Losing Money on Arbitrage: Optimal Dynamic Portfolio Choice in Markets with Arbitrage Opportunities

Jun Liu
UCLA

Francis A. Longstaff
UCLA and NBER

We derive the optimal investment policy of a risk-averse investor in a market where there is a textbook arbitrage opportunity, but where liabilities must be secured by collateral. We find that it is often optimal to underinvest in the arbitrage by taking a smaller position than collateral constraints allow. Even when the optimal policy is followed, the arbitrage portfolio typically experiences losses before the final conver- gence date. In fact, its initial performance may be indistinguishable from that of a conventional portfolio with a poor track record. These results have important impli- cations for the role of arbitrageurs in financial markets.

So there's an arbitrage. So what? This desk has lost a lot of money on arbitrages. Arbitrages aren't particularly great trades.
       — Treasury bond trader at a major Wall Street investment bank

One of the foundational principles of financial economics is that arbi- trages cannot exist in securities markets. The reasoning is that if they did, investors could attain infinite wealth by taking unlimited positions in them. Economic theory implies that an arbitrage is an investment oppor- tunity that is literally too good to be true.

In actual financial markets, however, investors may not be able to attain infinite wealth even if arbitrage opportunities exist. The reason for this is inherent in the corporate finance of the investment industry in

which secured lending is the dominant type of debt contract. Recall that the textbook strategy for exploiting an arbitrage requires taking offsetting long and short positions and holding them until convergence. An investor who takes a short position, however, generates a liability which must be secured by collateral. This collateral requirement drives an important wedge between textbook arbitrage strategies and strategies that are actually feasible. For example, consider an investor who implements an arbitrage strategy. If the arbitrage were then to widen rather than narrow, the investor would experience mark-to-market losses on the position. If the losses were severe enough, the investor might not have sufficient collateral to meet margin calls and be forced to liquidate some or all of the position at a loss before it had converged to its theoretical no-arbitrage value. Because of this, even the simplest strategies to exploit arbitrages could actually result in losses, a lesson painfully learned recently by many highly leveraged hedge funds. This fundamental risk in taking arbitrage positions is discussed in recent articles by Shleifer and Vishny (1997) and Loewenstein and Willard (2000b).

If arbitrages are actually risky investments from the perspective of an investor or hedge fund manager facing collateral constraints, then a number of interesting economic issues arise. For example, what is the optimal investment strategy when markets have arbitrage opportunities? Similarly, how do arbitrages compare with other investments in terms of their risk and return characteristics? To address these issues, this article studies a continuous-time model in which there are explicit arbitrage opportunities. To capture the spirit of standard textbook examples, we model the arbitrage opportunity as a security whose price converges to zero at some specified future time. In this setting, an investor could make arbitrage profits with certainty if he could hold the position until convergence at maturity. In the short run, however, the arbitrage may widen and force the investor to liquidate positions at a loss. Thus there is no guarantee that the investor can hold the position until it converges.

The results are surprising. We find that it is often optimal for the investor to underinvest in the arbitrage opportunity. Specifically, the investor often will not take the largest arbitrage position allowed by the collateral constraint. This contrasts with the popular view that an investor should take the largest position possible in any arbitrage opportunity. We also show that an investor may prefer a strategy that may underperform the riskless asset over a strategy that dominates the riskless asset.

Even when the investor follows the optimal investment strategy, the returns from investing in the arbitrage may not be as attractive as those from conventional assets. For example, we demonstrate that the investor can experience substantial losses on his portfolio prior to the convergence date of the arbitrage. In some cases, these losses can be more than 75% of

the value of the portfolio. Because of this, the arbitrage strategy may frequently look like an underperforming conventional portfolio. In some situations, it is even possible for the investor to have a loss at the convergence date of the arbitrage. In this situation, the investor ends up worse off than if he had invested only in the riskless asset. We show that the return distributions from following the optimal strategy are highly skewed toward negative values during the early stages of the arbitrage, and that the arbitrage portfolio is usually worth less than its initial value at some point during the life of the arbitrage. Finally, we find that the Sharpe ratio from investing in the arbitrage generally only averages about two in our numerical examples.

Our results demonstrate that experiencing large losses during the early stages of an arbitrage strategy is almost a hallmark of the optimal strategy. From this perspective, the real problem during the hedge fund crisis of 1998 may not have been that arbitrage funds used too much leverage or that they were speculating, but rather that many market participants had unrealistic expectations about how arbitrage strategies should perform over time. These results also have important implications for the role of arbitrageurs in financial markets. Standard economic theory takes as given the notion that arbitrages cannot exist in the markets, since if they did, arbitrageurs would immediately buy and sell the cheap and rich securities until the prices came back into line. Our analysis calls this simplistic view into question since it is not clear that an investor would actually choose to take a position in a specific arbitrage. If investors found it optimal to take only a very limited position in an arbitrage opportunity, or to avoid taking any position at all, then there is no reason why the arbitrage could not persist or even become wider. Given that collateral requirements are pervasive in all financial markets, these results suggest that many theoretical valuation arguments based on the absence of arbitrage principles may need to be reexamined.

Our research complements an important recent literature focusing on whether arbitrage opportunities can exist in equilibrium. Key examples of this literature include Basak and Croitoru (2000) and Loewenstein and Willard (2000a–c). These articles demonstrate that arbitrage or mispricing may be sustained in general equilibrium when financial markets have frictions or imperfections. In Basak and Croitoru, however, the arbitrageur always takes the maximum possible position allowed by the financial market constraints. Furthermore, Loewenstein and Willard do not identify the optimal portfolio strategy for an arbitrageur in their model. Our article contributes to the literature by demonstrating that when the real-world feature of collateral constraints is introduced, arbitrages become risky and agents may actually choose to take smaller positions than allowed by constraints. Furthermore, we show that returns on the arbitrage portfolio may be observationally equivalent to those

resulting from the "meltdown" of a conventional portfolio. Other important related work includes Brennan and Schwartz (1988, 1990), De Long et al. (1990), Duffie (1990), Dumas (1992), Tuckman and Vila (1992), Delgado and Dumas (1994), Dow and Gorton (1994), Yadev and Pope (1994), Chen (1995), Detemple and Murthy (1997), Zigrand (1997), Willard and Dybvig (1999), and Xiong (2001). Our article also extends the literature on margin constraints in financial markets. Important examples of this literature include Heath and Jarrow (1987), Hindy (1995), and Cuoco and Liu (2000). Finally, our results corroborate Shleifer and Vishny (1997), who show that arbitrage can be risky when there are margin constraints. Unlike Shleifer and Vishny, however, we explicitly study the optimal portfolio strategy for a risk-averse investor in a market with arbitrage opportunities.

The remainder of this article is organized as follows. Section 1 presents the dynamic portfolio choice problem in markets with arbitrage opportunities. Section 2 discusses the optimal portfolio strategy. Section 3 examines the return distributions resulting from following the optimal strategy. Section 4 summarizes the results and makes concluding remarks.

1. The Dynamic Portfolio Choice Problem

In this section we describe the continuous-time framework, explain how we model arbitrage opportunities, and then solve for the optimal portfolio strategy. To make the intuition as clear as possible, we focus on the simplest version of the model.

We model a simple two-investment financial market in which trading takes place continuously in time. The first investment is a riskless asset with value R_t, which earns a constant rate of interest r. The dynamics of the riskless asset are given by

$$dR = rRdt, \tag{1}$$

where $R_0 = 1$. Solving this equation for the value of the riskless asset gives $R_t = e^{rt}$. The second investment is an arbitrage opportunity with value A_t, where $0 \le t \le T$. Intuitively A_t can be thought of as the value of a textbook arbitrage portfolio that converges to zero at time T. To illustrate this type of portfolio, consider the case where there are two bonds with identical cash flows in all states of the world, but where the two bonds have different market prices. An example of this might be two Treasury STRIPS with identical maturity dates but different prices [see Daves and Ehrhardt (1993) and Grinblatt and Longstaff (2000)]. In this case, A_t can be thought of as the value of a portfolio that is long $100 notional amount of the cheaper bond and short $100 notional amount of the richer bond.

Note that the value of this portfolio must converge to zero at the maturity date of the bonds.[1]

To capture the intuitive notion of a textbook arbitrage as a portfolio with a value converging to zero at some future point in time T, we assume that the dynamics of A follow the Brownian bridge process

$$dA = \frac{-\alpha A}{T-t} dt + \sigma \, dZ, \tag{2}$$

where α and σ are positive constants, $0 \leq t \leq T$, and Z is a standard Brownian motion.[2] As $t \to T$, the drift of this process approaches $+\infty$ when $A_t < 0$, and $-\infty$ when $A_t > 0$. Thus, as $t \to T$, the mean reversion of the process toward zero becomes stronger and stronger, forcing A_T to converge to zero with probability one.[3] The parameter α governs the speed at which the arbitrage opportunity converges to zero. The parameter σ represents the volatility of the arbitrage and determines the distribution of possible arbitrage opportunities. Solving this stochastic differential equation results in the following expression for A_s, where $0 \leq t \leq s \leq T$,

$$A_s = \left(\frac{T-s}{T-t}\right)^\alpha A_t + \sigma \int_t^s \left(\frac{T-s}{T-\tau}\right)^\alpha dZ_\tau. \tag{3}$$

It is easily seen that A_s is normally distributed for $s < T$. We denote the expected value of A_s conditional on the value of A_t by M_s, where

$$M_s = \left(\frac{T-s}{T-t}\right)^\alpha A_t. \tag{4}$$

Similarly, the conditional variance of A_s, which we denote V_s^2, is given by

$$V_s^2 = \frac{\sigma^2(T-t)}{1-2\alpha} \left[\left(\frac{T-s}{T-t}\right)^{2\alpha} - \left(\frac{T-s}{T-t}\right) \right], \tag{5}$$

for $\alpha \neq 1/2$, and by

$$V_s^2 = \sigma^2(T-s)\ln\left(\frac{T-t}{T-s}\right), \tag{6}$$

for $\alpha = 1/2$. As $s \to T$, both M_s and V_s^2 converge to zero.

---

[1] Other examples of arbitrages that must converge to zero at a future time include put/call parity violations [see Longstaff (1995)], differences between market and cost-of-carry-model stock index futures prices [see Brennan and Schwartz (1988, 1990), MacKinlay and Ramaswamy (1988), and Duffie (1990)], and differences between the prices of otherwise identical on-the-run and off-the-run Treasury securities [see Amihud and Mendelson (1991), Kamara (1994), and Longstaff (2002)].

