

Optimal Portfolio Choice and the Valuation of Illiquid Securities

Francis A. Longstaff

University of California, Los Angeles

Traditional models of portfolio choice assume that investors can continuously trade unlimited amounts of securities. In reality, investors face liquidity constraints. I analyze a model where investors are restricted to trading strategies that are of bounded variation. An investor facing this type of illiquidity behaves very differently from an unconstrained investor. A liquidity-constrained investor endogenously acts as if facing borrowing and short-selling constraints, and one may take riskier positions than in liquid markets. I solve for the shadow cost of illiquidity and show that large price discounts can be sustained in a rational model.

The brass assembled at headquarters at 7 a.m. that Sunday. One after another, LTCM's partners, calling in from Tokyo and London, reported that their markets had dried up. There were no buyers, no sellers. It was all but impossible to maneuver out of large trading bets.—*Wall Street Journal*, November 16, 1998.

1. Introduction

A fundamental assumption underlying the traditional intertemporal portfolio choice paradigm of Merton (1969, 1971, 1973a), Dybvig and Huang (1988), Cox and Huang (1989), Karatzas et al. (1987), and others is that securities can be traded continuously in unlimited amounts. This assumption also underlies standard option pricing theory, such as Black and Scholes (1973), Merton (1973b), Harrison and Kreps (1979), and Harrison and Pliska (1981) where the number of shares of stock needed to replicate an option generally follows a stochastic process of unbounded variation, implying that infinite amounts of the stock must be traded.

In reality, however, investors face liquidity constraints in virtually all financial markets. Being unable to initiate or unwind a portfolio position instantly is a fact of life for traders in most financial markets, a lesson painfully learned by a number of highly leveraged hedge funds recently faced with the dilemma of raising cash to meet margin calls by unwinding positions in

I am grateful for the comments of Shlomo Benartzi, Antonio Bernardo, Michael Brennan, Vikram Kuriyan, Steven Lippman, Frank Milne, Alexander Reisz, Avaniidhar Subrahmanyam, and seminar participants at Indiana University, New York University, and the Federal Reserve Bank of Kansas City. I am particularly grateful for the comments and suggestions of the editor Bernard Dumas and of two anonymous referees. All errors are my responsibility. Address correspondence to Francis A. Longstaff, Anderson School, UCLA, Box 951481, Los Angeles, CA 90095-1481, or e-mail francis.longstaff@anderson.ucla.edu.

markets where liquidity had almost disappeared.¹ This inability is a subtle form of market incompleteness that exposes investors to additional risks not present in the traditional portfolio choice problem. To mitigate the effects of illiquidity, a risk-averse investor may select a portfolio very different from that which would be optimal if trading was unconstrained. This has important implications for asset pricing because the relative valuation of liquid and illiquid securities should reflect any welfare loss incurred by investors from their inability to trade in unlimited amounts.

The extent to which liquidity affects security prices has itself become a controversial issue in asset pricing. There is a widespread view on Wall Street that the liquidity of a security is a major determinant of its value. This view is strongly supported by recent empirical studies documenting that illiquid securities are priced at large discounts to otherwise identical liquid securities. For example, Amihud and Mendelson (1991) and Kamara (1994) show that the yield spread between illiquid Treasury notes and liquid Treasury bills of the same maturity averages more than 35 basis points. Boudoukh and Whitelaw (1991) find that the yield spread between the designated benchmark Japanese government bond and similar but less liquid Japanese government bonds averages more than 50 basis points. Silber (1992) shows that Rule 144 letter stock with a two-year liquidity restriction is privately placed at an average discount of 35% to otherwise identical registered stock. These and other similar results suggest that the market price of liquidity is very high.² In fact, the apparent price of liquidity is so high that critics of the efficient market paradigm argue that discounts for illiquidity are too large to be rational and view their size as clear evidence that investor sentiment drives prices in security markets.

In an effort to better understand the role that liquidity plays in security valuation, this article analyzes a continuous-time partial-equilibrium model in which an investor makes optimal portfolio decisions but is restricted to trading strategies that are of bounded variation. This is consistent with the characteristics of actual financial markets, where it may take an extended period of time to accumulate or unwind a specific portfolio position. In the academic literature, illiquidity has traditionally been measured in terms of bid-ask spreads or transaction costs. Among practitioners, however, illiquidity is often viewed as the risk that a trader may not be able to extricate himself from a position quickly when need arises. This article models liquidity in a way that is consistent with this latter definition. To provide a concrete motivation for trading, the continuous-time framework allows the volatility of returns to be stochastic.

¹ Examples include Long Term Capital Management, Ellington Capital Management, D. E. Shaw & Co., MKP Capital Management, and Askin Capital Management.

² For additional empirical evidence about the market price of liquidity, see Trout (1977), Pratt (1989), Wruck (1989), Cornell and Shapiro (1990), Cornell (1993), Daves and Ehrhardt (1993), Longstaff (1995a, 1995b, 2000), and Grinblatt and Longstaff (2000).

I solve first for the optimal portfolio strategy of an investor in the presence of liquidity constraints and compare it with the unconstrained optimal strategy. I show that the investor's optimal strategy consists of trading as much as possible, whenever possible. This contrasts with the optimal strategy for an investor facing transaction costs who trades only when large changes in value occur.³ Because the investor in my model can trade only at a limited rate, he has less control over the support of his wealth distribution. An important implication of this is that the investor endogenously acts as if facing borrowing and short-selling constraints, even though these constraints are not imposed. Despite this cautious behavior, however, the investor may choose to hold more of the risky asset than would be optimal in the absence of liquidity restrictions. In general, a constrained investor must hedge against both expected and unexpected changes in portfolio weights. In contrast, portfolio weights are completely under the control of the investor in the unconstrained portfolio problem.

Given the optimal strategy, we then solve for the investor's derived utility of wealth. The shadow price of liquidity is determined by comparing the constrained and unconstrained utilities of wealth and solving for the discount in the price of the illiquid asset that compensates the investor for the liquidity restrictions. We present a variety of numerical examples of the illiquidity discounts generated by the model. These results show that the discount for illiquidity can be substantial, particularly for assets that are traditionally margined or leveraged, such as stock, partnership interests, derivatives, real estate, and hedge-fund holdings. Even when the endogenous borrowing constraint is not binding, implied discounts for illiquidity can be on the same order of magnitude as those observed by Amihud and Mendelson (1991), Boudoukh and Whitelaw (1991), Kamara (1994), and others. These results offer hope that empirically observed discounts for illiquidity may be explainable within a rational model of investor behavior.

These results also contribute to the asset pricing literature in several other ways. For example, they demonstrate that the usual assumption that securities can be traded in unlimited amounts fundamentally affects the optimal portfolio strategy. This has important implications for traditional models of intertemporal portfolio choice as well as option pricing theory. In addition, the results for the unconstrained case provide an original closed-form solution to the investor's portfolio problem in a stochastic volatility model, and complement the closed-form solutions recently obtained by Liu (1999).

Related literature on the effects of liquidity on portfolio choice and asset valuation includes Mayers (1972, 1973, 1976), Brito (1977), Stapleton and Subrahmanyam (1979), Amihud and Mendelson (1986), Constantinides (1986), Dumas and Luciano (1991), Boudoukh and Whitelaw (1993), Jouini

³ Examples of these types of models include Constantinides (1986), Grossman and Laroque (1990), and Jouini and Kallal (1998).

and Kallal (1998), Huang (1998), and Constantinides and Mehra (1998). These articles typically focus on the effects of exogenous transaction costs or borrowing and short-selling constraints. This article differs in that it focuses on the endogenous effects of illiquidity on trading strategies and security values.

The remainder of this article is organized as follows. Section 2 discusses the nature of illiquidity in financial markets. Section 3 describes the continuous-time framework in which the investor makes portfolio decisions. Section 4 describes the optimal portfolio strategy in a market with liquidity constraints. Section 5 presents numerical examples illustrating how the optimal portfolio strategy differs from the unconstrained strategy and providing estimates of discounts for lack of liquidity. Section 6 summarizes the results and makes concluding remarks.

2. What Is Illiquidity?

Most market participants probably have a general sense of what *liquidity* means. Despite this, however, it is often difficult to know exactly what is meant when a market or security is referred to as being *illiquid*. The reason for this is that there are several closely related but distinct ways in which markets or securities can be illiquid. In this section, I briefly review several of the implicit definitions of *illiquidity* found in the literature.

In virtually all of the academic literature, *liquidity* is defined in terms of the bid-ask spread and/or transaction costs associated with trading a security. For example, this notion of liquidity is implicit in Glosten and Milgrom (1985), Amihud and Mendelson (1986), Constantinides (1986), Easley and O'Hara (1987), Glosten (1987), Glosten and Harris (1988), Stoll (1989), Davis and Norman (1990), Grossman and Laroque (1990), Dumas and Luciano (1991), Jouini and Kallal (1998), and many others. From this perspective, illiquidity is the situation in which investors find that they face higher trading and execution costs than at other times or in other markets. In this view, an investor can usually trade whenever desired, albeit at some (potentially high) cost.

