International Journal of Theoretical and Applied Finance
Vol. 13, No. 6 (2010) 821–838

DOI: 10.1142/S0219024910006017

AN ANALYSIS OF THE SUPPLY CURVE FOR LIQUIDITY RISK THROUGH BOOK DATA

MARCEL BLAIS

Department of Mathematical Sciences
Worcester Polytechnic Institute, Worcester
MA, 01609-2280, USA

PHILIP PROTTER*

School of Operations Research
Cornell University, Ithaca, NY, 14853-3801, USA
pep4@cornell.edu

Received 14 January 2010
Accepted 9 April 2010

We use order book data[1] combined with tick data to analyze the supply curve models of liquidity issues in stock and option market trading. We show that supply curves really exist, and further that for highly liquid stocks they are linear. For slightly less liquid stocks the supply curve tends to be jump linear.

Keywords: Liquidity risk; supply curve; hedging risk; semimartingale; arbitrage; option.

1. Introduction

Issues of liquidity have long plagued stock and option traders. If one follows a given hedging strategy, one's profits can be greatly diminished by liquidity costs due to updating positions in the strategy; such liquidity costs should figure into the price one charges for the option, but for this to be the case, one has to estimate them in advance. The seminal paper of Grossman and Miller [7] in 1988 takes a supply and demand equilibrium approach to this problem, and it begins with one of the best opening paragraphs we have had the pleasure of reading. In short, we could provide our version by paraphrasing the famous remark of Supreme Court Justice Potter Stewart:[2] "Perhaps I could never succeed in precisely defining what

*Corresponding author.

[1]We are grateful to Robert Ferstenberg of Morgan Stanley for graciously providing us with order book data.

[2]Justice Potter Stewart's original quote concerns hard core pornography, and it contains the following sentence: "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it, and the motion picture involved in this case is not that." Jacobellis versus Ohio, 378 U.S. 184 (1964).

is liquidity in the stock market. But I know it when I see it." This has been long been the attitude of practitioners and academics alike. We prefer the more modern approach of arbitrage free pricing, which has the benefit of allowing more explicit and detailed calculations. Such an approach to this problem is presented in the work of Çetin *et al.* [2], where the existence of a supply curve is postulated as the key new feature of the theory. In Sec. 2 we present the basics of this model. While this model appears to be an attractive and relatively simple, straightforward solution to issues of liquidity, up to now there has been no direct evidence that such a supply curve actually exists: that is, that it is not just a line with slope zero, a situation that would render the advances of the theory of Çetin *et al.* vacuous. In this paper we tackle three issues:

(1) We show via a statistical analysis of data that the supply curve postulated in Çetin *et al.* [2] actually does exist.
(2) We find the structure of the supply curve, and in particular we show in Sec. 3 that in the most interesting case of highly liquid stocks that the supply curve has the pleasingly simple structure of a straight line, with a time varying slope. While the slope $M(t)$ is clearly randomly changing, it is nevertheless quite stable: The variance of $M(t)$ is small.
(3) The model of Çetin *et al.* has been criticized because the perfect liquid market price on a derivative such as a call option is the same as the price of the model without liquidity considerations. Indeed, an improvement to the model has been proposed by Çetin *et al.* [4]. This is because from a modeling perspective, one can arbitrarily approximate a perfect hedging strategy with continuous strategies with paths of finite variation. In practice, of course, continuous trading is both financially and physically impossible; all trading is of necessity discrete. Nevertheless, this result shows that one can get arbitrarily small liquidity charges by trading with high frequency in small amounts. We show in Sec. 5 however, that if one does not limit one's attention to only one option, but instead has a portfolio of more than two options, liquidity issues actually do arise in an important way, using the model of Çetin *et al.* with linear supply curves.

In a related paper [1], we use these techniques to study the accuracy of the famous algorithm of Lee and Ready.

2. The Model

This section presents the model. We are given a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ satisfying the usual conditions where T is a fixed time. $\mathbb{P}$ represents the statistical or empirical probability measure. We also assume that $\mathcal{F}_0$ is trivial, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

We consider a market for a security that we will call a stock with no dividends. Also traded is a money market account that accumulates value at the spot rate of interest. Without loss of generality, we assume that the spot rate of interest is zero, so that the money market account has unit value for all times.[3]

2.1. *Supply curve*

We consider an arbitrary trader who acts as a price taker with respect to an exogenously given supply curve for shares bought or sold of this stock within the trading interval. More formally, let $S(t, x, \omega)$ represent the stock price, *per share*, at time $t \in [0, T]$ that the trader pays/receives for an order of size $x \in R$ given the state $\omega \in \Omega$. A positive order ($x > 0$) represents a buy, a negative order ($x < 0$) represents a sale, and the order zero ($x = 0$) corresponds to the marginal trade.

By construction, rather than the trader facing a horizontal supply curve as in the classical theory (the same price for any order size), the trader now faces a supply curve that depends on his order size.[4] Note that the supply curve is otherwise independent of the trader's past actions, endowments, risk aversion, or beliefs. This implies that an investor's trading strategy has no lasting impact on the price process.

We now impose some structure on the supply curve.

Assumption 2.1 (Supply Curve). (1) $S(t, x, \cdot)$ is $\mathcal{F}_t$−measurable and non-negative.

(2) $x \mapsto S(t, x, \omega)$ is a.e. t non-decreasing in x, a.s. (i.e. $x \le y$ implies $S(t, x, \omega) \le S(t, y, \omega)$ a.s. $\mathbb{P}$, a.e. t).
(3) S is C^2 in its second argument, $\partial S(t, x)/\partial x$ is continuous in t, and $\partial^2 S(t, x)/\partial x^2$ is continuous in t.
(4) $S(\cdot, 0)$ is a semi-martingale.
(5) $S(\cdot, x)$ has continuous sample paths (including time 0) for all x.

