} (4.9)

Equation (4.9) is separable and the relevant integration needs again the hypergeometric integral. Discarding some tedious computations we find out that solution to (4.6) is defined implicitly by:

$$\frac{x}{v_0} _2F_1\left( \begin{array}{c} \frac{1}{2}, \frac{1}{k+1} \\ \frac{3}{2} \end{array} \right| -\frac{\lambda\sigma^2 x^2}{k\eta v_0^{k+1}} \right) = T - t$$  \hspace{1cm} (4.10)

In order to meet the second initial condition $x(0) = X$ we see that $v_0$ must satisfy:

$$\frac{X}{v_0} _2F_1\left( \begin{array}{c} \frac{1}{2}, \frac{1}{k+1} \\ \frac{3}{2} \end{array} \right| -\frac{\lambda\sigma^2 X^2}{k\eta v_0^{k+1}} \right) = T$$  \hspace{1cm} (4.11)

Note that, being assigned all parameters $\lambda$, $\sigma$, $X$, $\eta$, $k$ equation (4.11) is an equation in the sole unknown $v_0$. Of course such equation has to be treated numerically: once the value of $v_0$ is detected it has te be inserted in (4.3) to obtain solution to (3.1).

As an (easy) example take $\lambda = \sigma = X = \eta = T = 1$ and $k = 1/2$. In the plot below we represent the left hand side of (4.11) and the value of $T$ at the right hand side. Using Mathematica® we find numerically $v_0 = 0.671525$. 

6
Afterwards we put this value in (4.3) and we plot the relevant function $x(t)$.

We conclude with a graphic representation, with the same parameters for several values of $k$. 

7
4.1 The Almgren zero-speed case

When Almgren takes \( v_0 = 0 \) the initial value problem (4.6) has to be considered with zero initial conditions \( x(T) = \dot{x}(T) = 0 \). Observe that the quadrature formula arising from (4.9), which is equivalent to (4.6) reads in this case as:

\[
\int_{x_0}^{x} \frac{dz}{\left(v_0^{k+1} + \frac{\lambda \sigma^2}{k \eta} z^2 \right)} = T - t \quad \Rightarrow \quad \int_{x_0}^{x} \frac{dz}{\lambda^2 \eta k} \frac{1}{z^{k+1}} = T - t \tag{4.12}
\]

but this means that the convergence condition \( \frac{2}{k+1} < 1 \) has to be imposed. Moreover since we assume \( k \) to be positive this means that the zero speed assumption is well posed only if \( k > 1 \): for \( k \leq 1 \) there are not solutions of the optimization problem (3.1) with zero speed. Moreover if we evaluate the integral at the left hand side of (4.12) for \( k > 1 \) we find:

\[
\frac{k + 1}{k - 1} \left( \frac{\lambda \sigma^2}{k \eta} \right)^{-\frac{1}{k+1}} x^{\frac{k-1}{k+1}} = T - t \quad \Rightarrow \quad x(t) = \left( \frac{(k - 1)T}{k + 1} \right)^{\frac{k-1}{k+1}} \left( \frac{\lambda \sigma^2}{k \eta} \right)^{\frac{1}{k+1}} \tag{4.13}
\]

so that at \( t = 0 \) we have:

\[
x(0) = \left( \frac{(k - 1)T}{k + 1} \right)^{\frac{k-1}{k+1}} \left( \frac{\lambda \sigma^2}{k \eta} \right)^{\frac{1}{k+1}} \tag{4.14}
\]
this means that we are not free to consider an arbitrary value of \(x(0)\) having assigned the speed in \(t = T\).

4.2 The case \(k = 1\)

In the case \(k = 1\) we have provided two solutions of (3.1): the first follows from the straightforward integration of the linear case, see equation (3.6), while the second stems from the hypergeometric implicit solution expressed by equation (4.3). Of course the two solution to (3.1) are, as matter of fact, the same. This can be understood recalling the following property of \(2F_1\), (see Abramowit and Stegun, 1964, entry 15.1.7, page 556):

\[
2F_1 \left( \frac{1}{2}, \frac{1}{2} \middle| -z^2 \right) = \frac{\ln \left( z + \sqrt{1+z^2} \right)}{z} = \frac{\text{arcsinh} \ z}{z}. \tag{4.15}
\]

In such a way, in this particular case we can use this identity to solve (4.10) with respect to \(t\) obtaining:

\[
x(t) = v_0 \sinh \left( (T - t) \sqrt{\frac{\lambda \sigma^2}{\eta k}} \right) \tag{4.16}
\]