In the practitioner literature, however, a somewhat different meaning is often attached to the term *illiquidity*. Traders view illiquidity as the situation where their ability to buy or sell securities (at any price) is limited or restricted. In extreme situations, illiquidity may be so severe that markets temporarily disappear. This type of illiquidity has more to do with the quantity of trades that can be executed than with the costs of trading. This notion of thin markets or thinly traded securities is somewhat difficult to reconcile with the standard economic view that there should be an equilibrium market clearing price at which any desired quantity can be traded.

Whatever the underlying reason for thin markets, however, recent market events, as evidenced by the introductory quotation of this article, make clear that investors can find themselves in the situation where they are not able to

trade as much as they would prefer and cannot initiate or unwind positions instantly. Although this type of illiquidity appears to be an important factor in financial markets, it has not yet received much attention in the academic literature. In this article, I develop a simple model of illiquidity that is more in the spirit of this thin-trading interpretation. In particular, I develop a model in which investors may only trade a limited amount of securities per period. The objective in doing this is to attempt to more closely capture this real-world phenomenon and to study the effects that thin trading-induced illiquidity may have on portfolio decisions and the valuation of securities.

3. The Continuous-Time Framework

In this section, I describe the continuous-time framework used throughout this article in which the investor makes portfolio decisions. As a benchmark for comparison with later results, I first characterize the optimal portfolio strategy in the absence of liquidity restrictions. To provide a concrete motivation for trading, this framework allows the volatility of risky security returns to be stochastic. Because of this, an unconstrained investor trades frequently in response to market conditions and may switch back and forth between leveraged and unleveraged positions.

I assume a simple two-asset securities market in which trade takes place continuously. The first asset is a riskless money market account with price $B(t)$ which earns the riskless rate of interest $r(t)$. The dynamics of $B(t)$ are

$$dB(t) = r(t)B(t)dt. \quad (1)$$

Because the riskless rate plays no direct role in our analysis, we assume that $r(t) = 0$, which implies that $B(t) = 1$ for all t .⁴ The second asset is risky and has price dynamics given by

$$dS(t) = (\mu + \lambda V^2(t))S(t)dt + V(t)S(t)dZ_1(t), \quad (2)$$

where μ and λ are constants, $V(t)$ is the instantaneous volatility of returns, and $Z_1(t)$ is a standard Brownian motion. The term $\lambda V^2(t)$ in the drift allows for the possibility of a volatility risk premium in the expected return of the risky asset. A similar risk premium is found in the stochastic volatility models of Merton (1980) and Cox et al. (1985). The instantaneous volatility of returns follows the dynamic process

$$dV(t) = \sigma V(t)dZ_2(t) \quad (3)$$

where σ is a constant and $Z_2(t)$ is a standard Brownian motion independent of $Z_1(t)$. This model is closely related to the stochastic volatility model

⁴ This assumption is equivalent to using the price of the money market account as the numeraire and has no effect on the results.

of Hull and White (1987) and is chosen for its simplicity.⁵ Note that from the properties of geometric Brownian motion, $V(t)$ cannot reach infinity for $t < \infty$. Because of this, it is easily shown by expressing the solution for $S(t)$ in exponential form that $S(t)$ cannot reach zero for $t < \infty$.

The investor is endowed with strictly positive initial wealth $W(0)$ and has a finite horizon T . To simplify the exposition, we assume that the investor only consumes at time T , although this assumption can be relaxed without affecting the basic results. In particular, the investor maximizes an expected utility function defined over the logarithm of his terminal wealth $W(T)$,

$$E[\ln W(T)]. \quad (4)$$

We assume logarithmic preferences to be able to focus more directly on the effects of illiquidity because the unconstrained optimal strategy is myopic and the investor does not hedge even though the investment opportunity set is stochastic.⁶ Thus, any hedging behavior in the presence of trading constraints is directly attributable to the effects of illiquidity.

Let $N(t)$ and $M(t)$ denote the number of shares of the risky and riskless securities held by the investor. The investor's wealth at time t is given by

$$W(t) = N(t)S(t) + M(t). \quad (5)$$

We require that portfolio strategies be chosen from the set of self-financing strategies. This implies the wealth dynamics

$$dW(t) = N(t) dS(t). \quad (6)$$

Following Dybvig and Huang (1988) and Cox and Huang (1989), we restrict the set of admissible trading strategies to those that imply $W(t) > 0$ for all t , $0 \leq t \leq T$. This entails little loss of generality because any strategy that allows the possibility of zero wealth cannot be optimal because $\ln(0) = -\infty$.

Because admissible strategies require $W(t) > 0$, the portfolio weight $w(t) = N(t)S(t)/W(t)$ is well defined.⁷ Substituting this into Equation (6) gives

$$dW(t) = (\mu + \lambda V^2(t))w(t)W(t)dt + V(t)w(t)W(t)dZ(t). \quad (7)$$

⁵ This model could easily be extended in several ways. For example, we could allow the volatility to follow an Ornstein-Uhlenbeck process as in Stein and Stein (1991) or a square-root process as in Heston (1993). The resulting implications for the portfolio choice problem, however, are very similar to those implied by this model.

⁶ If the investor has constant relative risk aversion (CRRA) preferences, for example, the investor chooses a portfolio that hedges against unexpected shifts in the volatility of the risky asset's return. See Merton (1971, 1973a).

⁷ If $W(t)$ can become zero, then the portfolio weight $w(t)$ can become infinite. In addition to allowing the portfolio weight to be well defined, Dybvig and Huang (1988) show that requiring $W(t) > 0$, $0 \leq t \leq T$ is sufficient to rule out arbitrage opportunities of the type discussed by Harrison and Kreps (1979).

Thus, from Equations (3) and (6), the volatility $V(t)$ and the controlled diffusion $W(t)$ follow a joint Markov process and the current values of $V(t)$ and $W(t)$ completely describe the state of the economy. Define the derived utility of wealth $J(W, V, t)$ by the following expression

$$J(W, V, t) = \max_{w(t)} E[\ln W(T)], \quad (8)$$

subject to the budget constraint (7) and where $w(t)$ is a member of the set of admissible strategies implying strictly positive wealth.

The Hamilton-Jacobi-Bellman equation for this problem is

$$\max_{w(t)} \left(\frac{w^2 V^2 W^2}{2} J_{WW} + \frac{\sigma^2 V^2}{2} J_{VV} + (\mu + \lambda V^2) w W J_W + J_t = 0 \right), \quad (9)$$

with the boundary condition $J(W, V, T) = \ln W(T)$. The first-order condition for optimality with respect to the control $w(t)$ is

$$w V^2 W^2 J_{WW} + (\mu + \lambda V^2) W J_W = 0, \quad (10)$$

which implies the optimal strategy $w^*(t)$

$$w^*(t) = - \left(\frac{\mu + \lambda V^2}{V^2} \right) \frac{J_W}{W J_{WW}}. \quad (11)$$

As in Merton (1971), we conjecture the following functional form for the derived utility of wealth:

$$J(W, V, t) = \ln W(t) + H(V, t). \quad (12)$$

Differentiating and substituting back into Equation (11) implies that the optimal portfolio weight $w^*(t)$ is given by

$$w^*(t) = \frac{\mu + \lambda V^2(t)}{V^2(t)}. \quad (13)$$

Because volatility is not constant, the optimal portfolio weight is time varying. When both λ and μ are positive, the investor holds a strictly positive amount of the risky asset. When μ is positive and λ is less than zero, however, the investor could choose to hold a leveraged position in the risky asset, an unleveraged long position in the risky asset, or even a short position in the risky asset, depending on the level of volatility. This portfolio behavior differs significantly from the constant portfolio weight strategy followed by the investor when the volatility of the risky asset is constant.