Except for the second condition, these restrictions are self-explanatory. Condition 2 is the situation where the larger the purchase (or sale), the larger the price impact that occurs on the share price. This is the usual situation faced in asset pricing markets, where the quantity impact on the price is due to either information effects or supply/demand imbalances (see Kyle [10], Glosten and Milgrom [6], Grossman and Miller [7]). It includes, as a special case, horizontal supply curves.[5]

Example 2.1 (Supply Curve). To present a concrete example of a supply curve, let $S(t, x) \equiv f(t, D_t, x)$ where D_t is an n-dimensional, $\mathcal{F}_t$-measurable semimartingale, and $f : R^{n+2} \to R^+$ is Borel measurable, C^1 in t, and C^2 in all its other arguments. This non-negative function f can be viewed as a reduced form supply

[3] A numéraire invariance theorem is proved in Çetin *et al.* [2].
[4] In contrast, the trader is assumed to have no quantity impact due to his trades in the money market account.
[5] This structure can also be viewed as a generalization of the model in Jouini [9] where the traded securities have distinct selling and buying prices following separate stochastic processes.

curve generated by a market equilibrium process in a complex and dynamic economy. Under this interpretation, the vector stochastic process D_t represents the state variables generating the uncertainty in the economy, often assumed to be diffusion processes or at least Markov processes (e.g. a solution to a stochastic differential equation driven by a Levy process).

2.2. *Trading strategies*

We start by defining the investor's trading strategy.

Definition 2.1. A trading strategy is a triplet $((X_t, Y_t : t \in [0,T]), \tau)$ where X_t represents the trader's aggregate stock holding at time t (units of the stock), Y_t represents the trader's aggregate money market account position at time t (units of the money market account), and τ represents the liquidation time of the stock position, subject to the following restrictions: (a) X_t and Y_t are predictable and optional processes, respectively, with $X_{0-} \equiv Y_{0-} \equiv 0$, and (b) $X_T = 0$ and τ is a predictable $(\mathcal{F}_t : 0 \leq t \leq T)$ stopping time with $\tau \leq T$ and $X = H1_{[0,\tau)}$ for some predictable process $H(t,\omega)$.

We are interested in a particular type of trading strategy — those that are self-financing. By construction, a self-financing trading strategy generates no cash flows for all times $t \in [0,T)$. That is, purchase/sales of the stock must be obtained via borrowing/investing in the money market account. This implies that Y_t is uniquely determined by (X_t, τ). The goal is to define this self-financing condition for Y_t given an arbitrary stock holding (X_t, τ).

Definition 2.2. A self-financing trading strategy (*s.f.t.s.*) is a trading strategy $((X_t, Y_t : t \in [0,T]), \tau)$ where (a) X_t is càdlàg if $\partial S(t,0)/\partial x \equiv 0$ for all t, and X_t is càdlàg with finite quadratic variation ($[X,X]_T < \infty$) otherwise, (b) $Y_0 = -X_0 S(0, X_0)$, and (c) for $0 < t \leq T$,

$$Y_t = Y_0 + X_0 S(0, X_0) + \int_0^t X_{u-} dS(u,0) - X_t S(t,0)$$
$$- \sum_{0 \leq u \leq t} \Delta X_u [S(u, \Delta X_u) - S(u,0)] - \int_0^t \frac{\partial S}{\partial x}(u,0) d[X,X]_u^c. \tag{2.1}$$

Condition (a) imposes restrictions on the class of acceptable trading strategies. Under the hypotheses that X_t is càdlàg and of finite quadratic variation, the right side of expression (2.1) is always well-defined although the last two terms (always being non-positive) may be negative infinity. The classical theory, under frictionless and competitive markets, does not need these restrictions. An example of a trading strategy that is allowed in the classical theory, but disallowed here, is $X_t = 1_{\{S(t,0)>K\}}$ for some constant $K > 0$ where $S(t,0)$ follows a Brownian motion. Under the Brownian motion hypothesis this is a discontinuous trading strategy that jumps infinitely often immediately after $S(t,0) = K$ (the jumps are not square summable), and hence Y_t is undefined.

Condition (b) implies the strategy requires zero initial investment at time 0. When studying complete markets in a subsequent section, condition (b) of the s.f.t.s. is removed so that $Y_0 + X_0 S(0, X_0) \neq 0$.

Condition (c) is the self-financing condition at time t. The money market account equals its value at time 0, plus the accumulated trading gains (evaluated at the marginal trade), less the cost of attaining this position, less the price impact costs of discrete changes in share holdings, and less the price impact costs of continuous changes in the share holdings. This expression is an extension of the classical self-financing condition when the supply curve is horizontal. To see this note that using condition (b) with expression (2.1) yields the following simplified form of the self-financing condition:

$$\begin{aligned} Y_t + X_t S(t,0) = & \int_0^t X_{u-} dS(u,0) \\ & - \sum_{0 \le u \le t} \Delta X_u [S(u, \Delta X_u) - S(u,0)] \\ & - \int_0^t \frac{\partial S}{\partial x}(u,0) d[X,X]_u^c \quad \text{for } 0 \le t \le T. \end{aligned} \tag{2.2}$$

The left side of expression (2.2) represents the classical "value" of the portfolio at time 0. The right side gives its decomposition into various components. The first term on the right side is the classical "accumulated gains/losses" to the portfolio's value. The last two terms on the right side capture the impact of illiquidity, both entering with a negative sign.

2.3. *The marked-to-market value of a s.f.t.s. and its liquidity cost*

This section defines the marked-to-market value of a trading strategy and its liquidity cost. At any time prior to liquidation, there is no unique value of a trading strategy or portfolio. Indeed, any price on the supply curve is a plausible price to be used in valuing the portfolio. At least three economically meaningful possibilities can be identified: (i) the immediate liquidation value (assuming that $X_t > 0$ gives $Y_t + X_t S(t, -X_t)$), (ii) the accumulated cost of forming the portfolio (Y_t), and (iii) the portfolio evaluated at the marginal trade ($Y_t + X_t S(t,0)$).[6] This last possibility is defined to be the *marked-to-market value* of the self-financing trading strategy (X, Y, τ). It represents the value of the portfolio under the classical price taking condition.

