
INDEX ARBITRAGE AS CROSS-SECTIONAL MARKET MAKING

CRAIG W. HOLDEN

1. INTRODUCTION

Index arbitrage¹ is conducted on a large scale. According to New York Stock Exchange (NYSE) statistics,² an average of 4.6 million shares per day are traded as part of index arbitrage trades. This represents 2.4% of total volume (see Table I). The Brady commission³ investigating the 1987 market crash estimated that \$1.1, \$2.0, and \$1.9 billion worth of stock were traded as part of index arbitrage transactions on October 15, 16, and 19, respectively. This represented 4.9%, 6.8%, and 4.5% of total volume, respectively.

Program trading⁴ in general and index arbitrage in particular have become the focus of public controversy in Japan and in the U.S. over their alleged contributions to market volatility. In Japan, the Finance

I would like to thank my UCLA dissertation committee, Michael Brennan (chairman), David Butz, Hashem Pesaran, Eduardo Schwartz, and Walter Torous. I thank *The Journal of Futures Market's* editor and referees, Brian Betker, Kent Daniel, Silverio Foresi, Bruno Gerard, David Hirshleifer, Tim Opler, James Pellicane, Richard Roll, Jim Shapiro, Avaniidhar Subrahmanyam, and Sheridan Titman, for helpful comments. I gratefully acknowledge the support of an Allstate Dissertation Fellowship. Any errors are my sole responsibility.

¹Defined as trading to exploit arbitrage opportunities between stock index portfolios and synthetic stock index portfolios, where the latter is constructed using derivative securities such as stock index futures.

²The NYSE requires member investment banking firms to self-report every simultaneous trade of 15 or more securities with a portfolio value of more than \$1 million and to separately identify which ones are index arbitrage trades. Statistics are for the period July, 1988 to December, 1993.

³See the Presidential Task Force on Market Mechanisms (1988).

⁴Defined as simultaneous trading in multiple securities.

■ *Craig W. Holden is an Assistant Professor of Finance at Indiana University.*

TABLE I

NYSE Statistics on Program Trading and Index Arbitrage Around November, 1989 Publicity Adverse to Index Arbitrage (Millions of Shares/Day)

	<i>Before Publicity 7/88–10/89</i>	<i>During Publicity 11/89</i>	<i>After Publicity 12/89–12/93</i>	<i>Total Period 7/88–12/93</i>
NYSE volume	160.5	144.4	199.6	189.1
Program trading ^a	8.3	4.7	11.0	10.2
Program trading/volume ^b	5.1%	3.2%	5.5%	5.4%
Index arbitrage	4.2	1.3	4.7	4.6
Index arbitrage/volume ^b	2.6%	0.9%	2.4%	2.4%
Index arbitrage/program trading	51.5%	26.9%	42.8%	44.4%

^aFor statistical purposes, NYSE defines program trading as simultaneous trading in 15 or more stocks with a portfolio value of \$1 million or more.

^bThe NYSE calculates percent of volume figures by summing buy orders and sell orders and dividing by the total volume of shares traded. This double-counting has been removed.

Ministry and the Tokyo Stock Exchange pressured the Osaka Securities Exchange (OSE) to scrap its lucrative stock index futures contract on the price-weighted, Nikkei 225 index, because they believe the index includes thinly traded issues that are subject to price swings when index arbitrage programs hit the market.⁵ Since then the Nikkei 300, a new index that is broader and value-weighted, has been developed and has started trading.

In the U.S. this public controversy developed initially in the wake of the 1987 market crash and reached a peak in reaction to the 190 point drop in the Dow Jones Industrial Average during the last 90 minutes of trading on October 13, 1989. At the peak of the negative publicity about index arbitrage during November, 1989, index arbitrage plummeted to a low of 1.3 million shares per day. Since then, index arbitrage has recovered to 4.7 million shares per day, which is slightly *above* its level prior to the negative publicity (see Table I). Also at the peak of the negative publicity, nine of the 14 investment banking firms that were the heaviest index arbitrage traders announced a halt in index arbitrage trading for their own accounts and six announced a halt in index arbitrage trading for clients as well (see Table II). Since then, four firms (Merrill Lynch, Dean Witter, Oppenheimer, and Shearson Lehman) have permanently abstained from index arbitrage. Several firms (Kidder Peabody, Morgan Stanley, etc.) have reactivated their index arbitrage trading. Many new firms (Nomura,

⁵See Business Week on 3/8/93.

TABLE II
Index Arbitrage by Active NYSE Member Firms Around November,
1989 Publicity Adverse to Index Arbitrage (Millions of Shares/Day)

	Before Publicity 7/88-10/89	During Publicity 11/89	After Publicity 12/89-12/93	Total Period 7/88-12/93	Rank before Publicity 7/88-10/89	Rank after Publicity 12/89-12/93	Announced Index Arb Halt for Accounts
Kidder Peabody	0.69	0.16	0.59	0.61	1	3	Own & clients
Morgan Stanley	0.64	0.14	0.35	0.42	2	6	Own
Susquehanna	0.56	0.12	0.69	0.65	3	2	
Merrill Lynch	0.51	0.04	0.00	0.13	4	NA	Own & clients
Bear Sterns	0.44	0.06	0.25	0.29	5	8	Own
LIT America	0.42	0.27	0.25	0.30	6	7	
First Boston	0.18	0.14	0.40	0.34	7	4	
Salomon Bros.	0.16	0.08	0.09	0.11	8	12	
Dean Witter	0.13	0.00	0.00	0.03	9	NA	Own & clients
Paine Webber	0.10	0.00	0.04	0.06	10	15	Own & clients
Oppenheimer	0.09	0.00	0.00	0.03	11	NA	Own
Shearson Lehman	0.06	0.00	0.00	0.02	12	NA	Own & clients
Goldman Sachs	0.06	0.04	0.08	0.08	13	13	Own & clients
Miller Tabak	0.03	0.14	0.04	0.04	14	16	Own & clients
Nomura	0.00	0.00	0.71	0.52	NA	1	
UBS Securities	0.00	0.00	0.39	0.29	NA	5	
Thomas L. Williams	0.00	0.00	0.14	0.10	NA	9	
Daiwa Securities	0.00	0.00	0.14	0.10	NA	10	
Walsh Greenwood	0.00	0.00	0.11	0.08	NA	11	
Cooper Neff	0.00	0.00	0.05	0.04	NA	14	
Total most active	4.08	1.18	2.80	3.09			
Total all firms	4.24	1.25	4.66	4.50			

NA = not active.

UBS Securities, etc.) have become active in index arbitrage for the first time and have filled the void left by those who exited (see Table II). At the peak of the negative publicity, the NYSE announced a series of voluntary and mandatory steps to restrict program trading. NYSE chairman John J. Phelan Jr. said, "We have, in effect, declared war on excessive volatility."⁶ Since then, the controversy has eased, but the controversy may well reemerge under certain market conditions.⁷

This article addresses four fundamental questions about arbitrage trading.⁸ The first question addressed is: what is the role of arbitrageurs in the financial markets? A paradigm is developed of arbitrageurs as *cross-section* market makers. Traditional market makers, such as specialists, independent floor brokers, scalpers, etc., are *time-series* market makers. They hold nonzero inventories and obtain *risky* profits by supplying liquidity to traders in a single security who arrive at different times. Arbitrageurs hold net zero inventories, in the sense that their long and short positions in risky securities net out to have zero value at maturity. Arbitrageurs obtain *riskless* profits by supplying liquidity to traders in different securities who arrive at the same time. Arbitrage trading and traditional market making are alternative technologies for providing market making services. Arbitrageurs and traditional market makers are competitive rivals.

The second question is: how does arbitrage trading exist in a financial market equilibrium? This question is answered by developing a model in which arbitrageurs optimally exploit arbitrage opportunities that are generated by a clientele effect and by risk-averse market makers in a financial market equilibrium. To construct the model, three major assumptions are made.⁹ First, there exists a clientele effect which disconnects the markets for two securities with the same terminal payoffs (see discussion in the next paragraph). In particular, one liquidity trader trades in stocks (a stock index portfolio) and a second liquidity

⁶See the Los Angeles Times on 11/02/89. Holden (1991) criticizes the reporting practices of most media coverage that routinely cite index arbitrage as a major factor in stock market movements. Implicitly, they are treating index arbitrage as if it affects the common direction of both the stock market and stock index futures rather than simply reducing the fair value difference between them.

⁷"But if the market declines sharply in coming months, as some analysts fear it might, 'people will clamor that it's bad' again ..." See the Wall Street Journal on 1/12/93.

⁸The paradigm case to which the model in this study is applied is index arbitrage. However, the model would apply just as well to covered interest arbitrage, triangular arbitrage, etc.

⁹Grossman (1988) provided the inspiration for this specific formulation with the following description of arbitrage, "It is crucial to recognize that the market makers on futures markets combine with market makers on the NYSE to enhance the overall liquidity of the equity market. A given amount of institutional selling faces the buying power of market makers on both markets. Index arbitrageurs take positions that unify both markets."

trader trades in "synthetic stocks" (stock index futures and bonds which yield the same payoffs as the stock index portfolio). Second, market makers are risk-averse. Market makers charge a liquidity premium proportional to the size of the trade¹⁰ to provide compensation for the price risk that they bear on the inventory they are carrying.¹¹ For example, if a customer can sell 1000 shares at a 1% premium, then alternatively that customer can sell 2000 shares at a 2% premium. This makes intuitive sense, because the later trade causes the market makers to carry twice as much inventory as the former trade.¹² With different liquidity shocks in each market, there are different *realized* prices (or liquidity premia) in each market. Hence, there is a price difference between the two securities. Finally, there exists a third group of traders, called arbitrageurs, who are the most efficient low cost traders between the two markets. They are in position to take advantage of the price difference. The arbitrageurs are modeled as identical oligopolists. Hence, they calculate the optimal quantity of arbitrage trading to maximize their individual profits in a symmetric Nash equilibrium. Their arbitrage trading reconnects the two markets and reduces the price difference.