To solve for the function $H(V, t)$, we substitute the solution for $w^*(t)$ into the dynamics of wealth given in Equation (7),

$$dW(t) = \frac{(\mu + \lambda V(t)^2)^2}{V(t)^2} W(t) dt + \frac{(\mu + \lambda V(t)^2)}{V(t)} W(t) dZ_1(t). \quad (14)$$

Solving this stochastic differential equation gives

$$W(T) = W(t) \exp \left(\int_t^T \frac{(\mu + \lambda V^2(s))^2}{2V^2(s)} ds + \int_t^T \frac{(\mu + \lambda V^2(s))}{V(s)} dZ_1(s) \right). \quad (15)$$

Given the exponential form of this expression, it is straightforward to verify that the investor's wealth is strictly positive for all t , $0 \leq t \leq T$. Thus, the portfolio strategy is admissible. We then substitute this expression for $W(T)$ into Equation (8), which results in

$$J(W, V, t) = \ln W(t) + \lambda\mu(T - t) + \frac{\mu^2}{2} \int_t^T E \left[\frac{1}{V^2(s)} \right] ds + \frac{\lambda^2}{2} \int_t^T E[V^2(s)] ds. \quad (16)$$

Evaluating the indicated expectations and then integrating gives the solution for the derived utility of wealth function

$$J(W, V, t) = \ln W(t) + \lambda\mu(T - t) + \frac{\mu^2}{6\sigma^2} \frac{1}{V^2(t)} (e^{3\sigma^2(T-t)} - 1) + \frac{\lambda^2}{2\sigma^2} V^2(t) (e^{\sigma^2(T-t)} - 1). \quad (17)$$

This expression clearly satisfies the conjectured functional form in Equation (12). In addition, differentiation verifies that this expression solves the Hamilton-Jacobi-Bellman equation with the associated boundary condition.

The expressions for the optimal portfolio weight in (13) and the derived utility of wealth in Equation (17) provide a complete solution to the investor's portfolio choice problem in this stochastic volatility framework. To my knowledge, the only other closed-form solution in a stochastic volatility model is given by Liu (1999), who solves the investor's portfolio choice problem when volatility follows either an Ornstein-Uhlenbeck or square-root process. These results contribute to the literature by providing a solution when volatility follows a geometric Brownian motion.

Finally, because the number of shares $N(t)$ of the risky asset held by the investor is equal to $w^*(t)W(t)/S(t)$, Itô's lemma can be applied to solve for the dynamics of $N(t)$. The resulting expression shows that the number of shares held by the investor follows a process of unbounded variation.⁸ This implies that the optimal trading strategy in this unconstrained framework requires trading the risky security in unlimited amounts.

⁸ The variation of a function $f(x)$ defined on $[a, b]$ is the supremum over all partitions $a = x_0 < x_1 < x_2 < \dots < x_k = b$ of the sum $\sum_{i=1}^k |f(x_i) - f(x_{i-1})|$. For a function of unbounded variation, this supremum is infinite.

4. Illiquidity and Portfolio Choice

In this section, we characterize the investor's optimal portfolio strategy in the presence of liquidity restrictions. The investor is allowed to choose an initial portfolio, but can then only make limited revisions to the portfolio. Specifically, we model the liquidity restrictions by requiring that the number of risky shares held by the investor follows the dynamics

$$dN(t) = \gamma(t) dt, \quad (18)$$

where $-\infty < -\alpha \leq \gamma(t) \leq \alpha < \infty$, and $\alpha \geq 0$ is a constant.⁹ These dynamics have a number of important implications for the way in which the investor can trade. For example, the restriction on the $\gamma(t)$ term implies that there are upper and lower bounds on the number of shares of the risky asset that the investor can trade per period.¹⁰

This is directly in the spirit of the thin-trading type of illiquidity considered in Section 2; modeling illiquidity in this way closely parallels the real-world situation in which investors find that they can only execute trades for a limited quantity of a security.¹¹ In addition, the bounds on $\gamma(t)$ imply that $N(t)$ is a function of bounded variation.¹² Intuitively, this means that the sample path of $N(t)$ follows a smooth process that is differentiable almost everywhere and locally deterministic. In contrast, the number of shares held by an investor in traditional models of dynamic portfolio choice are typically of unbounded variation.¹³

⁹ In a more general model, the liquidity parameter α could be allowed to be stochastic. For example, an investor could typically face liquid markets with a large value of α , but then periodically experience "flights to quality" during which the value of α might temporarily drop to zero. I am grateful to Bernard Dumas for this insight. Because the primary results of this section depend only on the boundedness of α , the assumption that α is constant could be relaxed significantly without affecting the results.

¹⁰ This definition of illiquidity parallels Longstaff (1995a, 1995b) in the sense that restrictions on liquidity are investor-specific rather than security-specific. There are many examples of this type of illiquidity, such as Rule 144 letter stock, pension assets, or real estate, and there are also institutional or regulatory reasons why some investors might at times face greater constraints than other investors. This definition, however, makes clear the partial equilibrium nature of my analysis. I am assuming that a specific investor may face trading restrictions even though the asset is traded continuously in the market by other investors who may face fewer constraints. Although beyond the scope of this article, a more extensive analysis might consider the dynamics of prices and investor behavior in a general equilibrium framework in which all investors face trading constraints. For discussions of other definitions of liquidity, see Lippman and McCall (1986), Amihud and Mendelson (1986), Constantinides (1986), Boudoukh and Whitelaw (1993), and Huang (1998).

¹¹ Extreme liquidity constraints, such as those imposed on Rule 144 letter stock where the investor cannot trade the stock for a period of two years after the stock is acquired in a private placement, can be nested within this framework by the restriction $\alpha = 0$. For a discussion of Rule 144 letter stock, see Silber (1992).

¹² This follows from Equation (18), which implies that $N(t)$ is an absolutely continuous function. An absolutely continuous function is a function $f(x)$ defined on $[a, b]$ such that, given ϵ , there is a $\delta > 0$ such that $\sum_{i=1}^n |f(x'_i) - f(x_i)| < \epsilon$ for every finite collection $\{(x'_i, x_i)\}$ of nonoverlapping intervals with $\sum_{i=1}^n |x'_i - x_i| < \delta$. A function is absolutely continuous if, and only if, it can be expressed as an integral. Absolutely continuous functions are continuous and differentiable almost everywhere. The bounded variation of an absolutely continuous function follows from Lemma 5.4.10 of Royden (1968).

¹³ For example, the number of shares needed to replicate an at-the-money call option in the Black-Scholes model when the volatility of returns on the underlying asset is strictly positive is of unbounded variation.

With these preliminaries, we can now examine how the optimal portfolio strategy differs when liquidity is constrained. Recall that in the unconstrained case, the optimal strategy involves a leveraged position in the risky asset if $(\mu + \lambda V^2) > V^2$. When his wealth declines, the investor can sell shares of the risky asset to maintain the optimal portfolio weight. The key point here is that the investor can control the fraction of wealth held in the form of the risky asset; as his wealth approaches zero, the number of shares of the risky asset held also approaches zero and negative wealth cannot occur.

When liquidity is constrained, however, the investor no longer has complete control over the fraction of wealth held in the form of the risky asset. To see this, imagine that the investor takes a leveraged position in the risky asset. Now consider what happens when the price of the risky asset drops rapidly and approaches zero. The investor may not be able to sell shares quickly enough to unwind the leveraged position before bankruptcy occurs. Similarly, when the investor holds a short position in the risky asset and the price of the risky asset increases rapidly, the investor may not be able to cover the short position quickly enough to avoid bankruptcy. The following proposition shows that positive wealth can be guaranteed if, and only if, the investor avoids taking a leveraged or short position in the risky asset.

Proposition 1. *$W(t) > 0$ for all $t, 0 \leq t \leq T$ almost surely if, and only if, $N(t) \geq 0$, $M(t) \geq 0$, and $N(t) + M(t) > 0$ for all $t, 0 \leq t \leq T$.*

Proof. See Appendix. ■

An immediate implication of this result is that the portfolio weight $w(t)$ must satisfy the condition $0 \leq w(t) \leq 1$ for the constrained portfolio strategy to be admissible. This is true even if the probability of ruin is very small; any positive probability of bankruptcy is sufficient to make a leveraged or short position inadmissible. The optimal strategy is clearly significantly affected by the liquidity restrictions; the desire to hedge against the perhaps remote possibility of bankruptcy has a first-order effect on how an investor behaves in the presence of thin-trading constraints.

Because the dynamics of wealth depend on $W(t)$, $N(t)$, $S(t)$, and $V(t)$, all four state variables are necessary for a Markovian representation of the economy. Consequently, the investor's derived utility of wealth depends functionally on each of these state variables and is defined as

$$J(W, N, S, V, t) = \max_{w(0), \gamma(t)} E[\ln W(T)], \quad (19)$$

subject to the boundary condition $J(W, N, S, V, T) = \ln W(T)$. Observe that there are two ways in which the investor controls the stochastic evolution of his wealth. First, the investor chooses the initial number of shares $N(0)$ of the risky asset, or equivalently, the initial fraction of wealth $w(0)$ invested in the risky asset. The investor then chooses the rate $\gamma(t)$ at which to

rebalance holdings of the risky asset, subject to liquidity constraints. Again, to be admissible, $w(0)$ and $\gamma(t)$ must be such that $0 \leq w(t) \leq 1$, for all t , $0 \leq t \leq T$.