[6]These three valuations are (in general) distinct except at one date, the liquidation date. At the liquidation time τ, the value of the portfolio under each of these three cases are equal because $X_\tau = 0$.

Motivated by expression (2.2), we define the liquidity cost to be the difference between the accumulated gains/losses to the portfolio, computed as if all trades are executed at the marginal trade price $S(t, 0)$, and the marked-to-market value of the portfolio.

Definition 2.3. The liquidity cost of a *s.f.t.s.* (X, Y, τ) is

$$L_t \equiv \int_0^t X_{u-} dS(u, 0) - [Y_t + X_t S(t, 0)].$$

The following lemma follows from the preceding definition.

Lemma 2.1 (Equivalent Characterization of the Liquidity Costs).

$$L_t = \sum_{0 \le u \le t} \Delta X_u [S(u, \Delta X_u) - S(u, 0)] + \int_0^t \frac{\partial S}{\partial x}(u, 0) d[X, X]_u^c \ \ge 0$$

where $L_{0-} = 0$, $L_0 = X_0[S(0, X_0) - S(0, 0)]$ *and* L_t *is non-decreasing in* t.

We refer the reader to [8] for a (simple) proof.

We see here that the liquidity cost is non-negative and non-decreasing in t. It consists of two components. The first is due to discontinuous changes in the share holdings. The second is due to the continuous component. This expression is quite intuitive. Note that because $X_{0-} = Y_{0-} = 0$, $\Delta L_0 = L_0 - L_{0-} = L_0 > 0$ is possible.

3. Inferring the Structure of the Supply Curve

3.1. *The data set*

The theory sketched in the previous section, and established in the papers [2, 3, 8], is well and nice, but one needs to show that the supply curve actually exists (that is, that is not a horizontal line, which would imply the preceding theory is vacuous), and that, if it exists, what its structure is.

The obvious tool to infer whether or not a supply curve exists and to describe it is to use massive amounts of tick data. Unfortunately this obvious and readily available tool is not well suited to the task. The reason is that with tick data the information as to *who initiated the trade*, *the buyer or the seller*, is not recorded. This has long been recognized as a problem in the literature devoted to studying liquidity and illiquidity, and various remedies have been proposed to resolve it. Perhaps the most well known is the 1991 algorithm due to Lee and Ready [11], which is thought to have an accuracy rate between 74% and 85%. (We find, however, in a related paper [1] that the actual success rate even for highly liquid stocks is only around 61%.) The basic idea is that if the price increases, the trade must be buyer initiated, and if the price decreases, it is seller initiated. For our purposes, using this argument to establish the liquidity model of Çetin *et al.* would be close to a circular proof: assuming a key hypothesis of the theory as true in order to establish the theory. It is problems of this type that render tick data not useful for our purposes.

Fortunately, however, there is another type of data, known as order book data (also called limit book data). A typical data element of the limit book will include an order not yet executed, either to buy or to sell a certain amount at a given price. When for example a large order to "buy at market" arrives, it is fulfilled by climbing up the "limit book ladder," and therefore a buy order will tend to raise the market price. A sell order reverses the preceding, by symmetry. The problem is that order book data is difficult to obtain for research purposes. The authors are very grateful to Robert Ferstenberg of Morgan Stanley for having graciously provided the authors with order book data [5], which was key to this study.

In addition to the standard tick data, the data set of Morgan Stanley provides the complete order book for selected stocks. In other words, this data set contains each posting of the top 10 bids and the top 10 asks, including both prices and shares available. Further, the top 10 bids and asks in this data set are aggregated. For example, at a fixed time t_0 suppose the highest bid price for a stock is P and there are N traders in the market each willing to buy x_i shares at P. Then the top bid entry at time t_0 would be for $\sum_{i=1}^{N} x_i$ shares at a price of P. This is a dramatic improvement over standard tick data alone.

The Morgan Stanley data tells us how many shares the market is willing to buy or sell at prices near the quoted price as well as tick data for trades. Unlike standard tick data, this gives us information about supply and demand for a stock. In fact, it gives us actual points on the supply curve. Consider the above example about the top bid entry at time t_0. Since there are $\sum_{i=1}^{N} x_i$ shares available for sale at price P, a trade will occur at a price of P if a seller comes to the market and initiates a trade for X shares at price P, where $X \leq \sum_{i=1}^{N} x_i$. This bid entry corresponds to a potential seller-initiated trade. Similarly, the top ask entry corresponds to a potential buyer-initiated trade.

In general, suppose there are N_B aggregate bids at prices P_{B_i} for x_{B_i} shares at time t_0. If a seller comes to market at time t_0 and sells K shares, then the first x_{B_1} shares will sell at a price of P_{B_1}. The seller will then sell x_{B_2} shares at a price of P_{B_2}, and so on until he has sold all K shares. Thus the average price that the seller pays at t_0 per share is $\frac{\sum_{i=1}^{n-1} x_i P_i + X_n P_n}{\sum_{i=1}^{n-1} x_i + X_n}$ where $n = \inf\{k : \sum_{i=1}^{k-1} x_i \leq K < \sum_{i=1}^{k} x_i\}$ and $X_n = K - \sum_{i=1}^{n-1} x_i$.

If we let $K_m = \sum_{i=1}^{m} x_i$, and we define $P_m = \frac{\sum_{i=1}^{m} x_i P_i}{\sum_{i=1}^{m} x_i}$, then $(-K_m, P_m)$ is a point on the supply curve. Note that we set the trade size to be negative because trades for K_m shares at a price of P_m would be seller initiated. Similarly, the aggregate asks give rise to buyer-initiated points on the supply curve.

This effectively gives us the "inner" portion of the supply curve. We say "inner portion" because the data only gives us the top 10 *best* bids and the top 10 *best* asks. Thus bids and asks that are far from being in the top 10 will not show up in this data set. Moreover, for many stocks, since there are often 10 bids and 10 asks in the data set and never less than 10, it seems reasonable to assume that at any given time there are more (perhaps many more) than the 20 best bids and asks. Thus it is reasonable to assume that we are dealing with truncated curves and only seeing the “middle” of them.

