THE UNIVERSITY OF CHICAGO

SELF-IMPOSED LIMITS TO ARBITRAGE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE GRADUATE SCHOOL OF BUSINESS
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
PHILIP Z. MAYMIN

CHICAGO, ILLINOIS
DECEMBER 2007

Dedicated to my family.

ABSTRACT

I document a multi-billion dollar discrepancy that lasted from 1992 to 1999 between two otherwise identical share classes of HSBC that did not suffer from the external limits to arbitrage that traditionally explain such other mispriced pairs as 3Com/Palm or Royal Dutch-Shell. Instead, I describe how self-imposed, internal limits to arbitrage such as restrictions on position size can result in persistent mispricings.

I also show that relatively more trading volume coincides with relatively lower prices in HSBC, and the same effects holds for 3Com/Palm, Royal Dutch-Shell, and other large mispriced pairs. An increase of one standard deviation in their relative volume coincides with a decrease of about one quarter of a standard deviation in their relative price. Self-imposed limits to arbitrage explain this phenomenon as well, so long as the more expensive security also tends to have greater daily volume: a higher relative price, or a wider discrepancy, leads to more arbitrage activity, and arbitrageurs trade equal volume in both securities, bringing the relative volume down closer to one.

Finally, I distinguish self-imposed limits to arbitrage from limits imposed as a result of market transactions costs or risk measures by calculating the market implied overall maximum position size of arbitrageurs for mispriced pairs spanning different time periods and countries and having different volume characteristics. The results are roughly constant at about one hundred days of typical trading volume, consistent with self-imposed limits.

ACKNOWLEDGMENTS

I thank my committee members Nicholas Barberis, Tobias Moskowitz, Ľuboš Pástor, and Richard Thaler for their detailed comments and continuous support. I also thank John Cochrane for comments on earlier versions of this paper.

I especially thank my committee chairman Richard Thaler. I am deeply grateful to him for his generous guidance, advice, insight, and encouragement.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

A doubly puzzling discrepancy between two identical but non-fungible share classes of HSBC lasted from 1992 to 1999 with the mispricing reaching highs of eight percent in each direction. Self-imposed limits to arbitrage explains and links both puzzles, with implications for other mispriced pairs.

Price discrepancies in pairs of securities typically fall into two broad classes: twins and stubs. Twins are dual-listing structures such as Royal Dutch-Shell in which two fundamentally identical share classes trade on different exchange in different countries. Typically neither listing faces short-selling constraints but noise trader risk in the form of home bias drives the discrepancy (de Jong et al. (2005); Froot and Dabora (1999)). Stubs, or equity carve-outs, particularly in the technology sector such as 3Com/Palm, arise when a parent stock is in the process of spinning off its publicly traded subsidiary in the same country and on the same exchange. The value of the "stub," or the parent stock less its holdings of the subsidiary, should be non-negative. Typically the discrepancy is short-lived and the high-priced subsidiary is impossible to borrow (Lamont and Thaler (2003)).

HSBC is a serendipitous example combining the best of both twins and stubs. The two share classes are pure twins, but listed in the same country, on the same exchange. Neither home bias nor short selling constraints apply. I document that none of the traditional, external limits to arbitrage bind in the case of HSBC, making it a rare example of a genuine, exploitable arbitrage opportunity, one that was in fact exploited: from 1997 to 1999 I was a trader at Long-Term Capital Management and one of our publicly unreported arbitrage positions in the fall of 1998 was a bet that the prices of the two HSBC share classes would converge.

1.1 Pricing Puzzle

The first puzzle is why the HSBC discrepancy existed at all, given that the traditional limits to arbitrage did not bind.

Over the seven years of the existence of the two HSBC share classes, noise traders committed £60 billion in capital to purchasing the more expensive share class (computed as the sum of the daily traded values of the more expensive share class), and it cost them £2 billion to do so, relative to the less expensive share class they could have purchased instead (computed as the sum of the product of the daily volume of the more expensive share class and its price premium to the other class).

Perhaps there are *self-imposed* limits to arbitrage? An arbitrageur might choose to limit his trading activity and overall position to certain percentages of the overall daily trading volume. For example, arbitrageurs might not want to participate in more than fifteen percent of daily volume or hold a position larger than five days average trading volume. These two limits are related through the amount of time it would take to liquidate a position in the market: five days of average volume liquidated at fifteen percent of daily volume takes 33 days to unwind.

I introduce a model of self-imposed limits to arbitrage, motivate it for behavioral, rational, and structural reasons, differentiate it from a simple reaction to market transactions costs and risk measures, explain why competition would not alleviate it, and develop its two key implications. One is that among arbitrage opportunities where one security in the pair tends to have both a higher price and higher volume, the volume discrepancy will be negatively related to the price discrepancy. The other is that the parameters of the model implied by market prices will tend to be relatively constant across different opportunities. The evidence supports both predictions.

Self-imposed limits to arbitrage can only be distinguished from the traditional, external limits to arbitrage when the latter do not bind, so HSBC is a touchstone for other discrepancies. Additionally, self-imposed limits to arbitrage may still be the effective restriction in other discrepancies where external limits to arbitrage do exist. For example, perhaps self-imposed limits apply to other twins and the home bias is a secondary effect. One way to test this possibility is to look at the volume predictions made by self-imposed limits to arbitrage.

1.2 Volume Puzzle

The second HSBC puzzle is the negative correspondence between the relative price (the natural logarithm of the ratio of the prices of the two share classes) and the relative volume (the natural logarithm of the ratio of the daily trading volumes in shares). When one share class's relative price was higher, its relative volume that same day was usually lower.

Furthermore, this same effect holds in 3Com/Palm, Royal Dutch-Shell, and two other large twins (Unilever and Reed-Elsevier), and the coefficients are approximately the same order of magnitude. In regressing the standardized relative price on the standardized relative shares volume, the coefficient is approximately -0.25, so when the relative volume is one standard deviation higher, the relative price is approximately one-quarter of a standard deviation lower. Standardization means demeaning, descaling, and detrending. Because the coefficient in a regression of standardized variables always equals the correlation, it is equivalent to say that the correlation between the standardized relative price and standardized relative volume is approximately -0.25.

Self-imposed limits to arbitrage explain this volume puzzle as well. Arbitrageurs essentially trade equal share volume for both securities in the pair, thus their trading activity tends to drive the relative volume towards one. If there are self-imposed limits to arbitrage, then arbitrageurs establish positions at a faster rate when the price discrepancy is wider, because they risk foregoing more profits if the opportunity converges before they have established their maximum position. Thus the increased trading of arbitrageurs at higher relative prices dampens the relative volume. This explanation requires the more expensive share class to also be the one with greater trading volume, as is typically the case in mispriced pairs, and it requires the securities to be relatively large stocks so that the maximum positions arbitrageurs choose to establish results in material profits relative to their capital.

Of the four other pairs mentioned above, all are comprised of large stocks. However, Reed-Elsevier has far smaller volume than the two other cross-country twins, Royal Dutch-Shell and Unilever. A model of self-imposed limits to arbitrage, calibrated using the mean and standard deviation of the relative volume of a given pair and the correlation between its standardized relative price and its standardized relative volume, implies a daily convergence probability of about one percent for HSBC, Royal Dutch-Shell, and Unilever, three percent for 3Com/Palm, which indeed was more likely to converge, and zero for Reed-Elsevier, which appears to have been largely ignored by arbitrageurs. Five other stubs like 3Com/Palm described by Lamont and Thaler (2003) are not large enough to attract arbitrageurs with self-imposed limits, each being an order of magnitude smaller than 3Com/Palm, and indeed the correlation of each between relative price and relative volume is not significantly different from zero.

CHAPTER 2

THE HSBC MULTIPLE SHARE CLASSES

HSBC's two London-listed share classes from 1992 to 1999 were each among the most highly traded and highly capitalized stocks of the entire UK stock market. Each was listed in the index with high weights and each had high average trading volume. They even traded next to each other in the same exchange. HSBC intended for the two to be identical in every respect, from dividends to voting rights to tax treatment. There was not even a multiplier between them: they were one-for-one equivalent. Each share class was easy to borrow and short. Analysts at the time recommended buying the cheaper one, saying that whoever was buying the more expensive share was "wasting the money." The only difference between the two share classes was the par value of the shares, a technical result of a merger with no effect on tax or other implications, other than rendering the two share classes non-fungible and thus preventing instantaneous riskless arbitrage.

Table 2.1 summarizes the two share classes. Price and volume data is from Bloomberg.[1]

2.1 History

On June 25, 1992, Hong Kong and Shanghai Bank (HSBC) acquired the entire outstanding share capital of Midland Bank for a combination of HSBC shares and cash or bonds. Prior to the acquisition, HSBC had been listed primarily in Hong Kong where its shares were denominated in Hong Kong dollars but had a secondary listing on the London Stock Exchange (LSE). These "Old" shares had a par value of HK$10

1. Two dates with obviously erroneous offer prices for the Old share (9/14/98 and 10/20/97) were removed, for the calculation of the average bid-offer spreads only.

Table 2.1: Summary of the Two HSBC Share Classes

Standard errors where appropriate are in brackets. Retail ownership was only available for 1992 and 1993. Turnover is annual shares traded divided by shares outstanding. Weekly beta to FTSE-100 is calculated by Bloomberg using weekly data. The Fama-French rows represent the risk loadings each share class exhibited on the three Fama-French (1993) benchmark portfolios: the value-weighted return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate (Mkt-Rf), the average excess return of small stocks to large stocks (SMB), and the average excess return of high value stocks to low value stocks (HML). The regression is run without allowing a free constant term.

	New	**Old**
Par Value	75GBp	HK$10
Index Membership (FTSE 100, All Share)	Yes	Yes
Retail Ownership by Individuals (1992)	14.36%	24.14%
Retail Ownership by Individuals (1993)	9.01%	20.05%
Average Bid-Offer Spread (basis points)	40.3 [0.7]	41.2 [0.7]
Average Daily Traded Value	£39,150,000	£23,870,000
Average Annual Turnover	320%	84%
Correlation of Log Price with Log Volume	-0.12	+0.33
Standard Deviation of Log Daily Returns	1.97% [0.03]	1.99% [0.03]
Weekly Beta to FTSE 100 Index	1.62 [0.09]	1.60 [0.09]
Fama-French: Loading on Mkt-Rf	1.33 [0.09]	1.38 [0.09]
Fama-French: Loading on SMB	0.60 [0.11]	0.69 [0.11]
Fama-French: Loading on HML	0.75 [0.13]	0.79 [0.14]

but traded in sterling on the LSE.

To consummate the acquisition and give the former Midland shareholders new HSBC shares, a new sterling-denominated share class was created. These "New" shares with a par value of 75p started trading on July 10, 1992 on the LSE. HSBC was granted dual primary listings in Hong Kong and in London. Because these New shares had different par values, they were not fungible with either the Hong Kong-listed or London-listed Old shares.

Soon after the acquisition, HSBC replaced the cash or bond element of the purchase price with freshly issued New shares. From then on, the New shares represented about one-third of the total number of shares and the Old shares two-thirds. HSBC became a UK tax resident by the end of 1992.

HSBC intended the two share classes to be identical.[2] The May 8, 1992 Listing Particulars of the New shares in connection with the acquisition of Midland Bank specifically states (HSBC, 1992, Paragraph 2):

> The new HSBC Holdings shares to be issued pursuant to the Offer will be issued credited as fully paid and will rank pari passu in all respects and have identical rights with the existing HSBC Holdings shares, including the right to receive in full all dividends and other distributions declared, made or paid hereafter, save for the recommended final dividend[3] of HK$1.31... in respect of the year ended 31 December 1991 payable on 26 May 1992 to holders on the register on 1 May 1992.