To compare this hypergeometric solution with (3.6) we have to evaluate \(v_0\) from (4.11):

\[
v_0 = \frac{\sqrt{\lambda \sigma X}}{\sqrt{\eta k^2 } \sinh \left( \frac{\sqrt{\lambda \sigma T}}{\sqrt{\eta k}} \right)} \tag{4.17}
\]

and substitute in (4.16) getting:

\[
x(t) = X \frac{\sinh \left( \frac{\sqrt{\lambda \sigma (T-t)}}{\sqrt{\eta k^2}} \right)}{\sinh \left( \frac{\sqrt{\lambda \sigma T}}{\sqrt{\eta k}} \right)} \tag{4.18}
\]

which is noting else but (3.6).

5 Conclusion

Rebalancing large portfolios of stocks requires taking into account two peculiar issues. The first one is related to the market impact of trades, which generates transaction costs. The second issue arises from the risk of finding counterparties willing to trade at the desired price. Both empirical and theoretical considerations suggest that the market impact of trades is typically nonlinear. Almgren (2003) proposes an optimal execution strategy that minimizes the tradeoff between volatility risk and transaction costs while taking into account this form of nonlinearity.
In this paper, we review the optimal liquidation strategy of Almgren (2003). We show that the solution method used by Almgren (2003) is ill-posed. The reason is that it is based on the assumption that no shares per unit of time are traded at the beginning of the period. We use an approach based on the Gaussian hypergeometric function to solve for the case of positive initial trades. Our results differ strongly from those of Almgren (2003). First of all, the problem admits a solution with no trading in the opening period only if additional parametric restrictions are imposed. Second, with positive initial trading, the optimal execution time depends on trading activity in the initial period.
A Solution of second order differential equations with a Gauss hypergeometric function

The linear second order differential equation in the unknown $u = u(t)$

$$t(1 - t) \ddot{u} + [c - (a + b + 1) t] \dot{u} - ab u = 0 \quad (A.1)$$

is known as Gauss hypergeometric equation. Parameters $a, b, c$ are not functions of the independent variable $t$ and can be in general complex number. Searching for a power series solution of (A.1) it can be seen that

$$2F_1 \left( \begin{array}{c} a, b \\ c \end{array} \mid t \right) := \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n} \frac{t^n}{n!} \quad (A.2)$$

where we use the Pochhamer symbol $(x)_n$, $n \in \mathbb{N}$ defined as:

$$\begin{cases} (x)_0 := 1 \\ (x)_n = x(x+1)(x+2) \cdots (x+n-1) \end{cases} \quad (A.3)$$

is the solution of (A.1) such that $u(0) = 1$, $\dot{u}(0) = ab/c$. Power series defining $2F_1$ converges for $|t| < 1$ and to continue the hypergeometric function $2F_1$ it is useful the integral representation ascribed to Leonhard Euler but really due to Adrien Marie Legendre:

$$2F_1 \left( \begin{array}{c} a, b \\ c \end{array} \mid t \right) = \frac{\Gamma(c)}{\Gamma(c-a)\Gamma(a)} \frac{1}{\Gamma(1)} \int_0^1 s^{a-1}(1-s)^{c-a-1} (1-ts)^b ds, \quad (A.4)$$

where $\text{Re } c > \text{Re } a > 0$, $|t| < 1$, and the Euler-Legendre integral (Gamma function) is defined for $x > 0$ by:

$$\Gamma(x) = \int_0^\infty e^{-u}u^{x-1}du. \quad (A.5)$$

A proof of the integral representation theorem and a good presentation of the Gauss hypergeometric function can be found at Seaborn (1991), the integral representation theorem is treated at section 10.7 pages 184-185 formula (10.39). Integral representation theorem provides an extension to the region where the complex hypergeometric function is defined, namely for its analytical continuation, to the (almost) whole complex plane excluding the half-straight line $(1, \infty)$. This function was first introduced in Dynamical Economics in a generalization of the Solow Swan model due to Mingari Scarpello and Ritelli, Mingari, Scarpello and Ritelli (2003), while Boucekkine and Ruiz-Tamarit Boucekkine and Ruiz-Tamarit (2008) use it in the Lucas-Uzawa model.

\footnote{Exercices de calcul intégral, II, quatrième part, sect. 2, Paris 1811}
References