3.2. *The supply curve for liquid stocks*

In this section we only consider the most liquid stocks traded on the New York Stock Exchange. This is more or less coincidental with the OEX 100. These are the stocks use to comprise the S&P 100, a subset of the S&P 500, which consists of the more liquid stocks. We made movies of this using time dependent data. Visual inspection of the movies indicated that the supply curves were linear, with no nonnegligible bid/ask jump, with slopes that moved randomly, but with a very small variance. Since the slope of these lines was small, it is important to see if they differ significantly from a horizontal line, which would render the theory vacuous, as previously mentioned.

The first step in investigating the supply curve is first to confirm statistically that it is non-trivial. In other words, we want to reject the classical case that $S(t,x) = S(t,0)$.

We assume a linear supply curve of the form

$$S(t,x) = Mx + b(t) \tag{3.1}$$

The parameter we are concerned with is the slope, M. We observe n supply curves, $\{\hat{S}_i(x) = m_i x + b_i\}_{i=1}^n$, and we assume that the slope residuals $M - m_i \sim N(0,\sigma^2)$. Thus $Z = \frac{m_i - M}{(\sigma_{m_i - M})/\sqrt{n}} \sim N(0,1)$.

To show that the supply curve is non-trivial, we test the null hypothesis $H_0 : M = 0$ against the alternative $H_1 : M \neq 0$. Thus $Z = \frac{m}{\sigma_m/\sqrt{n}} \sim N(0,1)$, and our test statistic, which has the t-distribution with $n-1$ degrees of freedom, is $t = \frac{\bar{m}}{S/\sqrt{n}}$, where S is the sample variance of the observed m_i. The data used is order book data for the week of 7-14-2003–7-18-2003.[7]

Theorem 3.1. *Assuming a supply curve of the linear form* (3.1), *with the null hypothesis* $H_0 : M = 0$ *and the alternative* $H_1 : M \neq 0$, *we are able to reject* H_0 *at the* 99% *confidence level for* 1937 *out of the* 2066 *stocks*[8] (93.8%) *in our data set. Moreover*, 92.3% *of the p-values are smaller than* 0.001.

The results of this test strongly indicate that a non-trivial supply curve does exist for most stocks.

[7]There is nothing particularly special about this time period. It was arbitrarily chosen for these trials.

[8]2066 out of the 2181 stocks in our data set had sufficient market activity during the specified time period to perform this test.

Table 1. Some p-values for rejection of the trivial case.

Ticker	p-value
azn	0.0
bp	0.0
emg	0.0
cost	5.4×10^{-14}
hps	1.6×10^{-8}
lac	1.1×10^{-5}
sey	1.5×10^{-3}
lngo	0.612

Visual analysis of the data strongly suggests in many cases,[9] namely those of highly *liquid*[10] stocks, that the supply curve is linear in x. This leads us to assume a supply curve of the form

$$S(t, x) = M(t)x + b(t) \tag{3.2}$$

Note that $b(t) = S(t, 0)$, so we can think of $b(t)$ as the stock price in the classical case, such as a geometric Brownian motion. $M(t)$ can be thought of as a differentiable function; however estimation of it as a function we leave to subsequent research. It can also be thought of as a mean reverting stochastic process, and that is currently being investigated by Protter and Talay [12]. We sample it at discrete times. Intuitively, we should think of $M(t)$ as a liquidity adjustment. It represents how much more we must pay per share for an order of size x.

3.3. *Goodness of fit using splines*

Using the Morgan Stanley data, we do a goodness of fit analysis for $S(t, x)$ (3.2). For data in a time interval $[0, T]$ we consider two supply curves,

$$\hat{S}_i^L(x) = m_i x + b_i \tag{3.3}$$

$$\hat{S}_i^S(x) = m_i x + b_i + \sum_{k=1}^{M_2} \alpha_{i,k}(x - \lambda_{i,k})_+^2 + \sum_{k=1}^{M_3} \beta_{i,k}(x - \theta_{i,k})_+^3 \tag{3.4}$$

Equation (3.3) is the linear model, and (3.4) is a cubic spline which contains the linear terms in (3.3) as well as M_2 quadratic spline terms and M_3 cubic spline terms. For both the linear and spline models we construct n supply curves, $\{\hat{S}_i^L(x)\}_{i=1}^n$ and $\{\hat{S}_i^S(x)\}_{i=1}^n$, where each curve is constructed using at most 20 data points from the order book[11] $\{(x_{i,j}, p_{i,j})\}_{j=1}^{n_i}$, $n_i \leq 20$.

[9] Using Matlab we have created several movies of our supply curve models. They give a nice intuition for the models and how well they fit the data. Some of these are currently available at the URL http://people.orie.cornell.edu/∼protter/Movie.wmv.
[10] *Liquid* in the sense that the stocks are traded frequently.
[11] At any given time the order book never has more than 10 best bids and 10 best asks.

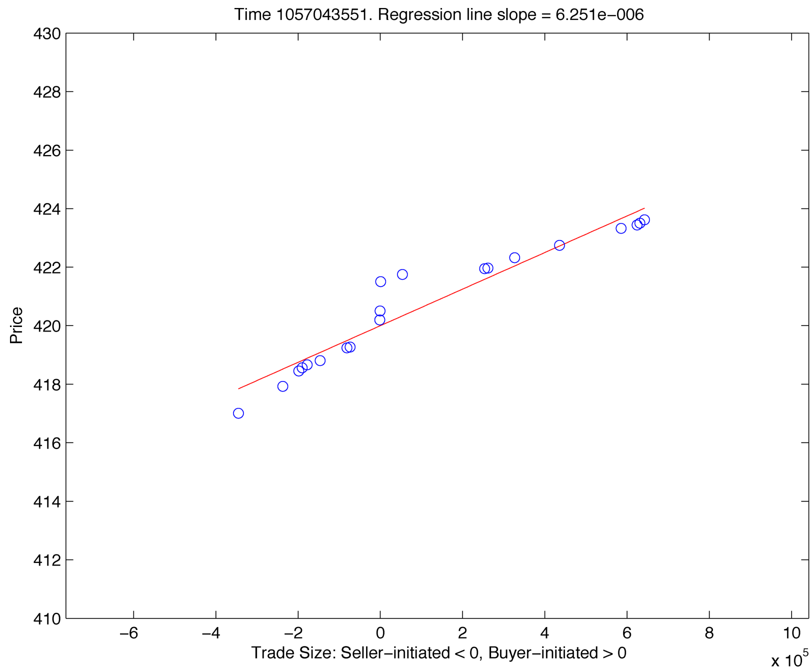

Fig. 1. British Petroleum: Linear supply curve.