There are several possible explanations for a clientele effect between stocks and synthetic stocks. First, Amihud and Mendelson (1980) developed the concept of a horizon clientele effect. Some traders may have a short investment (or performance evaluation) horizon such that they trade in synthetics which have the lowest one-time transaction cost. Other traders may have a long investment (or performance evaluation) horizon such that they trade in stocks which have a higher one-time transaction cost, but which never need to be rolled over and so have lower cumulative transaction costs. Second, Titman (1985) pointed out that the interest on the bond in the synthetic stock is tax-deductible and hence there could be a tax clientele effect. Third, some portfolio managers may not be permitted to trade in synthetic stocks. Fourth, some investors may prefer to hold portfolios other than the S&P 500 index that can only be constructed from individual stocks. Market

¹⁰A liquidity premium (or discount) is the difference between the market clearing price of the security and the shadow price of the security in the absence of liquidity trading demands.

¹¹Appendix B develops an alternative model of arbitrage trading where informed traders, uninformed traders, and risk-neutral market makers face a clientele effect and arbitrageurs trade in both markets. The adverse selection problem causes the risk-neutral market makers to charge a liquidity premium proportional to the size of the trade (i.e., a linear price function). The alternative model yields the identical arbitrage trading results as the main model yields in Theorem 1.

¹²In other words, the liquidity premia are generated by the market microstructure effect of *inventory*, not by macroeconomic risk factors.

makers are assumed to face legal, geographical, or cost restrictions that prevent trading in both markets.

The third question addressed is: what are the effects of arbitrage trading on market volatility and other equilibrium properties? Specifically, this study examines the effects of arbitrage trading on market volatility, on the hedging effectiveness of synthetic securities (measured by the intermarket price correlation), and on the liquidity of the markets. A comparison is made between a restricted economy where arbitrage is prohibited and an unrestricted economy where arbitrage is permitted. It is shown that permitting arbitrage causes *no net change in volatility*, an increase in hedging effectiveness of synthetic securities, and an increase in the liquidity of the markets.

The fourth question addressed is: what happens to arbitrage trading if discretionary liquidity trading is allowed? Specifically, this article adds discretionary liquidity traders and discretionary market makers to the model and permits them to make a choice of which market to trade in while facing transaction costs. In the context of covered interest arbitrage, Deardorff (1979) and Callier (1981) showed that when both discretionary liquidity traders and potential arbitrageurs have the *same* transaction costs in each market, then a choice of market by discretionary liquidity traders always makes it unprofitable to engage in arbitrage trading.¹³ They concluded that observed deviations from interest rate parity can only be "explained" by transaction costs if the deviations fall within the narrower bounds of the discretionary liquidity trader's optimal strategy, rather than the wider bounds of the arbitrageurs optimal strategy. This study extends their analysis by considering what happens when arbitrageurs have *lower* round trip transaction costs than other traders. In contrast to the previous literature, this study shows that profitable arbitrage trading *can* take place even when discretionary liquidity traders can choose which market to trade. Further, this article shows that the bounds of the arbitrageur's optimal strategy are narrower than the discretionary liquidity trader's optimal strategy, but are nonbinding.

This study ignores execution and tracking risk.¹⁴ If the execution risks and tracking risks of a particular trade cause variation in the

¹³Deardorff (1979) coins the term "one-way arbitrage" for the choice of market by discretionary liquidity traders.

¹⁴Execution risk is the cost uncertainty due to such factors as nonsynchronous prices, execution delay, etc. Tracking risk is the payoff uncertainty due to such factors as stochastic dividends, reinvestment of cash flows before maturity at stochastic interest rates, etc. This study does not address so-called "risk arbitrage" such as speculating in merger targets.

size of the realized profit, but do not cause net losses, then it still fits within the strict definition of arbitrage.¹⁵ Also, according to James Pellicane, Jr. of Merrill Lynch Capital Markets, it is fairly standard for investment banking firms to offer clients arbitrage programs which guarantee relative execution prices and, hence, insure them against execution risks. In other words, the investment banking firm takes on the extra risk and builds the ex-ante cost of this insurance into their fees.

The genesis of this study is a model by Brennan and Schwartz (1988) of arbitrage between stocks and synthetic stocks. They *assume* that there exists a price difference between these two securities which follows a Brownian bridge process that goes to zero as the futures contract nears maturity. They consider the problem of an arbitrageur who incurs a transaction cost to initiate an arbitrage position, a smaller transaction cost to unwind the arbitrage position, and a position limit on the quantity of arbitrage.¹⁶ They numerically solve for the optimum strategy of the arbitrageur and for the economic value of the arbitrage opportunities. This research extends their work by constructing a model at a more fundamental level to derive an *endogenous* price difference.

The model developed by Grossman and Miller (1988) is also influential to the development of this study. In their model, market makers purchase (sell) shares from a liquidity trader who arrives on date 1 and then sell (purchase) shares to an offsetting liquidity trader who arrives on date 2. In doing so, market makers bear price risk from date 1 to date 2 for which they charge a liquidity premium on date 1. They also make the entry of individual market makers endogenous to the model.

Kumar and Seppi (1989) developed a model of index arbitrage based on heterogeneous information sets and lags in the transmission and processing of price information. In their model, market makers in futures receive contemporaneous signals about the value of *basket* securities such as stock index futures; market makers in stocks receive contemporaneous signals about the value of *individual* stocks; and both types of market makers receive *lagged* price information from other markets. Because they are conditioning on different information sets, market makers quote prices that permit arbitrage opportunities between

¹⁵Specifically, it fits within the definition of an *arbitrage opportunity of the second type* as defined on page 53 of Ingersoll (1987). It is a limited liability investment (meaning no net loss at maturity) with a current *negative* commitment (meaning a positive initial profit even if there is uncertainty in the size of realized profit). In other words, an investment that yields either \$1 or \$2 now and has zero payoff at maturity is still an arbitrage opportunity even though the size of the profit is random.

¹⁶A position limit is a limitation on the quantity of long or short arbitrage positions that the arbitrageur can hold at any point in time.

markets. Kumar and Seppi define arbitrageurs as traders who invest in technologies to monitor and to trade on inter-market prices *faster* than market makers.

It should be emphasized that the Kumar and Seppi model and the model in this article are *not mutually exclusive*. In other words, it is possible that part of the profits that arbitrageurs obtain is derived from monitoring and trading on inter-market prices faster than market makers (as in Kumar and Seppi) and part of their profits come from exploiting the price differences that arise due to differential liquidity trading and risk averse market makers in the two markets (as in this study). There is nothing that is logically inconsistent between the two stories.

The plan of this article is as follows. The basic model is developed in Section 2. Section 3 illustrates the competition between arbitrageurs and market makers, when entry is endogenous. Section 4 contains an analysis of the effects of arbitrage trading on market volatility and other equilibrium properties. Section 5 illustrates what happens when transaction costs are added and liquidity traders and market makers are permitted to make a choice of which market to trade in. Appendix A contains the proofs. Appendix B develops an alternative model of arbitrage trading where informed traders, uninformed traders, and risk-neutral market makers face a clientele effect and arbitrageurs trade in both markets.

2. THE MODEL

Liquidity Trading and Time-Series Market Making

The object here is to construct a single-period, two-date model with three assets and no transaction costs. One asset is a riskless bond that pays an interest rate r . The other two assets are risky securities that pay a liquidating dividend D_2 on date 2. Let S_t and Z_t be the cum-dividend price per share of a stock and of a synthetic, respectively, on date t , $t = 1, 2$. One obtains: $D_2 = S_2 = Z_2$.

Assumption 1: D_2 is normally distributed.

Assumption 2: There are two liquidity traders. One liquidity trader trades only in stocks and the other liquidity trader trades only in synthetics. Individual market makers can trade in only one market.

On date 1, investor LS receives a liquidity shock (modeled as an exogenous stochastic endowment) of L shares of stock. Then the

investor trades with market makers in stocks who offset most of the shock by taking the opposite position and holding it until date 2. The market makers in stocks are *time-series* market makers, in the sense that they supply liquidity to traders in only one market (stocks) and they bear risk on the nonzero inventory that they carry from date 1 to date 2. Simultaneously, investor LZ receives a liquidity shock of I shares of synthetic stock,¹⁷ and, therefore, trades with market makers in synthetics who offset most of the shock by taking the opposite position and holding it until date 2. Analogously, the market makers in synthetics are *time-series* market makers as well. Arbitrageurs may also trade on date 1. Their trading is described later. On date 2, all traders liquidate their holdings.¹⁸ Let W_t be the wealth of a trader on date, t , $t = 1, 2$.

Assumption 3: Liquidity traders and market makers have the same utility function, $U(W_2) = -e^{-\gamma W_2}$. Arbitrageurs have *any* von Neumann–Morganstern utility function so long as they prefer more to less.