[2] Strictly speaking, the process A is a Brownian bridge only when $\alpha = 1$. Thus we are using the term Brownian bridge in a more general sense. The Brownian bridge process has been applied to security prices by Ball and Torous (1983), Brennan and Schwartz (1988, 1990), Duffie (1990), and Cheng (1991). Loewenstein and Willard (2000b) study the viability of a Brownian bridge as a return process.

[3] Rather than converging to zero at time T, the arbitrage process could be generalized to converge to some other fixed value by a simple modification of the drift term.

The Brownian bridge process allows A_t to take on both positive and negative values. When A_t is positive, the investor receives a positive cash flow of A_t by investing in a negative number of units of the arbitrage, and vice versa. By taking a position in the arbitrage and receiving a positive cash flow at time t, the investor simultaneously creates a liability, since the investor would need to pay the same amount to immediately unwind the arbitrage position. It is easily shown that $|A(t)|$ can exceed any fixed value with strictly positive probability. Thus this specification implies that there is always a risk that the arbitrage can widen further before its final convergence date. This risk plays both a major role in this model as well as in actual financial markets. As has been shown by Back and Pliska (1990), Cheng (1991), Basak and Croitoru (2000), Loewenstein and Willard (2000b), and others, the properties of the Brownian bridge process as $t \to T$ implies that the Brownian bridge describing the dynamics of A_t in this market does not admit the existence of an equivalent martingale measure.

Let N_t and P_t denote the number of units of the arbitrage and the riskless asset held by the investor. The investor's total wealth at time t is given by

$$W_t = N_t A_t + P_t R_t. \tag{7}$$

Following standard portfolio choice theory, we assume that the investor follows a self-financing strategy.[4] Applying the self-financing condition results in the following dynamics for W_t,

$$
\begin{aligned}
dW &= N dA + rPR \, dt \\
&= N dA + r(W - NA) dt \\
&= \left(rW - \left(r + \frac{\alpha}{T - t} \right) NA \right) dt + \sigma N \, dZ.
\end{aligned} \tag{8}
$$

This equation, along with the dynamics of A_t in Equation (2), implies that W_t and A_t follow a joint Markov process. Thus the state of the economy is completely specified by the current values of the state variables W_t and A_t. From Equation (8), W_T can also be expressed as

$$W_T = W_t \exp \left(\int_t^T \left(r - \left(r + \frac{\alpha}{T - s} \right) \frac{NA}{W} - \frac{\sigma^2}{2} \frac{N^2}{W^2} \right) ds + \sigma \int_t^T \frac{N}{W} \, dZ \right). \tag{9}$$

Harrison and Kreps (1979) and Harrison and Pliska (1981) show that restrictions on trading strategies are necessary to rule out unrealistic arbitrages arising from doubling strategies. In actual financial markets, however, even stronger restrictions on trading strategies are imposed

---

[4] This standard assumption rules out the possibility of later capital injections, credit support, or bailouts for the investor. This involves little loss of generality, since initial wealth can be defined to include the value of these contingent cash flows.

through the standard requirement that investors provide collateral as security for their short positions. Specifically, whenever an investor generates a liability by either shorting an asset or borrowing funds (which is the same as shorting the riskless asset), financial institutions generally require collateral exceeding the amount of the liability as protection against mark-to-market losses.[5] As described in the *Guidelines for Collateral Practitioners* (1999) issued by the ISDA, the mechanics of a standard collateral agreement require that collateral (typically in the form of cash or liquid securities such as Treasury or agency bonds) be provided by the counterparty with the liability. As the market value of the liability fluctuates, counterparties either provide additional collateral or receive collateral back.[6] The ISDA *Collateral Survey* (2000) indicates that most collateral agreements require that security positions be revalued and collateral transferred on a daily basis; the survey reports that 74% of collateral calls are made at a daily frequency, and a further 17% are made at a weekly frequency. The amount by which the required amount of collateral exceeds the liability is referred to as the "haircut," or the investor's equity in the position. The size of the required haircut generally varies with the type of collateral. For example, the haircut for shorting Treasury bonds might be $1 to $2 per $100 notional amount, while the haircut could be $10 to $20 per $100 notional amount for corporate bonds. Under standard collateral agreements, if the trade goes against the investor and generates losses, or if the value of the collateral itself decreases, investors may be forced to either provide additional collateral in response to a margin call or have their short positions liquidated.[7]

Although beyond the scope of this study, one interesting question that emerges from our analysis is why collateralization or secured lending is the dominant form of debt contract in the financial markets. Certainly the possibility of informational asymmetries and moral hazard problems between lenders and arbitrageurs may play some role in explaining the

---

[5] A recent survey of collateral practices among major over-the-counter (OTC) market participants conducted by the International Swaps and Derivatives Association (ISDA) reported that large-scale established financial institutions had collateral agreements with virtually all significant counterparties. Furthermore, many of the firms in the survey reported that virtually all OTC derivative trading was conducted on a collateralized basis. This indicates that collateral requirements are a pervasive feature of the financial markets and are thus likely to be important in equilibrium.

[6] Johannes and Sundaresan (2001) provide an excellent discussion of the economic implications of collateralization in the swaps market.

[7] There are other possible ways to deal with counterparty credit risk, although these are not as widely used as collateralization. For example, from the *Guidelines for Collateral Practitioners* (1999), "As discussed above, other credit enhancement techniques are available to market practitioners such as netting, third-party guarantees, establishing a specialized derivatives subsidiary (a Derivatives Products Company (DPC) or Special Purpose Vehicle (SPV)) and cash-settlement provisions. ... Nevertheless, collateral does appear to be the most widely used credit enhancement tool in the industry today. It is not as capital intensive as a DPC or SPV, does not require cumbersome liquidation of positions as do early terminations provisions or involve third parties as do guarantees." We are grateful to the referee for raising this point.

prevalence of collateralization.[8] The ISDA *Collateral Survey* (2000) reports that the primary reasons offered by market participants for collateralization are credit risk management and the fact that banks incur lower regulatory capital charges for collateralized transactions.

When an investor invests in N_t units of the arbitrage, the investor receives an immediate cash flow of $-N_t A_t$ and has a current mark-to-market liability of the same amount. For example, imagine that two bonds with identical future cash flows have prices of 100 and 101. Define the arbitrage as a long position in the first and a short position in the second. Thus $A_t = 100 - 101 = -1$, and taking a position in 10 units of the arbitrage generates an immediate cash flow of $-N_t A_t = 10$ and a liability of $N_t A_t = 10$. To capture the economics of collateralization in a simple, yet realistic way, we assume that the investor is required to hold liquid securities in the amount of the liability plus a margin of λ per unit (per notional amount) of the arbitrage held, where λ is a nonnegative constant. Thus we require that

$$P_t R_t \geq |N_t A_t| + \lambda |N_t|. \tag{10}$$

Assuming $N_t A_t \leq 0$ (which we show later to be satisfied by the optimal strategy), the collateral constraint can be expressed as a simple wealth constraint,

$$W_t \geq \lambda |N_t|. \tag{11}$$

Since λ is nonnegative, satisfying this constraint generally satisfies the less-restrictive condition that $W_t > 0$ for all t, $0 \leq t \leq T$.

This form of the collateral requirement closely follows actual OTC derivatives market practice. For example, the ISDA *Collateral Survey* states that, "The amount by which the value assigned to the collateral is less than full face value is termed the 'haircut', usually expressed as a percentage of face value." Since a unit of the arbitrage should be interpreted as being relative to a fixed face or notional amount in our framework, λ is directly a percentage of the face value or notional amount. Similarly margin requirements for exchange-traded financial futures at major futures exchanges such as the Chicago Board of Trade and the Chicago Mercantile Exchange are defined in terms of a fixed amount per contract or per notional amount, which is consistent with our specification. Repo lending for Treasury, agency, and mortgage-backed securities is typically done by requiring a haircut equal to a fixed fraction of the face amount of the bond. Thus our model of the haircut as a fixed amount or fraction per unit or notional amount of the arbitrage is also consistent

---

[8] For discussions of the role that secured lending plays in mitigating informational asymmetries and moral hazard problems, see Johnson and Stulz (1985), James (1988), Rajan and Winton (1995), Demarzo and Duffie (1999), Habib and Johnsen (1999), and Morellec (2001).

with market practice in the primary fixed-income markets.[9] Finally, we observe that our basic results are robust to the specific form of the collateral requirements; similar results are obtained when the required haircut is assumed to be proportional to $|A_t|$.[10]

The collateral requirement ensures that the counterparty taking the other side of the arbitrage position has collateral at least equal to the amount owed by the investor. Thus, if $\lambda = 1$, the investor in the above example who invested in $N_t = 10$ units of the arbitrage would have a long position of 1,000 in the first bond and a short position of 1,010 in the second bond, implying $N_t A_t = -10$. The investor would need to have collateral of $P_t R_t = 20$ to cover the net liability of $|N_t A_t| = 10$ generated by the arbitrage and to post the additional $\lambda|N_t| = 10$ collateral required. In this example, the investor has a liability of 1,010 and needs total collateral of 1,020 consisting of a long bond position with value 1,000 and 20 of the riskless asset.

It is important to observe that requiring collateral is fundamentally different from short-selling restrictions. In fact, as we show later, the optimal portfolio strategy has the property that the portfolio weight for the arbitrage can take on any negative value. Collateral constraints are also fundamentally different from transaction costs. Intuitively this is because investors receive all of the interest, dividends, and appreciation on the securities held as collateral in margin accounts. Thus the investor incurs no direct economic costs or losses from holding securities in margin accounts. Finally, collateral requirements differ from position limits such as those in futures markets.

The investor is endowed with strictly positive initial total wealth W_0 and has a finite investment horizon T corresponding to the date at which the arbitrage converges to zero. 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 dynamically chooses a portfolio N_t to maximize an expected utility function defined over the logarithm of his terminal wealth W_T,

$$E_t[\ln W_T]. \tag{12}$$

We use this simple preference structure to focus more directly on the intuition of how the arbitrage opportunity affects the portfolio problem.

---

[9] By requiring that collateral be held against the net value of the arbitrage, we are making the conservative assumption that netting across both legs of the arbitrage is possible. This is feasible when both legs of a position are executed with the same counterparty. Gross margining could easily be modeled within this framework by setting λ to a larger value. Margin requirements where the total required collateral is proportional to the value of an asset are considered in Cuoco and Liu (2000).