Although the investor has two controls over which to maximize expected utility, only $\gamma(t)$ is a continuous control; the control $w(0)$ is chosen at time zero. Accordingly, we find the optimal strategy by solving the Hamilton-Jacobi-Bellman equation for the continuous control $\gamma(t)$ conditional on a given $w(0)$, and then select the optimal initial portfolio weight $w(0)$ by maximizing over the conditional values of $J(W, N, S, V, t; w(0))$. When $t > 0$, the Hamilton-Jacobi-Bellman equation for the investor's problem is

$$\begin{aligned} \max_{\gamma(t)} \left(\frac{N^2 S^2 V^2}{2} J_{WW} + \frac{S^2 V^2}{2} J_{SS} + \frac{\sigma^2 V^2}{2} J_{VV} + N S^2 V^2 J_{WS} \right. \\ \left. + (\mu + \lambda V^2) N S J_W + (\mu + \lambda V^2) S J_S + \gamma J_N + J_t = 0 \right), \quad (20) \end{aligned}$$

subject to the boundary condition. Because the control $\gamma(t)$ is constrained, the first-order condition for optimality need not be satisfied. However, the control $\gamma(t)$ appears only as a coefficient in the J_N term. Thus, the Hamilton-Jacobi-Bellman equation is maximized by choosing $\gamma(t)$ to maximize the term γJ_N . This term is clearly maximized by selecting $\gamma(t) = \alpha$ if $J_N > 0$, and $\gamma(t) = -\alpha$ if $J_N < 0$, whenever this strategy is admissible, and $\gamma(t) = 0$ otherwise.¹⁴

This optimal strategy is very intuitive. When the investor's derived utility of wealth is an increasing function of $N(t)$, the investor chooses to buy additional shares of the risky asset as aggressively as possible, and vice versa. When the investor reaches $w(t) = 0$ or $w(t) = 1$, the investor cannot short or acquire more shares. Thus, the optimal strategy is to trade as aggressively as possible, whenever possible. This contrasts with the optimal trading strategy in models where investors face fixed transaction costs or similar frictions when trading securities. In those models, the typical optimal strategy is to trade only when the price of the risky asset has changed significantly.¹⁵ Because the investor's optimal portfolio strategy is essentially the bang-bang control of his trading rate, this stochastic optimal control problem parallels those described in Benes et al. (1980), Shreve (1981), and Karatzas and Shreve (1988), chapter 6.

Having determined the optimal portfolio strategy, I now turn to the problem of solving for the derived utility of wealth function. Though I cannot provide a closed-form solution, I do offer a formal solution that conveys much of

¹⁴ Note that this implies that $\gamma(t) = 0$ when $J_N = 0$. This is arbitrary, however, as $\gamma J_N = 0$ for any admissible value of $\gamma(t)$ when $J_N = 0$.

¹⁵ As an example of this type of trading strategy, see Constantinides (1986) and Grossman and Laroque (1990).

the intuition. Because $w(t)$ is well defined under the optimal strategy, the dynamics of wealth can again be expressed as

$$dW(t) = (\mu + \lambda V^2(t))w(t)W(t)dt + V(t)w(t)W(t)dZ_1(t). \quad (21)$$

Solving for $W(T)$ gives

$$W(T) = W(t) \exp \left(\int_t^T (\mu + \lambda V^2(s))w(s) - \frac{V^2(s)}{2}w^2(s)ds + \int_t^T V(s)w(s)dZ_1(s) \right). \quad (22)$$

Substituting this expression into Equation (19) results in the following formal solution for the derived utility of wealth function $J(W, N, S, V, t; w(0))$:

$$J(W, N, S, V, t; w(0)) = \ln W(t) + E \left[\int_t^T (\mu + \lambda V^2(s))w(s) - \frac{V^2(s)}{2}w^2(s)ds \right]. \quad (23)$$

This derived utility of wealth function is similar to that for the unconstrained case in that it can be separated into a $\ln W(t)$ term and an additional function.

This formal solution provides some insights into how the portfolio problem is affected by illiquidity. Equation (23) shows that the derived utility of wealth depends linearly on $w(t)$ through the first term in the integral, but also quadratically on $w(t)$ through the second term in the integral. Thus, in selecting an optimal constrained trading strategy, the investor faces a problem similar to that of a standard mean-variance maximization problem. What is different is that the trade-off is essentially between the mean and variance of the portfolio weight rather than the mean and variance of the returns of the portfolio. Intuitively, this is because the portfolio weight itself becomes a random variable when the investor faces liquidity constraints since the portfolio weight is no longer fully under his control. Thus, the presence of illiquidity has the effect of introducing a second level of mean-variance analysis into the investor's portfolio choice problem. Of course, the mean-variance problem is complicated by the fact that $V(t)$ also appears in the integral in Equation (23) and is random when $\sigma > 0$.

5. Numerical Results

To illustrate how liquidity restrictions affect optimal portfolio choice and asset valuation, this section presents a variety of numerical examples.

5.1 The numerical methodology

As shown in the previous section, the derived utility of wealth function depends on the four state variables W , N , S , and V , as well as time, when there are liquidity restrictions. In theory, standard finite difference techniques could be applied to solve the Hamilton-Jacobi-Bellman equation in Equation (20) numerically. In actuality, the dependence of the derived utility of wealth function on four state variables makes this traditional approach virtually intractable from a computational perspective.

A number of recent articles, however, demonstrate that dynamic programming problems similar to this constrained portfolio choice problem can be solved numerically using simulation techniques. Examples of this rapidly growing literature include Bossaerts (1989), Tilley (1993), Keane and Wolpin (1994), Barraquand and Martineau (1995), Carriere (1996), Broadie and Glasserman (1997a, 1997b, 1997c), Broadie et al. (1997, 1998), Raymar and Zwecher (1997), Averbukh (1997), Ibanez and Zapatero (1998), Carr (1998), Garcia (1999), Longstaff et al. (2000), Tsitsiklis and Van Roy (1999), Brandt and Santa-Clara (2000), and Longstaff and Schwartz (2001). In this section, I use the Longstaff and Schwartz (2001) technique to solve for the derived utility of wealth function in the presence of liquidity constraints; this technique is also used by Brandt and Santa-Clara (2000) in solving similar dynamic portfolio-choice problems with large numbers of state variables.

Briefly, the Longstaff and Schwartz (2001) approach, termed the LSM algorithm, is motivated by observing that the value function in a typical dynamic programming problem can be expressed as a conditional expectation function. For example, the continuation value or value of keeping an American option alive at an exercise date can be expressed as the expectation of its discounted future cash flows under the optimal exercise strategy, conditional on the current stock price. In the LSM algorithm, this conditional expectation function is estimated by regressing the realized ex post cash flows from the option on a basis set of functions of the current or ex ante stock price, where each path in the simulation is an observation in this cross-sectional regression. Using the fitted value from the regression as the estimated conditional expectation function, it is then straightforward to compare the values of continuation and immediate exercise at each exercise date along each path, resulting in a complete specification of a stopping strategy. Longstaff and Schwartz (2001) demonstrate that the LSM algorithm is accurate, converges rapidly to the option values implied by finite difference techniques, and is robust to the specification of the regression function.

I apply the LSM algorithm to this problem in the following way. First, for convenience, I redefine the state variables to be N , M , S , and V rather than W , N , S , V . This is without loss of generality because $W = NS + M$. Next, I discretize the investment horizon $[0, T]$ into equal intervals. For notational simplicity, I assume that these intervals are of length one and that the investment horizon T is expressed in units of the discretization interval. The

numerical results in this section are based on 20 periods per year; the discretization period used is .05 years. Normalizing the initial values of W and S to one, I then simulate 100,000 paths of S and V using the standard Euler approximations to the dynamics given in Equations (2) and (3). At time $T - 1$, I then draw 100,000 independent values of N and M from a uniform distribution on $[0, 1]$ and assign them randomly to the 100,000 paths of S and V . By drawing N and M from a uniform distribution, I provide the variation in the data needed to efficiently estimate a conditional expectation function while guaranteeing that the bounds imposed by Proposition 1 are satisfied.

Now observe that the derived utility of wealth function $J(N, M, S, V, T - 1; w(0))$ can be expressed as the expected value of $\ln W_T$, conditional on the state variables at time $T - 1$. To approximate this conditional expectation, I regress the ex post values of $\ln W_T$, where $W_T = N_{T-1}S_T + M_{T-1}$ on a set of basis functions of the values of the state variables at time $T - 1$.¹⁶ The fitted value from this regression now provides an efficient estimator of the conditional expectation function or derived utility of wealth $J(N, M, S, V, T - 1; w(0))$. I then differentiate the closed-form conditional expectation function with respect to N , holding W , S , and V fixed, to obtain an explicit functional approximation for J_N at time $T - 1$. With this approximation of J_N , I can then determine whether it is optimal to increase N by αdt or to decrease N by αdt at time $T - 1$, given the value of the state variables at time $T - 1$ along any simulated path.