Let liquidity traders and market makers in each market have the same initial endowments of θ shares of stock (or correspondingly synthetics) and B dollars in bonds. Consider the portfolio problem of liquidity trader LS who receives a liquidity shock of L shares on date 1. Let θ^{LS} shares and B^{LS} dollars be the after-trade holdings of stocks and bonds on date 1. On date 1, the expected utility of terminal wealth is maximized:

$$\max_{\theta^{LS}} E[-\exp\{-\gamma(W_1 + \theta^{LS}(\tilde{D}_2 - S_1) + rB^{LS})\}] \quad (1)$$

where $W_1 \equiv (\theta + L)S_1 + B$, and s.t.

$$B^{LS} \leq B - (\theta^{LS} - \theta - L)S_1 \quad (2)$$

Substituting in the objective function, evaluating the expectation using the normal distribution, and taking the derivative with respect to θ^{LS} yields the liquidity trader's demand:

$$\theta^{LS} = \frac{E[\tilde{D}_2] - (1 + r)S_1}{\gamma \text{Var}[\tilde{D}_2]} \quad (3)$$

A second group of traders in the stock market includes market makers who supply liquidity to the market. By assumption, they do not

¹⁷One share of synthetic stock is the quantity of stock futures and bonds that would replicate the payoffs and risk of one share of stock in a perfect market.

¹⁸Grossman and Miller (1988) have a date 3 as well. A date 3 could be added to this model to match the models more closely, but it is not essential for the purposes of this study.

receive a liquidity shock on date 1. The optimal demand of a market maker in stocks is $\theta^{MS} = \theta^{LS}$. The synthetics market has a similar set-up. Liquidity trader LZ receives a shock of I synthetic shares on date 1. The optimal demand θ^{LZ} is the same as in eq. (3) except for being in synthetics. By assumption, market makers in synthetics MZ do not receive a liquidity shock on date 1. Their optimal demand is $\theta^{MZ} = \theta^{LZ}$.

Cross-Sectional Market Making and Market Clearing

Define one long (short) arbitrage position as the purchase (sale) of one share of stock and the simultaneous sale (purchase) of one share of synthetics.¹⁹

Let θ^A be the total quantity of arbitrage positions taken by *all* arbitrageurs. For simplicity the arbitrageurs are restricted to trading only in arbitrage positions and in the riskless asset.²⁰

Let N^S , N^Z , and N^A be the number of market makers in stocks, market makers in synthetics, and arbitrageurs, respectively. Market clearing in stocks and in synthetics on date 1 requires that the sum of the optimal demands of all traders on date 1 equal the supply

$$\theta^{LS} + N^S \theta^{MS} + \theta^A = (1 + N^S)\theta + L \quad (4)$$

$$\theta^{LZ} + N^Z \theta^{MZ} - \theta^A = (1 + N^Z)\theta + I \quad (5)$$

Substituting the optimal demands into the respective market clearing equations and solving for the market clearing price of stocks and synthetics, one obtains:

$$S_1 = aE[D_2] - b \frac{[(1 + N^S)\theta + L - \theta^A]}{1 + N^S} \quad (6)$$

$$Z_1 = aE[D_2] - b \frac{[(1 + N^Z)\theta + I + \theta^A]}{1 + N^Z} \quad (7)$$

¹⁹The model does not distinguish between short-sales vs. regular sales from inventories. In practice, this distinction is very important because the NYSE tick rule constrains short-sales, but not regular sales. The "natural clientele" for short index arbitrage programs are traders who already own the underlying stocks, such as index funds, so that the index trades can be executed as quickly as regular sales.

²⁰In other words, they cannot place different size orders in the two risky asset markets and all arbitrage profits are invested in bonds.

where

$$a \equiv \frac{1}{1+r}$$

and

$$b \equiv \frac{\gamma \text{Var}[D_2]}{1+r}$$

Define $\hat{\varepsilon}$ as the price difference as a consequence of liquidity trading (with zero arbitrage). Using eqs. (6) and (7), one obtains:

$$\hat{\varepsilon} \equiv (Z_1 - S_1) |_{\theta^A=0} = b \left(\frac{L}{1+N^S} - \frac{I}{1+N^Z} \right) \quad (8)$$

Arbitrage Trading

Define ε as the price difference as a consequence of *both* liquidity trading and arbitrage trading. Using eqs. (6)–(8), one obtains:

$$\varepsilon = Z_1 - S_1 = \hat{\varepsilon} - c\theta^A \quad (9)$$

where

$$c = b \left(\frac{1}{1+N^Z} + \frac{1}{1+N^S} \right).$$

Let $\hat{\theta}^A$ be the quantity of arbitrage positions traded by an *individual* arbitrageur. Let $\bar{\theta}_1^A$ be the individual's *conjecture* of the average quantity of arbitrage positions traded by all other arbitrageurs. The *individual* arbitrageur must decide what quantity of arbitrage to trade. In a Nash equilibrium, each arbitrageur sets $\hat{\theta}^A$ to maximize individual profits while holding fixed the *conjecture* about $\bar{\theta}_1^A$. In equilibrium, these conjectures must be *correct*.

Let U^A be the utility function of the individual arbitrageur. The individual arbitrageur maximizes the expected utility of terminal wealth:

$$\max_{\hat{\theta}^A} E[U^A(\varepsilon \hat{\theta}^A (1+r))] \quad (10)$$

There is *no* budget constraint since there is no cost of increasing $\hat{\theta}^A$.

On date 1 L and I are known, so that eq. (10) reduces to maximizing the individual arbitrage profits. Substituting for ε from eq. (9), one obtains:

$$\max_{\hat{\theta}^A} \{ \hat{\varepsilon} - c(N^A - 1)\bar{\theta}^A - c\hat{\theta}^A \} \hat{\theta}^A \quad (11)$$

Taking the partial derivative of eq. (11) with respect to $\hat{\theta}^A$, one obtains the equilibrium arbitrage quantity $\hat{\theta}^{A*}$ as a function of the conjecture $\bar{\theta}^A$. Assuming that all arbitrageurs are identical, the only symmetric Nash equilibrium is when $\bar{\theta}^A = \hat{\theta}^A$. Substituting the correct conjecture, one obtains the following theorem for the equilibrium.

Theorem 1: Given Assumptions 1–3, the price difference ε , the equilibrium arbitrage quantity of each arbitrageur $\hat{\theta}^{A*}$, and the equilibrium profit of each arbitrageur $\hat{\pi}^A$ are

$$\begin{aligned}\varepsilon &= \frac{\hat{\varepsilon}}{1 + N^A} \\ \hat{\theta}^{A*} &= \frac{\hat{\varepsilon}}{(1 + N^A)c} \\ \hat{\pi}^A &= \frac{(\hat{\varepsilon})^2}{(1 + N^A)^2 c}\end{aligned}\tag{12}$$

There are a number of interesting comparative statics. $\hat{\varepsilon}$ is increasing in the differential liquidity shock $((L/(1 + N^S) - I/(1 + N^Z)))$. In turn, $|\varepsilon|$, $|\theta^A|$, and $\hat{\pi}^A$ are increasing in $|\hat{\varepsilon}|$. Intuitively, the differential liquidity shock scales the magnitude of the arbitrage opportunity. The price difference is increasing in the risk aversion of the traders γ and in the variance of the liquidating dividend $\text{Var}[D_2]$ and is decreasing in the riskless rate r . All three variables in Theorem 1 are decreasing in N^A . Intuitively, as N^A increases, the arbitrageurs behave more competitively. In the limit, as $N^A \rightarrow \infty$, all three variables go to zero.

Figure 1 shows that the individual arbitrageur's profit function, holding the behavior of others fixed, is a parabola. Beyond the apex of the parabola, an increase in the individual arbitrageur's quantity of arbitrage decreases the arbitrage profit because an increase in the quantity of arbitrage reduces the price difference. Indeed, a large enough quantity of arbitrage can drive the price difference to zero. Therefore, the profit maximizing arbitrage quantity is always finite, even without position limits.²¹

The model demonstrates two points. First, it shows how arbitrage trading can exist in a financial market equilibrium. Specifically, it shows that arbitrage trading is a consequence of the following assumptions: a clientele effect, risk-averse market makers, and arbitrageurs who

²¹This is also a feature of the Kumar and Seppi model for exactly the same reason.

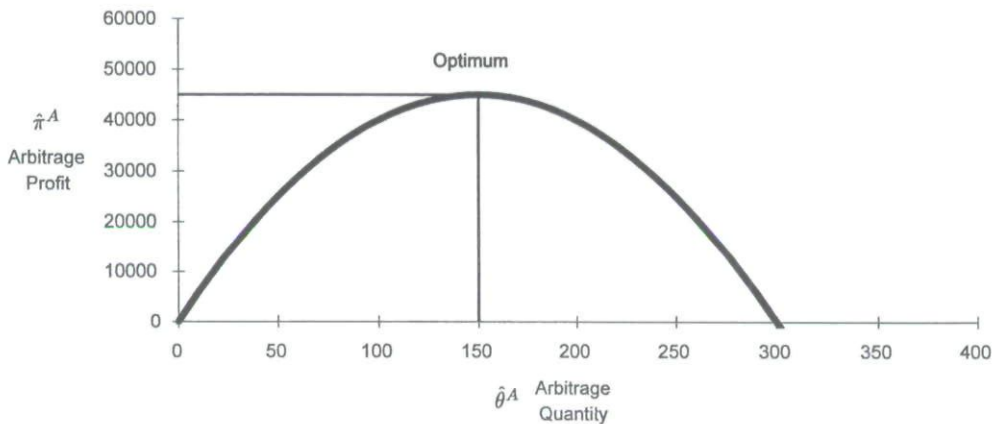


FIGURE 1.