[10] Specifically, we solve for the optimal portfolio strategy in the case where the haircut per unit of the arbitrage $\lambda(|A_t|)$ is an increasing function of $|A_t|$, where $\lambda(0) > 0$. An underinvestment result similar to that presented in the next section holds for this more general (but less common) form of the collateral restriction (proof available upon request).

Define the derived utility of wealth function $J(W, A, t)$ by the following expression

$$J(W, A, t) = \max_{N} E_t[\ln W_T],\tag{13}$$

subject to the budget constraint in Equation (9) and where N is a member of the set of admissible strategies satisfying the collateral constraint. Because the problem and collateral constraint are homogeneous in W_t, we demonstrate in the following proposition that N_t must be of the form $F_t W_t$, where F is a function of A and t only. Substituting this into Equation (9) implies

$$
\begin{aligned}
J(W, A, t) &= \ln W_t + \max_{F} E_t\left[\int_t^T r - \left(r + \frac{\alpha}{T-s}\right)FA - \frac{\sigma^2}{2}F^2 ds\right] \\
&= \ln W_t + r(T-t) - \max_{F} E_t\left[\int_t^T \left(r + \frac{\alpha}{T-s}\right)FA + \frac{\sigma^2}{2}F^2 ds\right].
\end{aligned}
\tag{14}
$$

Because of the quadratic form of the integrand in F and the fact that the dynamics of A are independent of F, the optimal portfolio strategy can be determined in closed form by a state-by-state minimization.

Proposition 1. *The Optimal Arbitrage Position. The optimal portfolio strategy for the investor is*

$$
N_t = \begin{cases}
\dfrac{1}{\lambda} W_t, & \text{if } A_t < -\dfrac{1}{\lambda}\dfrac{\sigma^2}{\left(r + \frac{\alpha}{T-t}\right)}, \\[2ex]
-\dfrac{r + \frac{\alpha}{T-t}}{\sigma^2} A_t W_t, & \text{if } |A_t| \leq \dfrac{1}{\lambda}\dfrac{\sigma^2}{\left(r + \frac{\alpha}{T-t}\right)}, \\[2ex]
-\dfrac{1}{\lambda} W_t, & \text{if } A_t > \dfrac{1}{\lambda}\dfrac{\sigma^2}{\left(r + \frac{\alpha}{T-t}\right)}.
\end{cases}
\tag{15}
$$

Proof of Proposition 1. See the appendix.

This optimal portfolio strategy has many interesting features that are discussed in the next section. By substituting the optimal portfolio strategy into Equation (14) and evaluating the expectations, it can be shown that if $|A_t| < \infty$, $W_t < \infty$, and $\lambda > 0$, then $J(W, A, t) < \infty$ for all t, $0 \leq t \leq T$ (proof available upon request). This result that the derived utility of wealth is finite depends critically on the condition that $\lambda > 0$. If $\lambda = 0$, then it is easily shown that the strategy

$$N_t = -\frac{r + \frac{\alpha}{T-t}}{\sigma^2} A_t W_t\tag{16}$$

implies that $E[\ln W_T] = \infty$. Thus the collateral constraint fundamentally changes the economics of the arbitrage opportunity in this financial market.

2. The Optimal Portfolio Strategy

In this section we examine in more detail the optimal portfolio strategy. Several key properties of the optimal strategy are immediately apparent from Proposition 1. First, the optimal strategy always requires taking a position in the arbitrage opposite in sign from the value of the arbitrage A_t. Thus taking a position in the arbitrage generates an immediate cash inflow to the investor. Since there are collateral requirements, however, the investor must keep liquid assets at least equal in amount to the cash inflow plus $\lambda |N_t|$. Thus the investor is constrained in the way this cash can be used.

Since the investor faces collateral constraints, it is perhaps not surprising that the investor only takes a finite position in the arbitrage. For example, if $A_t < 0$, the maximum value of N_t that the collateral restriction allows is $\frac{1}{\lambda} W_t$. Thus if $W_t = 100$ and $\lambda = 1$, the maximum number of units of the arbitrage the investor can hold is $N_t = 100$, independent of how large the arbitrage opportunity A_t becomes. Note, however, that this does not limit the leverage that the investor can utilize in his portfolio. In particular, since the portfolio weight for the arbitrage is $N_t A_t / W_t$, the maximum portfolio weight for the arbitrage in this example is $-100 A_t / 100 = -A_t$, which is unbounded. Thus, while the number of units of the arbitrage that can be held is bounded for a given W_t, the portfolio weight invested in the arbitrage is not.

What is surprising, however, is that the investor often finds it optimal to take a smaller position in the arbitrage opportunity than the collateral restrictions allow. For example, when

$$-\frac{1}{\lambda}\frac{\sigma^2}{\left(r + \frac{\alpha}{T-t}\right)} < A_t < \frac{1}{\lambda}\frac{\sigma^2}{\left(r + \frac{\alpha}{T-t}\right)}, \tag{17}$$

the optimal N_t is less in absolute value than the maximum number of units of the arbitrage that could be held while satisfying the collateral constraint. In fact, when $|A_t|$ is close to zero, the optimal N_t may only be a small fraction of the maximum allowable number of units of the arbitrage. To illustrate this, we simulate how often an investor following the optimal strategy will reach the collateral constraint. In particular, we simulate paths of A_t and report the percentage of paths where the bounds shown in Equation (17) are exceeded for different values of $t < T$.

In doing this, our approach is to report the results for a variety of realistic parameter values. Admittedly there is little empirical evidence in the literature about the dynamic properties of arbitrages. Despite this, there are a few recent empirical papers documenting the properties of apparent arbitrages or deviations from the law of one price. For example, Amihud and Mendelson (1991) and Kamara (1994) study the differences between the prices of Treasury bills and Treasury bonds that have paid

their next to last coupon and have essentially become Treasury bills. Longstaff (2004) studies the differences between the prices of Treasury zero coupon bonds or STRIPS and the prices of Treasury-guaranteed zero coupon bonds issued by government agencies. Cornell and Shapiro (1989) and Boudoukh and Whitelaw (1991) document differences in the pricing of Treasury and government bonds with identical cash flows. The type of pricing anomalies studied in these articles, where two bonds with identical cash flows in all states of the world trade at different prices, closely matches the notion of a textbook arbitrage in this article. In particular, the arbitrage would consist of taking a long $100 notional position in the cheaper of the two bonds and taking an offsetting short $100 notional position in the richer of the two bonds.

For simplicity, we assume a one-year horizon in all of our numerical examples. To parameterize the model, we first note that the distribution of the maximum value x of the arbitrage during the investment horizon is given by $(4/\sigma^2)x \exp(-2x^2/\sigma^2)$ when $A_0 = 0$ and $\alpha = 1$. The expected value of the maximum is $E[x] = (\sigma/2)\sqrt{\pi/2}$. Thus the expected maximum value is 0.6267 and 1.2533 for $\sigma = 1$ and $\sigma = 2$, respectively. Given this, we consider initial values of the arbitrage of $A_0 = 0$ and $A_0 = 1$. The value $A_0 = 1$ represents the 86th percentile that the maximum can attain when $\sigma = 1$, and the 39th percentile when $\sigma = 2$.

Turning first to the estimation of the speed of the mean reversion parameter α, Table 1 of Longstaff (2004) reports serial correlations for

Table 1
Percentage of times that margin constraint is binding

A_0	λ	α	σ	$t = .25$	$t = .50$	$t = .75$	$t = .999$
0	1	1	1	9.44	32.40	56.58	97.32
0	1	1	2	.09	5.17	25.26	94.70
0	1	2	1	33.18	51.82	66.26	98.14
0	1	2	2	5.11	19.40	38.03	95.93
0	10	1	1	86.18	92.41	95.83	99.77
0	10	1	2	73.18	84.93	91.37	99.50
0	10	2	1	91.86	94.86	96.51	99.86
0	10	2	2	84.23	90.05	93.05	99.59
1	1	1	1	53.27	53.43	62.20	97.28
1	1	1	2	.66	7.72	26.47	94.47
1	1	2	1	70.77	59.94	66.72	98.13
1	1	2	2	11.78	21.68	37.98	95.93
1	10	1	1	96.96	95.51	95.98	99.78
1	10	1	2	81.82	85.57	90.68	99.47
1	10	2	1	97.49	95.31	96.80	99.86
1	10	2	2	88.22	90.52	93.16	99.59

This table reports the percentage of 10,000 simulated paths for which the margin constraint is binding at the indicated horizons. The initial value of the arbitrage is set equal to zero, the final convergence date for the arbitrage is one year and the simulation uses 10,000 discretization points per year in modeling the arbitrage process. The initial value of the arbitrage is A_0. The parameter λ represents the margin requirement. The parameters α and σ represent the speed of convergence and the volatility of the arbitrage process. The riskless rate is 6%.

the price differences between Treasury-guaranteed bonds with identical cash flows. Mapping from these serial correlations into the speed of the mean reversion parameter implies an estimate of $\alpha = 2.33$ for the one-year zero-coupon case, roughly corresponding to a five-month half life. Similarly Cornell and Shapiro (1989) document an apparent arbitrage involving on- and off-the-run Treasury bonds following the auction of May 1986. This arbitrage lasted for several months, but was gone after 12 months. Based on this evidence, we believe that values of α in the range of one to two appear reasonable for this calibration exercise. To parameterize σ, we observe that Kamara (1994) finds that the maximum apparent arbitrage for three-month Treasury bills and notes during a seven-year sample period is on the order of 160 basis points, implying a maximum price difference of about $0.40 per $100 notional amount. Similarly Longstaff reports a maximum size for differences between one-year zero-coupon bonds for a 10-year sample period of 80 basis points, implying a maximum price difference of $0.80 per $100 notional amount. Based on this, in conjunction with the results for the expected maximum value of a Brownian bridge mentioned above, we chose to use values of σ in the range of one to two. These values are likely conservative given that Cornell and Shapiro (1989) and Boudoukh and Whitelaw (1991) find maximum values for the differences between bonds with identical cash flows on the order of $7.00 per $100 notional.