Even on matters that are not explicitly spelled out, the company made it clear that the two share classes were to be treated as equals (HSBC, 1992, Paragraph

2. There was one economic difference. The Old HK-par shares have an embedded conversion option in that holders of the London-listed shares can convert them into Hong Kong-listed shares and vice versa. However, such an option should certainly not be negative, and it is the Old, HK-par shares which typically traded at a discount to the New, GBP-par shares.

3. The "recommended final dividend" was a dividend for the preceding calendar year which the New shares could not enjoy because they had not yet been issued. From the date the New shares first traded, July 10, 1992, and on, the two shares received exactly the same dividends.

8(A)(a)):

> The existing HSBC Holdings shares of HK$10 each and the new HSBC Holdings shares of 75p each rank *pari passu* in all respects. Fully paid Ordinary Shares confer identical rights in respect of capital, dividends..., voting and otherwise notwithstanding that they are denominated in different currencies and shall be treated as if they are one single class of shares.

The Listing Particulars then immediately goes on with extraordinary detail to ensure that any new resolutions regarding rights issues, splits, reverse splits, cancellations, extraordinary dividends, and the like will only be allowed insofar as the "resolution shall affect all the Ordinary Shares in issue in like manner and to like extent."

According to Ellerton (1997), even the taxation treatment of the two shares do not differ:

> Some institutions appear to believe that there are dividend, tax, or currency complications, although this is not the case... Hong Kong and London registered shareholders can elect to receive the dividends in HK$s or in sterling; in either event, the dividend tax credit will come attached to the dividend cheque, whether or not it is of any economic benefit to the holder. The economic value of the tax credit depends entirely on the tax domicile and status of the shareholder, but has no bearing at all on the class of share.

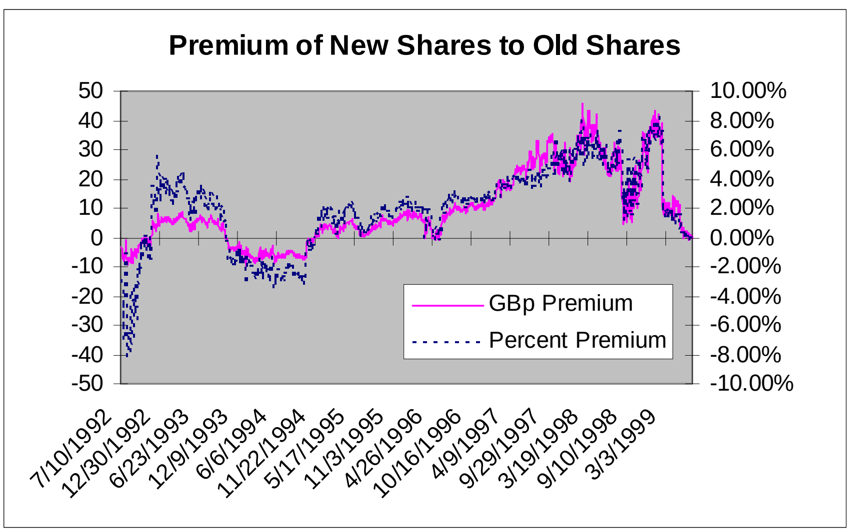
The two share classes finally merged at parity when HSBC decided to list its shares on the New York Stock Exchange (NYSE) in July 1999. The company issued new shares with par values in US dollars to replace each of the other share classes on a equivalent exchange ratio: each share, regardless of whether it was Old or New, received three dollar-denominated shares.

2.2 Pricing Discrepancy

The returns of the two share classes had identical sample properties. The correlation between the daily returns of the two share classes was 0.97. Table 2.1 lists for each share class the standard deviation of daily returns, the beta to the FTSE 100 index, and the risk loadings using the Fama-French three-factor model. All estimates are within one standard error across the two share classes.

Figure 2.1 shows the historical discrepancy between the Old and New shares as the premium of the New shares to the Old shares, both in pence and as a percentage of the concurrent Old share price.

Figure 2.1: HSBC Historical Premium of New Shares to Old Shares. The graph represents the historical premium that the New HSBC shares had over the Old HSBC shares. The solid line represents the premium in pence and uses the left axis; the dashed line represents the premium as a percentage of the Old share price and uses the right axis. The two peaks occur in February, 1998 and February, 1999.



The average absolute value of the percentage premium of the New shares to the Old shares was 2.89 percent, peaking on February 2, 1999 at GBp 42 or 8.26% of the value of the Old shares. The average absolute value of the pence premium was GBp

11.6, peaking on February 25, 1998 at GBp 45.7 or 8.01% of the value of the Old shares.

Given that the total market value of HSBC on each of those peak dates was approximately £40 billion according to Bloomberg data, a discrepancy on those dates suggests the firm could have been misvalued by as much as £3.2 billion, ranging from valuing the entire firm at the cheaper price to valuing the entire firm at the more expensive price.

Each shareholder who purchased a single stock of the New share class when it was at a premium could have received a higher dividend yield without the loss of fundamental exposure to the overall performance of the company by purchasing the Old share instead.

Various investment bank analysts issued reports in 1997 when the premium widened again after a three-year lull. This discrepancy was substantial enough, and visible enough, to cause reports to be written until the convergence in 1999.

Chris Ellerton of SBC Warburg wrote on February 14, 1997:

> The current 3.8% differential between the two classes of HSBC's share capital is close to the widest since 1993. Institutions buying the HSBC STG stock [New shares] in preference to the HK$10 shares [Old shares] are in effect sacrificing 1.8p of net dividend, equivalent to 15bp of yield... Institutions choosing to pay the 56p premium for the STG stock appear to be **wasting the money** and we recommend they buy the HK$ stock and enhance the yield. [emphasis added]

BZW Securities wrote on September 4, 1997:

> The two share listings, which are a relic of the November 1991 merger... are identical in dividends, rights, voting, and all other material matters. However, the shares are not fungible because the nominal "par" value of one listing is 10 Hong Kong dollars, while the other is 75 British pence. This difference is entirely nominal, and both listings trade in pence.

2.3 Volume Discrepancy

Table 2.1 lists summary statistics about the volume discrepancies. The New shares experienced average daily traded value about 60% greater than that of the Old shares, but, because there were about half as many New shares outstanding as Old shares, the New shares experienced turnover (annual shares traded divided by shares outstanding) nearly four times that of the Old shares.

The volume differential was often low in periods when the price discrepancy was widest.[4] In particular, during the two peak discrepancy times of February 1998 and February 1999, the two share classes had average trade volume within five percent of each other.

Ownership by insiders was not equal. Board members tended to hold the Old shares since that is what they were issued originally. Retail individual investors also tended to own a larger portion of the Old shares, though that data was only available for 1992 and 1993.

The average bid-offer spreads of the two share classes were within two standard errors of each other.

Figure 2.2 shows how well the relative price and the opposite relative shares volume move together. I use the opposite relative volume in the time series graph to make it is easier to spot comovement. The relative price, drawn in gray, seems to move almost in lockstep with the opposite relative volume, drawn in black: as Old volume rises relative to New volume, the New price rises relative to the Old price. Figure 2.3 shows the same data smoothed with monthly (21-day) moving averages where the effect is even more pronounced.

4. This is exactly what the self-imposed limits to arbitrage theory predicts: with the discrepancy at its widest, arbitrage activity would be at its peak, thus the relative volume would be driven closest to one.

Figure 2.2: HSBC Time Series of Log Price Ratio and Opposite Log Volume Ratio. The black line (left-hand axis) is the log of the ratio of the Old volume to the New volume. The gray line (right-hand axis) is the log of the ratio of the New price to the Old price. The opposite ratios are graphed to make the comovement easier to spot.

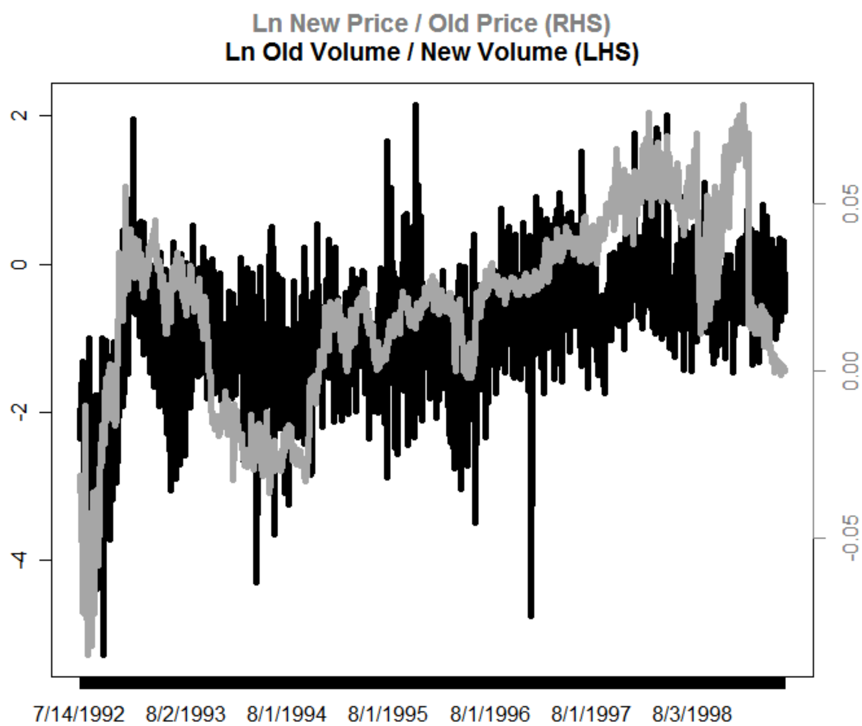
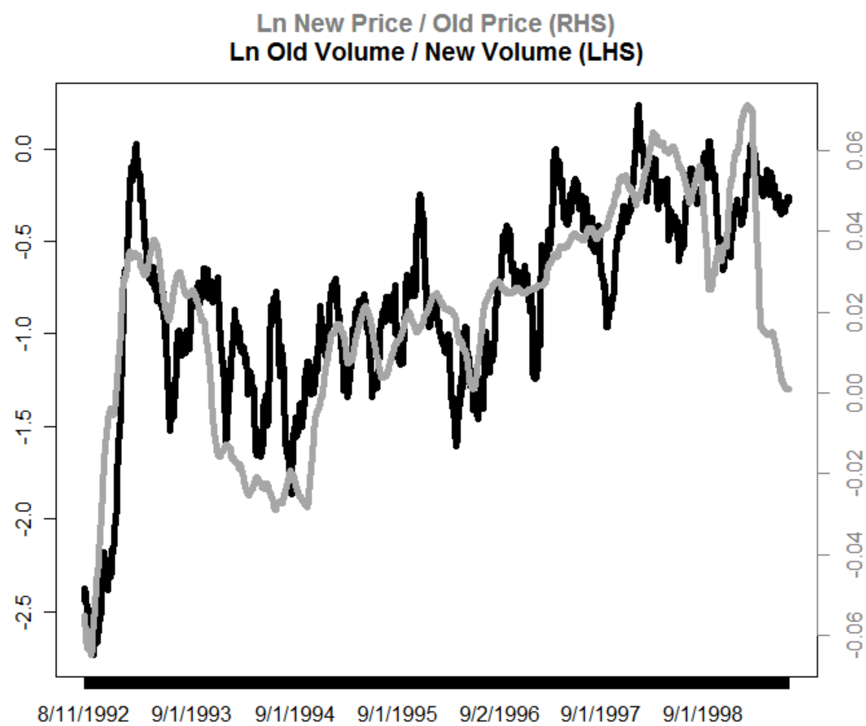
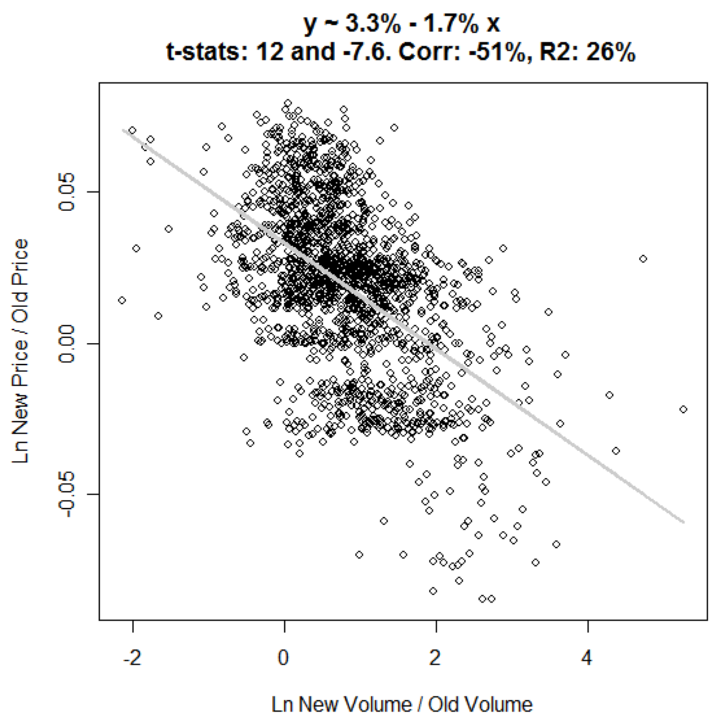