The optimal arbitrage quantity. $\hat{\pi}^A$ as a function of $\hat{\theta}^A$, holding fixed the conjecture about $\bar{\theta}^A$, is a parabola. The optimal $\hat{\theta}^A$ is always finite because of the impact of $\hat{\theta}^A$ on ε .

have a cost advantage over the liquidity traders. Second, it shows that identical oligopolistic arbitrageurs do not drive arbitrage opportunities out of existence, but optimally exploit arbitrage opportunities in a symmetric Nash equilibrium. Thus, the price difference is reduced, but not eliminated.

3. COMPETITION TO PROVIDE MARKET MAKING SERVICES

To explain the role of arbitrageurs in the financial markets, a paradigm of arbitrageurs as *cross-sectional* market makers is developed. Arbitrageurs hold net zero inventories and obtain *riskless* profits by supplying liquidity to traders in different securities who arrive at the same time. This section shows the competition between arbitrageurs and market makers by expanding the model to make the entry of both types of agents *endogenous*.

First, the events on date 1 are enriched. Assume that individuals can make alternative investments costing C^A , C^S , or C^Z to become an arbitrageur, a market maker in stocks, or a market maker in synthetics, respectively. The following sequence of events takes place on date 1. First before the liquidity shocks are known, each individual must decide whether to become a market maker or an arbitrageur. Next, the liquidity shocks are realized. Finally, everyone trades.

Define $E_B[\cdot]$ as the expectations operator *before* the liquidity shocks are known. In equilibrium, the expected utility of wealth from entering W^E must be equal to the expected utility of wealth W^{NE} from not

entering. N^S , N^Z , and N^A are determined by the following entry indifference condition for each market:

$$E_B[U(\tilde{W}^E)] = E_B[U(\tilde{W}^{NE})] \tag{13}$$

To evaluate these equations, this study needs to make an explicit assumption about the ex-ante distribution of the dividends and liquidity shocks.

Assumption 4: The distributions of $\tilde{D}_2$, $\tilde{L}$, and $\tilde{I}$ are multivariate normal with zero means, except $E_B[D_2] > 0$ and correlations given by:

$$\text{Corr}(\tilde{L}, \tilde{I}) = \rho \quad \text{Corr}(\tilde{L}, \tilde{D}_2) = 0 \quad \text{Corr}(\tilde{I}, \tilde{D}_2) = 0$$

Substituting eq. (3), the results of Theorem 1, and Assumption 4 into eq. (13), canceling the common deterministic terms, and rearranging yields the following lemma.

Lemma 1: Given Assumptions 1–4, N^S , N^Z , and N^A are determined by the following system of simultaneous equations:

$$E_B[\exp\{-\gamma((\tilde{\theta}^{MS} - \theta)(\tilde{D}_2 - (1 + r)\tilde{S}_1) + \theta\tilde{D}_2)\}] = \exp\{-\gamma\hat{C}^S\} \tag{14}$$

$$E_B[\exp\{-\gamma((\tilde{\theta}^{MZ} - \theta)(\tilde{D}_2 - (1 + r)\tilde{Z}_1) + \theta\tilde{D}_2)\}] = \exp\{-\gamma\hat{C}^Z\} \tag{15}$$

$$E_B\left[U^A\left(\left(B - C^A + \frac{(\hat{\varepsilon})^2}{(1 + N^A)^2c}\right)(1 + r)\right)\right] = E_B[U^A(B(1 + r))] \tag{16}$$

where

$$\begin{aligned} \hat{C}^S &= C^S + \theta E[D_2] - \frac{\gamma}{2} \text{Var}[D_2], \\ \hat{C}^Z &= C^Z + \theta E[D_2] - \frac{\gamma}{2} \text{Var}[D_2] \end{aligned}$$

and where N^A , N^S , and N^Z are truncated to a nonnegative integer.

Note that the number of participants in each market depends upon the utility function of the arbitrageurs $U^A(\cdot)$ and this includes depending on their risk preferences. Intuitively, an individual must

tradeoff a *certain* investment cost C^A against an *uncertain* magnitude of arbitrage profit. Hence, the risk preferences of the arbitrageur matter to the endogenous-entry equilibrium even though the realized arbitrage opportunity will provide riskless profits.

Unfortunately, this system of simultaneous equations does not have a general analytic solution. However, it can be solved analytically in at least one special case. Consider the case of when the arbitrageur is risk neutral, $U^A(W) = W$, the risky security is a purely speculative security that is held in zero net supply, $\theta = 0$ [Grossman and Miller (1988) make this assumption throughout their article], and when the two markets are similar, meaning that $C^S = C^Z$ and $\text{Var}[L] = \text{Var}[I]$. The latter assumption implies that the number of market makers in each market is the same, $N^S = N^Z$. Substituting this special case into the entry indifference equations in Lemma 1 and solving the resulting system of simultaneous equations (see Appendix A), one obtains the following theorem:

Theorem 2: Given Assumptions 1–4 and $U^A(W) = W$, $\theta = 0$, $C^S = C^Z$, and $\text{Var}[L] = \text{Var}[I]$, the endogenous-entry equilibrium is characterized by:

$$\begin{aligned}
 N^S &= N^Z = f(C^A, C^S) \\
 N^A &= g(C^A, C^S) \\
 \frac{\partial N^S}{\partial C^A} &= \frac{\partial N^Z}{\partial C^A} > 0 \\
 \frac{\partial N^A}{\partial C^A} &< 0 \\
 \frac{\partial N^S}{\partial C^S} &= \frac{\partial N^Z}{\partial C^S} < 0 \\
 \frac{\partial N^A}{\partial C^S} &> 0 \\
 \frac{\partial |\varepsilon|}{\partial C^A} &> 0 \\
 \frac{\partial |\hat{\theta}^{A*}|}{\partial C^A} &> 0 \\
 \frac{\partial \hat{\pi}^A}{\partial C^A} &> 0
 \end{aligned}$$

$$\begin{aligned}\frac{\partial |\varepsilon|}{\partial C^S} &> 0 \\ \frac{\partial |\hat{\theta}^{A*}|}{\partial C^S} &> 0 \\ \frac{\partial \hat{\pi}^A}{\partial C^S} &> 0\end{aligned}$$

See Appendix A for the proof of this theorem and all remaining lemmas and theorems.

The key point this theorem highlights is that the number of participants in each market is a function of *both* C^A and C^S . A decrease in C^A , increases N^A and decreases N^S and N^Z . A decrease in C^S increases N^S and N^Z and decreases N^A . Hence, this theorem directly shows the cost competition between the *cross-sectional* market makers and the *time-series* market makers. Ultimately, liquidity traders benefit from the heightened cost competition that alternative technologies provide by obtaining the lowest cost market making services.

Further intuition about the paradigm comes from examining the cost extremes. In the limit as $C^A \rightarrow \infty$, $N^A \rightarrow 0$ and N^S and $N^Z \rightarrow$ an upper limit that would exist in each market where isolated.²² In the limit as $C^A \rightarrow 0$, $N^A \rightarrow \infty$ and N^S and $N^Z \rightarrow$ a lower limit that would exist in a single integrated market. Intuitively, as $N^A \rightarrow \infty$, their profits go to zero and hence they costlessly transfer demands from one market to the other. It is interesting that as C^A is varied from infinity down to zero, the model makes a smooth transition from completely segmented markets to completely integrated markets.

The second group of partial derivatives in Theorem 2 shows that the costs of entry C^A and C^S scale the magnitude of the arbitrage opportunity. Larger costs imply larger $|\varepsilon|$, $|\hat{\theta}^{A*}|$, and $\hat{\pi}^A$.

4. THE EFFECTS OF ARBITRAGE TRADING

The Setup

The question addressed in this section is: what is the effect of arbitrage trading on market volatility, on the hedging effectiveness of synthetic securities (as measured by the intermarket price correlation), and on

²²The upper limit is determined by imposing the nonnegative integer constraint and substituting $N^A = 0$ into eq. (29) in Appendix A.

market liquidity? Before describing these three properties in detail, two concepts must be defined. Define the *fundamental price of stocks*,²³ as its value *with zero liquidity shocks and zero arbitrage*, $\hat{S} \equiv S|_{(L=I=\theta^A=0)} = aE[D_2] - b\theta$. Note that $\hat{S}$ is not stochastic. Define the liquidity premium for stocks as $\phi^S \equiv \hat{S} - S$. When arbitrage connects the two markets, ϕ^S is a linear combination of L and I [see eq. (26) in the Appendix A].

The first effect analyzed is the volatility of prices. Volatility is measured by the variance of stock prices, which is related to the variance of liquidity premium in stocks as follows:

$$\text{Var}[S] = \text{Var}[\hat{S} + \phi^S] = \text{Var}[\phi^S] \quad (17)$$

where the second equality follows from the fact that $\hat{S}$ is not stochastic.

The second effect is the hedging effectiveness of synthetic securities as measured by the intermarket price correlation. In other words, how tightly the synthetic security tracks the underlying security. The intermarket price correlation depends on the liquidity premia as follows:

$$\begin{aligned} \text{Corr}[S, Z] &= \frac{\text{Cov}[(\hat{S} - \phi^S - E[S]), (\hat{Z} - \phi^Z - E[Z])]}{\sqrt{\text{Var}[\hat{S} - \phi^S - E[S]]} \sqrt{\text{Var}[\hat{Z} - \phi^Z - E[Z]]}} \\ &= \frac{E[\phi^S \phi^Z]}{\sqrt{\text{Var}[\phi^S]} \sqrt{\text{Var}[\phi^Z]}} \end{aligned} \quad (18)$$

where the second equality follows from $E[\phi^S] = E[\phi^Z] = 0$.