Table 1 shows that the collateral constraint often does not bind. Thus the underinvestment region shown in Equation (17) is far from trivial. The intuition for why the investor does not always take the largest possible position in the arbitrage is directly related to the risk of the arbitrage widening (combined with the usual trade-off between risk and expected return). When A_t differs only slightly from zero, it is almost as likely that the arbitrage will widen as narrow, since the drift is close to zero. Furthermore, when A_t is close to zero, the potential loss from the arbitrage widening can be much larger than the possible gain from the arbitrage converging, at least in the near term. Specifically, the investor can realize a small gain per unit of the arbitrage if it converges to zero over the next short interval, but can experience a large loss if it widens to several times its current value. If the investor suffers large losses in the early stages, he clearly has less wealth to exploit arbitrages at a later stage. By being too aggressive with small arbitrages, the investor risks finding himself in a state of the world where there is a large arbitrage, but his ability to exploit the arbitrage is severely reduced because of losses suffered as the arbitrage widened. It is important to stress that while the percentages shown in Table 1 depend on the parameter values chosen, the basic underinvestment result is robust to the choice of parameters. Specifically, the existence of the middle underinvestment region given in Proposition 1 holds for all positive values of λ, α, and σ. Thus the qualitative nature of the results in

Table 1 (and all subsequent tables in this article) are generic properties of the model and do not depend on the specific parameterization used.

Table 1 also shows that the lower the collateral requirement λ, the less frequently the collateral constraint binds. Similarly, the riskier the arbitrage as measured by σ^2, the less frequently the investor finds it optimal to take the maximum position. As the speed of convergence α increases, the investor takes a more aggressive position and the collateral constraint is more likely to be binding. Finally, while the collateral constraint often does not bind, it is important to observe that the probability that the constraint binds approaches one as $t \to T$. Thus, while the constraint does not bind at every instant, it binds in the global sense of Loewenstein and Willard (2000b).

It is useful to put these results into perspective with those in the literature. In an interesting recent article, Basak and Croitoru (2000) demonstrate that arbitrage opportunities or mispricing can exist within a general equilibrium model in which investors face constraints on their portfolio weights. In their framework, however, investors always take the largest possible position in the arbitrage whenever there is mispricing. Thus our results differ fundamentally from theirs. Loewenstein and Willard (2000b) present an example of an economy in which there is a risky asset where the logarithm of its price follows a Brownian bridge process similar to ours. They show that there is an optimum for an investor who has limited credit capacity. In one sense, this parallels our result, since we also find that the investor's problem has an optimal solution. Loewenstein and Willard, however, do not provide a characterization of the optimal portfolio strategy for the agent in their example, focusing instead on the important issue of the viability of the process. Thus our emphasis on optimal portfolio strategies and the resulting implications for the nature of return differs in a fundamental way from their focus.

Finally, our results complement those of Shleifer and Vishny (1997), who study a model in which a risk-neutral arbitrageur can trade a security at a price which may deviate temporarily from its fundamental value. Capital is provided to the arbitrageur by investors on the basis of past return performance (rather than future investment opportunities). This creates an agency conflict for the arbitrageur who attempts to maximize the amount of funds under management. In their equilibrium, the total amount of funds invested in the arbitrage can be less than the total funds available to arbitrageurs for some sets of parameters. It is important to observe, however, that a price-taking arbitrageur in their model has first-order conditions that are unrelated to the size of his position in the arbitrage.[11]

---

[11] This follows directly from Shleifer and Vishny [1997; Equation (8)]. Their first-order condition for the price-taking arbitrageur does not depend on the amount D_1 invested in the arbitrage, and is satisfied for specific combinations of the market prices p_1, p_2, and the probability q, independent of the individual atomistic arbitrageur's choice of D_1.

Thus the arbitrageur in their equilibrium is indifferent between taking a fully invested position or a partially invested position in the arbitrage. Because of this, their underinvestment result is not due to optimizing behavior by the arbitrageur, but rather to the special structure of the equilibrium demands of the nonstrategic noise traders and investors in their model. In fairness to Shleifer and Vishny, however, their focus is on the important issue of the effects of agency conflicts between portfolio managers and investors rather than on the optimal portfolio choices of the investors in their model.

In some sense, one of the most striking features of the optimal portfolio strategy is that the investor essentially treats the arbitrage opportunity as if it were simply a conventional investment opportunity. For example, when the inequality in Equation (17) holds, the optimal portfolio strategy is proportional to the instantaneous expected return on the arbitrage divided by its instantaneous variance. This closely parallels the standard Merton (1971) result for a logarithmic investor who allocates his wealth between a risky and a riskless asset. Thus the investor in our model does not have any special "arbitrage" motive for taking a position in the arbitrage; the investor takes a position in the arbitrage that is essentially the same as the usual "hedging and speculative" position he would take in an asset with the same instantaneous risk and return trade-off. What is different in this framework, of course, is that the risk and return trade-off for the arbitrage portfolio tends to get progressively better as $t \to T$.

To further explore this, we also solve for the optimal portfolio when there is a third investment available to the investor which follows a standard geometric Brownian motion and can be interpreted as a stock index fund. We find that the investor acts as if there were simply two conventional risky assets available to him. Furthermore, while his portfolio holding in the arbitrage is affected by the instantaneous correlation between the arbitrage and the stock index fund, the investor often fails to take the largest position in the arbitrage opportunity allowed by the collateral requirements. Thus the underinvestment result holds even when there are additional risky assets in the economy.[12]

From Proposition 1 it is clear that the optimal N_t is continuous even at the boundary where the collateral constraint becomes binding. Thus there are no abrupt changes in the size of the arbitrage position when the boundary is reached. Over time, however, the absolute value of N_t tends to decrease after the boundary is reached. To see the intuition for this,

---

[12] The exception, of course, is when the additional risky asset is perfectly correlated with the dynamics of the arbitrage. In this special case, the investor can form a riskless portfolio that earns a return that differs from the riskless rate. We abstract from this type of arbitrage since there is no clear sense in which there is a convergence date for the arbitrage, prior to which the arbitrage could widen. We are grateful to the referee for suggesting the inclusion of additional assets in the analysis. See also Cheng (1991), Föllmer and Imkeller (1993), Pikovsky and Karatzas (1996), and Loewenstein and Willard (2000b).

consider the case where the boundary is just reached and the value of the arbitrage then moves back toward zero. The constraint is no longer binding and the size of the arbitrage position is reduced since the arbitrage is no longer as large. On the other hand, if the boundary is reached and the arbitrage widens, the investor then suffers a decrease in his wealth. Because of this decline in wealth, the constraint $W_t \geq \lambda |N_t|$ can only be satisfied by reducing the absolute value of N_t in this self-financing framework. Thus the investor must partially liquidate his position in the arbitrage at a loss.

Since the optimal strategy involves taking a position in the arbitrage opposite in sign to A_t, the portfolio weight for the arbitrage position,

$$w_t = \frac{N_t A_t}{W_t}, \tag{18}$$

is less than or equal to zero. Because w_t is the ratio of the investor's liabilities to his total wealth, w_t can also be interpreted as a leverage ratio. To give a sense of the distribution of portfolio weights that results from following the optimal portfolio, Table 2 provides summary statistics for the percentage portfolio weights for different values of t and of the parameters. As shown, the optimal portfolio strategy can be highly leveraged even when there are collateral constraints. For example, when $A_0 = 0$ and $\sigma = 2$, the portfolio leverage ratio $|w_t|$ can be greater than four.

3. The Returns from Arbitrage

In this section we examine the wealth distributions obtained from following the optimal investment strategy in a market with arbitrage opportunities. Specifically, the investor's wealth at time t can be expressed as

$$W_t = W_0 e^{rt} \exp\left(\int_0^t - \left(r + \frac{\alpha}{T-s} \right) FA - \frac{\sigma^2}{2} F^2 ds + \sigma \int_0^t F \, dZ \right). \tag{19}$$

It can be shown that the investor's wealth is strictly positive for all t, $0 \leq t \leq T$. Thus the optimal investment strategy satisfies the positive wealth constraint of Dybvig and Huang (1988). Because of the boundedness of F, it follows from Proposition 1 that W_t is finite with probability one.

In specific cases, the range of possible returns that can be obtained from investing optimally in the arbitrage opportunity can be narrowed. The following proposition gives sufficient conditions for the optimal investment portfolio to dominate the riskless asset at time T (proof available upon request).

Proposition 2. Dominance of the Optimal Strategy. *If* $0 < \alpha \leq 1$, *then* $W_T \geq W_0 e^{rT}$ *a.s. when the optimal strategy is followed.*
Thus, when $0 < \alpha \leq 1$, the optimal investment strategy cannot achieve a lower return than the riskless rate. In this sense, the optimal strategy

Table 2
Summary statistics for the percentage portfolio weights invested in the arbitrage

A_0	λ	α	σ	t	Min.	Mean	Median	Max.	Std. Dev.
0	1	1	1	.250	−202.65	−23.20	−11.65	.00	23.81
				.500	−196.48	−34.09	−23.26	.00	33.07
				.750	−187.47	−32.50	−29.13	.00	27.40
				.999	−15.16	−2.60	−2.21	.00	1.97
0	1	1	2	.250	−405.30	−25.61	−11.65	.00	36.29
				.500	−392.95	−48.17	−23.26	.00	61.11
				.750	−374.95	−56.07	−34.45	.00	57.60
				.999	−30.33	−5.19	−4.42	.00	3.94
0	1	2	1	.250	−179.28	−25.91	−17.72	.00	25.23
				.500	−150.16	−28.35	−25.56	.00	24.36
				.750	−115.83	−21.98	−19.15	.00	17.78
				.999	−7.83	−1.54	−1.30	.00	1.17
0	1	2	2	.250	−358.55	−36.69	−17.72	.00	46.53
				.500	−300.32	−47.03	−26.53	.00	50.58
				.750	−231.66	−40.13	−29.56	.00	37.61
				.999	−15.67	−3.08	−2.61	.00	2.34
1	1	1	1	.250	−241.25	−70.80	−75.11	.00	45.64
				.500	−246.48	−54.13	−52.36	.00	43.00
				.750	−201.68	−38.29	−34.01	.00	31.04
				.999	−15.06	−2.60	−2.21	.00	1.97
1	1	1	2	.250	−407.51	−44.89	−23.56	.00	56.06
				.500	−442.95	−58.52	−29.55	.00	70.22
				.750	−378.36	−59.20	−37.54	.00	59.86
				.999	−30.23	−5.20	−4.41	.00	3.94
1	1	2	1	.250	−201.96	−56.70	−56.65	.00	36.19
				.500	−174.46	−34.78	−31.78	.00	28.36
				.750	−115.76	−22.49	−19.49	.00	18.20
				.999	−7.83	−1.54	−1.30	.00	1.17
1	1	2	2	.250	−347.68	−54.13	−29.91	.00	60.25
				.500	−323.93	−50.68	−30.03	.00	53.30
				.750	−225.41	−40.41	−29.74	.00	37.84
				.999	−15.67	−3.08	−2.61	.00	2.34

This table reports summary statistics for the percentage portfolio weights for the indicated horizons based on 10,000 simulated paths. The final convergence date for the arbitrage is one year and the simulation uses 10,000 discretization points per year in modeling the arbitrage process. The initial value of the arbitrage is A_0. The parameter λ represents the margin requirement. The parameters α and σ represent the speed of convergence and the volatility of the arbitrage process. The riskless rate is 6%.

generates a pure arbitrage at time T. When $\alpha > 1$, however, the returns from the arbitrage do not dominate the riskless asset. To illustrate this, Table 3 provides summary statistics for the wealth distributions at different horizons obtained from following the optimal portfolio strategy. Table 3 confirms the dominance result that when $\alpha = 1$, the optimal arbitrage portfolio ends up doing better than the riskless asset at time T. When $\alpha > 1$, however, the final value of the arbitrage portfolio can be less than the value of the riskless portfolio, and can even be less than the initial value of the investor's wealth. Thus the arbitrage portfolio is no longer an arbitrage, even in the classic sense.