In order to measure how well the linear model fits the data, we compute the test statistic

$$F_i = \frac{(SSR_i^L - SSR_i^S)/(M_2 + M_3)}{SSR_i^S/(n_i - M_2 - M_3 - 3)}. \tag{3.5}$$

F_i has an F-distribution with M_2+M_3 numerator degrees of freedom and $n_i-M_2-M_3-3$ denominator degrees of freedom. In order to test the goodness of fit for each i, we formulate the following hypothesis test using data for the week of 7/14/2003–7/18/2003.[12] We test the null hypothesis $H_0 : \alpha_{i,1} = \alpha_{i,2} = \cdots = \alpha_{i,M_2} = \beta_{i,1} = \beta_{i,2} = \cdots = \beta_{i,M_3} = 0$ against the alternative $H_1 : H_0$ is not true. By accepting H_0 we show that the linear model fits the data and the extra variability provided by the spline fit is not adding significant information to the model.

Theorem 3.2. *Using the supply curve models specified in* (3.3) *and* (3.4), $M_2 = 5$, $M_3 = 5$, *with the null hypothesis* $H_0 : \alpha_{i,1} = \alpha_{i,2} = \cdots = \alpha_{i,M_2} = \beta_{i,1} = \beta_{i,2} = \cdots = \beta_{i,M_3} = 0$, *and the alternative* $H_1 : H_0$ *is not true, we are able to not reject* H_0 *at the* 95% *confidence level for* 19 *out of* 23 *liquid stocks*[13] (82.6%) *for more than* 90% *of the fitted supply curves* $\hat{S}_i^L(x)$ *and* $\hat{S}_i^S(x)$.

[12]There is nothing particularly special about this time period. It was arbitrarily chosen for these trials.
[13]"Liquid" in the sense that the order book is updated frequently and has many entries.

This hypothesis test indicates that the model (3.2) fits the data closely in the majority of cases.[14]

3.4. *Stock splits*

Suppose 1 unit of a given commodity splits into k units at time τ. This will affect the shape of the supply curve because a position of x shares in the stock immediately before the split needs to have the same value at the time of the split. We have a no arbitrage condition

$$\lim_{t \to \tau^-} xS(x,t) = kxS(kx,\tau) \tag{3.6}$$

This condition says that the value of x shares does not change due to stock splitting.[15]

Theorem 3.3 (The Inverse Square Law for Liquidity). *Suppose* 1 *unit of a given commodity with a supply curve as in* (3.2) *splits into k units at time τ, then*

$$S(x,\tau) = \frac{1}{k^2} \lim_{t \to \tau^-} M(t)x + \frac{1}{k} \lim_{t \to \tau^-} b(t). \tag{3.7}$$

Proof. Setting $u = kx$, using (3.2), and substituting into the no arbitrage condition (3.6) gives

$$\frac{u}{k} \lim_{t \to \tau^-} S\left(\frac{u}{k}, t\right) = uS(u,\tau) \tag{3.8}$$

and $S(u,\tau) = \frac{1}{k} \lim_{t \to \tau^-} [M(t)\frac{u}{k} + b(t)]$. Thus

$$S(kx,\tau) = \frac{1}{k} \lim_{t \to \tau^-} S(x,t) \tag{3.9}$$

and

$$M(\tau)kx + b(\tau) = \frac{1}{k} \lim_{t \to \tau^-} (M(t)x + b(t)) \tag{3.10}$$

Since this holds for all x, plugging in $x = 0$ gives

$$b(\tau) = \lim_{t \to \tau^-} \frac{1}{k} b(t) \tag{3.11}$$

Now using (3.11) and (3.10) gives

$$M(\tau)kx = \frac{1}{k} \lim_{t \to \tau^-} M(t)x \tag{3.12}$$

(3.11) and (3.12) give the desired result. □

[14] Note that we do expect the spline fit to be a better fit for beginning and end-of-day data.

[15] Suppose $\lim_{t \to \tau^-} xS(x,t) < kxS(kx,\tau)$, then buy x shares before τ. Once the stock splits, sell kx shares, and an instantaneous profit is made. If the inequality is reversed, short x shares of the stock before time τ and buy kx shares at time τ.

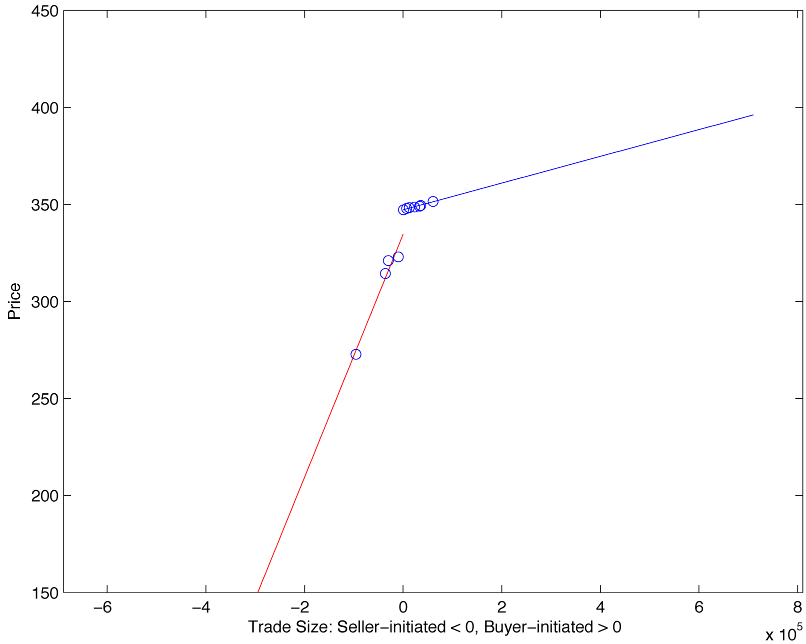

Fig. 2. Allied Domecq: Jump linear supply curve.