The third effect is the liquidity of the markets. Liquidity is defined as the slope of the price schedule. Equivalently, it is the order flow necessary to induce prices to rise by one dollar. Letting $\mathcal{L}^S$ denote the liquidity of the stock market and letting ω denote the order flow, one obtains:

$$\mathcal{L}^S \equiv \frac{\omega}{-\phi^S} = \frac{\frac{(L - \theta^A)}{1 + N^S} - L}{-\phi^S} = \frac{L}{\phi^S} - \frac{1}{b} \quad (19)$$

Partial Equilibrium Analysis

Consider the effects of exogenously increasing the number of arbitrageurs while holding constant the number of market makers.

²³The analysis in this entire section is done using the case of the stock market. A corresponding analysis with corresponding results can be applied to the synthetics market.

Theorem 3: Given Assumptions 1–4 and assuming similar markets (meaning $\text{Var}[L] = \text{Var}[I]$ and $C^S = C^Z$) when analyzing market volatility, an exogenous increase in the number of arbitrageurs while holding the number of market makers fixed (i) decreases $\text{Var}[S_1]$ and $\text{Var}[Z_1]$; (ii) increases $\text{Corr}[S, Z]$; and (iii) increases $E[\mathcal{L}^S]$ and $E[\mathcal{L}^Z]$.

This is in contrast to the Kumar and Seppi model where arbitrageurs exploit an information advantage at the expense of other traders and, thus, *reduce* expected liquidity.

Full Equilibrium Analysis

Next, consider the effects of arbitrage trading in full equilibrium, where the entry of individuals to become arbitrageurs and market makers is endogenous. This study assesses these effects by comparing a *restricted economy*, where arbitrage is banned by law, to an *unrestricted economy*, where arbitrage is permitted. Let the subscript $*$ denote a variable or expression in the restricted economy. In the restricted economy, $N_*^A = 0$ and N_*^S and N_*^Z are determined by the entry indifference equations (14) and (15) in Lemma 1. In the unrestricted economy, N^A , N^S , and N^Z are determined by the system of entry indifference equations in Lemma 1.

To compare the two economies, start from the restricted economy ($C^A = \infty$) and then consider lifting the prohibition against arbitrage to shift to the unrestricted economy ($C^A < \infty$). If $C^A < \infty$ is not too large, then N^A will increase from zero to a positive number. From Theorem 3, one knows that an increase in N^A while holding the number of market makers fixed decreases market volatility, increases hedging effectiveness, and increases market liquidity.

However, from Theorem 2, one knows that reducing C^A reduces the number of market makers. This reduction causes offsetting effects on volatility and other equilibrium properties. It is of interest, therefore, to examine the *net effect* of arbitrage trading in full equilibrium.

First, the variance of the liquidity premium in the two economies is compared in the following lemma (see Appendix A).

Lemma 2: Given Assumptions 1–4, one obtains:

$$\begin{aligned}\text{Var}[\phi^S] &= \text{Var}[\phi_*^S] \\ \text{Var}[\phi^Z] &= \text{Var}[\phi_*^Z]\end{aligned}\tag{20}$$

The intuition for this result is that the expected utility of *not entering* the unrestricted and of *not entering* the restricted economies are equal. Combined with eq. (13), one obtains:

$$E_B[U(\tilde{W}^E)] = E_B[U(\tilde{W}^{NE})] = E_B[U(\tilde{W}_*^{NE})] = E_B[U(\tilde{W}_*^E)] \quad (21)$$

Since the expected utility of *entering* is strictly a function of the variance of the liquidity premium, it follows that the variance of the liquidity premium must be the same in the unrestricted and restricted economies. In other words, individuals keep becoming market makers until this condition is satisfied. Lemma 2 leads to the following theorem (see Appendix A).

Theorem 4: Given Assumptions 1–4, the effects of arbitrage trading are shown by the following comparison of two economies in full equilibrium:

- (i) $\text{Var}[S_1] = \text{Var}[S_{1*}]$ and $\text{Var}[Z_1] = \text{Var}[Z_{1*}]$
- (ii) $\text{Corr}[S, Z] > \text{Corr}[S_*, Z_*]$
- (iii) $E[\mathcal{L}^S] > E[\mathcal{L}_*^S]$ and $E[\mathcal{L}^Z] > E[\mathcal{L}_*^Z]$

The intuition for these results is that permitting arbitrage has two offsetting effects. First, N^A goes from zero to a positive number which leads to the Theorem 3 effects. Second, as the arbitrageurs crowd out some of the (time-series) market makers, one obtains offsetting effects. The net effect on volatility is that permitting arbitrage causes *no net change in volatility*. Permitting arbitrage causes an increase in the hedging effectiveness of synthetic securities and an increase in the liquidity of the markets, even net crowding out market makers, because arbitrageurs *combine* the market making capacity of the two markets.

5. ARBITRAGE TRADING AND DISCRETIONARY LIQUIDITY TRADING

The question addressed in this section is: what happens to arbitrage trading if we allow discretionary liquidity traders? Specifically, what happens if one adds discretionary liquidity traders and discretionary market makers to the model and permits them to make a choice of which market to trade in while facing transaction costs? Although permitting the choice of which market to trade in, this study maintains the assumption of a clientele effect by allowing the discretionary liquidity traders and market makers in stocks to have *zero* transaction costs

to trade in stocks and transaction costs k to “switch” to synthetics. Similarly, discretionary liquidity traders and market makers in synthetics have zero transaction costs to trade in synthetics and transaction costs k to “switch” to stocks. Arbitrageurs are assumed to have a round-trip transaction cost k^A to initiate one long or short arbitrage position. The idea that arbitrageurs are the most efficient, low cost traders is modeled by specifying $k^A < k$.

The ability of traders to choose which market to trade in bounds the price difference ε to an interval determined by the cost of switching, $\varepsilon \in [-k, k]$ (all of the details for this section are given in Appendix A). The intuition for this result is easy to demonstrate. Suppose traders anticipate that $\varepsilon > k$. Then, the discretionary liquidity traders and market makers in stocks can switch some of their *selling* demand from stocks to synthetics to obtain a *higher* price net of transaction costs k . Alternatively, the discretionary liquidity traders and market makers in synthetics can switch some of their *buying* demand from synthetics to stocks to obtain a *lower* price net of k .

Correspondingly the price difference with zero arbitrage $\hat{\varepsilon}$ is bound to a larger interval $\hat{\varepsilon} \in [-\hat{k}, \hat{k}]$ where $\hat{k} \equiv k^A + k(1 + N^A)$.²⁴ The bounds of this larger interval are determined by the optimal arbitrage strategy. At the upper bound, when $\hat{\varepsilon} = \hat{k}$, adding optimal arbitrage strategy yields $\varepsilon = k$. Similarly, at the lower bound, $\hat{\varepsilon} = -\hat{k}$ corresponds with $\varepsilon = -k$.

Formally, the arbitrageur's problem with transaction costs is to maximize profits:

$$\begin{aligned} \text{for } \hat{\varepsilon} \geq 0 \quad & \max_{\hat{\theta}^A \geq 0} (\hat{\varepsilon} - c(N^A - 1)\bar{\theta}^A - c\hat{\theta}^A - k^A)\hat{\theta}^A \\ \text{for } \hat{\varepsilon} < 0 \quad & \max_{\hat{\theta}^A \leq 0} (\hat{\varepsilon} - c(N^A - 1)\bar{\theta}^A - c\hat{\theta}^A + k^A)\hat{\theta}^A \end{aligned}$$

Taking the partial derivative with respect to $\hat{\theta}^A$, substituting the correct conjecture $\bar{\theta}^A = \hat{\theta}^A$, and solving leads to the following theorem.

Theorem 5: Given that liquidity traders and market makers can switch at cost k and arbitrageurs have a round trip cost $k^A < k$, then the arbitrageur's optimal quantity of arbitrage $\hat{\theta}^{A*}$ and the price difference ε are:

²⁴This comes from solving for the value of $\hat{\varepsilon}$ in eq. (23) that causes $\varepsilon = k$.

$$\hat{\theta}^{A*} = \begin{cases} \frac{\hat{\varepsilon} - k^A}{(1 + N^A)c} & \text{if } \hat{\varepsilon} > k^A \\ 0 & \text{if } k^A \geq \hat{\varepsilon} \leq -k^A \\ \frac{\hat{\varepsilon} + k^A}{(1 + N^A)c} & \text{if } -k^A > \hat{\varepsilon} \end{cases} \quad (22)$$

$$\varepsilon = \begin{cases} \frac{\hat{\varepsilon} - k^A}{(1 + N^A)} & \text{if } \hat{\varepsilon} > k^A \\ \hat{\varepsilon} & \text{if } k^A \geq \hat{\varepsilon} \leq -k^A \\ \frac{\hat{\varepsilon} + k^A}{(1 + N^A)} & \text{if } -k^A > \hat{\varepsilon} \end{cases} \quad (23)$$

In contrast to Deardorff (1979) and Callier (1981), Theorem 5 shows that profitable arbitrage trading can take place even when discretionary liquidity traders and discretionary market makers can choose which market to trade in. Deardorff concluded that empirical deviations from interest rate parity can only be "explained" by transaction costs if the deviations fall within the bounds of the liquidity trader's optimal strategy, rather than the (wider) bounds of the arbitrageur's optimal strategy.

