The Korean Communications in Statistics Vol. 15, No. 1, 2008, pp. 117–124

Estimation of Liquidity Cost in Financial Markets†

Johan Lim[1], Kiseop Lee[2], HyunSeok Song[3]

Abstract

The liquidity risk is defined as an additional risk in the market due to the timing and size of a trade. A recent work by Cetin *et al.* (2003) proposes a rigorous mathematical model incorporating this liquidity risk into the arbitrage pricing theory. A practical problem arising in a real market application is an estimation problem of a liquidity cost. In this paper, we propose to estimate the liquidity cost function in the context of Cetin *et al.* (2003) using the constrained least square (LS) method, and illustrate it by analyzing the Kellogg company data.

Keywords: Constraint least square; liquidity cost; semi-parametric model.

1. Introduction

Classical theories of financial markets assume the frictionless and the competitive market paradigms. A frictionless market is one without transaction costs and trade restrictions, and a competitive market assumes that traders act as price takers so that their trades do not have any impact on the price process. This idealized state is well applied to highly liquid stocks and is useful to create tractable models and clarify ideas. However, since transaction costs do of course exist and since unlimited liquidity is only reasonably approximated by no more than 100 stocks or so on the New York Stock Exchange, one is naturally led to question of how these assumptions are allowed to be weakened and how this more closely approximates reality.

A relaxation of the frictionless and competitive market introduces the notion of liquidity risk which is often defined as the additional risk in the market due to the timing and size of a trade. Recently, Cetin *et al.* (2003) propose a rigorous model incorporating the liquidity risk into the arbitrage pricing theory. They establish a mathematical

† Johan Lim's research is supported by Basic Research Fund of the College of Business and Economics, Yonsei University.

1) Assistant Professor, Department of Applied Statistics, Yonsei University, Seoul 120-749, Korea. Correspondence: johanlim@yonsei.ac.kr

2) Assistant Professor, Department of Mathematics, University of Louisville, Louisville, KY 40292, U.S.A. E-mail: kiseop.lee@louisville.edu

3) Researcher, Yurie Asset Management Company, Seoul 150-600, Korea.

formulation of a liquidity cost, admissible strategies, self-financing strategies, and an approximately complete market. Also, they show that the two fundamental theorems of finance hold under the existence of the liquidity risk, approximately modified; the first fundamental theorem states that under the competitive market assumption, the market is arbitrage-free if and only if there exists an equivalent martingale measure; the second theorem tells us that if the equivalent martingale measure is unique, then the market is complete. In particular, they study an extension of the Black-Scholes economy incorporating liquidity risk as an illustration of the theory.

A practical problem arising in a real market application is an estimation problem of a liquidity cost. Many people both in academia and industry have studied how to measure the liquidity itself; for example, Houweling *et al.* (2003), Fleming (2002), Datar *et al.* (1998), and Muranaga and Ohsawa (2003), using various statistics such as bid-ask spread, volume, turnover ratio and impact cost. In this paper, instead of the liquidity itself, we are interested in estimating the supply curve in the context of Cetin *et al.* (2003). A supply curve $S(t, x, \omega)$ represents the stock price per share at time $t \in [0, T]$ that the trader pays/receives for an order of size $x \in \mathbf{R}$ given the state $\omega \in \Omega$. A positive order($x > 0$) represents a buy, a negative order($x < 0$) represents a sale, and the zeroth order($x = 0$) corresponds to the marginal trade. In a perfectly liquid market, all orders can be considered marginal. For the detailed structure of the supply curve, we refer to Section 2 of Cetin *et al.* (2003).

As a special case, we consider a multiplicative model

$$S(t, x) = e^{f(x)}\, S(t, 0), \tag{1.1}$$

where $S(t, 0)$ is the price of liquid market by Black-Scholes stock price equation:

$$S(t, 0) = S_0 \exp\Big\{ \big(\mu - \sigma^2/2\big)t + \sigma W_t \Big\}, \tag{1.2}$$

where W_t is the standard Brownian motion and S_0 is arbitrary initial value. We make the assumption of monotone $S(t, x)$ in x and, hence, the assumption of monotone $f(x)$ in x, which tells that the larger volume of the purchase (or sale), the larger the price impact that occurs on the share price. We also assume $f(0) = 0$.