4. The Illiquid Case

The data suggests that for highly illiquid commodities,[16] there is a jump at $x = 0$. For this case, we introduce the *jump-linear supply curve*

$$S(t, x) = [M_{-}(t)x + b_{-}(t)]1_{\{x<0\}} + [M_{+}(t) + b_{+}(t)]1_{\{x\geq 0\}} \tag{4.1}$$

We fit two regression lines to the data in this case. One line is fitted to the best bids and another is fitted to the best asks. Using $r_i^2 = \frac{SSR}{SST}$ yields two correlation coefficients for each line, r_B and r_A, for the bids and asks, respectively.

We do not perform a goodness of fit test for this model. This model has more freedom in fitting the data than does the linear supply curve model (3.2) because two line segments will fit a set of data points better than one line segment. Doing a goodness of fit test in the same manner as in Theorem 3.2 would thus result in an even better fit, and as we have already seen, the linear model fits extremely well. What we are concerned with is when this model fits the data significantly better than the linear model. To do this we define

$$\overline{\rho^2} = \alpha_A \overline{r_A^2} + \alpha_B \overline{r_B^2} \tag{4.2}$$

[16]Again, see the movies currently available at the URL http://www.people.orie.cornell.edu/~protter/Movie.wmv for a dramatic illustration of this point.

where $\alpha_A + \alpha_B = 1$. $\overline{\rho^2}$ is similar to $\overline{r^2}$ in that it is between zero and one, and a value close to one indicates a good fit for the model.

The weights α_A and α_B can be chosen according to different conventions. $\alpha_A = \alpha_B = \frac{1}{2}$ is a natural choice. For our analysis, we choose $\alpha_A = \frac{N_A}{N}$ and $\alpha_B = \frac{N_B}{N}$, where N is the total number of regressions done,[17] N_B and N_A are the numbers of regression lines fit to the bids and to the asks, respectively. This gives more weight to the side of the market that changes more frequently.

We want to see if the jump-linear model is more appropriate than the linear model on a case by case basis. In doing so, there are three things that we should take into account:

(1) The degree to which the jump-linear model better fits the data compared to the linear model.
(2) The size of the average bid-ask spread.
(3) The average number of points in the order book.[18] If there are fewer than 20 points, then the order book is complete, and the market for that stock is not as active as stocks that average more points in their order books.

The better the jump-linear model outperforms the linear model, the larger we can expect $\overline{\rho^2} - \overline{r^2}$ to be. Based on the previously computed values of $\overline{r^2}$ we expect $1 - (\overline{\rho^2} - \overline{r^2})$ to be close to 1 in the majority of cases. We define

$$\phi(\overline{\rho^2}, \overline{r^2}) = \min[1 - (\overline{\rho^2} - \overline{r^2}), 1]. \tag{4.3}$$

Suppose we are given K order book updates in a time interval $[0, T]$. At time t, let δ_t be the width of the bid-ask spread, and let p_t denote the midpoint of the bid-ask spread at time t (the market price). We define $\overline{\Delta}$ as the normalized average relative width of the bid-ask spread over $[0, T]$

$$\overline{\Delta} = \frac{1}{K}\sum_{i=1}^{K} \frac{\delta_{t_i}}{p_{t_i}}. \tag{4.4}$$

We define $\overline{\theta}$ as the average size of the order book over $[0, T]$. Suppose each order book update has η_i points, then

$$\overline{\theta} = \frac{1}{K}\sum_{i=1}^{K} \frac{\eta_i}{20}. \tag{4.5}$$

In the cases of (4.3) and (4.5), the statistics defined will be smaller for illiquid stocks and larger for liquid stocks. Because liquid stocks usually have a smaller bid-ask spread, $\overline{\Delta}$ (4.4) should be small for liquid stocks, thus $1 - \overline{\Delta}$ should be be close to 1. For illiquid stocks with larger bid-ask spreads, $1 - \overline{\Delta}$ will be significantly

[17] N is also the number of order book updates, as only one side of the market (either the best bids or the best asks) is updated at a time.
[18] At any given time, there are at most 10 best bids and 10 best asks. Fewer than 20 such points would mean that the order book shows all of (or most of) the standing orders in the market.

below 1 and in some cases ≤ 0. Since illiquid stocks do not trade as actively as liquid stocks, the order book usually contains fewer points, thus $\overline{\theta}$ will be smaller for illiquid stocks. To get a rough measure of which model to use, we define

$$\psi(\overline{\Delta}, \overline{\theta}, \overline{\rho^2} - \overline{r^2}) = \overline{\theta} \cdot \phi(\overline{\rho^2}, \overline{r^2}) \cdot (1 - \overline{\Delta}). \tag{4.6}$$

The further ψ is from 1, the more the jump-linear model is preferable to the linear model. This provides us with a heuristic measure of liquidity for a stock because ψ contains information about both the linear and jump-linear fits, the average size of the bid-ask spread, and the average number of entries in the order book. Deciding on a critical value ψ^* such that stocks with $\psi < \psi^*$ are considered illiquid and $\psi^* \leq \psi$ are considered liquid is a heuristic matter. Though it is a statistic computed from our two models, it serves as a backwards parameter for choosing the appropriate model for a set of stocks.