The next step is to roll backward to time $T - 2$. Again, I draw 100,000 independent values of N and M from a uniform distribution on $[0, 1]$ and assign them randomly to the 100,000 paths of S and V . I again regress the ex post values of $\ln W_T$, this time assuming that the estimated optimal portfolio strategy is followed at time $T - 1$, on the same set of basis functions of the state variables at time $T - 2$. The fitted value from this regression now provides an estimate of the conditional expectation function or derived utility of wealth $J(N, M, S, V, T - 2; w(0))$, which is differentiated to approximate the derivative J_N at time $T - 2$. The process is repeated for each time period recursively (at each period assuming that optimal strategies are pursued at later dates), until the J_N function has been approximated for each time $1, 2, 3, \dots, T$.

With these estimates of the J_N functions and for a particular choice of $w(0)$, I can then solve for $J(W, N, S, V, 0; w(0))$ by simply taking the average value over all paths of $\ln W_T$ as given by following the optimal portfolio strategy $\gamma^*(t)$ implied by the estimated functions J_N . I then solve for

¹⁶ I use a set of 22 basis functions in estimating the conditional expectation function, consisting of a constant, $\ln W_{T-1}$ and its square, the first three powers of S_{T-1} , V_{T-1} , N_{T-1} , and M_{T-1} , and the cross products $S_{T-1}V_{T-1}$, $S_{T-1}N_{T-1}$, $S_{T-1}M_{T-1}$, $V_{T-1}N_{T-1}$, $V_{T-1}M_{T-1}$, $N_{T-1}M_{T-1}$, and $S_{T-1}^2N_{T-1}^2$. For the case where $\sigma = 0$, the terms involving V_{T-1} are omitted because they are not stochastic. I also examine a variety of alternative functional forms for the conditional expectation function with an equal or greater number of terms than this specification. The numerical results are virtually indistinguishable from those given by this specification. This is consistent with Longstaff and Schwartz (2001), who find that the algorithm is very robust to the specification of the conditional expectation function.

the optimal initial portfolio $w^*(0)$ by finding the value of $w(0)$ that maximizes the value of $J(W, N, S, V, 0; w(0))$. The derived utility of wealth $J(W, N, S, V, 0)$ is then given as $J(W, N, S, V, 0; w^*(0))$.

Because I cannot solve this portfolio choice problem using finite difference techniques, I cannot directly compare the results from the LSM algorithm with those obtained by the traditional finite difference approach. Despite this, however, there are a number of diagnostics that provide support for the reliability of the results from the LSM algorithm. First, when $\alpha = .00$, the value of derived utility of wealth can be obtained by simply optimizing the expected value of $\ln W_T$ over all values of $w(0)$. In this case, the value of the derived utility of wealth can be estimated directly, and the results indicate that the LSM algorithm converges to this estimated value.

Second, when $\sigma = .00$, the volatility of risky asset returns V is no longer stochastic, and only three state variables appear in the problem. In this special case, it is feasible to solve for the derived utility of wealth function using finite difference techniques and then compare the results with those obtained from the LSM algorithm. I do this by solving for the derived utility of wealth function using a standard finite difference algorithm with 100 grid points each for the state variables S , N , and M and requiring that the third derivatives of the derived utility function with respect to the relevant state variables equal zero on the boundaries of the grid. Though the granularity of the finite difference grid and the random noise inherent in a simulation-based algorithm lead to small differences between the two approaches, numerical tests indicate that the LSM algorithm generally gives results that agree within 1% of those obtained by the finite difference algorithm. For example, when $W = 1$, $S = 1$, $T = 1$, and $\alpha = .10$, the LSM and finite difference estimates of the derived utility of wealth are .00934 and .00927, respectively, for $V = .7071$; .02474 and .02469, respectively, for $V = .4472$; and .07496 and .07495, respectively, for $V = .2236$.

Finally, the economics of the problem imply that as α increases from zero to larger values, the constrained derived utility of wealth should increase and approach the unconstrained derived utility of wealth given in Equation (17). Extensive numerical tests indicate that the estimated LSM value of the derived utility of wealth increases smoothly from the $\alpha = .00$ value to the unconstrained value as α increases; the LSM value is almost always between the completely illiquid $\alpha = .00$ value and the unconstrained value. This feature can be seen in the numerical results that follow.

5.2 Optimal portfolio choice

Perhaps the most direct way to identify how liquidity restrictions affect portfolio decisions is to compare the investor's initial portfolio when there are restrictions to the initial portfolio chosen in the absence of restrictions. Recall that the investor is not constrained in choice of the initial portfolio; liquidity restrictions only affect the ability to rebalance the portfolio subsequently.

Table 1
Optimal initial portfolio weight for the risky asset in the presence of liquidity restrictions

Horizon	α	Volatility	Unconstrained	$\sigma = .00$	$\sigma = .20$	$\sigma = .40$	$\sigma = .60$
1	.00	.7071	.20	.182	.176	.163	.144
1	.00	.4472	.50	.499	.487	.458	.417
1	.00	.3536	.80	.800	.782	.734	.663
1	.00	.3162	1.00	.994	.968	.903	.813
1	.00	.2236	2.00	1.000	1.000	1.000	1.000
1	.00	.1414	5.00	1.000	1.000	1.000	1.000
1	.10	.7071	.20	.199	.191	.182	.168
1	.10	.4472	.50	.495	.489	.470	.427
1	.10	.3536	.80	.789	.773	.741	.678
1	.10	.3162	1.00	1.000	.934	.879	.817
1	.10	.2236	2.00	1.000	1.000	1.000	1.000
1	.10	.1414	5.00	1.000	1.000	1.000	1.000
2	.00	.7071	.20	.162	.155	.135	.114
2	.00	.4472	.50	.497	.482	.437	.380
2	.00	.3536	.80	.809	.780	.699	.604
2	.00	.3162	1.00	.998	.956	.848	.723
2	.00	.2236	2.00	1.000	1.000	1.000	.985
2	.00	.1414	5.00	1.000	1.000	1.000	.999
2	.10	.7071	.20	.169	.177	.182	.170
2	.10	.4472	.50	.466	.492	.466	.411
2	.10	.3536	.80	.837	.782	.720	.634
2	.10	.3162	1.00	1.000	.923	.855	.756
2	.10	.2236	2.00	1.000	1.000	1.000	1.000
2	.10	.1414	5.00	1.000	1.000	1.000	1.000

Horizon denotes the investor's horizon in years; α is the maximum number of shares that can be traded per year where the total number of shares that could be initially purchased is normalized to one; volatility is the current volatility of returns on the risky asset; unconstrained is the initial optimal portfolio weight for the risky asset in the absence of liquidity constraints; and σ is the volatility of volatility parameter. The expected return parameter μ is set equal to .10 and the market price of volatility risk λ equals zero.

Table 1 reports the optimal initial portfolio weights for the unconstrained and constrained cases for different values of T , V , and σ . The values of the expected return and the current volatility of the risky asset are chosen to imply unconstrained portfolio weights ranging from .20 to 5.00. The values of σ are chosen to be consistent with the historical behavior of stock index volatility. The stock index volatility is typically very high. For example, the Bloomberg system reports that the annualized standard deviation of monthly percentage changes in the Chicago Board Options Exchange VIX index of implied volatility for S&P 500 index options during the 1994–1998 period is in excess of 50%.

In general, the investor chooses a lower initial portfolio weight in the presence of liquidity constraints, even when the unconstrained weight is admissible in the constrained case. For example, when the unconstrained weight is .20, the constrained weight ranges from .181 to .144 for $T = 1$. Similarly, when the unconstrained weight is .50 or .80. Intuitively, the reason for this is that the unconstrained investor only needs to consider the current value of $V(t)$ in determining his optimal portfolio. In contrast, a constrained investor has to consider the expected value of future optimal portfolio weights as well as the current value. For example, note that the expected value of

the portfolio weight increases over time as the expected value of the stock increases relative to the value of the riskless asset. Thus, the constrained investor hedges against the expected trend in the portfolio weight by taking a smaller initial position in the risky asset. This investment behavior differs significantly from that implied by traditional models of dynamic portfolio choice. In Merton (1969, 1971) for example, investors hedge only against unexpected changes in the state of the economy. In the presence of liquidity constraints, both expected and unexpected changes in the state variables affect the optimal decision.