This latter result is particularly interesting given that it is actually possible to find an investment strategy that dominates the riskless asset

Table 3
Summary statistics for the value of the optimal portfolio

A_0	λ	α	σ	t	Min.	Mean	Median	Max.	Std. Dev.
0	1	1	1	.250	20.63	105.27	111.33	119.64	15.39
				.500	20.79	118.32	128.70	151.15	28.82
				.750	23.78	144.88	149.04	220.03	44.69
				1.000	113.53	278.85	266.45	763.86	88.83
0	1	1	2	.250	8.30	105.38	111.38	119.64	15.45
				.500	6.67	120.39	131.62	151.15	30.13
				.750	5.69	158.30	176.55	223.09	53.44
				1.000	131.45	482.73	453.52	1413.09	186.92
0	1	2	1	.250	22.02	110.98	118.60	139.39	22.99
				.500	28.58	135.44	137.98	214.04	40.09
				.750	39.94	183.14	176.55	428.22	65.85
				1.000	98.83	420.94	387.01	1457.41	179.84
0	1	2	2	.250	5.94	112.72	122.03	139.39	24.70
				.500	7.87	147.22	162.17	214.04	48.42
				.750	11.79	228.79	230.84	430.42	97.04
				1.000	90.10	995.53	873.80	5488.36	583.91
1	1	1	1	.250	22.37	126.10	123.85	201.28	43.63
				.500	19.43	155.30	156.46	252.65	58.47
				.750	28.49	199.18	195.50	366.68	76.16
				1.000	184.98	390.99	366.18	1049.28	141.19
1	1	1	2	.250	9.36	111.29	119.89	136.56	24.49
				.500	4.35	131.71	145.59	172.88	38.60
				.750	5.53	176.57	197.01	251.04	63.33
				1.000	139.82	541.30	512.30	1664.83	214.64
1	1	2	1	.250	33.05	150.73	142.36	278.51	52.08
				.500	39.56	201.94	197.95	414.53	72.05
				.750	65.25	278.91	261.88	739.87	109.86
				1.000	182.21	640.59	580.94	2669.69	287.75
1	1	2	2	.250	7.38	129.38	142.73	179.41	40.08
				.500	6.85	174.35	189.56	273.14	65.14
				.750	12.45	272.57	270.88	547.72	123.60
				1.000	88.47	1183.89	1029.33	6326.61	717.29

This table reports summary statistics for the value of the optimal portfolio at the indicated horizons based on 10,000 simulated paths. The initial value of the portfolio is 100. The final convergence date for the arbitrage is one year and the simulation uses 10,000 discretization points per year in modeling the arbitrage process. The initial value of the arbitrage is A_0. The parameter λ represents the margin requirement. The parameters α and σ represent the speed of convergence and the volatility of the arbitrage process. If the initial wealth of 100 was invested in the riskless asset only, its value in one year would be 106.18.

even when $\alpha > 1$. Specifically, let D_t denote the portfolio strategy given by

$$D_t = \begin{cases} \dfrac{1}{\lambda} W_t, & \text{if } A_t < -\dfrac{\sigma^2}{2r\lambda}, \\[2mm] 0 & \text{if } |A_t| \le \dfrac{\sigma^2}{2r\lambda}, \\[2mm] -\dfrac{1}{\lambda} W_t, & \text{if } A_t > \dfrac{\sigma^2}{2r\lambda}. \end{cases} \tag{20}$$

We call this strategy the barrier strategy since it is zero until the arbitrage reaches a specific level. Because $|D_t|$ is always less than or equal to the collateral constraint, this portfolio strategy is always feasible. It can be shown that for all $\alpha > 0$, $W_T \geq W_0 e^{rT}$ *a.s.* when the barrier strategy is

followed, implying that the investor can achieve a wealth distribution that dominates the riskless asset. Even though this strategy is available to the investor when $\alpha > 1$, the investor finds it optimal to ignore it. Surprisingly, the optimal strategy is to invest in a way that runs the risk of under-performing the riskless asset even though there is a strategy available that guarantees the investor's return cannot be less than the riskless asset.

Another interesting feature relates to the shape of the distribution of investment returns. During the early stages of the investment horizon, the mean value of the portfolio is often substantially lower than the median value. As the final convergence date approaches, the distribution typically becomes skewed toward higher values and the mean exceeds the median. To illustrate this, Figure 1 graphs the distribution of values for the optimal arbitrage portfolio at times $t = .25$, $t = .50$, $t = .75$, and $t = 1.00$.
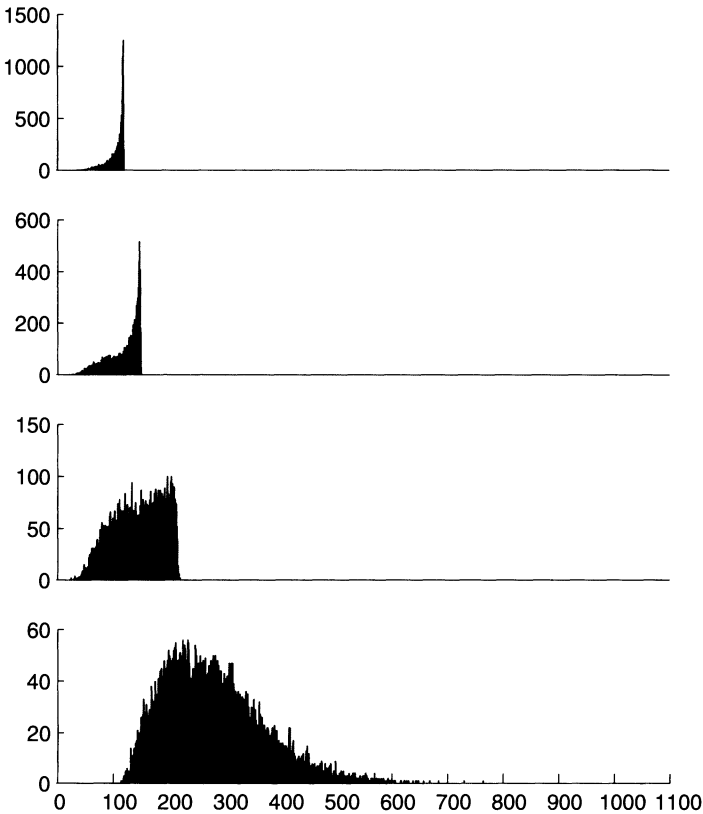

Figure 1
The distribution of wealth
From top down, the graph shows the distribution of the value of the optimal portfolio at time $t = .25$, .50, .75, and 1.00. The initial value of the portfolio is 100. The initial value of the arbitrage A_0 is zero. The parameter values are $\alpha = 1$, $\lambda = 1$, and $\sigma = 1$.

The distribution at $t = .25$ is highly skewed toward the left. This is also true for $t = .50$. At $t = .75$, however, the nature of the distribution begins to change and a more symmetrical pattern appears. At the final maturity date T, the distribution now becomes highly skewed toward higher values, and values less than the riskless value of 106.18 disappear since $\alpha = 1$ in this example. This clearly has important implications for the typical value-at-risk analysis currently widely used among practitioners.

The value of the optimal arbitrage portfolio is highly variable over time. Starting from an initial value of 100, the arbitrage portfolio can actually lose more than 75% of its value by $t = .25$. Analyzing these particular paths reveals that the investor takes a large position in the arbitrage portfolio at an early date, but then loses significant amounts as the arbitrage continues to widen. When the investor reaches the collateral constraint, the investor is forced to unwind his position at a loss in order to satisfy the collateral constraint as the arbitrage widens further. Although the arbitrage ultimately converges to zero at time T, the investor is unable to fully participate at later stages since his wealth is now much lower. Thus investors who experience large losses early during the life of the arbitrage end up with lower returns at time T. This can be seen in Figure 2, which graphs the final value of the portfolio at time T against the minimum value of the portfolio during the life of the arbitrage. Thus early losses due to the widening of the arbitrage do not entirely "come back" later on as the arbitrage ultimately converges to zero.

The mean values of the arbitrage portfolios display an interesting pattern. Initially they tend to be somewhat larger than the value of the riskless portfolio. Over time, however, the means grow rapidly and ultimately far exceed the value of the riskless portfolio. The farther the initial value of the arbitrage is from zero, the higher the final expected value of the optimal portfolio. This is intuitive, since when $A_0 \neq 0$, the investor immediately has the opportunity to invest in an arbitrage. The distribution of returns is typically shifted toward higher values when the value of α increases. Intuitively, a higher value of the speed of mean reversion implies that an arbitrage tends to converge more rapidly. On the other hand, the investor finds it optimal to take a larger position in the arbitrage for any given value of A_t. Because of this latter effect, there can be paths where the investor does worse than would be the case for a smaller value of α. The standard deviation of the value of the arbitrage portfolio demonstrates that uncertainty about the ultimate value of the portfolio is not resolved evenly over time. In the early stages in the life of the arbitrage, the standard deviation of the value of the optimal portfolio is fairly small. As the final convergence date T is approached, however, the standard deviation of the value of the arbitrage portfolio grows rapidly.