We apply the constrained least square (LS) method to estimate the model (1.1) studied in Geyer (1991). Geyer (1991) studies nonparametric function estimation with monotone or convex shape constraints. He suggests to use the LS method with becomes a quadratic programm with linear inequality constraints for shape constraints. In our problem, for the adjacent observed points $S(t_i, x_i)$ and $S(t_{i+1}, x_{i+1})$, the model (1.1) implies that

$$\log\left(\frac{S(t_{i+1}, x_{i+1})}{S(t_i, x_i)}\right) = f(x_{i+1}) - f(x_i) + u_i,$$

where u_i follows the Gaussian distribution with mean $(t_{i+1} - t_i)\big(\mu - \sigma^2/2\big)$ and variance

$(t_{i+1} - t_i)\sigma^2$. Thus, the main problem to be solved is, for $s_i = t_{i+1} - t_i$,

$$\text{minimize} \ \sum_{i=1}^n \frac{1}{s_i}\left[\log\left(\tfrac{S(t_{i+1},x_{i+1})}{S(t_i,x_i)}\right) - \left(s_i(\mu - \sigma^2/2) + f(x_{i+1}) - f(x_i)\right)\right]^2, \tag{1.3}$$

subject to f is monotone and $f(0) = 0$.

We illustrate the model and the constraint LS by analyzing the trading data of the company Kellogg during a week in the year 2000.

We organize the paper as follows. We study details of the model by Cetin *et al.* (2003) in Section 2, and analyze the Kellogg company trading data in Section 3. Section 4 concludes the paper with some additional issues not covered in the main body.

2. Semi-parametric Model for Liquidity Cost

Cetin *et al.* (2003) propose the multiplicative model (1.1) where $S(t,0)$ is the Black-Scholes price and linear function $f(x) = \alpha x$. They made several assumptions to $S(t,x)$ including that $f(x)$ is twice differentiable in x; (i) it is measurable to the filtered probability space; (ii) for almost every t, $S(t,x)$ is a non–decreasing function of x a.s. $\mathbf{P}$; (iii) it is twice continuously differentiable with respect to x, and both $\partial S(t,x)/\partial x$ and $\partial^2 S(t,x)/\partial x^2$ are continuous in t; (vi) $S(\cdot,x)$ is a semi–martingale; and (v) $S(t,x)$ has a continuous sample path for every x.

Unlike Cetin *et al.* (2003), in this paper, we only assume that $f(x)$ is monotone in x and $f(0) = 0$ as addressed in the Introduction.

The interest of the paper is to estimate the model (1.1) from the observed transaction data $S(t_i, x_i)$ for $i = 1, 2, \ldots, n$. For the regularly observed data (for the adjacent time points t_{i+1} and t_i, $t_{i+1} - t_i$ is constant Δ over i), the model (1.1) implies that

$$\log\left(\frac{S(t_{i+1},x_{i+1})}{S(t_i,x_i)}\right) = f(x_{i+1}) - f(x_i) + \left(\mu - \frac{\sigma^2}{2}\right)\Delta + \epsilon_i,$$

where ϵ_i follows the Gaussian distribution with mean 0 and variance $\Delta\sigma^2$ which does not dependent on t_i and t_{i+1}.

The main problem to be solved, thus, is formulated into

$$\text{minimize} \ \sum_{i=1}^n \left[\log\left(\frac{S(t_{i+1},x_{i+1})}{S(t_i,x_i)}\right) - \left(\alpha + f(x_{i+1}) - f(x_i)\right)\right]^2, \tag{2.1}$$

subject to f is monotone and $f(0) = 0$,

where $\alpha = \Delta(\mu - \sigma^2/2)$.