In some cases, however, the constrained investor selects a higher initial portfolio weight than the unconstrained investor. In particular, when $T = 2$, $\sigma = 0$, and the unconstrained optimal portfolio weight is .80, the constrained investor selects an initial portfolio weight of .809 for $\alpha = .00$, and .837 for $\alpha = .10$. Thus, the presence of liquidity constraints can result in the investor taking a more aggressive investment position. The intuition for this result is related to the distribution of $w(t)$. When $\sigma = 0$, $w(t) = 0$ and $w(t) = 1$ become absorbing states, and the variance of the portfolio weight process becomes zero at these endpoints. Thus, in trading off the expected value of $w(t)$ against the variance of $w(t)$ in Equation (23), the constrained investor has an incentive to move toward the closest endpoint to minimize the variance of $w(t)$; when the unconstrained portfolio weight is greater than .50, the constrained investor may choose to hold a more aggressive initial portfolio.

These two factors, however, do not fully explain the portfolio strategy of the constrained investor. For example, when the unconstrained optimal portfolio is $w(t) = 1$, there is no upward drift in the portfolio weight and the variance of $w(t)$ is minimized. Despite this, however, the constrained investor still selects an initial portfolio $w(0) < 1$ when $\sigma > 0$. The reason for this is due to the fact that when $\sigma > 0$, the investor must take into account the correlation between $V(t)$ and $w(t)$ in making portfolio decision, because the term $V^2(t)w^2(t)$ appears in the integral in Equation (23). Numerical tests show that there is a negative correlation between $V(t)$ and $w(t)$ that is maximized at some value of $w(t)$ that is generally less than $w(t) = 1$. Thus, even when the unconstrained optimal is $w(t) = 1$, the constrained investor may hold a less aggressive initial portfolio to benefit from the negative correlation between $V(t)$ and $w(t)$.

Table 1 also shows that as α increases from 0 to .10, the optimal initial portfolio weight generally approaches the unconstrained optimal portfolio weight. This is intuitive because increasing α relaxes the liquidity constraint faced by the investor. Finally, Table 1 shows that when the unconstrained investor holds a leveraged portfolio, the constrained investor holds a portfolio satisfying $0 \leq w(t) \leq 1$ to avoid the possibility of bankruptcy. Thus, liquidity constraints effectively impose endogenous short-selling and leverage constraints on investors.

These numerical results illustrate just how different the portfolio problem is for an investor who faces liquidity constraints. In particular, a constrained investor must now consider the portfolio weight $w(t)$ as a random variable and solve a mean-variance problem in the portfolio weight. In addition, the investor hedges against expected changes in the portfolio composition, in sharp contrast with traditional continuous-time portfolio choice results. Finally, the risk of bankruptcy endogenously imposes significant restrictions on the investor's admissible trading strategies.

5.3 Discounts for illiquidity

Imposing restrictions on an investor's ability to trade the risky asset clearly reduces his derived utility of wealth, $J(W, N, S, V, t) \leq J(W, V, t)$. To make an investor facing trading restrictions as well off as he would be in their absence, additional wealth needs to be given to the investor. The amount of additional wealth required can be viewed as the shadow cost of illiquidity. Given the logarithmic form of the derived utility functions, the investor's wealth would need to be increased by the simple scale factor $R = \exp(J(W, V, t) - J(W, N, S, V, t))$ to compensate for the effect of illiquidity. Because the investor's optimal portfolio strategy is independent of his wealth level, increasing his wealth by a factor of R results in the investor purchasing R times as many shares of the risky asset initially. Thus, increasing his wealth by a factor of R can be interpreted as reducing the price per share by a factor of $1/R$, implying a percentage price discount for illiquidity of $1 - 1/R$. Table 2 presents numerical estimates of the discounts for illiquidity for the same sets of parameters as in Table 1.

Table 2 shows that the discounts for lack of liquidity can be fairly substantial. The largest discounts occur when the endogenous constraint on borrowing is binding. This is intuitive because the differences between the constrained and unconstrained portfolios are largest in this case. For realistic combinations of μ and V , we would expect that the investor would prefer to hold a leveraged portfolio and that the endogenous borrowing constraint would be binding. For example, the average rate of return on the S&P 500 over the short-term interest rate has been roughly 10% during the past decade, while the standard deviation of returns has been approximately 15 to 20%. These values imply an unconstrained optimal portfolio weight of in neighborhood of three or four.

Table 2 shows that for a two-year period of complete illiquidity with $\alpha = .00$, and when the unconstrained portfolio weight is 2.00, the discount for illiquidity is 7.502% for $\sigma = .20$, and 17.658% for $\sigma = .40$. These discounts are somewhat lower but still on the same order of magnitude as those observed for illiquid letter stock. Recall that discounts for illiquid letter or Rule 144 stock can be measured directly by comparing the price at which the stock is privately placed, which is observable at the time of the private placement, with the simultaneous price of unrestricted registered shares. The

Table 2
Percentage liquidity discounts for the risky asset

Horizon	α	Volatility	Unconstrained	$\sigma = .00$	$\sigma = .20$	$\sigma = .40$	$\sigma = .60$
1	.00	.7071	.20	.105	.207	.509	1.132
1	.00	.4472	.50	.033	.265	.977	2.480
1	.00	.3536	.80	.000	.364	1.488	3.853
1	.00	.3162	1.00	.001	.455	1.858	4.795
1	.00	.2236	2.00	2.473	3.158	5.452	10.509
1	.00	.1414	5.00	14.798	16.150	20.721	30.428
1	.10	.7071	.20	.067	.115	.338	.877
1	.10	.4472	.50	.026	.191	.819	2.240
1	.10	.3536	.80	.000	.303	1.340	3.628
1	.10	.3162	1.00	.003	.426	1.744	4.576
1	.10	.2236	2.00	2.473	3.158	5.451	10.497
1	.10	.1414	5.00	14.798	16.150	20.721	30.428
2	.00	.7071	.20	.493	.838	2.152	5.984
2	.00	.4472	.50	.229	1.041	4.198	13.348
2	.00	.3536	.80	.056	1.355	6.327	20.210
2	.00	.3162	1.00	.023	1.662	7.845	24.583
2	.00	.2236	2.00	4.894	7.502	17.658	44.244
2	.00	.1414	5.00	27.409	32.027	48.463	79.911
2	.10	.7071	.20	.269	.517	1.407	4.904
2	.10	.4472	.50	.178	.677	3.412	12.275
2	.10	.3536	.80	.043	1.048	5.613	19.275
2	.10	.3162	1.00	.021	1.435	7.163	23.717
2	.10	.2236	2.00	4.894	7.502	17.608	43.915
2	.10	.1414	5.00	27.409	32.027	48.463	79.898

Horizon denotes the investor's horizon in years; α is the maximum number of shares that can be traded per year where the total number of shares that could be initially purchased is normalized to one; volatility is the current volatility of returns on the risky asset; unconstrained is the initial optimal portfolio weight for the risky asset in the absence of liquidity constraints; and σ is the volatility of volatility parameter. The expected return parameter μ is set equal to .10 and the market price of volatility risk λ equals zero.

difference between these two prices provides a direct measure of the market valuation of the two-year nonmarketability period imposed on the purchaser of the private placement. Silber (1992) shows that the average price discount on letter stock is about 35%.

Even when the unconstrained optimal portfolio weight is less than or equal to one, however, Table 2 shows that there can be a substantial discount for lack of liquidity. For example, when $T = 1$, $\alpha = .00$, and the unconstrained portfolio weight is .50, the discount for lack of liquidity is .265% for $\sigma = .20$, and .977% for $\sigma = .40$. These clearly represent economically significant discounts and are on the same order of magnitude as those observed by Amihud and Mendelson (1991), who compare the difference in prices between off-the-run Treasury notes and bonds with Treasury bills with the same maturity. For example, Amihud and Mendelson find that the average yield difference between Treasury notes and bills with the same maturity is about 42.8 basis points, which implies a pricing difference of roughly .21% for a six-month Treasury note. Similar results are obtained by Kamara (1994).

The discounts for lack of liquidity in Table 2 are all increasing functions of the volatility of volatility parameter σ . This is intuitive because variation in $V(t)$ is the primary motivation for trading in this economy. Specifically,

when the investor does not face trading constraints, Equation (13) shows that the variation in his optimal portfolio holding is driven entirely by the variation in $V(t)$. As σ increases, the variation in the unconstrained investor's optimal portfolio weight increases, resulting in a greater need to rebalance the portfolio substantially. Thus, as σ increases, the welfare loss to an investor facing liquidity constraints becomes larger.

The relationship between discounts for lack of liquidity and asset volatility is more complex. As the volatility of returns on the risky asset increases, the unconstrained portfolio weight decreases and the endogenous borrowing constraint becomes less binding. On the other hand, as the volatility of returns on the risky asset increases, there is more risk that the constrained investor's portfolio weight will deviate significantly from the unconstrained optimal value. The combination of these two effects explains why there is a nonmonotonic relationship between the discounts and the volatility of the risky asset. Finally, as the value of α increases, the discount for lack of liquidity decreases. Intuitively, this makes sense because increasing the value of α relaxes the trading restriction imposed on the investor.