Figure 2.3: HSBC Time Series of Monthly Moving Averages of Log Price Ratio and Opposite Log Volume Ratio. The black line (left-hand axis) is the 21-day moving average of the log of the ratio of the Old volume to the New volume. The gray line (right-hand axis) is the 21-day moving average of the log of the ratio of the New price to the Old price. The opposite ratios are graphed to make the comovement easier to spot.



The scatterplot and regression in figure 2.4 of the relative price on the relative shares volume confirms the strong negative relation. The New price is at an average premium of 3.3% with a t-statistic of 12 when its volume matches that of the Old shares, and increases by another 1.7% with a t-statistic of 7.6 for each unit increase of the log ratio of the New volume to the Old volume. The correlation is -0.51. All reported t-statistics are corrected using Newey and West (1994) as implemented and described by Zeileis (2004).

Figure 2.4: HSBC Log Price Ratio and Log Shares Volume Ratio. The scatterplots show the relation between the log ratio of the New price to the Old price (y-axis) versus the log ratio of the New shares volume to the Old shares volume (x-axis). The regression results reported use non-parametric Newey-West t-statistics.



Replacing the relative shares volume, $\log \frac{V_{New}}{V_{Old}}$, with the relative shares turnover, $\log \frac{V_{New}/SO_{New}}{V_{Old}/SO_{Old}}$, makes no difference because the ratio of the amount of shares outstanding for each class differed by less than one percent throughout the seven-year period, where V_i and SO_i are the time series of daily shares volume traded and shares outstanding, respectively, of share class i.

Replacing the relative shares volume with the relative notional volume

$$\log \frac{P_{New} V_{New}}{P_{Old} V_{Old}} = \log \frac{P_{New}}{P_{Old}} + \log \frac{V_{New}}{V_{Old}}$$

where P_i is the price time series of share class i, makes essentially no difference in the case of HSBC because it simply rewrites the regression equation. If:

$$\log \frac{P_{New}}{P_{Old}} = \alpha + \beta \log \frac{V_{New}}{V_{Old}}$$

where $\alpha = 3.3\%$ and $\beta = -1.7\%$ as reported above, then by algebra:

$$\log \frac{P_{New}}{P_{Old}} = \frac{\alpha}{1+\beta} + \frac{\beta}{1+\beta} \log \frac{P_{New} V_{New}}{P_{Old} V_{Old}}$$

and because $1 + \beta = 98.3\%$ is so near one, the regression results will differ hardly at all.

Replacing both the log price ratio and the log shares volume ratio with their standardized counterparts does make a difference, halving the correlation. Table 2.2 summarizes the regression results.

An increase of one standard deviation in the relative volume coincides with an approximately one quarter of a standard deviation decrease in the relative price.

2.4 Transactions Costs

Because of taxes, not every arbitrageur could exploit this discrepancy. It required tax-free status to avoid capital gains tax and an ability to transact in over-the-counter

Table 2.2: Price-Volume Regressions for HSBC

This table shows four regressions measuring the correspondence between the relative price and the relative volume: $y = \alpha + \beta x + \epsilon$, where $y = \log \frac{P_{New}}{P_{Old}}$ is the relative price and $x = \log \frac{V_{New}}{V_{Old}}$ is the relative shares volume or $x = \log \frac{V_{New} P_{New}}{V_{Old} P_{Old}}$ is the relative notional volume (as specified in the first column). The third row demeans and descales, and the fourth row also detrends, y and x before running the regression. The correlation ρ between y and x, the R^2 of the regression, and Newey-West corrected t-statistics (in brackets) are also reported.

Volume	Standardized?	α	β	ρ	R^2
Shares	No	3.3% [+12.0]	−1.7% [−7.6]	−51%	26%
Notional	No	3.3% [+11.5]	−1.7% [−7.2]	−48%	24%
Shares	Demeaned and Descaled	0% [0.0]	−23% [−3.8]	−23%	5%
Shares	Demeaned, Descaled, and Detrended	0% [0.0]	−22% [−4.2]	−22%	5%

swaps with counterparties to avoid the stamp tax of 50 basis points charged on all stock purchases on the London Stock Exchange. Swaps also allow the leverage required to make the trade attractive from a return on capital perspective by requiring lower initial margins than Regulation T.

For arbitrageurs with top-tier credit, financing spreads were on the order of 50 basis points per year, comprised of the general collateral borrow fee of 30 basis points and a long funding cost of about 20 basis points.

Stock commissions were on the order of 10 basis points per side. The bid-offer spread was 40 basis points per share class, but trading in the spread rarely cost the full spread on both sides when putting on the position because an arbitrageur could place offers near the mid on one share class and, upon being filled, immediately pay half the bid-offer spread on the other to establish the position. Such a strategy would cost between 20 and 40 basis points total. Liquidation of the position, on the other hand, could incur as much as the full bid-offer spread of both shares and would cost between 40 and 80 basis points total. de Jong et al. (2005) cite commissions of 25 basis points per side and bid-offer spreads of 40 basis points in their study of 13 dual-listed companies but their estimates are based on "conversations with a large number of investment firms" and likely reflect averages for arbitrageurs. Top-tier arbitrageurs could negotiate lower commissions than the average. In addition, de Jong et al. (2005) follow Mitchell et al. (2002) in imposing special borrow rates for the shorts despite the fact that shares in twins are often available at general collateral rates. Mitchell et al. (2002) confirms that the average borrow fee for general collateral stocks, albeit in the U.S. market, was between 25 and 50 basis points.

Convergence can occur either through price movement, thus requiring costly liquidation in the market, or a structural change allowing fungibility, at which point the long shares could be returned to the lender of the borrowed shares.

Table 2.3: Round-Turn Transactions Costs for Arbitraging HSBC Discrepancy

Assuming the costs described in the text and that price convergence will occur within one year, this table lists the costs borne by an arbitrageur exempt from capital gains and stamp taxes in establishing, maintaining, and ultimately liquidating the position. (If structural convergence instead of price convergence occurs within one year, the liquidating costs would be zero.) All numbers are in basis points.

	Cost
Establishing Position: Brokerage Fees	20
Establishing Position: Bid-Offer Spreads	30
Maintaining Position: Financing Spread	50
Liquidating Position: Brokerage Fees	20
Liquidating Position: Bid-Offer Spreads	60
Total Costs	180

The round-turn costs assuming price convergence within one year are 180 basis points, as summarized in table 2.3. (The round-turn costs assuming structural convergence within one year would be 100 basis points.) Ex post, the price discrepancy converged to zero six times over the seven years, so an ex ante expectation of one year is not unreasonable. Furthermore, a one-year horizon is often used as a convenient rule of thumb in such ambiguous situations where even the distribution of convergence time is unknown. Such a rule of thumb may be rational, since arbitrageurs often receive incentive bonuses based on realized profit once a year, or it may be a behavioral framing bias in the same spirit as Benartzi and Thaler (1995).

Leverage on swaps of 10 times or more were common for top-tier arbitrageurs during this period. Assuming a position long £100 and short £100 required posting interest-bearing collateral worth £10, and assuming the arbitrageur allocated an additional £10 of risk capital to support the position, the return on capital to a top-tier arbitrageur for a one-year price convergence from a 5% initial discrepancy would be

$\frac{100}{20} \cdot (5.0\% - 1.8\%) = 16\%$[5]. de Jong et al. (2005) and Mitchell et al. (2002) use far more conservative leverage numbers by assuming Regulation T applies, under which initial margins allow only 2 times leverage. However, arbitrageurs actually use swaps and contracts-for-differences to exempt themselves from Regulation T, Regulation X, and similar restrictions on leverage.

The absolute value of the log ratio of the New share to the Old share exceeded 180 basis points more than 70% of the time during the seven year existence of the HSBC pricing discrepancy.

Is this an arbitrage opportunity or merely a risky spread trade within transactions costs bounds? Because the position needs to be maintained, it is not arbitrage in the textbook sense of riskless profit. But it is a standard example of a practical arbitrage opportunity, as first described by Shleifer and Vishny (1997), and as the term is commonly used in contexts such as statistical arbitrage, risk arbitrage, convertible bond arbitrage, etc. Furthermore, the transactions costs described above do not include possibilities such as arranging simultaneous block trades of both securities in establishing the position, continually hedging the spread on an equal notional basis and thus potentially earning extra income as the spread varies, or selling outperformance options on the spread to receive both upfront premium and costless entry should the premium widen to the target level.

5. This is conservative. At a 5% price discrepancy, an arbitrageur who bought £100 of the cheap security and sold £100 of the expensive security would have purchased more shares of the cheap stock than he sold of the expensive stock, and would receive the full dividend on the excess position of £5. The dividend yield averaged 3.5% so the arbitrageur would receive an extra 17.5 basis points per year from the excess dividends, partially offsetting his carry costs and thus increasing his profits further.

CHAPTER 3

SELF-IMPOSED LIMITS TO ARBITRAGE

"Limits to arbitrage" theories state that pricing discrepancies persist because of restrictions placed on arbitrageurs. Such restrictions include fundamental risk, an inability to raise additional capital in times of need (Shleifer and Vishny (1997)), transactions costs including the cost of carry (Dow and Gorton (1994)), short selling constraints (Lamont and Thaler (2003)), and noise trader risk (de Jong et al. (2005)). These restrictions are all external, beyond the control of the arbitrageur; the implication is that when arbitrageurs are not constrained by such external restrictions, they would immediately establish arbitrage positions large enough to eliminate the pricing discrepancy.