This study obtains the same conclusion but for different reasons. Here, the arbitrageurs' transaction cost k^A is less than the discretionary traders cost k . However, because the arbitrageurs are imperfectly competitive, they do *not* drive the price difference into the *narrower* bounds, $[-k^A, k^A]$.²⁵ Instead, the price difference is contained in the *wider* bounds of the discretionary traders, $\varepsilon \in [-k, k]$. In this sense, the narrower bounds of the arbitrageurs are *nonbinding*. This study obtains the same conclusion as Deardorff, that empirical deviations from a no-arbitrage parity condition can only be explained by the bounds of the discretionary traders $[-k, k]$. However, the reason is not that the bounds of the arbitrageurs are wider, but rather that they are nonbinding despite being narrower.

6. CONCLUSION

The idea of pricing assets by *no arbitrage* is so deeply ingrained in finance that it is difficult to entertain the idea of arbitrage trading in equilibrium. Yet, this study shows that arbitrage trading can exist in equilibrium. Nobody is surprised to find market makers in real security markets,

²⁵However, *perfectly* competitive arbitrageurs would drive the price difference to the interval $[-k^A, k^A]$ and make this narrower bound the relevant bound for explaining empirical deviations.

as opposed to Walrasian auctioneers. This study suggests that nobody should be surprised to find arbitrage traders engaged in cross-sectional market making.

APPENDIX A: PROOFS

Proof of Theorem 2

Substituting for θ^{MS} from eq. (3), for S_1 from the definition of $\hat{S}$ and ϕ^S , and $\theta = 0$ in the entry indifference equation for market makers in Lemma 1 and using the definitions of a and b , one obtains:

$$E_B \left[\exp \left\{ -\gamma \left(\frac{\tilde{\phi}^S}{b} (\tilde{D}_2 - E[\tilde{D}_2]) + (1+r)\tilde{\phi}^S \right) \right\} \right] = \exp\{-\gamma\hat{C}^S\} \quad (24)$$

Integrating with respect to $\tilde{D}_2$ and $\tilde{\phi}^S$ and solving for the variance of the liquidity premium, one obtains:

$$\text{Var}[\phi^S] = \frac{\exp\{2\gamma\hat{C}^S\} - 1}{2\gamma(1+r)b - \gamma^2\text{Var}[D]} \equiv \alpha(\hat{C}^S) \quad (25)$$

The liquidity premium for stocks is calculated by substituting from eqs. (8) and (12) as follows:

$$\begin{aligned} \phi^S &\equiv \hat{S} - S = \left(\frac{b}{1 + N^S} \right) (L - N^A \hat{\theta}^{A*}) \\ &= \left(\frac{b}{1 + N^S} \right) ((1 - c_L)L + c_I I) \end{aligned} \quad (26)$$

where

$$\begin{aligned} c_L &\equiv \frac{N^A b}{(1 + N^A)(1 + N^S)c} \\ c_I &\equiv \left(\frac{N^A b}{(1 + N^A)(1 + N^Z)c} \right) \end{aligned}$$

Substituting from eq. (9) and setting $N^S = N^Z$, one obtains:

$$c_L = c_I = \frac{N^A}{2(1 + N^A)} \quad (27)$$

Substituting $c_L = c_I$ in eq. (26) and using $\text{Var}[L] = \text{Var}[I]$ one obtains:

$$\text{Var}[\phi^S] = \left(\frac{b}{1 + N^S} \right)^2 [(1 - c_L)^2 + (c_L)^2 + 2(1 - c_L)c_L\rho] \text{Var}[L] \quad (28)$$

Equating the RHS of eqs. (25) and (28), substituting from eq. (27), solving for N^S , one obtains:

$$N^S = N^Z = \left(\left(\frac{b^2 \text{Var}[L]}{\alpha(\hat{C}^S)} \right) F(N^A) \right)^{\frac{1}{2}} - 1 \quad (29)$$

where

$$F(N^A) \equiv \frac{1}{2(1 + N^A)} + \frac{1}{2} + \frac{N^A \rho}{1 + N^A}$$

Substituting $U^A(W) = W$, $c = (2b/(1 + N^S))$, and $\hat{\varepsilon} = (b/(1 + N^S))(L - I)$ into eq. (16) and solving for N^A yields:

$$N^A = \left(\frac{b \text{Var}[L - I]}{2C^A(1 + N^S)} \right)^{\frac{1}{2}} - 1 \quad (30)$$

The proof is completed in two phases. Phase one is based on the special case of $\rho = 0$ and phase two is based on the alternative case of $\rho \neq 0$. First consider $\rho = 0$. Substituting eq. (30) and $\text{Var}[L - I] = \text{Var}[L] + \text{Var}[I] = 2\text{Var}[L]$ into eq. (29) and rearranging, yields a quadratic equation. Applying the quadratic formula, solving for N^S , and selecting the positive root yields:

$$N^S = N^Z = \frac{\gamma b h C^A}{4(e^{2\gamma \hat{C}^S} - 1)} + \frac{1}{2} \left[\left(\frac{\gamma b h C^A}{2(e^{2\gamma \hat{C}^S} - 1)} \right)^2 + \frac{2\gamma^2 b \text{Var}[L] h}{e^{2\gamma \hat{C}^S} - 1} \right]^{\frac{1}{2}} - 1 \quad (31)$$

where N^S and N^Z are truncated to a nonnegative integer and $h \equiv (2b(1 + r) - \gamma \text{Var}[D])$.

Substituting eq. (31) and $\text{Var}[L - I] = 2 \text{Var}[L]$ into eq. (30) yields:

$$N^A = \left[\frac{\text{Var}[L]}{\frac{\gamma h(C^A)^2}{4(e^{2\gamma\hat{C}^S} - 1)} + \frac{1}{2} \left[\left(\frac{\gamma h(C^A)^2}{2(e^{2\gamma\hat{C}^S} - 1)} \right)^2 + \frac{2\gamma^2 b \text{Var}[L] h(C^A)^2}{e^{2\gamma\hat{C}^S} - 1} \right]^{\frac{1}{2}}} \right]^{\frac{1}{2}} - 1 \quad (32)$$

where N^A is truncated to a nonnegative integer.

Now turn to phase two based on $\rho \neq 0$. Substitute $\text{Var}[L - I] = 2 \text{Var}[L](1 - \rho)$ into eq. (30). Add 1 to both sides of eqs. (29) and (30). Then square both sides of eq. (29) and take both sides of eq. (3) to the fourth power. Then use the RHS of eq. (29) to substitute out the $(1 + N^S)^2$ term in eq. (30). By rearranging eq. (30), one obtains a function $G(N^A)$ that implicitly defines N^A :

$$G(N^A) \equiv (1 + N^A)^2 \frac{F(N^A)}{(1 - \rho)^2} = \frac{\text{Var}[L] \alpha(\hat{C}^S)}{(C^A)^2} \quad (33)$$

where $\alpha(\hat{C}^S)$ comes from eq. (25). Note that the function $G(N^A)$ is monotonic in N^A and ρ . That is, $(\partial G(N^A))/(\partial N^A) > 0$ and $(\partial G(N^A))/(\partial \rho) > 0$. Fixing all of the parameters except for N^A and ρ and starting from the solution given above for the case of $\rho = 0$, increase (decrease) ρ from 0 to the specific parameter value of ρ and then decrease (increase) N^A by a corresponding amount such that $G(N^A)$ is held constant. This will satisfy the equation above and thus is an equilibrium value for N^A . Substitute this value for N^A into the RHS of eq. (29) to obtain the equilibrium value of N^S . The partial derivatives in Theorem 1 follow directly from eqs. (31) and (32) for $\rho = 0$ and follow by the implicit function theorem from eqs. (33) and (29) for $\rho \neq 0$.

Proof of Theorem 3

Take the partial derivative of eq. (28) with respect to N^A using the chain rule

$$\frac{\partial \text{Var}[\phi^S]}{\partial N^A} = \left(\frac{b}{1 + N^S} \right)^2 [-2(1 - \rho)(1 - 2c_L)] \text{Var}[L] \frac{\partial c_L}{\partial N^A} < 0$$

The final inequality follows from the fact the second multiplier is negative because c_L is strictly less than $\frac{1}{2}$ when $N^S = N^Z$, the first and third multipliers are positive by inspection, and the last multiplier is positive from the equation which defines c_L . Combining this with eq. (17) yields (i).

Define the following convenience variables

$$X(N^A) \equiv \frac{N^A b[(1 + N^S) + (1 + N^Z)]}{(1 + N^A)(1 + N^S)(1 + N^Z)c} = \frac{N^A}{(1 + N^A)}$$

$$d \equiv \frac{1 + N^Z}{(1 + N^S) + (1 + N^Z)}$$

$$V \equiv \frac{\text{Var}[I]}{\text{Var}[L]}$$

Substituting eq. (26) and these convenience variables into eq. (18), one obtains:

$$\text{Corr}[S, Z] = \frac{(1 - Xd)(Xd) + (1 - X(1 - d))(X(1 - d))V}{\sqrt{(1 - Xd)^2 + (X(1 - d))^2 V} \sqrt{(1 - X(1 - d))^2 V + (Xd)^2}} \quad (34)$$

Taking the derivative of the RHS of eq. (34) with respect to N^A , one obtains:

$$\frac{\partial \text{Corr}[S, Z]}{\partial N^A} = \frac{[(X - 1)^2(d^2V + (d - 1)^2) + dV(1 - d) + (1 - d)(2 - d)](1 - X)V \left(\frac{\partial X}{\partial N^A} \right)}{((1 - X(N^A)d)^2 + (X(N^A)(1 - d))^2 V)^{\frac{3}{2}} ((1 - X(N^A)(1 - d))^2 V + (X(N^A)d)^2)^{\frac{3}{2}}} \quad (35)$$

Over the ranges of the convenience variables of $X(N^A) \in (0, 1)$, $d \in (0, 1]$, and $V \in (0, +\infty)$, it is apparent by inspection that the sign of the RHS of eq. (35) is positive which yields (ii).