3. How to Solve

In this section, we shortly discuss how to solve the main problem (2.1).

The problem to be solved is approximated by a quadratic program (QP) with inequality constraints. The minimization problem (2.1) is an infinite dimensional problem, and we approximate the function f as a piecewise constant function that

$$f(x) = f_k, \quad \text{if} \quad x \in [c_{k-1}, c_k),$$

where $-\infty = c_0 < c_1 < \cdots < c_p < c_{p+1} = \infty$. Here, the function f is monotone, and we assume $f_1 \leq \cdots \leq f_p$. Then, the problem becomes

$$\sum_{i=1}^{n} \left\{ y_i - \alpha - a_i \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{p-1} \\ f_p \end{pmatrix} \right\}^2,$$

where

$$y_i = \log\left(\frac{S(t_{i+1}, x_{i+1})}{S(t_i, x_i)}\right)$$

and a_i is $1 \times p$ vector indicating the interval having x_i. Equivalently, with matrix notation, the problem (2.1) becomes a QP:

$$\begin{array}{ll} \text{minimize} & \mathbf{f}^T A^T A \mathbf{f} - 2\mathbf{y} A \mathbf{f} \\ \text{subject to} & C\mathbf{f} \preceq 0, \quad 0 \preceq \mathbf{f} \end{array} \tag{3.1}$$

where $\mathbf{f} = (\alpha, f_1, \ldots, f_p)^T$ and $\mathbf{y} = (y_1, \ldots, y_n)$, A is defined as

$$\begin{pmatrix} \alpha + & f(x_2) - f(x_1) \\ \alpha + & f(x_3) - f(x_2) \\ \vdots & \\ \alpha + & f(x_n) - f(x_{n-1}) \end{pmatrix} = A \begin{pmatrix} \alpha \\ f_1 \\ \vdots \\ f_p \end{pmatrix},$$

and C is

$$C = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}.$$

The two most common methods to solve a constrained QP are the simplex method and the interior point methods. The primal dual simplex method, originally developed by Dantzig (1963), solves the KKT conditions by reformulating the problem to a linear programming problem. The Karush-Kuhn-Tucker (KKT) conditions of a linearly constrained QP is a set of linear equations and solved analytically (Van de Panne, 1975;

Boyd and Vandenberghe, 2003). An alternative to the simplex method is the interior point method by Karmarkar (1984). The interior point method easily utilizes the sparse structure of a quadratic form and has shown to be very competitive to the simplex method especially to solve a large scale QP (Boggs *et al.*, 1996). A high-quality implementation of a primal-dual interior-point method for QPs is available from solvers that are already available. Examples are MOSEK software package (MOSEK ApS, 2002), and YALMIP (Löfberg, 2003), and CVX (Grant *et al.*, 2005), which have a simple interface that recognizes and solves QPs.

4. Analysis of Kellogg Company Data

The data set we use in this section records transactions of the company Kellogg for 5 business days (a week) in the year 2000. Three variables, transaction time, volume, and price, are observed for each transaction, where the negative volume implies the sale-initiated trade and the positive volume implies buy-initiated trade. Total 7644 transactions are observed and the price trajectory of day 1 is plotted in Figure 4.1.

We filter the original transaction data for the next analysis. We sample the data at every five minutes using the nearest data to the given time points from the beginning of the market so that we have 79 observed points in each day. The filtering procedure is illustrated by Figure 4.1.