In practice, arbitrageurs often impose limitations on themselves beyond what would be prudent given market transactions costs and risk measures, and these limitations prevent them from immediately establishing enormous arbitrage positions. Some examples anecdotally common to various hedge funds are restricting position size to a certain portion of fund capital, e.g., five percent, restricting trading activity to a certain percentage of average daily trading volume, e.g. fifteen percent, and restricting overall position size to a certain number of average daily trading volume, e.g. five days. Other examples include country risk limits (World Bank (2005)) and defining maximum positions in single stocks or market segments (Drachter et al. (2006)).

In principle such restrictions could merely be a rational reaction to market transactions costs or risk measures. However, self-imposed limits distinguish themselves through *constancy*, by having round numbers that are constant across opportunities with differing transactions costs and risks. For example, some stocks are inherently riskier than others. Setting a limit of five percent of fund capital for any stock ignores these differences in risks. Therefore such a limit is more likely to be self-imposed.

These self-imposed restrictions could be *behavioral*. Arbitrageurs may impose these constraints on themselves as a self-control mechanism to prevent themselves from being carried away by the latest arbitrage opportunity. Such constraints may be imposed because of an agency problem within an organization conducting arbitrage or even within the same person (Thaler and Shefrin (1981)).

(Anecdotally, just such reasoning may be behind a variety of price-independent constraints that investment banks and brokerage houses impose on internal trading desks. Often such an institution will impose strict risk limits on certain products or categories, such as overall correlation or volatility exposure, regardless of the attractiveness of the opportunity. Indeed, it is these strict limits that trigger the investment banks and brokerage houses to approach hedge funds to offload some of the risk and participate in the arbitrage opportunity. In such situations, hedge funds and brokerages trade with each other but hold similar overall positions, a relatively common occurrence that would be far less common without the self-control constraints on position size on the banks and brokerages.)

These self-imposed restrictions could also be *rational* and *structural*. The rationale for the position restriction is to allow orderly and timely liquidation in case the position needs to be unwound. It is related to the trading volume restriction through the number of days it would take to liquidate. Five days of average trading volume would take 33 days of trading fifteen percent of daily volume to unwind. The number of days to liquidation is a constraint on the structure of the arbitrageur. For hedge funds, a substantial portion, e.g. 90 percent, of a client's assets are typically guaranteed to be returned within a relatively short time period, e.g. 30 days, with the remaining 10 percent held back until the year-end audit is completed (Joseph

G. Nicholas (2004)). Therefore, the position restriction is a combination of the trading restriction and the liquidity guarantees of the arbitrageur to its investors. Unlike assets under management, which can fluctuate based on opportunity (Shleifer and Vishny (1997)) and are therefore an external restriction, the liquidity guarantees are self-imposed constraints because they are fixed at fund formation.

Different arbitrageurs may have different motivations but regardless of whether they impose limits on themselves for reasons of self-control or investment mandate, so long as they set a maximum position size, they must determine how to optimally trade a price discrepancy. This is where the model of self-imposed limits to arbitrage links the limits themselves to the actions of the arbitrageur, with implications for both the price and volume discrepancies of the mispriced pair.

In particular, arbitrageurs will seek to minimize market impact on entry while maximizing the eventual position size and, ultimately, profit. Consider a pricing discrepancy of ten percent in which the marginal arbitrageur wishes to establish a position of five days volume. What would be his net entry price if he attempts to establish the entire position in one day? What would be his net entry price if he spreads his trading over the next thirty days?

Waiting to establish the position carries the risk that the discrepancy will become less attractive, but it also allows for the possibility that the discrepancy will widen. However, the trading restriction is not merely a type of noise trader risk. It is a self-imposed restriction that would exist even in the absence of noise trader risk. Suppose the future distribution of the pricing discrepancy is, absent any trades by the arbitrageur, constant at 10 percent with a small daily probability of collapsing to zero permanently. While noise traders must continue to exist to maintain the price discrepancy at 10 percent, there is no *risk* to the arbitrageur because the discrepancy will never widen. Nevertheless, so long as the price impact of trading five days worth of average trading volume in a single day substantially exceeds the cumulative price impact of establishing the same position over the subsequent month, and the likelihood of fundamental convergence over the subsequent month is relatively small, arbitrageurs will choose to impose this trading restriction on themselves, and the pricing discrepancy will persist. This argument is similar in spirit to Kondor (2006) in which convergence trading by arbitrageurs allows a price discrepancy to persist despite the absence of noise trader risk. In that model, a pricing discrepancy can be constant with a small probability of collapse, absent arbitrageur activity, but trading by arbitrageurs mitigates the pricing discrepancy while increasing its risk. In a sense that model is a type of self-imposed limit to arbitrage, but it makes no predictions about volume. There, arbitrageurs trade essentially infinite volume until the discrepancy drops to an equilibrium price still above zero but which has a positive probability of widening each day, at which point arbitrageurs are indifferent to further trading. Here, by contrast, I consider arbitrageurs who more smoothly manage their trading activity because of self-imposed constraints on position size and market impact costs of trading.

Formally, if the marginal arbitrageur has chosen to limit his overall position to N days of average trading volume, the current pricing discrepancy in dollars is D_0, the average market impact in dollars of trading N days worth of trading volume over t days is $f(N,t)$, the current expected value of the pricing discrepancy over the subsequent t days absent convergence is $E_0(D_t)$, and the constant daily probability of convergence is p, then an arbitrageur will choose t^* to maximize expected profit:

$$\max_t \sum_{s=1}^{t-1} \frac{s}{t} Np(1-p)^{s-1}\left(E_0(D_s) - f\left(\frac{s}{t}N, s\right)\right) + N(1-p)^t \left(E_0(D_t) - f(N,t)\right) \tag{3.1}$$

In cases where the probability of overnight convergence is zero and the expected discrepancy is constant, the solution is simply the value of t that minimizes $f(N,t)$. If waiting lowers average market impact, $\frac{\partial}{\partial t}f(N,t) < 0$, the arbitrageur will choose to trade over t^* days such that trades of size N/t^* experience the minimal, irreducible amount of market impact. In practice this amount is typically on the order of 1-5% of daily volume.

In cases where the probability of overnight convergence is one, the solution t^* is the value of t that maximizes the total instantaneous convergence profit:

$$\max_t \frac{N}{t}\left(D_0 - f\left(\frac{N}{t},1\right)\right) \tag{3.2}$$

Here, the arbitrageur is choosing t as a way to scale the one-day position. If the one-day market impact function is a step function in the size traded $f(x,1) = I_{x>x^*}D_0$, for example, representing existing bid and offer sizes with no further depth behind them, then the arbitrageur will choose $t^* = N/x^*$, trading all of the available size but not further. In practice this strategy applies to orders placed to clear the market on the close, where a marginal order impacts the average price of the entire trade. When intraday trading is available prior to the close such that further marginal trades do not impact previous trades, arbitrageurs will treat it as a sequence of mini-days, e.g. minutes, with a constant probability of instantaneous convergence as per equation 3.1.

In more general cases where the market impact function $f(N,t) = \sqrt{N/t}$ is the square root of the portion of average daily volume traded and $E_0(D_s) = D_0$ for all times s, the derivative of equation 3.1 can be set to zero to obtain an equation that can be solved numerically for the optimal t^* for given values of N, p, and D_0:

$$\frac{N}{(1-p)pt}\left((A(t,p)+C(t,p))\,D_0-(B(t,p)+C(t,p))\sqrt{\frac{N}{t}}\right)=0 \tag{3.3}$$

where:

$$A(t,p) \equiv 2\left(1-(1-p)^t\right)(1-p) \tag{3.4}$$

$$B(t,p) \equiv 3(1-p)-(1-p)^t\,(3+p\,(pt-3)) \tag{3.5}$$

$$C(t,p) \equiv 2(1-p)^t t\,(1+p(pt-1))\ln(1-p) \tag{3.6}$$

Would competition between arbitrageurs eliminate the self-imposed restriction on trading volume? A more aggressive arbitrageur could establish a larger position faster than his competitors if he is willing to sustain larger market impact. However, even the more aggressive arbitrageur will ultimately be limited by his self-imposed restriction on his overall position. What we should then observe when aggressive arbitrageurs enter the market is a drop in the pricing discrepancy as they establish their position with large price impact, followed by a period of lower average pricing discrepancy as the aggressive arbitrageur continues to build his position, followed by a sharp rise back as they attain their maximum position and cease trading. This cessation causes their market impact to dissipate and the less aggressive arbitrageurs do not fully replace the lost selling pressure. Exactly such a pattern appears in the last two years of the HSBC pricing discrepancy. Figure 2.1 shows that the HSBC discrepancy reached an all-time high in February, 1998, collapsed quickly to a low near parity in August, 1998 where it remained until November, 1998, then spiked back up to its highs in February, 1999.

What impact would self-imposed limits to arbitrage have on relative volume? Let

V_i be the daily shares volume of share class $i = 1, 2$, composed of a "standard" or "normal" volume S_i, i.e., the volume that share class i would have experienced had there been no activity by arbitrageurs, and k_i, the number of shares traded by arbitrageurs. As per above, arbitrageurs trade some percentage, $\theta = N/t^*$, of the standard volume of the less traded share class:

$$k \equiv k_1 = k_2 = \theta \min (S_1, S_2) \tag{3.7}$$

Note that k, the number of shares traded by arbitrageurs, is the same for each share class. This is exactly true for arbitrageurs establishing positions on an equal share basis and is approximately true for arbitrageurs establishing positions on an equal notional basis. Note also that arbitrageurs limit themselves to a certain portion of average or standard volume, not the actual volume of the day; this allows θ to exceed one if necessary. The alternative would implicitly cap the market impact of any trades once the arbitrageur were responsible for all of that day's volume whether the arbitrageur traded 100% or 200% or any other multiple of the average daily volume.

How does greater arbitrage activity affect the relative volume? In absolute terms, it drives the relative volume closer to one:

$$\lim_{k \to \infty} \frac{V_1 + k}{V_2 + k} = 1 \tag{3.8}$$

Whether greater arbitrage activity increases or decreases relative volume, however, depends on whether V_1/V_2 is above or below one. Because the discrepancy D_0 is assumed to be positive in equation 3.1 as arbitrageurs trade to drive it towards zero, we have implicitly assumed that $P_1 > P_2$, where P_i is the price of share class i. There are then two cases with respect to which share class has greater volume.