Substituting eq. (26) in eq. (19) and evaluating the expectation, one obtains:

$$E[\mathcal{E}^S] = \frac{1 + N^S}{b(1 - c_L)} - \frac{1}{b} \quad (36)$$

Taking the partial derivative eq. (36) with respect to N^A using the chain rule,

$$\frac{\partial E[\mathcal{E}^S]}{\partial N^A} = \left(\frac{1 + N^S}{b} \right) (-1) (1 - c_L)^{-2} (-1) \left(\frac{\partial c_L}{\partial N^A} \right) > 0$$

The final inequality follows from the fact that the first multiplier is positive. By inspection, the middle multiplier is positive because c_L is strictly less than 1, and the last multiplier is positive from the definition of c_L .

Proof of Lemma 2

Substituting for θ^{MS} from eq. (3) and for S_1 from the definition of $\hat{S}$ and eq. (26) in the entry indifference equation for market makers in Lemma 1 and integrating with respect to $\tilde{D}_2$, one obtains:

$$E[e^{A(\phi^S)^2 + B\phi^S + C}] = \exp\{-\gamma \hat{C}^S\} \quad (37)$$

where

$$\begin{aligned} A &\equiv \frac{\gamma^2}{2b^2} \text{Var}[D_2] - \frac{\gamma(1+r)}{b} \\ B &\equiv \frac{\gamma^2 \theta^2}{b} \text{Var}[D_2] - \theta \gamma (1+r) \\ C &\equiv \frac{\gamma^2 \theta^2}{2} \text{Var}[D_2] - \gamma(1+r) \theta a E[D_2] \end{aligned} \quad (38)$$

Evaluating the expectation, one obtains:

$$\frac{1}{\sqrt{1 - 2A \text{Var}[\phi^S]}} \left(\exp \left\{ \frac{B^2 \text{Var}[\phi^S]}{2(1 - 2A \text{Var}[\phi^S])} \right\} \right) (\exp\{C\}) = \exp\{-\gamma \hat{C}^S\} \quad (39)$$

Repeating this analysis for the restricted economy, one obtains the counterpart to eq. (39):

$$\frac{1}{\sqrt{1 - 2A \text{Var}[\phi_*^S]}} \left(\exp \left\{ \frac{B^2 \text{Var}[\phi_*^S]}{2(1 - 2A \text{Var}[\phi_*^S])} \right\} \right) (\exp\{C\}) = \exp\{-\gamma \hat{C}^S\} \quad (40)$$

where A , B , and C are the identical functions as before. Equating the LHS of eqs. (39) and (40), the lemma follows immediately.

Proof of Theorem 4

Combining eq. (17) with Lemma 2 yields (i) immediately. In the unrestricted economy, $c_L |_{N^A > 0} > 0$ and $c_I |_{N^A > 0} > 0$, which from eqs. (18) and (26) implies $\text{Corr}[S, Z] > 0$. In the restricted economy, $c_L |_{N^A = 0} = c_I |_{N^A = 0} = 0$, which implies $\text{Corr}[S^*, Z^*] = 0$. Hence, (ii) follows. Define the expected liquidity of the restricted economy as:

$$E[\mathcal{L}_*^S] = \frac{(1 + N^S |_{N^A = 0})}{b(1 - c_L |_{N^A = 0})} - \frac{1}{b} = \frac{1 + N_*^S}{b} - \frac{1}{b} \quad (41)$$

Comparing the expected liquidity in the two economies, one obtains:

$$E[\mathcal{L}^S] - E[\mathcal{L}_*^S] = \frac{1 + N^S}{b(1 - c_L)} - \frac{1 + N_*^S}{b} \quad (42)$$

The RHS of eq. (42) can be signed by substituting eq. (26), evaluated for both the restricted and the unrestricted economies in eq. (20), and rearranging, to obtain:

$$\begin{aligned} \left(\frac{1 + N_*^S}{b} \right) &= \left(\frac{1 + N^S}{b} \right) \left(\frac{1}{\sqrt{(1 - c_L)^2 + (c_I)^2 \frac{\text{Var}[I]}{\text{Var}[L]}}} \right) \\ &< \left(\frac{1 + N^S}{b} \right) \left(\frac{1}{1 - c_L} \right) \end{aligned} \quad (43)$$

where the inequality follows from the fact that c_I , $\text{Var}[I]$, and $\text{Var}[L]$ are all positive. The inequality in eq. (43) implies that the RHS of eq. (42) is positive and, hence, (iii) follows.

Proof of Theorem 5

Let ψ^i be the quantity of switching to trade in the other market by trader i , where $i = LS$ is the liquidity trader in stocks, $i = MS$ is a market maker in stocks, $i = LZ$ is the liquidity trader in synthetics, and $i = MZ$ is a market maker in synthetics. Proceed as in Section 2, except

that the amount of switching to trade in the other market is calculated last. Modify eq. (1) to obtain the objective function of trader LS :

$$\max_{\theta^{LS}} E[\exp\{-\gamma(W_1 + \theta^{LS}(\tilde{D}_2 - S_1) + \psi^{LS}(\tilde{D}_2 - Z_1 - I_S k) + rB^{LS})\}]$$

where

$$W_1 \equiv (\theta + L)S_1 + B$$

and s.t.

$$B^{LS} \leq B - (\theta^{LS} - \theta - L)S_1 - \psi^{LS}(Z_1 + I_S k)$$

and where I_S is an indicator variable defined as $I_S = 1$ when $\psi^{LS} > 0$ and $I_S = -1$ when $\psi^{LS} < 0$.

From the first order conditions, the optimal demands are:

$$\begin{aligned}\theta^{LS} &= \frac{E[\tilde{D}_2] - (1 + r)S_1}{\gamma \text{Var}[\tilde{D}_2]} - \psi^{LS} \\ \psi^{LS} &= \frac{E[\tilde{D}_2] - (1 + r)(Z_1 + I_S k)}{\gamma \text{Var}[\tilde{D}_2]} - \theta^{LS}\end{aligned}\quad (44)$$

Calculating the optimal demands for the other traders, plugging these demands into expanded market clearing equations, and solving for the market clearing prices, one obtains:

$$S_1 = aE[D_2] - b \frac{[(1 + N^1)\theta + L + \psi - \theta^A]}{1 + N^1} \quad (45)$$

$$Z_1 = aE[D_2] - b \frac{[(1 + N^2)\theta + I - \psi + \theta^A]}{1 + N^2} \quad (46)$$

where $\psi \equiv (\psi^{LS} + N^S \psi^{MS}) - (\psi^{LZ} + N^Z \psi^{MZ})$ is the total quantity of shares switched by all traders. Given that the cost k of switching is the same in both directions for all liquidity traders and market makers, then only the optimal total amount of switching ψ is determined, not the specific levels of ψ^{LS} , ψ^{MS} , ψ^{LZ} , and ψ^{MZ} .²⁶

²⁶In equilibrium when ψ is nonzero, it is shown later that ε is equal to k or $-k$. Hence, liquidity traders and market makers are indifferent as to how the total amount of switching is divided up. In other words, they are indifferent to the specific levels of ψ^{LS} , ψ^{MS} , ψ^{LZ} , and ψ^{MZ} .

Hence, the price difference with zero arbitrage is:

$$\begin{aligned}\hat{\varepsilon} &= (Z_1 - S_1) |_{\theta^A=0} = b \left(\frac{L + \psi}{1 + N^S} - \frac{I - \psi}{1 + N^Z} \right) \\ &\equiv \varepsilon^* + b \left(\frac{\psi}{1 + N^S} + \frac{\psi}{1 + N^Z} \right)\end{aligned}\quad (47)$$

where

$$\varepsilon^* \equiv b \left(\frac{L}{1 + N^S} - \frac{I}{1 + N^Z} \right)$$

The optimal strategy of the arbitrageur

$$\hat{\theta}^{A*} = \begin{cases} \frac{\hat{\varepsilon} - I_\varepsilon k^A}{(1 + N^A)c} & \text{if } |\hat{\varepsilon}| > k^A \\ 0 & \text{if } k^A \geq |\hat{\varepsilon}| \end{cases} \quad (48)$$

where I_ε is defined as an indicator variable that equals 1 when $\hat{\varepsilon} \geq 0$ and equals -1 when $\hat{\varepsilon} < 0$.