We solve the constrained LS problem (2.1) using MOSEK ApS optimization software

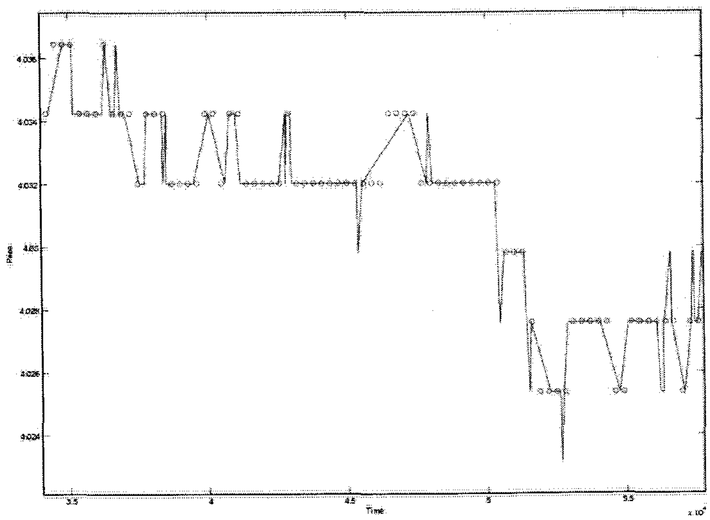

Figure 4.1: Sampling of raw transaction data. The solid line is the trajectory of the original transaction price, and the circles represent the sampled prices.

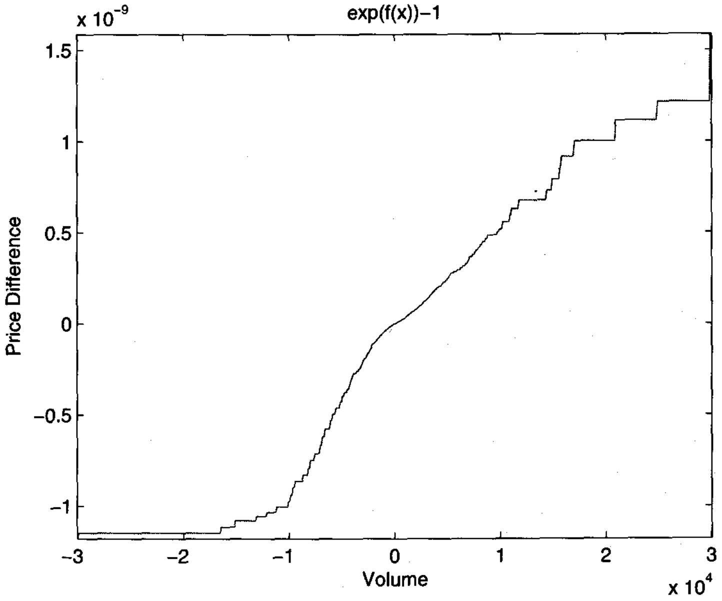

Figure 4.2: Estimated liquidity cost of Kellogg company

which uses a state-of-the-art interior point method, whose trial version is freely given to researchers. The estimated liquidity cost curve is plotted in Figure 4.2.

The liquidity cost per share, when trading x shares, is

$$S(t,x) - S(t,0) = \{\exp(f(x)) - 1\}S(t,0).$$

In other words, we pay $(e^{f(x)} - 1)S(t,0)$ more per share if we trade x shares at one time. If we do block trading of n blocks of x/n shares, then the cost per share we pay more becomes

$$S\big(t,(x/n)\big) - S\big(t,0\big) = \{\exp\big(f(x/n)\big) - 1\}S\big(t,0\big).$$

Thus, to investigate the relationship between trading volume x and its liquidity cost, it would appropriate to study the curve $e^{f(x)} - 1$ instead of $f(x)$ itself.

Figure 4.2 shows that, in a buy order, the graph slowly increases up to 10,000 shares but then increases more rapidly after 10,000. In a sale order, the graph decreases moderately up to 5,000 shares but then decreases more rapidly after 5,000 shares. This supports a widely studied asymmetric liquidity effect, which indicates that block sellers pay more liquid premium than block buyers. Roughly speaking, a buy order can tolerate up to about 10,000 shares, but a sale order can tolerate only up to 5,000. In addition, even when we compare the intervals $[-5,000,\ 0]$ and $[0,\ 10,000]$, the sale side has a steeper slope, which also supports the asymmetry. Interested readers can consult Frino *et al.* (2003), Tse and Xiang (2005) and references therein.