In the first case, the more expensive share class also has more standard volume, so $S_1 > S_2$. This is the case with HSBC and it is perhaps a reasonable case in general as it implies noise traders are purchasing the more expensive share class, and adding to its volume, rather than selling the less expensive share class. Let $S_1 = S_2(1+\chi)$ where $\chi > 0$ measures the excess standard volume, so that $V_1 = S_1 + \theta S_2 > S_2 + \theta S_2 = V_2$. Then:

$$\frac{V_1}{V_2} = \frac{S_1 + \theta S_2}{S_2 + \theta S_2} = \frac{S_1/S_2 + \theta}{1+\theta} = 1 + \frac{\chi}{1+\theta} \tag{3.9}$$

or:

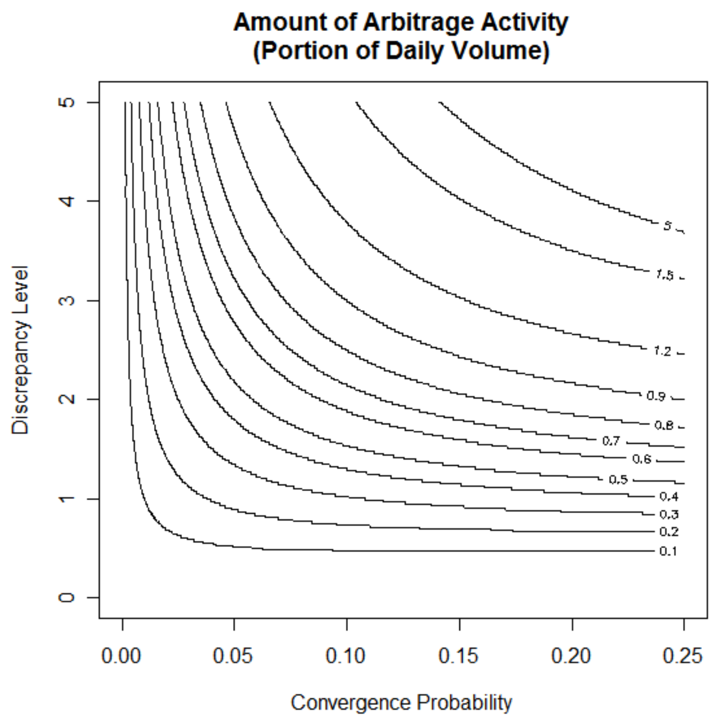
$$\ln \frac{V_1}{V_2} = \ln\left(1 + \frac{\chi}{1+\theta}\right) \approx \frac{\chi}{1+\theta} \tag{3.10}$$

So in the first case, the relative volume $\ln V_1/V_2$ decreases in the amount of arbitrage activity θ and increases in the excess normal volume χ. When the more expensive share class also typically has greater volume, then increased arbitrageur activity will drive the relative volume down towards one.

In the second case, the more expensive share class has less standard volume, so $S_1 < S_2$. Again let $S_1 = S_2(1+\chi)$ where now $\chi < 0$. Then equation 3.10 holds with the new, negative χ. So in the second case, the relative volume $\ln V_1/V_2$ increases in the amount of arbitrage activity θ and decreases in the absolute value of the excess normal volume $|\chi|$. When the more expensive share class typically has lower volume, then increased arbitrageur activity will drive the relative volume up towards one.

How does the amount of arbitrage activity θ vary with the relative price? Figure 3.1 displays the optimal amount of arbitrage activity θ for a variety of probabilities p and constant expected and current discrepancy values D, assuming $N = 5$, a common position limit set by hedge funds, and $f(N,t) = \sqrt{N/t}$, a commonly used market impact function that increases costs in the square root of the amount of arbitrage

Figure 3.1: Amount of Arbitrage Activity. This contour map shows the amount of arbitrage activity as a portion of the daily trading volume for a given discrepancy level and daily probability of convergence for an arbitrageur establishing a position of five days average trading volume and facing square-root market impact costs. The same essential shape appears for other position limits.



activity $\theta = N/t$. The same essential shape appears for other values of N.

For any given daily probability of convergence p, figure 3.1 shows that the amount of arbitrage activity θ always increases in the discrepancy level D. I have also tried $f(N,t) = \theta^{3/5}$, an improvement to the standard square-root function suggested by Almgren et al. (2005), with the same qualitative results.

The discrepancy level D and the market impact function $f(N,t)$ were defined in terms of dollars but in fact they can be in any units so long as they are consistent. For example, they could both be scaled to be expressed in terms of the percentage discrepancy. It is particularly useful to interpret both in terms of "sigmas," or units of standard deviation above the mean, because it obviates recalibration for each new mispriced pair.

With this interpretation, the discrepancy level D is the standardized relative price. How does it relate to the standardized relative volume? Figure 3.1 shows that θ increases for higher D for all probability levels. What is the effect of a higher θ on relative volume? Equation 3.10 shows that relative volume decreases for higher θ when $\chi > 0$. Furthermore, standardizing $\chi (1+\theta)^{-1}$ will not affect the sign of the relation but will render irrelevant the absolute value of the excess standard volume, $|\chi|$. So self-imposed limits to arbitrage predict a negative coefficient $b < 0$ in the linear relation between standardized relative price and standardized relative volume:

$$Z\left[\ln \frac{P_1}{P_2}\right] = a + bZ\left[\ln \frac{V_1}{V_2}\right] \tag{3.11}$$

where $Z[\cdot]$ is the standardization function that demeans, descales, and detrends its input. Detrending a variable means regressing it on time and then subtracting the product of the time and the resultant regression coefficient from the original variable. Detrending cleans empirical data of any time dependence but does not affect the simulated model results. I use the phrase "standardized" by itself in the text as a shorthand to mean a variable has been demeaned, descaled (divided by its standard deviation), and detrended.

This negative volume-price link holds only when the two share classes have substantially different volume and the more expensive share class tends to have more volume, so that $\chi > 0$, as is the case with HSBC. In principle, the relation could be zero if the two share classes have approximately equal volume, so that $\chi = 0$, and it could even be positive in the unlikely case the more expensive share class has less volume, so that $\chi < 0$.

At this point we have seen that self-imposed limits to arbitrage, specifically restrictions on overall position size and trading activity, allow, first, a pricing discrepancy to persist and, second, predict that the standardized relative price will have a negative relation with the standardized relative volume, so long as the more expensive share class also has greater volume.

We can go further and calculate both the market implied convergence probability and the amount of arbitrage activity over time, assuming for the remainder of this section, as is the case for HSBC, that $\chi > 0$. The first step is to calibrate the value of the regression coefficient b. As above, interpret the discrepancy level D as the standardized relative price:

$$D = Z\left[\ln \frac{P_1}{P_2}\right] \tag{3.12}$$

Substitute from equation 3.10 to obtain:

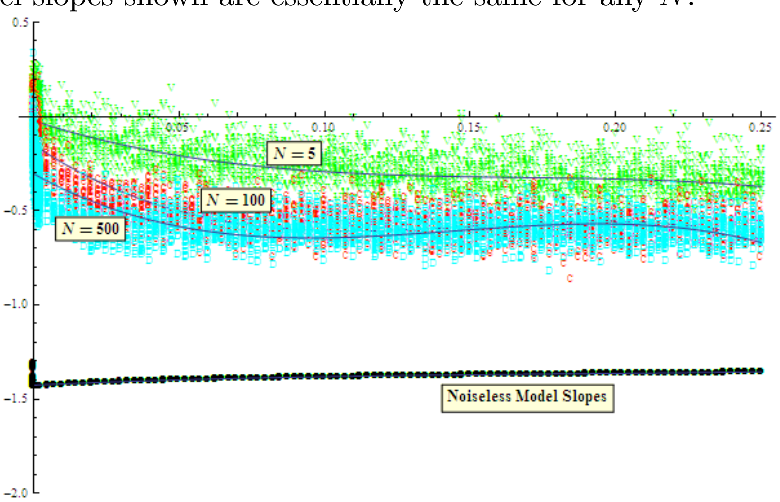
$$D = a + bZ\left[\ln\left(1 + \frac{\chi}{1+\theta}\right)\right] \approx a + bZ\left[\ln\left(1 + \frac{1}{1+\theta}\right)\right] \tag{3.13}$$

with the last approximation holding because of the effect of standardization.

The bottom of figure 3.2 shows the ordinary least squares (OLS) slopes of the regressions of D on the standardized log of $1+(1+\theta)^{-1}$ for each of the 50 probability values of p from one-half of a basis point through 25 basis points in steps of one-half of a basis point and for each of the 99 probability values of p from 50 basis points through 25 percent in steps of 25 basis points, each having 200 values of D ranging from 0.025 to 5 in steps of 0.025. The smaller, more important values of p are sampled more frequently.

The noiseless model slope has a mean and a median of -141% and lies in a tight range between -145% and -139%, essentially independent of the level of the conver-

Figure 3.2: Noisy and Noiseless Model Slopes. The top part of this chart plots the noisy model slope b in the OLS regression of the discrepancy level D on the standardized natural log of one plus the ratio of the excess standard volume χ to one plus the amount of arbitrage activity θ in the equation: $D = a + bZ\left[\ln\left(1 + \chi\left(1+\theta\right)^{-1}\right)\right] + \varepsilon$ where χ is distributed lognormally with mean $\mu = 1.01$ and standard deviation $\sigma = 0.81$ to mimic the distribution of HSBC's volume ratio V_{New}/V_{Old} during those periods when its price ratio does not exceed the transactions costs bounds, and $\theta = N/t^*$, with t^* being the solution to equation 3.1 for an arbitrageur establishing a position of N days of typical trading volume and facing square-root market impact costs, where $N = 5$ for the topmost plot (shown in green, marked by the Roman numeral V), $N = 100$ for the second plot (shown in red, marked by the Roman numeral C), and $N = 500$ for the third plot (shown in cyan, marked by the Roman numeral D). Those three plots are each overlayed with the best-fit cubic. The bottom part of this chart plots the noiseless model slopes using the same method but with χ set to one. The noiseless model slopes shown are essentially the same for any N.



gence probability. But from table 2.2, the actual slope between standardized relative price and standardized relative volume for HSBC is -23%. Why the difference?

We have assumed that the excess standard volume χ is constant but of course there is daily noise. Adding noise to χ will reduce the slope. Though the standard volumes S_1 and S_2 are unobservable, as a first approximation, we can let the standard volume ratio $S_1/S_2 = 1 + \chi$ mimic the distribution of the total volume ratio V_1/V_2 during those times when the relative price $\ln(P_1/P_2)$ is within the transactions costs bounds, the presumption being that arbitrageurs do not trade within the transactions costs bounds.

For HSBC, the log volume ratio $\ln(V_{New}/V_{Old})$, has mean 1.01 and standard deviation 0.81 when the absolute value of the log price ratio $\ln(P_{New}/P_{Old})$ is less than 180 basis points, the critical level of transactions costs documented in section 2.4. The top part of figure 3.2 shows the OLS slope estimate of b computed in the regression of equation 3.13 between the standardized relative price D and the standardized log of $1+\chi(1+\theta)^{-1}$ for the same values of p and D as for the noiseless model slopes described above, where $N = 5$ and $1 + \chi$ is drawn from a lognormal distribution with log mean $\mu = 1.01$ and log standard deviation $\sigma = 0.81$. The figure also plots a fitted cubic. Furthermore, the top part of figure 3.2 also plots the results for $N = 100$ and $N = 500$.

Unlike the noiseless model slopes of the bottom of figure 3.2, the noisy model slopes of the top of figure 3.2 vary with the convergence probability. From table 2.2, HSBC exhibits an actual slope of -0.22 between standardized relative price and standardized relative volume. we can calculate what convergence probability in the model corresponds to that slope by using the fitted cubic. For $N = 5$, the implied convergence probability is 6%. For $N = 100$, it is 0.7%. For $N = 500$, it is zero.