A detailed analysis of the returns from following the optimal strategy reveals that returns have three primary sources. First, the investor benefits
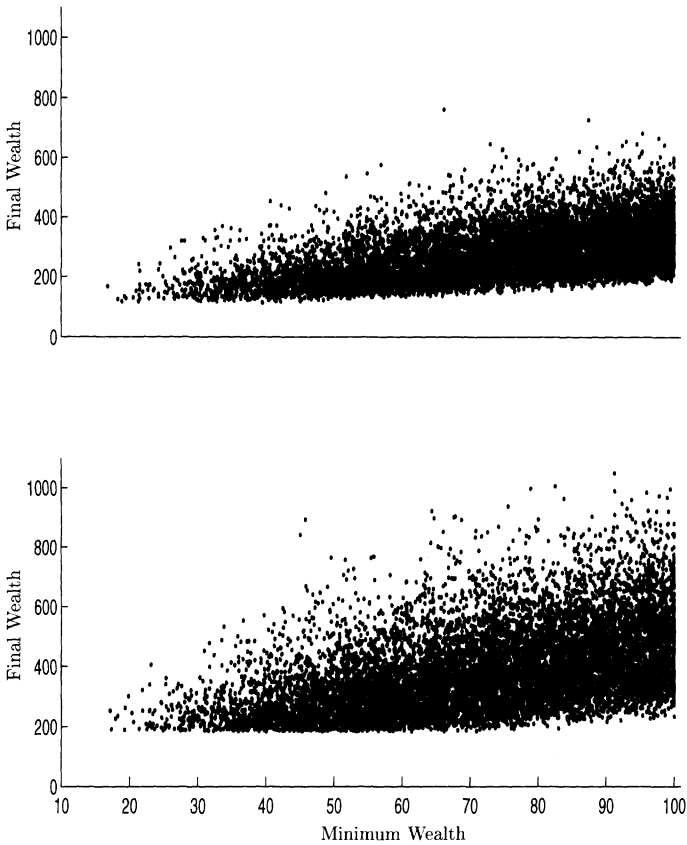
Figure 2
Graph of wealth versus the minimum value of wealth
The graph shows the relation between the final value and the minimum value of the optimal portfolio.
The initial value of the portfolio is 100. The initial value of the arbitrage A_0 is zero and one, respectively,
in the the top and bottom graph. The parameter values are $\alpha = 1$, $\lambda = 1$, and $\sigma = 1$.

by investing directly in an arbitrage which then eventually converges. The
more frequently there is an arbitrage which then converges, the higher the
value of the portfolio at the final maturity. Second, the final value of
the portfolio is adversely affected by reaching the collateral constraint.
This can be seen in Figure 3, which graphs the final value of the portfolio
against the percentage of times that the collateral constraint is binding
along a path. There is a strong negative relation between the final value of
the portfolio and the frequency with which the collateral constraint is
binding. Intuitively this is because when the collateral constraint is bind-
ing and there is an increase in the size of the arbitrage, the investor is
forced to reduce his position at a loss rather than more aggressively
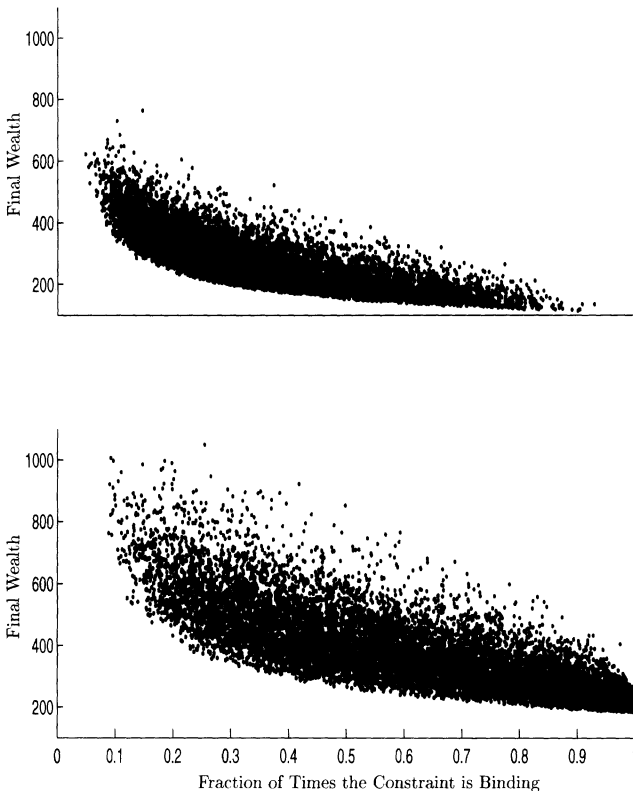exploiting the wider arbitrage.

631

Figure 3
Graph of wealth versus the fraction of times the constraint is binding
The graph shows the relation between the final value of the optimal portfolio and the fraction of times
that the collateral constraint is binding. The initial value of the portfolio is 100. The initial value of the
arbitrage A_0 is zero and one, respectively, in the top and bottom graph. The parameter values are $\alpha = 1$,
$\lambda = 1$, and $\sigma = 1$.

Given these two effects, the investor does best when the value of the
arbitrage tends to return frequently to zero and stays away from larger
values which would then cause the collateral constraint to be binding more
frequently. This surprising implication is illustrated in Figure 4, which
plots the final value of the portfolio against the average value of the
arbitrage during its life. When the arbitrage is initially zero, the highest
final value of the portfolio tends to be for those paths for which the
average value of the arbitrage is close to zero. Similarly, when $A_0 = 1$,
the highest final values of the arbitrage portfolio tend to occur for paths
where the arbitrage returns quickly to the neighborhood of zero, resulting
in average values of A_t of between zero and one. Thus the highest returns
occur along paths where there is a steady flow of small arbitrages that
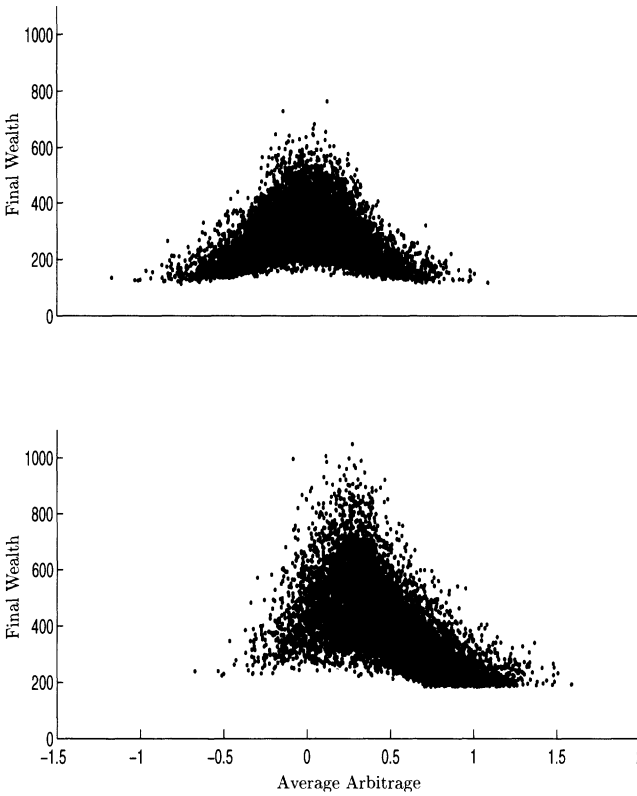converge rapidly and where large widenings in the value of A_t do not

Figure 4
Graph of wealth and the average of the arbitrage
The graph shows the relation between the final value of the optimal portfolio and the average of the arbitrage. The initial value of the portfolio is 100. The initial value of the arbitrage A_0 is zero and one, respectively, in the the top and bottom graph. The parameter values are $\alpha = 1$, $\lambda = 1$, and $\sigma = 1$.

occur. This is consistent with the well-known Wall Street description of the business of relative value or arbitrage investing as "picking up nickels in front of a steamroller." The third source of returns is more subtle. Because of the collateral constraint, the investor is forced to place any cash generated by taking a position in the arbitrage into the riskless asset. Since the arbitrage is a source of cash, the balance invested in the collateral account tends to be larger when the investor takes a position in the arbitrage. Over time, the excess funds in the riskless asset generate additional returns from the accrual of interest.

Although the eventual value of the optimal portfolio is on average much higher than the riskless asset, the intermediate values of the portfolio typically reflect losses at some point during the life of the arbitrage. Table 4 reports pathwise statistics from following the optimal portfolio. These pathwise statistics indicate that for a very high percentage of paths,

Table 4
Pathwise summary statistics for the value of a portfolio following the optimal strategy

A_0	λ	α	σ	Percent <100	Percent <R_t	Percent max. >W_T	Average t_{Min}	Average t_{Max}	Average min.	Average max.
0	1	1	1	99.89	100.00	.17	.3605	.9999	75.34	278.85
0	1	1	2	99.91	100.00	.49	.4366	.9999	67.09	482.73
0	1	2	1	99.94	99.98	.57	.2532	.9997	75.81	420.94
0	1	2	2	99.96	99.99	1.74	.3170	.9985	65.64	995.60
0	10	1	1	99.50	99.86	.00	.1722	1.0000	96.85	120.27
0	10	1	2	99.71	99.89	.00	.2271	1.0000	93.13	135.31
0	10	2	1	99.38	99.72	.00	.1269	1.0000	97.27	125.42
0	10	2	2	99.69	99.84	.01	.1625	1.0000	94.00	147.20
1	1	1	1	98.60	98.66	.14	.1980	.9999	72.17	390.99
1	1	1	2	99.18	99.33	.50	.3538	.9999	65.27	541.30
1	1	2	1	97.86	97.98	.53	.1182	.9999	80.04	640.59
1	1	2	2	98.77	98.83	1.80	.2272	.9980	63.67	1183.97
1	10	1	1	97.64	98.49	.00	.1036	1.0000	97.67	125.20
1	10	1	2	98.63	99.00	.00	.1877	1.0000	93.32	138.73
1	10	2	1	96.55	97.42	.00	.0547	1.0000	98.42	131.35
1	10	2	2	98.26	98.65	.01	.1273	1.0000	94.89	151.57

This table reports summary statistics taken over 10,000 paths for the value of a portfolio where the optimal strategy is followed. The final convergence date for the arbitrage is one year and the simulation uses 10,000 discretization points per year in modeling the arbitrage process. The initial value of the arbitrage is A_0. The parameter λ represents the margin requirement. The parameters α and σ represent the speed of convergence and the volatility of the arbitrage process. The riskless rate is 6%. Percent <100 is the percentage of paths for which the minimum value of the portfolio was less than 100. Percent <R_t is the percentage of paths for which the return on the portfolio was less than the riskless asset at some point. Percent max. > W_T is the percentage of paths for which the maximum value of the portfolio occurred prior to T. The values of Average t_{Min} and Average t_{Max} are the average of the times at which the minimum and maximum values of the arbitrage occurred. The values Average min. and Average max. are the average minimum and maximum values of the portfolio.

the minimum value of the optimal portfolio is actually less than its initial value of 100. Specifically, the percentage of paths for which there is an actual capital loss at some point during the life of the arbitrage is typically in excess of 96%. Note that this is also true for the case where $\alpha = 1$, which guarantees that the final value of the arbitrage is strictly greater than the riskless asset. Clearly, the probability of the value of the portfolio dropping below the value of the riskless asset at some point is even higher than the probability of dropping below 100; Table 4 shows that the probability of underperforming the riskless asset at some point during the investment horizon is typically greater than 97%.