Another important thing we can observe is the common trading strategy which divides a large block trade into small ones to hide the block trade. This trading strategy is justified by monotonicity of f, which means $|\exp(f(x/n)) - 1| < |\exp(f(x)) - 1|$ for all n. So, assuming that the market price does not move a lot during the time, you can pay less per share by dividing your block trade into small ones.

For example, suppose a trader wants to sell 20,000 shares. Then, dividing this into two 10,000 share orders does not help the case much, since a trade with size 10,000 is still a large trade and it still costs about same liquidity cost. If the trader wants to cut the liquidity cost to the half, she should divide about three 6,600 share orders roguhly (from Figure 4.2). Still, she may not be able to cut the liquidity cost to the half, since the price may move up during the time. But it is unlikely for the price to be doubled, which can offset the gain from from the smaller liquidity cost. Therfore, it is still a good strategy for the trader.

5. Conclusion

We have studied how to estimate the supply curve modeling a liquidity cost, using the constrained least square method. We analyzed Kellogg company data from 1994 to 2000. The estimated supply curve shows well studied asymmetry of liquidity premium, which implies that block selling usually cost more liquidity cost than block buying. An advantage of this approach over previous studies is that this modeling not only shows the asymmetry, but also gives an explicit mathematical equation which exactly calculates the liquidity premium.

Acknowledgement

We are grateful to associate editor and referee for several helpful suggestions.

References

Boyd, S. and Vandenberghe, L. (2003). *Convex Optimization.* Cambridge University Press.

Boggs, P. T., Domich, P. D. and Rogers, J. E. (1996). An interior point method for general large-scale quadratic programming problems. *Annals of Operations Research*, **62**, 419–437.

Cetin, U., Jarrow, R. A. and Protter, P. (2003). Liquidity risk and arbitrage pricing model. *Finance and Stochastics*, **8**, 311–341.

Dantzig, G. B. (1963). *Linear Programming and Extensions.* Princeton University Press, Princeton, New Jersey.

Datar, V., Naik, N. and Radcliffe, R. (1998). Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, **1**, 203–219.

Fleming, M. J. (2002). Are larger treasury issues more liquid? Evidence from bill reopenings. *Journal of Money, Credit, and Banking*, **3**, 707–735.

Frino, A., Mollica, V. and Walter, T. (2003). Asymmetric price behaviour surrounding block trades: A market microstructure explanation. `http://www.cls.dk/caf/wp/wp-154.pdf`.

Geyer, C. (1991). Constrained maximum likelihood exemplified by isotonic convex logistic regression. *Journal of the American Statistical Association*, **86**, 717–724.

Grant, M., Boyd S. and Ye, Y. (2005). CVX: Matlab software for disciplined convex programming. Available from `www.stanford.edu/~boyd/cvx/`.

Houeweling, P., Mentink, A. and Vorst, T. (2003). How to measure corporate bond liquidity. `http://www.tinbergen.nl/discussionpapers/03030.pdf`.

Karmarkar, N. (1984). A new polynomial-time algorithm of linear programming. *Combinatorica*, **4**, 374–495.

Löfberg, J. (2003). YALMIP. Yet Another LMI Parser. Version 2.4. Available from `http://control.ee.ethz.ch/~joloef/yalmip.php`.

MOSEK APS (2002). *The MOSEK Optimization Tools Version 2.5. User's Manual and Reference.* Available from `www.mosek.com`.

Muranaga, J. and Ohsawa M. (2003). Measurement of liquidity risk in the context of market risk calculation. `http://www.bis.org/publ/ecsc07j.pdf`.

Tse, Y. and Xiang, J. (2005). Asymmetric liquidity and asymmetric volatility. `www.fma.org/Chicago/Papers/Asymmetric_Liquidity_and_Asymmetric_Volatility.pdf`

Van de Panne, C. (1975). *Methods for Linear and Quadratic Programming.* North-Holland Publishing Company, Amsterdam.

[Received September 2007, Accepted December 2007]