6. Conclusion

This article solves the investor's intertemporal portfolio choice problem when the investor is limited to trading strategies of bounded variation. This more closely approximates the thin-trading illiquidity in actual markets where investors cannot trade unlimited amounts of securities instantaneously. The resulting optimal trading strategy endogenously imposes borrowing and short-selling constraints on the investor. Intuitively, this is because the inability to trade unlimited amounts exposes an investor with a leveraged position to the risk of bankruptcy, something that an unconstrained investor is able to avoid. To avoid this risk, however, the constrained investor may give up a large percentage of the welfare gains available from investing in the capital markets.

The resulting discounts for illiquidity can be substantial, even then the endogenous borrowing constraint is not binding. Although the discounts depend on the specific parameter values used, the numerical results suggest that the valuation effects of illiquidity constraints similar to those modeled here can be of the same order of magnitude as those observed empirically. Thus, these results suggest that observed discounts for illiquidity could be reconciled with market rationality. I note, however, that as these are only partial equilibrium results, they should be viewed as suggestive rather than definitive. One possible direction for future research might be to embed this analysis within a full general equilibrium framework with multiple agents and securities.

Appendix

Proof of Proposition 1. If $N(t) \geq 0$, $M(t) \geq 0$, and $N(t) + M(t) > 0$ for all t , $0 \leq t \leq T$, then since $S(t) > 0$ for $T < \infty$, $W(t) = N(t)S(t) + M(t) > 0$ for all t , $0 \leq t \leq T$. To prove the

only if part of the proposition, we need to show that there is a positive probability that $W(t) \leq 0$ for some t , $0 \leq t \leq T$ if $N(t) < 0$, $M(t) < 0$, or $N(t) + M(t) \leq 0$ for some t , $0 \leq t \leq T$. Note that $N(t) + M(t) \leq 0$ can only occur if $N(t) < 0$, $M(t) < 0$, or $N(t) = M(t) = 0$. If $N(t) = M(t) = 0$ for some t , then clearly $W(t) \leq 0$. Hence, it is sufficient to examine the two cases A $N(t) < 0$, and B $M(t) < 0$.

Case A. Assume that $N(t) < 0$ for some t , $0 \leq t \leq T$. Since $N(t)$ is continuous, the set of t where $N(t) < 0$ is open. Thus, $N(T) < 0$ implies that $N(t) < 0$ for some $t < T$. Thus, the assumption implies that $N(t) < 0$ for some t , $0 \leq t < T$. This case can now be partitioned into the two subcases (a) $N(t) < 0$ and $M(t) \leq 0$ for some t , $0 \leq t < T$, and (b) $N(t) < 0$ and $M(t) > 0$ for some t , $0 \leq t < T$.

Subcase (a). Since $S(t) > 0$ for $T < \infty$, $W(t) = N(t)S(t) + M(t) \leq 0$.

Subcase (b). Define

$$\epsilon = \min \left(T - t, -\frac{N_t}{4\alpha} \right) > 0.$$

Define C as the set of price paths on which the average value of $S(\tau)$, $t \leq \tau \leq t + \epsilon$ is less than or equal to $S(t + \epsilon)$. On this set,

$$\begin{aligned} M(t + \epsilon) &= M(t) - \int_t^{t+\epsilon} \gamma(\tau) S(\tau) d\tau \\ &\leq M(t) + \alpha \int_t^{t+\epsilon} S(\tau) d\tau \\ &\leq M(t) + \alpha \epsilon S(t + \epsilon) \\ &\leq M(t) - \frac{N(t)}{4} S(t + \epsilon). \end{aligned}$$

where the equality follows from the self-financing condition $dM(t) = -S(t)dN(t)$, and the first inequality follows from the liquidity constraints. Similarly,

$$N(t + \epsilon) \leq N(t) + \alpha \epsilon \leq \frac{3N(t)}{4},$$

where the first inequality follows from the upper bound on the rate at which shares can be traded. Together, these inequalities imply

$$\begin{aligned} W(t + \epsilon) &\leq \frac{3N(t)}{4} S(t + \epsilon) + M(t) - \frac{N(t)}{4} S(t + \epsilon) \\ &= \frac{N(t)}{2} S(t + \epsilon) + M(t). \end{aligned}$$

Let D be the set of price paths where $S(t + \epsilon) \geq -\frac{2M(t)}{N(t)}$. Then, on the set $C \cap D$, $W(t + \epsilon) \leq 0$. Because the transitional density of $S(t + \epsilon)$ conditional on $S(t)$ and $V(t)$ is continuous on $(0, \infty) \times (0, \infty)$, this set has strictly positive probability.

Combining the implications of the two subcases shows that $W(t) > 0$ almost surely for all t , $0 \leq t \leq T$ cannot be satisfied under the assumption of case A.

Case B. Assume that $M(t) < 0$ for some t , $0 \leq t \leq T$. Since $M(t)$ is continuous, the set of t where $M(t) < 0$ is open. Thus, $M(T) < 0$ implies that $M(t) < 0$ for some $t < T$. Thus, the assumption implies that $M(t) < 0$ for some t , $0 \leq t < T$. This case can now be partitioned into the two subcases (a) $M(t) < 0$ and $N(t) \leq 0$ for some t , $0 \leq t < T$, and (b) $M(t) < 0$ and $N(t) > 0$ for some t , $0 \leq t < T$.

Subcase (a). Since $S(t) > 0$ for $T < \infty$, $W(t) = N(t)S(t) + M(t) \leq 0$.

Subcase (b). Define

$$\epsilon = \min \left(T - t, -\frac{M_t}{4\alpha S_t} \right) > 0.$$

Define E as the set of price paths on which $S_\tau \leq 2S_t$, where $t \leq \tau \leq t + \epsilon$. On this set,

$$\begin{aligned} M(t + \epsilon) &= M(t) - \int_t^{t+\epsilon} \gamma(\tau) S(\tau) d\tau \\ &\leq M(t) + \alpha \int_t^{t+\epsilon} S(\tau) d\tau \\ &\leq M(t) + 2\alpha\epsilon S(t) \\ &\leq M(t) - \frac{M(t)}{2S(t)} S(t) \\ &= \frac{M(t)}{2}, \end{aligned}$$

Similarly,

$$N(t + \epsilon) \leq N(t) + \alpha\epsilon \leq N(t) - \frac{M(t)}{4S(t)}.$$

Together, these inequalities imply

$$W(t + \epsilon) \leq N(t)S(t + \epsilon) - \frac{M(t)}{4S(t)} S(t + \epsilon) + \frac{M(t)}{2}.$$

Let F be set of price paths where $0 < S(t + \epsilon) \leq \frac{2M(t)S(t)}{M(t) - 4N(t)S(t)}$. Then on the set $E \cap F$, $W(t + \epsilon) \leq 0$. Because the transitional density of $S(t + \epsilon)$ conditional on $S(t)$ and $V(t)$ is continuous on $(0, \infty) \times (0, \infty)$, this set has strictly positive probability.

Combining the implications of the two subcases shows that $W(t) > 0$ almost surely for all t , $0 \leq t < T$ cannot be satisfied under the assumption of case B. Combining cases A and B, the only if part of the proposition is proven.

References

- Amihud, Y., and H. Mendelson, 1986, "Asset Pricing and the Bid-Asked Spread," *Journal of Financial Economics*, 17, 223–249.
- Amihud, Y., and H. Mendelson, 1991, "Liquidity, Maturity, and the Yields on U.S. Treasury Securities," *Journal of Finance*, 46, 1411–1425.
- Averbukh, V., 1997, "Pricing American Options Using Monte Carlo Simulation," Ph.D. dissertation, Cornell University.
- Barraquand, J., and D. Martineau, 1995, "Numerical Valuation of High Dimensional Multivariate American Securities," *Journal of Financial and Quantitative Analysis*, 30, 383–405.
- Benes, V., E. Shepp, and H. Witsenhausen, 1980, "Some Solvable Stochastic Control Problems," *Stochastics*, 4, 39–83.
- Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.
- Bossaerts, P., 1989, "Simulation Estimators of Optimal Early Exercise," working paper, Carnegie-Mellon University.