These results have many interesting implications for performance expectations for hedge funds investing in arbitrage opportunities. These results indicate that experiencing capital losses prior to the final horizon is part of the inherent nature of investments in arbitrage opportunities in markets with collateral constraints. Thus there is a definite "darkest before dawn" nature to arbitrage investments. This contrasts dramatically with the widely held view that investors in arbitrage opportunities should never experience significant losses. An immediately corollary of this widely held view is that arbitrage funds can experience losses only if they

are not really investing in arbitrage opportunities, but speculating in conventional types of investments. Our analysis, however, demonstrates that this common wisdom is flawed. Losses during the early stages of an arbitrage opportunity are almost inevitable for an investor pursuing an optimal investment strategy in the arbitrage; during the early stages of the arbitrage strategy, its returns may be observationally indistinguishable from those of a severely distressed conventional portfolio. Finally, Table 4 shows that the average minimum ranges from about 60 to 98 for a portfolio initially worth 100. Thus an investor following an optimal strategy can expect to be down as much as 40% at some point for some parameter values. This again contrasts with the common view that true arbitrage positions should never show losses.

One popular measure of the attractiveness of a portfolio's return is the traditional Sharpe ratio. To make our analysis of Sharpe ratios compatible with the ratios typically reported by the financial industry, we do the following. For each simulated path, we compute the sample standard deviation of changes in the value of the portfolio. We do this for horizons of .25, .50, .75, and 1.00 year and take the excess return of the portfolio at the same horizons over that of the riskless asset. Dividing the annualized excess return by the annualized return gives the estimated Sharpe ratio. We repeat this process for 10,000 paths and provide summary statistics for the resulting distribution of Sharpe ratios at the various horizons. These summary statistics are reported in Table 5. Figure 5 graphs the distribution of Sharpe ratios for selected values of the parameters.

The Sharpe ratios for investing in the arbitrage are quite variable. This is particularly true at the early stages. At the convergence date, however, the average Sharpe ratio is about two for all of the examples shown in Table 5. Curiously, this is about the same as the average Sharpe ratio of 1.82 for the relative-value hedge funds reported as of December 3, 2001, by the Web site HedgeFund.net, which tracks the performance of more than 2,000 hedge funds.[13] Figure 5 shows that most of the Sharpe ratios at the final convergence date are between zero and four. Thus there is no guarantee that a hedge fund following the optimal investment strategy will have a Sharpe ratio even as large as that for the S&P 500.[14]

4. Conclusion

We examine the optimal investment policy of a risk-averse investor in a market where there are textbook arbitrage opportunities and where the

---

[13] On December 3, 2001, HedgeFund.net reports that the average Sharpe ratio for convertible arbitrage hedge funds is 2.57, for fixed-income arbitrage funds is 1.86, for options arbitrage hedge funds is 2.10, and for risk arbitrage hedge funds is 0.90.

[14] For evidence about the return performance and trading strategies of hedge funds, see Fung and Hsieh (1997, 1998) and Ackermann, McEnally, and Ravenscraft (1999).

Table 5
Summary statistics for the annualized sharpe ratio

A_0	λ	α	σ	t	Min.	Mean	Median	Max.	Std. Dev.
0	1	1	1	.250	−8.95	1.05	1.24	8.03	1.97
				.500	−4.00	.85	1.00	5.98	1.38
				.750	−2.51	.86	.85	5.13	1.03
				1.000	.12	1.66	1.60	5.13	.67
0	1	1	2	.250	−9.05	1.08	1.24	8.03	1.94
				.500	−3.98	.95	1.09	5.98	1.32
				.750	−2.81	.99	1.10	5.13	1.06
				1.000	.34	1.82	1.76	4.79	.63
0	1	2	1	.250	−7.70	1.07	1.20	7.99	2.00
				.500	−2.97	.95	.92	5.94	1.26
				.750	−1.64	1.06	1.01	5.02	.89
				1.000	−.18	2.19	2.18	5.17	.70
0	1	2	2	.250	−7.72	1.23	1.37	7.99	1.91
				.500	−3.96	1.12	1.23	5.94	1.29
				.750	−2.10	1.15	1.12	5.02	.98
				1.000	−.18	2.19	2.17	5.17	.70
1	1	1	1	.250	−6.36	.99	.86	8.14	2.06
				.500	−3.10	1.05	.99	6.91	1.33
				.750	−1.88	1.11	1.05	4.72	.95
				1.000	.68	1.83	1.75	4.68	.61
1	1	1	2	.250	−6.22	1.14	1.29	8.11	1.96
				.500	−3.91	1.06	1.20	5.48	1.33
				.750	−2.86	1.07	1.16	4.79	1.05
				1.000	.45	1.86	1.79	4.60	.62
1	1	2	1	.250	−4.45	1.56	1.35	8.11	1.80
				.500	−1.83	1.50	1.42	6.71	1.11
				.750	−.68	1.46	1.38	4.68	.80
				1.000	.60	2.29	2.25	5.27	.65
1	1	2	2	.250	−6.11	1.38	1.57	8.27	1.92
				.500	−3.07	1.26	1.32	5.58	1.28
				.750	−1.98	1.24	1.20	4.96	.97
				1.000	−.17	2.22	2.20	5.17	.70

This table reports summary statistics for the distribution of annualized Sharpe ratios based on 10,000 simulated paths. The final convergence date for the arbitrage is one year and the simulation uses 10,000 discretization points per year in modeling the arbitrage process. The Sharpe ratios are computed pathwise from the annualized mean and standard deviations of changes in the value of the optimal portfolio. The initial value of the arbitrage is A_0. The parameter λ represents the margin requirement. The parameters α and σ represent the speed of convergence and the volatility of the arbitrage process.

investor must post collateral against the risk of short positions. We find that the optimal policy often results in the investor underinvesting in the arbitrage by taking a smaller position than would be allowed by the collateral constraint. Even when the optimal policy is followed, the initial returns from the arbitrage strategy may be indistinguishable from those from an ordinary portfolio with a losing track record.

There are many possible extensions of this analysis. Alternative preference structures or objective functions could be used in solving the investor's or hedge fund manager's problem. A simple perturbation argument, however, suggests that our basic results hold for preference structures sufficiently close to logarithmic, and are not an artifact of the myopic nature of logarithmic preferences. In fact, we conjecture that our
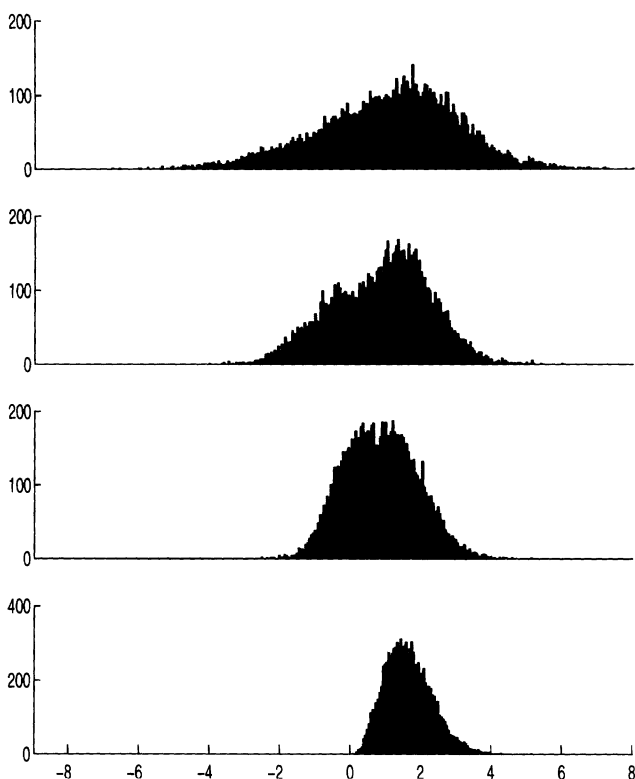
Figure 5
Distribution of the Sharpe ratio
From top down, the graph shows the distribution of the annualized Sharpe ratio at time $t = .25, .50, .75,$ and 1.00. The initial value of the portfolio is 100. The initial value of the arbitrage A_0 is zero. The parameter values are $\alpha = 1, \lambda = 1,$ and $\sigma = 1$.

results are true for virtually all risk-averse preferences. Furthermore, using a simple binomial tree example, we can show that our results hold in a discrete-time setting and are not an artifact of continuous-time modeling. The primary message of this article, however, is that when the real-world feature of collateral constraints is introduced, the economics of arbitrage become fundamentally different. In particular, arbitrages become risky investments, and the issue of whether there would be sufficient demand from investors to completely eliminate them becomes relevant.