Using Morgan Stanley data for the week of 3/14/2003–3/18/2003,[19] we computed ψ for 1218 stocks. We obtained the results in Table 2. The median ψ value was 0.136. Based on on visual inspection of the jump-linear fit, once ψ gets above 0.5, it seems that the linear model is the most appropriate model. For the stocks with $\psi < 0.25$, the jump-linear model seems to fit best. As ψ increases from 0.25 to 0.5, we see that the jump-linear model is still the best fitting model, but the bid-ask spreads tend to shrink significantly and the linear model becomes a nicer fit. Often in this range we will see jump-linear fits that have "kinks" at $x = 0$ because the slopes of the two regression lines differ significantly.

It's not until around $\psi = 0.5$ that the linear model looks like it should be used. This is particularly interesting because it shows that less than 12% of the stocks that we used would be considered liquid by this criteria. The 963 stocks we omitted from this simulation would almost all actually be considered illiquid. Indeed, we omitted them due to lack of market activity and lack of sufficient data to do the regressions, so realistically less than 8% of the stocks in the data set provided to us by Morgan Stanley would be considered liquid. This represents slightly less than half of the OEX S&P 100.

Table 2. 7/14/2003–7/18/2003 numerical results for 1218 stocks.

ψ Range	Number	Percentage
$0 < \psi < 0.1$	135	11.1
$0.1 \leq \psi < 0.25$	895	74.1
$0.25 \leq \psi < 0.5$	39	3.2
$0.5 \leq \psi < 0.75$	76	6.3
$0.75 \leq \psi < 1$	71	5.8

[19]There is nothing particularly special about this time period. It was arbitrarily chosen for these trials.

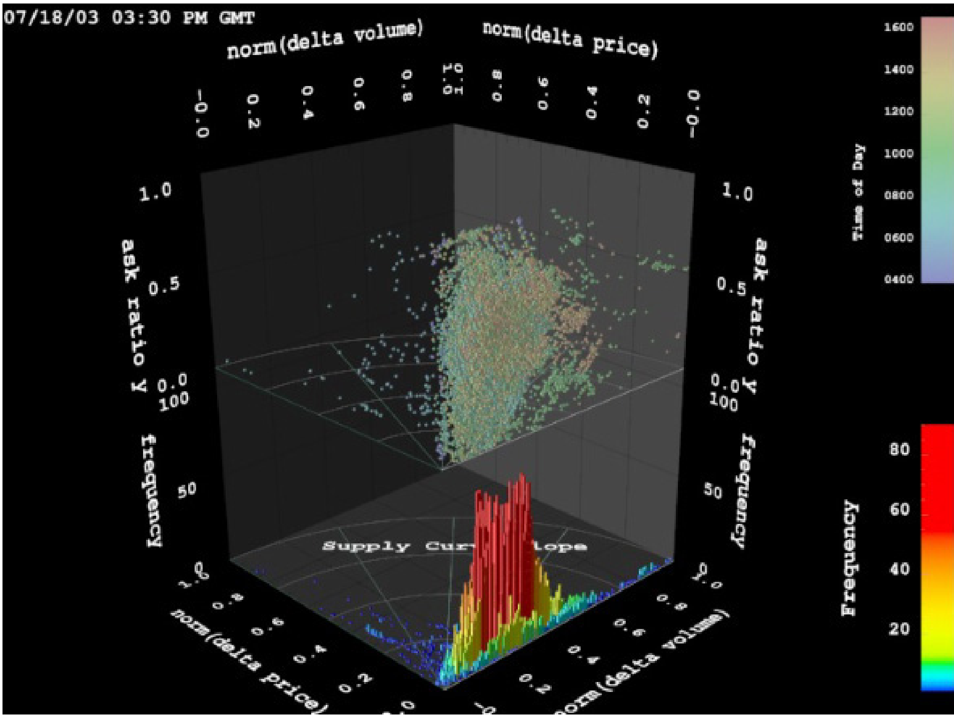

Fig. 3. AstraZeneca supply curve evolution.

We end this section with a three dimensional graph of supply curves (where the third dimension is time) of a week's data for the relatively illiquid stock, AstraZeneca. We thank Chris Pelkie for help with this graphic.

5. Option Pricing and Liquidity

When pricing options in the Black-Scholes framework, one obtains the classical Black-Scholes formula for the option price because one prices for a trade of size zero. This is a natural way to price an option when trying to answer the question, "What is *the* price of a European call option?"

The problem is that in reality the classic Black-Scholes method is not an accurate way to value an option. If you have a long position in an option, then you are holding $x > 0$ of the options. If you have a short position in an option, then you are holding $x < 0$ of the options. The question now becomes, "What is the value of my position in these options?" Using this approach, we do not wipe out the non-trivial part of the supply curve when pricing the option, and things get more interesting: The price depends on our hedging strategy, which of necessity involves liquidity.

Consider pricing a European call option, C, with maturity T and strike price K. Here we consider the number of options in the pricing scheme. The natural question to address is, "What is the value of x such options?" Based on the answer to that question, we derive the option price which depends on x. For a position in x options, we denote the price per option as $C(x)$. We also assume that if an option lands in the money, then its payoff is constructed instantaneously at maturity time T.

Theorem 5.1. *Consider a European call option and assume a supply curve as in* (3.2) *for the underlying asset.* $C(x)$ *is monotone-decreasing in* $x \geq 0$. *In particular*, $C(x) \leq C(0)$ *for any* $x \geq 0$.