- Brandt, M., and P. Santa-Clara, 2000, "Dynamic Portfolio Choice with Simulation Methods," working paper, University of California, Los Angeles.
- Boudoukh, J., and R. F. Whitelaw, 1991, "The Benchmark Effect in the Japanese Government Bond Market," *Journal of Fixed Income*, September, 52–59.
- Boudoukh, J., and R. F. Whitelaw, 1993, "Liquidity as a Choice Variable: A Lesson from the Japanese Government Bond Market," *Review of Financial Studies*, 6, 265–292.
- Brito, N. O., 1977, "Marketability Restrictions and the Valuation of Capital Assets Under Uncertainty," *Journal of Finance*, 32, 1109–1123.
- Broadie, M., and P. Glasserman, 1997a, "Pricing American-Style Securities Using Simulation," *Journal of Economic Dynamics and Control*, 21, 1323–1352.
- Broadie, M., and P. Glasserman, 1997b, "Monte Carlo Methods for Pricing High-Dimensional American Options: An Overview," working paper, Columbia University.
- Broadie, M., and P. Glasserman, 1997c, "A Stochastic Mesh Method for Pricing High-Dimensional American Options," working paper, Columbia University.
- Broadie, M., P. Glasserman, and G. Jain, 1997, "Enhanced Monte Carlo Estimates for American Option Prices," *Journal of Derivatives*, 5, 25–44.
- Broadie, M., J. Detemple, E. Ghysels, and O. Torres, 1998, "American Options with Stochastic Dividends and Volatility: A Nonparametric Investigation," working paper, Columbia University.
- Carr, P., 1998, "Randomization and the American Put," *Review of Financial Studies*, 11, 597–626.
- Carriere, J., 1996, "Valuation of Early-Exercise Price of Options Using Simulations and Nonparametric Regression," *Insurance: Mathematics and Economics*, 19, 19–30.
- Constantinides, G. M., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy*, 94, 842–862.
- Constantinides, G. M., and R. Mehra, 1998, "Junior Can't Borrow, an Alternative Explanation for the Equity Premium Puzzle," working paper, Columbia University.
- Cornell, B., 1993, "Adverse Selection, Squeezes, and the Bid-Ask Spread on Treasury Securities," *Journal of Fixed Income*, June, 39–47.
- Cornell, B., and A. C. Shapiro, 1990, "The Mispricing of U.S. Treasury Bonds: A Case Study," *Review of Financial Studies*, 2, 297–310.
- Cox, J., and C.-F. Huang, 1989, "Optimal Consumption and Investment Policies when Asset Prices Follow a Diffusion Process," *Journal of Economic Theory*, 49, 33–83.
- Cox, J., J. Ingersoll, and S. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–408.
- Daves, P. R., and M. C. Ehrhardt, 1993, "Liquidity, Reconstitution, and the Value of U.S. Treasury Strips," *Journal of Finance*, 48, 315–329.
- Davis, M., and A. Norman, 1990, "Portfolio Selection with Transaction Costs," *Mathematics of Operations Research*, 15, 676–713.
- Dumas, B., and E. Luciano, 1991, "An Exact Solution to a Dynamic Portfolio Choice Problem under Transaction Costs," *Journal of Finance*, 46, 577–595.
- Dybvig, P., and C.-F. Huang, 1988, "Nonnegative Wealth, Absence of Arbitrage, and Feasible Consumption Plans," *Review of Financial Studies*, 1, 377–401.
- Easley, D., and M. O'Hara, 1987, "Price, Trade Size, and Information in Securities Markets," *Journal of Financial Economics*, 19, 69–90.

- Garcia, D., 1999, "A Monte Carlo Method for Pricing American Options," working paper, University of California, Berkeley.
- Glosten, L., 1987, "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42, 1293-1307.
- Glosten, L., and L. Harris, 1988, "Estimating the Components of the Bid/Ask Spread," *Journal of Financial Economics*, 21, 123-142.
- Glosten, L., and P. Milgrom, 1985, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14, 71-100.
- Grinblatt, M., and F. L. Longstaff, 2000, "Financial Innovation and the Role of Derivative Securities: An Empirical Analysis of the U.S. Treasury's STRIPS Program," *Journal of Finance*, 55, 1415-1436.
- Grossman, S., and G. Laroque, 1990, "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58, 25-52.
- Harrison, J. M., and D. M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381-408.
- Harrison, J. M. and S. R. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 261-271.
- Heston, S., 1993, "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-344.
- Huang, M., 1998, "Liquidity Shocks and Equilibrium Liquidity Premia," working paper, Stanford University.
- Hull, J., and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatility," *Journal of Finance*, 42, 281-300.
- Ibanez, A., and F. Zapatero, 1998, "Monte Carlo Valuation of American Options through Computation of the Optimal Exercise Frontier," working paper, University of Southern California.
- Jouini, E., and H. Kallal, 1998, "Efficient Trading Strategies in the Presence of Market Frictions," working paper, New York University.
- Kamara, A., 1994, "Liquidity, Taxes, and Short-Term Treasury Yields," *Journal of Financial and Quantitative Analysis*, 29, 403-417.
- Karatzas, I., and S. Shreve, 1988, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.
- Karatzas, I., J. P. Lehoczky, and S. Shreve, 1987, "Optimal Portfolio and Consumption Decisions for a Small Investor on a Finite Horizon," *SIAM Journal of Control Optimization*, 25, 1557-1586.
- Keane, M., and K. Wolpin, 1994, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648-672.
- Lippman, S., and J. McCall, 1986, "An Operational Measure of Liquidity," *American Economic Review*, 76, 43-55.
- Liu, J., 1999, "Portfolio Selection in Stochastic Environments," working paper, University of California, Los Angeles.
- Longstaff, F., 1995a, "How Much Can Marketability Affect Security Values?" *Journal of Finance*, 50, 1767-1774.
- Longstaff, F., 1995b, "Placing No-Arbitrage Bounds on the Value of Nonmarketable and Thinly-Traded Securities," *Advances in Futures and Options Research*, 8, 203-228.
- Longstaff, F., 2000, "The Term Structure of Very Short-Term Rates: New Evidence for the Expectations Hypothesis," *Journal of Financial Economics*, 58, 397-415.
- Longstaff, F., and E. Schwartz, 2001, "Valuing American Options by Simulation: A Simple Least Squares Approach," *Review of Financial Studies*, 14, 113-147.

- Longstaff, F., P. Santa-Clara, and E. Schwartz, 2000, "Throwing Away a Billion Dollars: The Cost of Suboptimal Exercise Strategies in the Swaptions Market," working paper, University of California, Los Angeles.
- Mayers, D., 1972, "Nonmarketable Assets and Capital Market Equilibrium under Uncertainty," in M. C. Jensen (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York.
- Mayers, D., 1973, "Nonmarketable Assets and the Determination of Capital Asset Prices in the Absence of a Riskless Asset," *Journal of Business*, 46, 258–267.
- Mayers, D., 1976, "Nonmarketable Assets, Market Segmentation and the Level of Asset Prices," *Journal of Financial and Quantitative Analysis*, 11, 1–12.
- Merton, R., 1969, "Lifetime Portfolio Selection under Uncertainty: The Continuous Time Model," *Review of Economics and Statistics*, 51, 247–257.
- Merton, R., 1971, "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, 3, 373–413.
- Merton, R., 1973a, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–888.
- Merton, R., 1973b, "The Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.
- Merton, R., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.
- Pratt, S. P., 1989, *Valuing a Business*, Irwin, Homewood, Ill.
- Royden, H. L., 1968, *Real Analysis*, Macmillan, New York.
- Shreve, S., 1981, "Reflected Brownian Motion in the 'Bang-Bang' Control of Brownian Drift," *SIAM Journal of Control and Optimization*, 19, 469–478.
- Silber, W. L., 1992, "Discounts on Restricted Stock: The Impact of Illiquidity on Stock Prices," *Financial Analysts Journal*, July/August, 60–64.
- Stapleton, R. C., and M. G. Subrahmanyam, 1979, "Marketability of Assets and the Price of Risk," *Journal of Financial and Quantitative Analysis*, 14, 1–10.
- Stein, E., and J. Stein, 1991, "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4, 727–752.
- Stoll, H., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance*, 44, 115–134.
- Raymar, S., and M. Zwecher, 1997, "A Monte Carlo Valuation of American Call Options on the Maximum of Several Stocks," *Journal of Derivatives*, 5, 7–23.
- Tilley, J. A., 1993, "Valuing American Options in a Path Simulation Model," *Transactions of the Society of Actuaries*, 45, 83–104.
- Trout, R. R., 1977, "Estimation of the Discount Associated with the Transfer of Restricted Securities," *Taxes*, June, 381–385.
- Tsitsiklis, J., and B. Van Roy, 1999, "Optimal Stopping of Markov Processes: Hilbert Space Theory, Approximation Algorithms, and an Application to Pricing High-Dimensional Financial Derivatives," *IEEE Transactions on Automatic Control*, 44, 1840–1851.
- Wruck, K. H., 1989, "Equity Ownership Concentration and Firm Value: Evidence from Private Equity Financings," *Journal of Financial Economics*, 23, 3–28.