When $k^A \geq |\hat{\varepsilon}|$ it is possible to substitute eq. (47) into eq. (23) to obtain:

$$\begin{aligned}\varepsilon &= \frac{\varepsilon^* + b \left(\frac{\psi}{1 + N^S} + \frac{\psi}{1 + N^Z} \right) + I_\varepsilon k^A}{(1 + N^A)} \\ &\equiv \bar{\varepsilon} + \frac{b \left(\frac{\psi}{1 + N^S} + \frac{\psi}{1 + N^Z} \right)}{(1 + N^A)}\end{aligned}\quad (49)$$

where

$$\bar{\varepsilon} \equiv \frac{\varepsilon^* + I_\varepsilon k^A}{(1 + N^A)}$$

To determine ψ , there are three cases to consider. First, when $|\hat{\varepsilon}| \leq \hat{k}$, then for traders LS and MS (traders LZ and MZ) trading in stocks (synthetics) dominates trading in synthetics (stocks). Hence, $\psi = 0$. Second, when $\hat{\varepsilon} > \hat{k}$, then switching will take place until $\varepsilon = k$, such that there is no incentive for additional switching in equilibrium. Third, if $\hat{\varepsilon} < -\hat{k}$, then short switching will take place until $\varepsilon = -k$. Hence, total switching ψ is determined by evaluating the RHS of eq. (49) equal to $I_\varepsilon k$. Solving for ψ , one obtains:

$$\psi = \frac{-(\bar{\varepsilon} - I_\varepsilon k)(1 + N^A)(1 + N^S)(1 + N^Z)}{b[(1 + N^S) + (1 + N^Z)]} \quad (50)$$

APPENDIX B: AN ALTERNATIVE MODEL WITH INFORMED TRADERS AND ARBITRAGEURS

This appendix develops an alternative model of arbitrage trading where informed traders, uninformed traders, and risk-neutral market makers face a clientele effect and arbitrageurs trade in both markets. Start with a standard Kyle (1985) model (including his notation) in both markets. Let the superscript "*i*" stand for either stocks ($i = S$) or synthetics ($i = Z$). Let u^i denote the requested number of shares in the i th market by an *uninformed* trader in the i th market. Let x^i denote the requested number of shares in the i th market by an *informed* trader in the i th market. Let v be the terminal value of the stock and the synthetic. Assume the uninformed trader, informed trader, and market maker in the i th market are faced with a clientele effect that prohibits them from trading in the other market.

Assume all of the random variables are multivariate normally distributed with means, variances, and correlations given by:

$$u^S \sim N(0, \sigma_u^2)$$

$$u^Z \sim N(0, \sigma_u^2)$$

$$v \sim N(\bar{v}, \Sigma)$$

$$\text{Corr}(u^S, u^Z) = \rho$$

$$\text{Corr}(u^S, v) = 0$$

$$\text{Corr}(u^Z, v) = 0$$

Conjecture an arbitrage trading function of the form $\theta^A = k(u^Z - u^S)$, where k is a constant. Let y^i be the net uninformed trade (from both uninformed traders and arbitrageurs) in the i th market. Then one obtains:

$$y^S \equiv u^S + \theta^A = u^S + k(u^Z - u^S) = (1 - k)u^S + ku^Z$$

$$y^Z \equiv u^Z - \theta^A = u^Z - k(u^Z - u^S) = ku^S + (1 - k)u^Z$$

From the properties of normally distributed variables, one knows:

$$y^S \sim N(0, \sigma_y^2) \quad y^Z \sim N(0, \sigma_y^2)$$

$$\sigma_y^2 = [(1 - k)^2 + k^2 + 2(1 - k)k\rho]\sigma_u^2$$

Conjecture a linear price function in each market:

$$p^i = \mu^i + \lambda^i(x^i + y^i)$$

The informed trader in the i th market privately learns v and then maximizes:

$$\max_{x^i} E[x^i(v - p^i) | v] = \max_{x^i} x^i(v - \mu^i - \lambda^i(x^i + E[y^i | v]))$$

Evaluating $E[y^i | v] = 0$, taking the derivative w.r.t. x^i , and solving for x^i , one obtains:

$$x^i = \frac{v - \mu^i}{2\lambda^i} \equiv \beta^i(v - \mu^i)$$

where $\beta^i \equiv (1/(2\lambda^i))$ and the S.O.C. yields $\lambda^i > 0$.

Risk neutral, competitive market makers set the price equal to the conditional expectation of the terminal value given their information set:

$$p^i = E[v | x^i + y^i]$$

Calculating the conditional expectation and equating it to the conjectured form of the price function to verify the conjecture, one obtains:

$$\begin{aligned} E[v | x^i + y^i] &= \bar{v} + \left(\frac{\beta^i \Sigma}{(\beta^i)^2 \Sigma + \sigma_u^2} \right) (x^i + y^i - \beta^i(\bar{v} - \mu^i)) \\ &= \mu^i + \lambda^i(x^i + y^i) \end{aligned}$$

Solving for the unknown coefficients, one obtains:

$$\mu^i = \bar{v} \quad \text{and} \quad \lambda^i = \frac{\beta^i \Sigma}{(\beta^i)^2 \Sigma + \sigma_u^2}$$

Substituting for β^i and solving for λ^i , one obtains:

$$\lambda^i = \frac{1}{2} \sqrt{\frac{\Sigma}{\sigma_y^2}} \equiv \lambda$$

Having determined the price in each market, one can now calculate the price difference between the two markets:

$$\begin{aligned} \varepsilon &\equiv p^Z - p^S = \left[\bar{v} + \lambda \left(\left(\frac{1}{2\lambda} \right) (v - \bar{v}) + y^Z \right) \right] \\ &\quad - \left[\bar{v} + \lambda \left(\left(\frac{1}{2\lambda} \right) (v - \bar{v}) + y^S \right) \right] \\ &= \lambda(y^Z - y^S) = \lambda([u^Z - \theta^A] - [u^S + \theta^A]) \end{aligned}$$

By zeroing out the arbitrage trade (θ^A), one obtains the price difference with zero arbitrage:

$$\hat{\varepsilon} = \lambda(u^Z - u^S)$$

Substituting this back into the price difference equation, one obtains the same form as in eq. (9):

$$\varepsilon = \hat{\varepsilon} - c\theta^A$$

where $c \equiv 2\lambda$.

As before, the individual arbitrageur maximizes individual arbitrage profits:

$$\max_{\hat{\theta}^A} \{ \hat{\varepsilon} - c(N^A - 1)\bar{\theta}^A - c\hat{\theta}^A \} \hat{\theta}^A$$

Taking the partial derivative of the objective function with respect to $\hat{\theta}^A$, one obtains the equilibrium arbitrage quantity $\hat{\theta}^{A*}$ as a function of the conjecture $\bar{\theta}^A$. Assuming that all arbitrageurs are identical, the only symmetric Nash equilibrium is when $\bar{\theta}^A = \hat{\theta}^A$. Substituting the correct conjecture, one obtains the same equilibrium as in Theorem 1.

$$\begin{aligned} \varepsilon &= \frac{\hat{\varepsilon}}{1 + N^A} \\ \hat{\theta}^{A*} &= \frac{\hat{\varepsilon}}{(1 + N^A)c} \\ \hat{\pi}^A &= \frac{(\hat{\varepsilon})^2}{(1 + N^A)^2 c} \end{aligned}$$

Finally, equating the arbitrage trading function to its conjectured form, one verifies the conjecture:

$$\theta^A = N^A \left(\frac{\hat{\varepsilon}}{(1 + N^A)c} \right) = N^A \left(\frac{\lambda(u^Z - u^S)}{(1 + N^A)2\lambda} \right) = k(u^Z - u^S)$$

Solving for the unknown coefficient, one obtains:

$$k = \frac{N^A}{2(1 + N^A)}$$

which completes the analysis. Hence, this study shows that an alternative model of arbitrage trading, where informed traders, uninformed traders, and risk-neutral market makers face a clientele effect and arbitrageurs trade in both markets, can yield the identical arbitrage trading results as the main model yields in Theorem 1.

BIBLIOGRAPHY

- Amihud, Y., and Mendelson, H. (1980): "Dealership Markets: Market-Making with Inventory," *Journal of Financial Economics*, 8:31-53.
- Black, F. (1986): "Noise," *Journal of Finance*, 41:529-543.
- Brennan, M. J., and Schwartz, E. S. (1987): "Optimal Arbitrage Strategies under Basis Variability," in *Essays in Financial Economics*, Sarnat, M. (ed.), Amsterdam, The Netherlands: North Holland.
- Brennan, M. J., and Schwartz, E. S. (1988): "Arbitrage in Stock Index Futures," *Journal of Business*, 63:S7-S32.
- Callier, P. (1981): "One Way Arbitrage, Foreign Exchange and Securities Markets: A Note," *Journal of Finance*, 36:1177-1186.
- Deardorff, A. V. (1979): "One Way Arbitrage and Its Implications for the Foreign Exchange Markets," *Journal of Political Economy*, 87:351-364.
- Grossman, S. J. (1988): "An Analysis of the Implications for Stock and Futures Volatility of Program Trading and Dynamic Hedging Strategies," *Journal of Business*, 61:275-298.
- Grossman, S. J., and Miller, M. H. (1988): "Liquidity and Market Structure," *Journal of Finance*, 43:617-633.
- Holden, C. W. (1991, September/October): "Index Arbitrage and The Media," *Financial Analysts Journal*, pp. 8-9.
- Ingersoll, J. (1987): *The Theory of Financial Decision Making*. Totowa, NJ: Rowman and Littlefield.
- Kumar, P., and Seppi, D. J. (1989): "Information and Index Arbitrage," Working paper #88-89-90, Graduate School of Industrial Administration, Carnegie Mellon University.
- Kyle, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53:1315-1335.
- MacKinlay, C., and Ramaswamy, K. (1988): "Index-Futures Arbitrage and the Behavior of Stock Index Futures," *Review of Financial Studies*, 1:137-158.
- New York Stock Exchange. (7/88)-(12/93). *Program Trading Report*. New York, NY.
- The Presidential Task Force on Market Mechanisms (1988): *Report of The Presidential Task Force on Market Mechanisms*. Washington, D.C.: USPGO.
- Stoll, H. R., and Whaley, R. E. (1987): "Program Trading and Expiration-Day Effects," *Financial Analyst Journal*, 43:16-28.
- Titman, S. (1985): "The Effect of Forward Markets on the Debt-Equity Mix of Investor Portfolios and the Optimal Capital Structure of Firms," *Journal of Financial and Quantitative Analysis*, 20:19-27.

Copyright of Journal of Futures Markets is the property of John Wiley & Sons, Inc. / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.