Appendix

Proof of Proposition 1. We first prove that because of the homogeneity of the problem, the optimal portfolio strategy N_t must be of the form $F_t W_t$, where F_t is a function of t and A_t only. The agent's optimization problem is

$$\max_N E_t[\ln W_T], \tag{A.1}$$

637

subject to the constraints

$$dW = \left(rW - \left(r + \frac{\alpha}{T-t} \right) NA \right) dt + \sigma N dZ,$$

$$W_t \geq \lambda |N_t|,$$

(A.2)

where $W_t > 0$ for all t, $0 \leq t \leq T$. Since $W_t > 0$, $F_t = \frac{N_t}{W_t}$ is well defined and the original optimization problem is equivalent to the optimization problem

$$\max_F E_t[\ln W_T]$$

subject to the constraints

$$dW = \left(r - \left(r + \frac{\alpha}{T-t} \right) FA \right) W dt + \sigma F W dZ,$$

$$|F_t| \leq \frac{1}{\lambda}.$$

(A.3)

By an application of Itô's lemma,

$$\ln W_T = \ln W_t + r(T-t) - \int_t^T \left(\left(r + \frac{\alpha}{T-s} \right) FA + \frac{\sigma^2}{2} F^2 \right) ds + \sigma F dZ,$$

and the optimal F solves the following problem,

$$\min_F E_t \left[\int_t^T \left(\left(r + \frac{\alpha}{T-t} \right) FA + \frac{\sigma^2}{2} F^2 \right) ds \right],$$

(A.4)

subject to the constraint

$$|F_s| \leq \frac{1}{\lambda}.$$

However, since W_t does not appear in Equation (A.4) or the constraint, the optimal control F can only depend on A_t and t. Hence N_t is of the form $F_t W_t$.

Turning now to the optimal F, note that a realization of a path of A_t does not depend on the control F. Thus, minimizing the integral in Equation (A.4) pathwise for A_t clearly minimizes the conditional expectation in Equation (A.4). Given a path of A_t, the problem

$$\min_F \int_t^T \left(\left(r + \frac{\alpha}{T-s} \right) FA + \frac{\sigma^2}{2} F^2 \right) ds,$$

(A.5)

subject to the constraint

$$|F_s| \leq \frac{1}{\lambda},$$

can then be solved using standard calculus of variation techniques [see, e.g., Kamien and Schwartz (1991)]. Given the quadratic form of the integrand in Equation (A.5), it is now easily shown that the optimal portfolio strategy N_t is given by

$$N_t = \begin{cases} \dfrac{1}{\lambda} W_t, & \text{if } A_t < -\dfrac{1}{\lambda} \dfrac{\sigma^2}{(r + \frac{\alpha}{T-t})}, \\[2ex] -\dfrac{r + \frac{\alpha}{T-t}}{\sigma^2} A_t W_t, & \text{if } |A_t| \leq \dfrac{1}{\lambda} \dfrac{\sigma^2}{(r + \frac{\alpha}{T-t})}, \\[2ex] -\dfrac{1}{\lambda} W_t, & \text{if } A_t > \dfrac{1}{\lambda} \dfrac{\sigma^2}{(r + \frac{\alpha}{T-t})}. \end{cases}$$

(A.6)

References
Ackermann, C., R. McEnally, and D. Ravenscraft, 1999, "The Performance of Hedge Funds: Risk, Return, and Incentives," *Journal of Finance*, 54, 833–874.

Amihud, Y., and H. Mendelson, 1991, "Liquidity, Maturity, and the Yields on US Treasury Securities," *Journal of Finance*, 46, 1411–1425.

Back, K., and S. Pliska, 1990, "On the Fundamental Theorem of Asset Pricing with an Infinite State Space," *Journal of Mathematical Economics*, 20, 1–18.

Ball, C. A., and W. N. Torous, 1983, "Bond Price Dynamics and Options," *Journal of Financial and Quantitative Analysis*, 18, 517–531.

Basak, S., and B. Croitoru, 2000, "Equilibrium Mispricing in a Capital Market with Portfolio Constraints," *Review of Financial Studies*, 13, 715–748.

Boudoukh, J., and R. Whitelaw, 1991, "The Benchmark Effect in the Japanese Government Bond Market," *Journal of Fixed Income*, 1, 52–59.

Brennan, M., and E. S. Schwartz, 1988, "Optimal Arbitrage Strategies Under Basis Variability," in M. Sarnat (ed.), *Essays in Financial Economics*, North Holland, New York.

Brennan, M., and E. S. Schwartz, 1990, "Arbitrage in Stock Index Futures," *Journal of Business*, 63, S7–S31.

Chen, Z., 1995, "Financial Innovation and Arbitrage Pricing in Frictional Economies," *Journal of Economic Theory*, 65, 117–135.

Cheng, S. T., 1991, "On the Feasibility of Arbitrage-based Option Pricing when Stochastic Bond Prices are Involved," *Journal of Economic Theory*, 53, 185–198.

Cornell, B., and A. Shapiro, 1989, "The Mispricing of U.S. Treasury Bonds: A Case Study," *Review of Financial Studies*, 2, 297–310.

Cuoco, D., and H. Liu, 2000, "A Martingale Characterization of Consumption Choices and Hedging Costs with Margin Requirements," *Mathematical Finance*, 10, 355–385.

Daves, P. R., and M. C. Ehrhardt, 1993, "Liquidity, Reconstitution, and the Value of U. S. Treasury Strips," *Journal of Finance*, 48, 315–329.

Delgado, F., and B. Dumas, 1994, "How Far Apart Can Two Riskless Interest Rates Be?" working paper, University of Pennsylvania.

De Long, B., A. Shleifer, L. Summers, and R. Waldmann, 1990, "Noise Trader Risk in Financial Markets," *Journal of Political Economy*, 98, 703–738.

Demarzo, P., and D. Duffie, 1999, "A Liquidity-Based Model of Security Design," *Econometrica*, 67, 65–99.

Detemple, J., and S. Murthy, 1997, "Equilibrium Asset Prices and No Arbitrage with Portfolio Constraints," *Review of Financial Studies*, 10, 1133–1174.

Dow, J., and G. Gorton, 1994, "Arbitrage Chains," *Journal of Finance*, 49, 819–849.

Duffie, D., 1990, "The Risk-Neutral Value of the Early Arbitrage Option: A Note," *Advances in Futures and Options Research*, 4, 107–110.

Dumas, B., 1992, "Dynamic Equilibrium and the Real Exchange Rate in a Spatial Separated World," *Review of Financial Studies*, 5, 153–180.

Dybvig, P., and C.-F. Huang, 1988, "Nonnegative Wealth, Absence of Arbitrage, and Feasible Consumption Plans," *Review of Financial Studies*, 1, 377–401.

Föllmer, H., and P. Imkeller, 1993, "Anticipation Cancelled by a Girsanov Transformation: A Paradox on Wiener Space," *Annales de l'Institut Henri Poincare*, 29, 569–586.

Fung, W., and D. Hsieh, 1997, "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds," *Review of Financial Studies*, 10, 275–302.

Fung, W., and D. Hsieh, 1998, "Performance Attribution and Style Analysis: From Mutual Funds to Hedge Funds," working paper, Duke University.

Grinblatt, M., and F. A. 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.

Habib, M., and D. Johnsen, 1999, "The Financing and Redeployment of Specific Assets," *Journal of Finance*, 54, 693–720.

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. Pliska, 1981, "Martingales and Arbitrage in Multiperiod Securities Markets," *Stochastic Processes and Their Applications*, 11, 215–260.

Heath, D., and R. Jarrow, 1987, "Arbitrage, Continuous Trading, and Margin Requirements," *Journal of Finance*, 42, 1129–1142.

Hindy, A., 1995, "Viable Prices in Financial Markets with Solvency Constraints," *Journal of Mathematical Economics*, 24, 105–136.

International Swaps and Derivatives Association, 1999, *Guidelines for Collateral Practitioners*, ISDA, New York.

International Swaps and Derivatives Association, 2000, *Collateral Survey 2000*, ISDA, New York.

James, C., 1988, "The Use of Loan Sales and Standby Letters of Credit by Commercial Banks," *Journal of Monetary Economics*, 22, 395–422.

Johannes, M., and S. Sundaresan, 2001, "Collateralizing Swaps: Implications for Valuation and Zero Extraction," working paper, Columbia University.

Johnson, H., and R. Stulz, 1985, "An Analysis of Secured Debt," *Journal of Financial Economics*, 14, 501–521.

Kamara, A., 1994, "Liquidity, Taxes, and Short-Term Treasury Yields," *Journal of Financial and Quantitative Analysis*, 29, 403–417.

Kamien, M. I., and N. L. Schwartz, 1991, *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, Elsevier Science, Amsterdam.

Loewenstein, M., and G. Willard, 2000a, "Liquidity and Arbitrage," working paper, MIT.

Loewenstein, M., and G. Willard, 2000b, "Local Martingales, Arbitrage, and Viability: Free Snacks and Cheap Thrills," *Economic Theory*, 16, 135–161.

Loewenstein, M., and G. Willard, 2000c, "Rational Equilibrium Asset-Pricing Bubbles in Continuous Trading Models," *Journal of Economic Theory*, 91, 17–58.

Longstaff, F., 1995, "Option Pricing and the Martingale Restriction," *Review of Financial Studies*, 8, 1091–1124.

Longstaff, F., 2004, "The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices," forthcoming in *Journal of Business*.

MacKinlay, A., and K. Ramaswamy, 1988, "Index-Future Arbitrage and the Behavior of Stock Index Futures Prices," *Review of Financial Studies*, 1, 137–158.

Merton, R., 1971, "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *Journal of Economic Theory*, 3, 373–413.

Morellec, E., 2001, "Asset Liquidity, Capital Structure, and Secured Debt," *Journal of Financial Economics*, 61, 173–206.

Pikovsky, I., and I. Karatzas, 1996, "Anticipative Portfolio Optimization," *Advances in Applied Probability*, 28, 1095–1122.

Rajan, R., and A. Winton, 1995, "Covenants and Collateral as Incentives to Monitor," *Journal of Finance*, 50, 1113–1146.

Shleifer, A., and R. W. Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance*, 52, 35–55.

Tuckman, B., and J. L. Vila, 1992, "Arbitrage with Holding Costs: A Utility-Based Approach," *Journal of Finance*, 44, 1115–1153.

Willard, G., and P. Dybvig, 1999, "Empty Promises and Arbitrage," *Review of Financial Studies*, 12, 807–834.

Xiong, W., 2001, "Convergence Trading with Wealth Effects: An Amplification Mechanism in Financial Markets," *Journal of Financial Economics*, 62, 247–292.

Yadev, P., and P. Pope, 1994, "Stock Index Futures Mispricing: Profit Opportunities or Risk Premia?," *Journal of Banking and Finance*, 18, 921–953.

Zigrand, J., 1997, "Arbitrage and Endogenous Market Integration," working paper, University of Chicago.