Proof. If the option is in the money at time T, then the party with the long position in the option buys x shares of the underlying stock S at a price of K. When selling these shares of S at time T, a price of $S(T, -x)$ is paid per share. After selling x shares, the payoff per share of this position is

$$C_T(x) = e^{-rT} \max[S(T, -x) - K, 0] \tag{5.1}$$

To get the option price $C(x)$, we use the risk neutral measure Q and take the expectation. The risk neutral measure can come from the underlying stock price $S(t, 0)$ which is equal to $b(t)$. However if one models $M(t)$ as a nonnegative mean reverting diffusion, for example, then one introduces a new Brownian motion, and a new source of randomness, rendering the model incomplete. Q remains a risk neutral measure, but it is no longer unique, and merely one choice from an infinite number. If one introduces a volatility derivative to the market which is traded, however, such a derivative can render the market complete. In this case, we of course take Q be to the risk neutral measure obtained from this procedure. For any choice of risk neutral measure Q we have $C(x) = E^Q[\max(e^{-rT}(S(T, -x) - K, 0))]$. Thus

$$C(x) = e^{-rT} E^Q[\max(-xM(T) + S(T, 0) - K, 0)] \tag{5.2}$$

Observing that in (5.2) x only appears in the negative term $-xM(T)$, we get the desired result. $\square$

Further note that the result (5.2) is consistent with classical result.

For a position in x European put options, we denote the price per option as $P(x)$, and we get the following result:

Theorem 5.2. *Consider a European put option and assume a supply curve as in* (3.2) *for the underlying asset.* $P(x)$ *is monotone decreasing in* $x \geq 0$. *In particular*, $P(x) \leq P(0)$ *for any* $x \geq 0$.

We omit the proof of this result as it is very similar to the proof in the case of the European call.

In the classical theory, it is always optimal to exercise all options that land in the money. Now that the stock price is also a function of trade size, this no longer holds *a priori*. In creating the payoff of a position in a number of options by going to the market and trading the underlying at time T the option payoff changes with the size of the position. In fact it is *possible that it is not always optimal to exercise all options that land in the money*. Thus given a position of x options, since the option price varies with x, the next question to ask is: how many of the options should be exercised at time T?

Theorem 5.3. *Suppose we are given a long position of x European call options with strike price K and maturity time T that land in the money. We assume a supply curve as in (3.2) for the underlying asset. Assuming we want to maximize the payoff of this position at time T, the number of shares exercised should be* $\min\left(x, \frac{S(T,0)-K}{2M(T)}\right)$.

Proof. The goal is to maximize the payoff of the long position, thus we need to solve the optimization problem

$$\max_{0\le y\le x} y \cdot \max[S(T,-y) - K, 0] \tag{5.3}$$

Using (3.2), we get

$$\max_{0\le y\le x} [yS(T,0) - y^2 M(T) - Ky, 0] \tag{5.4}$$

Setting $\frac{\partial}{\partial y}[(S(T,0) - K)y - M(T)y^2] = 0$ yields the optimal solution of

$$x^* = \frac{S(T,0) - K}{2M(T)} \tag{5.5}$$

for the quadratic term.

Since the nonzero term in (5.4) is quadratic in y, and we have a long position in x options, if the option lands in the money $(S(T,0) > K)$, then the number of options exercised should be $\min(x, x^*)$. $\square$

Theorem 5.4. *Suppose we are given a long position of x European put options with strike price K and maturity time T that land in the money. We assume a supply curve as in (3.2) for the underlying asset. Assuming we want to maximize the payoff of this position at time T, the number of shares exercised should be* $\min\left(x, \frac{K-S(T,0)}{2M(T)}\right)$.

We omit the proof of this theorem as it is very similar to the proof in the case of the call option.

This tells us that if one has a long position in an option, it is not necessarily optimal to exercise all of the options if the option lands in the money. The size of the trades that must be made to create the option payoff at time T cause a reduction in the payoff. Note that (5.5) is also consistent with the classical case. In that case, $M(T) = 0$, thus we can think of x^* as ∞, and it would be optimal to exercise all options that land in the money.

6. Conclusion

We took as given the liquidity analysis of Çetin, Jarrow and Protter [2], which depends on a Supply Curve. Using order book data provided to us by Robert Ferstenberg of Morgan Stanley [5] we analyzed this model, first by showing that supply curves really do exist, and second, by exhibiting their structure. Through the use of methods to visualize the data, we assumed that for highly liquid stocks the supply curve structure was linear with a dynamically changing random slope, and we used the method of cubic splines to confirm this hypothesis. We also showed that a jump linear form was present for less liquid stocks.

Acknowledgments

Philip Protter is supported in part by NSF grant DMS-0906995. The authors gratefully acknowledge the help and wisdom supplied by Robert Jarrow throughout the research of this paper.

References

[1] M. Blais and P. Protter, Signing trades and an evaluation of the Lee-ready algorithm, Preprint (2010).
[2] U. Çetin, R. Jarrow and P. Protter, Liquidity risk and arbitrage pricing theory, *Finance and Stochastics* **8** (2004) 311–341.
[3] U. Çetin, R. Jarrow, P. Protter and M. Warachka, Pricing options in an extended black scholes economy with illiquidity: Theory and empirical evidence, *Review of Financial Studies* **19** (2006) 493–529.
[4] U. Çetin, M. Soner and N. Touzi, Option hedging for small investors under liquidity costs, to appear in *Finance and Stochastics* (2008).
[5] R. Ferstenberg and H. Alexander, Morgan stanley order book data, July 1, 2003–December 19 (2003).
[6] L. Glosten and P. Milgrom, Bid, Ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* **14**(March) (1985) 71–100.
[7] S. J. Grossman and M. H. Miller, Liquidity risk and option pricing theory, *Journal of Finance* **43** (1988) 617–633.
[8] R. Jarrow and P. Protter, Liquidity risk and option pricing theory, in *Handbook in Operation Research and Management Science: Financial Engineering*, J. Birge and V. Linetsky, eds. (North Holland, 2007), pp. 727–762.
[9] E. Jouini, Price functionals with Bid-Ask spreads: An axiomatic approach, *Journal of Mathematical Economics* **34**(4) (2000) 547–558.
[10] A. Kyle, Continuous auctions and insider trading, *Econometrica* **53** (1985) 1315–1335.
[11] C. Lee and M. Ready, Inferring trade direction from intraday data, *Journal of Finance* **46** (1991) 733–746.
[12] P. Protter and D. Talay, Viscosity stochastic control of liquidity in asset pricing, Working paper (2010).