It is impossible to know what ex ante convergence probability arbitrageurs assumed for the HSBC mispricing. As an estimate, we can determine the daily convergence probability p that would give a 50% cumulative probability of convergence within, say, six months, or 126 business days:

$$(1-p)^{126} = 0.50 \Rightarrow p = 0.55\% \tag{3.14}$$

Ex post, the discrepancy crossed parity after exceeding transactions costs seven times during the seven year period. Table 3.1 lists the seven periods. The average number of business days of each period is 232 days, or a little less than one year. If 232 days represented the expected median of the distribution of the duration of the trade for arbitrageurs, then by the same method as above the implied daily convergence probability would be 30 basis points. If instead we use only the average of the first six periods (146 business days), then the implied probability is 47 basis points. However, arbitrageurs faced with self-imposed limits would not establish their maximum position on the first day of each period, so perhaps the better estimate using ex post data would be to assume that their position would be established gradually, and on average would be at the halfway point of the period. This adjustment essentially doubles the earlier probability estimates, so using the average of all seven periods suggests a daily convergence probability of 60 basis points while using the average of just the first six periods, perhaps the better estimate, suggests a daily convergence probability of 95 basis points.

A daily convergence probability of 95 basis points is close to the 70 basis points implied by the model for $N = 100$. In fact it equals the model implied probability for $N = 77$.

How does $N = 77$ or $N = 100$ relate to the informal statement that arbitrageurs impose something like $N = 5$ as their position limit? Suppose there are 20 arbi-

Table 3.1: Length of HSBC Mispricings

This table lists the number of business days in each period that starts when the absolute value of the HSBC discrepancy first exceeds the transactions costs bounds and ends when the discrepancy crosses parity.

Period Start	**Period End**	**Business Days**
7/14/1992	10/23/1992	73
11/30/1992	10/18/1993	223
11/23/1993	11/17/1994	250
12/13/1994	3/2/1995	55
3/16/1995	3/27/1996	262
4/11/1996	4/29/1996	13
6/12/1996	5/26/1999	747

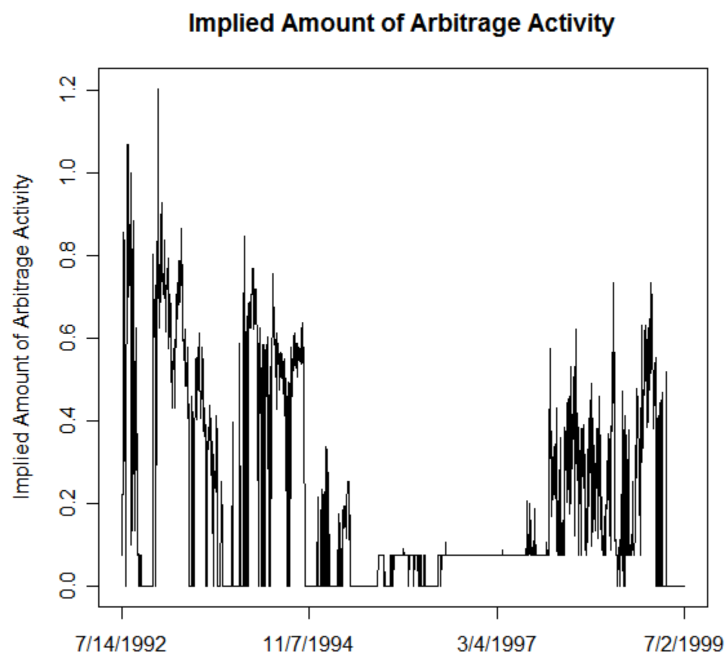
trageurs trading the pair. Because transactions costs and funding costs have to be minimized, this would likely just be the big banks and brokerages and the largest few of the hedge funds. Each of them knows the others are also trading so in their analysis of market impact and what portion to trade, and from the point of view of the pair itself, the equation being solved is how quickly to trade the total, i.e. the market impact assumes $N = 100$, rather than 20 separate market impacts of $N = 5$.

Using $p = 0.95\%$ as our estimate of ex ante daily convergence probability and $N = 77$ as our estimate of the total position size by arbitrageurs facing square-root market impact costs, figure 3.3 plots the time series of the implied amount of arbitrage activity θ by solving equation 3.1 for each day of the HSBC discrepancy where the current and expected future discrepancy is set to the absolute value of the standardized relative price.

The model suggests that arbitrageur activity amounted to 20 percent of daily volume on average during the seven years the HSBC discrepancy existed, with typical peaks in the range of 60 to 80 percent and some as high as 100 to 120 percent.

Self-imposed limits to arbitrage predicts both that the pricing discrepancy will

Figure 3.3: Amount of Arbitrage Activity in HSBC. This figure shows the time series of the implied amount of arbitrage activity calculated from the optimal solution to equation 3.1 for an arbitrageur establishing a maximum position of 77 days average trading volume and facing square-root market impact costs, assuming the daily convergence probability is 0.95%, and using the absolute value of HSBC's standardized relative price as the measure of the current and expected future discrepancy level. The arbitrage activity is identically zero when the discrepancy is below the transactions costs threshold of 1.8%.



persist and that the amount of arbitrage activity increases with relative price and decreases with relative volume, thus providing a link between the two puzzles. In short, arbitrageurs trade more to establish their maximum positions faster when the discrepancy between the two securities is wider, thus equalizing the relative volume between them.

CHAPTER 4

OTHER MISPRICED PAIRS

HSBC is a unique discrepancy because it has all the advantages of twins and stubs. But do self-imposed limits to arbitrage help explain other pairs, even those that are otherwise considered to have been allowed to persist only due to external limits to arbitrage?

4.1 3Com/Palm

The history of the 3Com/Palm discrepancy is from Lamont and Thaler (2003): Palm was a wholly-owned subsidiary of 3Com until March 2, 2000 when 3Com sold a fraction of its stake, intending to spin-off the remaining shares within a year once it received a favorable tax ruling. The expected approval came May 8, the discrepancy vanished, and 3Com indeed spun-off its remaining shares of Palm on July 27. Figure 4.1 shows how the relative price of 3Com to Palm seems to comove with the opposite relative volume. Because the 3Com/Palm discrepancy lasted for only about one hundred business days, I do not show moving averages. Unlike HSBC's New and Old shares, 3Com and Palm are not one-for-one equivalent: each 3Com share is entitled to 1.5 Palm shares. Because I do not divide by this factor, the point at which 3Com is exactly worth its holdings of Palm occurs when the log price ratio is ln(1.5)=0.4, which occurs in May, 2000. I use CRSP data for prices and volumes on all stubs in this paper to match Lamont and Thaler (2003).

The scatterplot in figure 4.2 likewise confirms the strong negative relation. According to the regression results, 3Com's price increases by 18% with a t-statistic of 6.4 relative to Palm when the log 3Com volume decreases by one relative to log Palm volume. The correlation is -0.63.

Figure 4.1: 3Com/Palm Time Series of Log Price Ratio and Opposite Log Volume Ratio. The black line (left-hand axis) is the log of the ratio of the Palm volume to the 3Com volume. The gray line (right-hand axis) is the log of the ratio of the 3Com price to the Palm price. The opposite ratios are graphed to make the visual comovement easier to spot. Each 3Com share is entitled to 1.5 Palm shares so 3Com is worth exactly its holdings of Palm when the ratio of prices is ln(1.5)=0.4, which occurs in May, 2000.

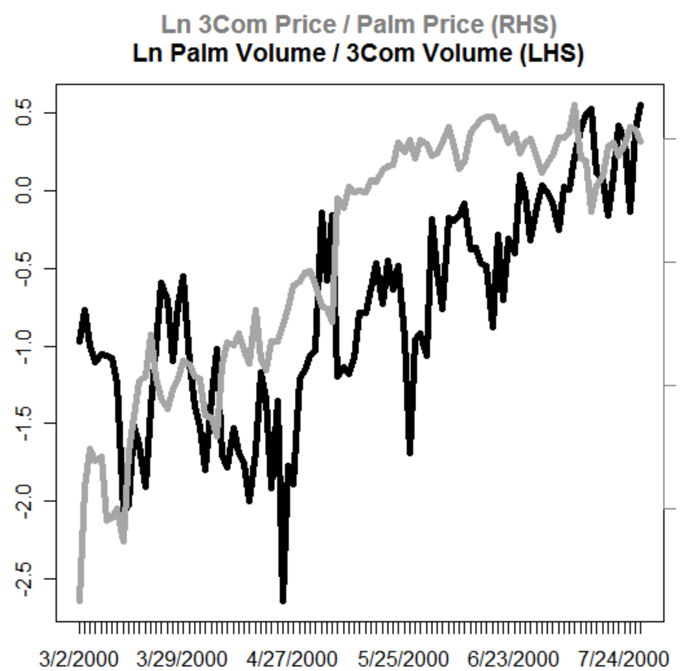
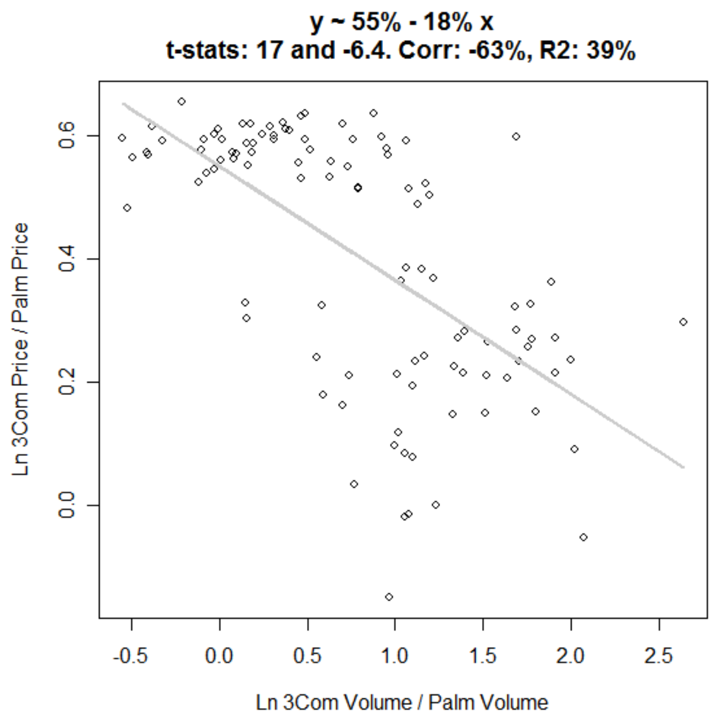


Figure 4.2: 3Com/Palm Log Relative Price vs. Log Relative Volume. The scatterplots show the relation between the log ratio of the 3Com price to the Palm price (y-axis) versus the log ratio of the 3Com volume to the Palm volume (x-axis). The regression results reported use non-parametric Newey-West t-statistics.



Lamont and Thaler (2003) suggest that some market participants are unable to add and subtract, while arbitrageurs are estopped with borrow constraints. On the other hand, Cochrane (2003) argues that the 3Com/Palm discrepancy is an example of convenience yield where market participants hold Palm shares for only a few days, and so are subject to only a small overpricing on average.

Cochrane (2003) notes a positive correlation between the price of Palm and its share volume, consistent with a theory of convenience yield. However, that correlation is between the absolute levels of Palm's price and Palm's volume. The negative correlation we see both here and in HSBC are between the relative price and the relative volume of the two pairs.

By themselves, the correlations between each stock's log price with its own log volume do not imply anything about the correlation between the log price ratio and the log volume ratio. Each of 3Com and Palm exhibit positive correlation between the log of their price and the log of their volume at 0.48 and 0.41, respectively, but the relative correlation is -0.63. In HSBC, the New shares exhibit -0.11 correlation between its log price and its log volume while the Old shares exhibit +0.33 correlation between its log price and its log volume, but the relative correlation is -0.51.

We can see why this can occur if we let:

$$p_1 = a_1 + b_1 v_1 + \epsilon_1 \tag{4.1}$$

$$p_2 = a_2 + b_2 v_2 + \epsilon_2 \tag{4.2}$$

denote the regression results of log prices $p_i = \log P_i$ on log volumes $v_i = \log V_i$ with residuals ϵ_i for each of the two share classes $i = 1, 2$. Then we can express the log relative price $p_1 - p_2 = \log{(P_1/P_2)}$ in terms of the log relative volume $v_1 - v_2 =$

Table 4.1: Price-Volume Regressions for 3Com/Palm

This table shows four regressions measuring the correspondence between the relative price and the relative volume: $y = \alpha + \beta x + \epsilon$, where $y = \log \frac{P_{3Com}}{P_{Palm}}$ is the relative price, $x = \log \frac{V_{3Com}}{V_{Palm}}$ is the relative shares volume or $x = \log \frac{V_{3Com} P_{3Com}}{V_{Palm} P_{Palm}}$ is the relative notional volume (as specified in the first column). The third row demeans and standardizes, and the fourth row also detrends, y and x before running the regression. The correlation (ρ) between y and x, the R^2 of the regression, and Newey-West corrected t-statistics in brackets are also reported.

Volume	Standardized?	α	β	ρ	R^2
Shares	No	55% [+16.6]	−18% [−6.4]	−63%	39%
Notional	No	57% [+9.6]	−14% [−3.6]	−39%	16%
Shares	Demeaned and Standardized	0% [0.0]	−82% [−9.7]	−82%	67%
Shares	Demeaned, Standardized, and Detrended	0% [0.0]	−43% [−2.8]	−43%	19%

$\log(V_1/V_2)$ as follows:

$$p_1 - p_2 = (a_1 - a_2) + \frac{b_1 v_1 - b_2 v_2}{v_1 - v_2}(v_1 - v_2) + (\epsilon_1 - \epsilon_2) \tag{4.3}$$

The coefficient in a regression of the log relative price $p_1 - p_2$ on the log relative volume $v_1 - v_2$ can be almost anything and depends among other things on the correlation between the log relative volume and the relative residuals $\epsilon_1 - \epsilon_2$. In the case of 3Com/Palm, that correlation is -0.85.

Table 4.1 summarizes the four regression results to ease comparison with the similar table 2.2 for HSBC.

4.2 Other Twins and Stubs

I also run the analysis for three other dual-listed companies for the 5 years of daily data from April 25, 2002 through April 24, 2007. Volume data prior to this time does not always agree between Datastream and Bloomberg so I focus only on the most recent data. Furthermore, I use the prices and volumes of the two local shares because that is where most trading occurred, and I convert all prices into a common currency using mid-point foreign exchange levels at the end of the day. Regressions are done on standardized variables only. Standardization, particularly detrending, not only makes comparisons more convenient across pairs, but is also helpful because of the simplifying assumption of the model that the future expected discrepancy level, barring convergence, equals the current discrepancy level.

The three other dual-listed companies are Royal Dutch-Shell, Unilever, and Reed-Elsevier. All had one share class trading in London and one in the Netherlands and long-standing equalization agreements among the two share classes. These last two pairs also highlight that it is not simply the country of residence that is important in determining the discrepancy since the two were often mispriced in opposite directions.

In addition, I also run the analysis for the five other mispriced tech stock carve-outs identified by Lamont and Thaler (2003). The main difference between these five and 3Com/Palm is size. While 3Com/Palm was on the order of $2.5 billion in market capitalization, these other five were far smaller, ranging from $150 million to $600 million. Furthermore, and more importantly for the self-imposed limits to arbitrage model, trading volume was commensurate with size. In the model, arbitrageurs limit their position size relative to the average daily volume of the less frequently traded security in the pair. 3Com/Palm averaged $230 million of daily trading volume in the less traded security (Palm) while the five other spin-offs ranged from $5 million to

Table 4.2: Market Regressions

The table below shows the results of regressing the demeaned, standardized, and detrended log price ratio on the demeaned, standardized, and detrended log shares volume ratio of the various pairs of securities. By construction, the free constant term is zero in all regressions and the coefficient equals the correlation. The Newey-West corrected standard errors and t-statistics are also reported, as is the R-squared.

Pair	$\boldsymbol{\beta = \rho}$	**Std. Err.**	**t-Stat**	$\mathbf{R^2}$
HSBC New-Old	-22%	5%	-4.2	5%
3Com/Palm	-43%	15%	-2.9	19%
RD-Shell '02-'07	-25%	7%	-3.4	6%
Unilever '02-'07	-23%	6%	-3.8	5%
Reed-Elsevier '02-'07	-9%	4%	-2.2	1%
Creative/Ubid	13%	19%	+0.7	2%
HNC/Retek	10%	9%	+1.1	1%
DaisyTek/PFSWeb	12%	7%	+1.8	1%
Metamor/Xpedior	-21%	13%	-1.6	4%
Methode/Stratos	-3%	10%	-0.3	0%

$33 million in the average daily dollar volume of the less frequently traded security in the pair. We should therefore expect the smaller pairs to be far less attractive, if at all, to arbitrageurs, and the effects of self-imposed limits to arbitrage to be negligible.

Table 4.2 shows the results of regressing the standardized relative price on the standardized relative volume, and reiterates the equivalent results from HSBC and 3Com/Palm for comparison.

Each of the coefficients is significantly negative except for the five small tech stock carve outs, consistent with self-imposed limits to arbitrage.

Of the four large pairs other than HSBC that are the ones most likely to draw the interest of arbitrageurs relative to other twins and stubs, we can calculate the market implied convergence probability and average amount of arbitrage activity as we did for HSBC, by calculating the model-implied correlation between standardized

Table 4.3: Implied Convergence Probabilities

The table below shows the market implied convergence probability $\hat{p}$ of the self-imposed limits to arbitrage model calibrated using the mean μ and the standard deviation σ of the log volume ratio and the regression slopes of the given pair, assuming the overall maximum position of arbitrageurs is $N = 100$. Also reported is the smaller of the average daily value traded for each of the two securities in the given pair.

Pair	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	$\widehat{p}\,(100)$	**Traded Value**
HSBC New-Old	1.01	0.81	70bps	£24 million
3Com/Palm	0.78	0.70	340bps	$230 million
RD-Shell '02-'07	0.27	0.38	120bps	€300 million
Unilever '02-'07	0.40	0.44	60bps	€130 million
Reed-Elsevier '02-'07	0.57	0.35	0bps	€34 million

relative volume and standardized relative price for a variety of probability values, then using the best-fit cubic to back out the implied probability value for which the model-implied correlation matches the market correlation. The distribution of the excess standard volume $1+\chi$ is again calibrated using only that portion of the volume that is within the transactions costs bounds of the given pair, which for simplicity I assume to be the same as for HSBC. The only exception is 3Com/Palm, for which the excess standard volume is calibrated using the entire history, because it is unclear if there was ever a time that the discrepancy was within the transactions costs bounds. Table 4.3 shows these results assuming the overall maximum position size by arbitrageurs is $N = 100$.

Reed-Elsevier is the only one to have a zero market implied convergence probability. It is as if arbitrageurs were not aware of the pricing discrepancy, were unable to trade it due to short selling or other external constraints, or, as suggested by the self-imposed limits to arbitrage model, chose not to trade it because of its relatively smaller size. As shown in table 4.3, Reed-Elsevier's trading volume was far lower than that of the two other London-Dutch multiple share classes, Royal Dutch-Shell and Unilever. Reed-Elsevier's average traded value was similar to that of HSBC, but the HSBC discrepancy occurred nearly ten years earlier, and volume has in general grown since then. For example, the Unilever share class listed in London doubled its average daily traded value over just the past five years. Further, HSBC was not a cross-country listing but rather a unique multiple share class opportunity where both share classes traded in the same country and on the same exchange, suggesting that it is not the best comparison for Reed-Elsevier in terms of critical volume.

Also, the far higher log mean of the volume ratio for Reed-Elsevier relative to the other pairs agrees with the hypothesis that arbitrageurs may have been simply absent from this pair. Though I traded the Reed-Elsevier discrepancy during this interval, informal feedback from brokers and counterparties at the time suggested it was rarely traded or even followed by other arbitrageurs.

The four other pairs have market implied convergence probabilities between 60 and 340 basis points, with the 3Com/Palm stub nearly four times as likely to converge overnight as the average of the other three. This result is reasonable because stubs have specific end dates while twins typically do not. What would have been a reasonable ex ante estimate of the daily convergence probability for 3Com/Palm? When the deal was announced, it was expected to almost surely converge within six months, pending a ruling by the IRS. If the probability of convergence within six months was 99 percent, then the implied daily probability p would be 359 basis points, close to the 340 basis points implied for $N = 100$.

A more illuminating approach would be to extend the implied probability results of table 4.3 for a variety of different values of N, then view the N as a function of the implied p. Then for consistent estimates of p, we can see how close the implied N's are to each other.

Figure 4.3: Maximum Position Size vs. Daily Convergence Probability. For each of the five pairs, the implied convergence probability p is calculated assuming the overall maximum position limit N and using the particular pair's parameters of standard volume from table 4.3. This figure shows the plot of N vs. p.

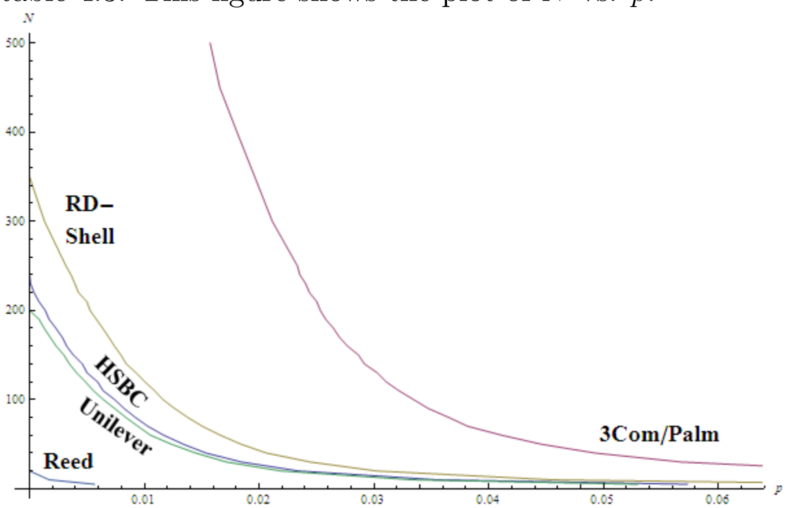


Figure 4.3 graphs this result. HSBC and Unilever are virtually identical, and Royal Dutch-Shell is only slightly higher than HSBC. (The extra RD-Shell premium may reflect the corporate restructuring RD-Shell experienced around the time of the sample period which likely increased the chance of further convergence relative to Unilever, which had remained unchanged.) 3Com/Palm has the same basic shape but lies much higher due to its higher probability of convergence.

We can calculate the implied N for a given p by interpolating with a cubic between every pair of points. Table 4.4 lists the results. (Not listed is 3Com/Palm because of its qualitatively different probabilities. For completeness, however, note that for $p = 0.0359$ as calculated above, the implied overall maximum position is 88 days of trading volume.) The final two rows of table 4.4 correspond to the ex post estimates of the ex ante probabilities of HSBC depending on whether all seven crossings are used ($p = 0.0060$) or only the first six are used ($p = 0.0095$), as per the discussion towards the end of section 3.

Table 4.4: Implied Maximum Positions

The table below shows the implied position sizes N of the self-imposed limits to arbitrage model calibrated using the mean μ and the standard deviation σ of the log volume ratio and the regression slopes of the given pair, assuming the given daily convergence probability p, and using a cubic fit between successive points of the numerically estimated results.

Probability p	**HSBC**	**RD-Shell**	**Unilever**
$p = 0.0100$	73	120	65
$p = 0.0075$	99	158	89
$p = 0.0050$	130	209	118
$p = 0.0060$	118	187	105
$p = 0.0095$	77	127	69

In all cases, the implied maximum positions seem to be around $N = 100$. The simple average of the numbers in the first three rows of table 4.4 is $N = 118$. Does $N = 100$ or $N \approx 120$ mean something? Are there certain "bins" or "buckets" that various N fall into in terms of approximate equivalence? They do, and it has to do with how quickly liquidation can happen.

Suppose the arbitrageur needs to liquidate. How long does he have to do so? Let's say he expects to liquidate aggressively, trading fifty percent of average daily volume. Some arbitrageurs must be out within a month; for them, $N = 10$ because it will take them $10/0.50 = 20$ business days, or one month, to liquidate the position aggressively. So the buckets basically are arbitrageurs that need to liquidate within a month ($N = 10$), between one month and one quarter ($N = 10$ to 30), between one quarter and six months ($N = 30$ to 60), and between six months and one year ($N = 60$ to 120). The last group, possibly extended out to nearly two years, is the one where most financial institutions and large hedge funds typically fall. Not only do they have longer lock-ups and longer notice periods for liquidation, but they also have a concept of "windows" where no more than e.g. 25% of the portfolio can be liquidated to investors each quarter. Furthermore, the SEC's registration requirements in effect from 2004 through 2006 used a two-year lockup as a distinguishing characteristic of private equity firms, implicitly assuming nearly all hedge funds have shorter lock-ups.

With these illustrative bins, all of the maximum positions from table 4.4 fall within this "six months to two years" bin, with HSBC and Unilever even closer to each other.

We can now address the *constancy* implication of self-imposed limits to arbitrage described in section 3. If the maximum position sizes were set because of transactions costs or risk measures relating specifically to the pair under consideration, one would expect there to be differences between a pair such as HSBC that traded in the same currency, in the same country, on the same exchange, from 1992 through 1999, with log standard volume having mean 1.01 and standard deviation 0.81, and a pair such as Unilever that traded in different currencies, in different countries, on different exchanges, from 2002 through 2007, with log standard volume having mean 0.40 and standard deviation 0.44. The fact that the implied maximum position is virtually identical suggests that it is indeed a model of self-imposed limits to arbitrage driving these discrepancies and not merely market transactions costs and risk measures.

Summarizing all of these results, as a general rule of thumb for a mispriced pair where one of the securities has substantially more volume than the other, a one standard deviation increase in a security's relative volume will coincide with a decrease of about one-quarter of a standard deviation in its relative price. Additionally, the implied overall maximum position size for arbitrageurs seems to be about one hundred days of typical trading volume.

CHAPTER 5

ALTERNATIVE EXPLANATIONS

Could traditional, external limits to arbitrage have also applied to HSBC? Or perhaps some liquidity or liquidity-related difference between the share classes explains the pricing discrepancy?

5.1 External Limits to Arbitrage

Mispriced pairs typically suffer from external limits to arbitrage, usually either short sales constraints or home bias. In the case of Palm/3Com, Lamont and Thaler (2003) shows that Palm shares were virtually unavailable to borrow, so arbitrageurs were not able to establish positions. In the case of twins such as Royal Dutch-Shell, de Jong et al. (2005) and Froot and Dabora (1999), whom they cite, show that cross-country twins are correlated with their domestic market indices.

In the case of HSBC, both share classes were in general widely available to borrow at general collateral fees. Both share classes were also listed on the same exchange, in the same country, and were members of the same indices. As documented in section 2.4, more than 70% of the time the discrepancy was wider than the transactions costs faced by arbitrageurs. Only non-fungibility prevented immediate riskless arbitrage. Arbitrageurs could have, and did, establish and maintain positions. Therefore, neither short selling constraints nor home bias nor other transactions costs were a binding external limit to arbitrage in HSBC.

The seminal limit to arbitrage is performance-based arbitrage (Shleifer and Vishny (1997)): arbitrageurs may have the most difficulty raising capital when arbitrage opportunities are greatest if investors chase performance. In such situations, the correlations between the returns of hedge funds and the returns of an arbitrage strategy

Table 5.1: Hedge Fund Returns and the HSBC Arbitrage

This table shows the correlations between the monthly returns of a variety of hedge fund indices and a frictionless arbitrage strategy for HSBC calculated assuming the arbitrageur is long the cheaper share class and short the more expensive share class at the end of each day, on an equal notional basis, and ignoring commissions, financing charges, or stock borrow fees. The hedge fund indices are taken from the CSFB/Tremont family of asset-weighted hedge fund indices (hedgeindex.com).

	Correlation
CSFB/Tremont Hedge Fund Index	-0.17
HEDG Convertible Arbitrage	-0.14
HEDG Dedicated Short	0.31
HEDG Emerging Markets	-0.19
HEDG Equity Mkt Neutral	-0.08
HEDG Event Driven	-0.25
HEDG Distressed	-0.30
HEDG E.D. Multi-Strategy	-0.18
HEDG Risk Arbitrage	-0.18
HEDG Fixed Income Arb	0.01
HEDG Global Macro	-0.05
HEDG Long/Short Equity	-0.29
HEDG Managed Futures	0.07
HEDG Multi-Strategy	-0.11

in the two HSBC share classes ought to be positive.

Table 5.1 shows the correlations between a frictionless equal-notional arbitrage strategy in HSBC with a variety of CSFB/Tremont hedge fund indices. The correlation with the overall hedge fund universe is -0.17. Only the Dedicated Short index shows a non-negligibly positive correlation, but relative value trading between two otherwise identical securities would not be a strategy that a Dedicated Short fund would follow. Long/Short Equity is the most likely category to arbitrage HSBC, and its correlation was -0.29. Therefore, capital requirements were not a limit to arbitrage in HSBC.

Indeed, this evidence suggests that hedge funds may have been long the pricing discrepancy, betting that it would continue to widen. Brunnermeier and Nagel (2004) find that hedge funds were on average long the technology sector during the internet bubble as well. In both instances, hedge funds not only failed to arbitrage away a pricing discrepancy, they may have exacerbated it. Brunnermeier and Nagel (2004), however, use actual hedge fund holdings to show they were long the technology sector, with few exceptions. No such data is available for HSBC and the only hedge fund I am aware of that actively traded in the pair (LTCM) was certainly not betting the discrepancy would widen.

5.2 Liquidity

Could the changes to relative volume and relative price actually reflect changes to relative liquidity or relative liquidity risk? I explore two specific alternatives:

1. Liquidity: Pastor and Stambaugh (2003) find that spikes in illiquidity correlate with spikes in volume. Perhaps higher relative volume in HSBC reflects lower relative liquidity. Then the relatively less liquid asset must generate higher future returns to compensate investors for holding a less liquid security. Those higher future returns are achieved by immediately lowering the price. Thus this hypothesis holds that higher relative volume coincides with lower relative price because both are driven by lower relative liquidity.

2. Liquidity Risk: Johnson (2007) proposes a model in which volume and liquidity are uncorrelated on average but volume is positively correlated with liquidity risk, the second moment of liquidity. Specifically, he predicts: "Even holding expected volume constant, higher realized volume should be associated with higher

> contemporaneous liquidity risk." Perhaps higher relative volume in HSBC is associated with higher relative liquidity risk. But how should higher liquidity risk, an increased probability of liquidity either increasing or decreasing, correlate with returns and prices? Johnson (2007)'s model implies that liquidity has a fat left tail, so it collapses faster than it expands. Therefore, it is reasonable to conclude that higher liquidity risk requires higher returns to compensate investors, and should therefore be associated with lower contemporaneous prices.

Any explanation of both HSBC puzzles based on differences in liquidity or liquidity risk inherently predicts that the more expensive share class should be the one with greater liquidity or less liquidity risk and therefore should have less relative volume. Specifically, whenever the relative price is positive, the relative volume ought to be negative. Figure 2.4 shows, however, that when the relative price is positive, the relative volume was far more often positive than negative.

In other words, liquidity-driven explanations of both puzzles predict that the constant term in the regression of non-standardized relative price on non-standardized relative volume should be statistically indistinguishable from zero. Table 2.2 shows, however, that the t-statistic on HSBC's constant estimate of 3.3% is 12.

The model of self-imposed limits to arbitrage, by contrast, makes no prediction about the constant term or the average level of relative volume except to require that the two share classes have different enough standard volume so that activity by arbitrageurs would bring the relative volume closer to one.

Additionally, neither model predicts the relation between relative price and relative volume will disappear for pairs with lower absolute volume. Meanwhile, self-imposed limits to arbitrage predicts, and the evidence shows, that smaller pairs do not exhibit the negative relation.

These facts weigh against both the liquidity and the liquidity risk alternatives.

There is further evidence against the liquidity explanation as well. The negative correlation between liquidity and volume appears to hold only in extreme events. Pastor and Stambaugh (2003) report that correlation in low-liquidity months is -0.27 but is 0.18 in high-liquidity months, or across all months in general. Johnson (2007) also reports that eight recent empirical studies have found that higher volume does not necessarily lead to more liquid markets. Furthermore, using the bid-offer spread as one measure of liquidity, table 2.1 shows that the two share classes had identical liquidity across the seven year period but the New share class nevertheless had far more volume on average.

There is also further evidence against the liquidity risk explanation. This alternative suggests that changes in liquidity risk explain both the pricing puzzle and the volume puzzle of HSBC, and that the activity of arbitrageurs is not involved in any way. Therefore, in this view, the slope of the regression of relative price on relative volume should not depend on the average relative price. Specifically, the slope for the subsample of data points occurring when the pricing discrepancy is within the transactions costs bounds should be the same as the slope for the complementary subsample when the pricing discrepancy is outside the transactions costs bounds.

By contrast, the model of self-imposed limits to arbitrage predicts that the slope should be far lower in the first subsample than in the second because within the transactions costs bounds there are no arbitrageurs to drive the result.

Table 5.2 lists the slopes for both subsamples for a variety of thresholds. In each case, the slopes in the "above" subsamples far exceed the slopes in the "below" subsamples. Specifically, in the threshold of 1.8% matching the transactions costs bounds for arbitrageurs documented in section 2.4, the regression coefficient of relative price on relative volume is only -31 basis points when the relative price is below the

Table 5.2: HSBC Slopes by Threshold

This table shows the coefficient ("Slope Below") from regressing the relative price on the relative volume of the two HSBC share classes for subsamples when the absolute value of the relative price is below a given threshold. It also reports the number of data points used ("# Below") and the similar results for the complementary subsamples when the absolute value of the relative price is above the same threshold. The threshold of 1.8% corresponds to the transactions costs bounds for arbitrageurs.

Threshold	**# Below**	**Slope Below**	**# Above**	**Slope Above**
0.5%	145	-9bps	1617	-186bps
1%	218	-16bps	1544	-189bps
1.8%	519	-31bps	1243	-216bps
2%	618	-35bps	1144	-224bps
3%	1126	-63bps	636	-258bps
4%	1367	–93bps	395	-339bps
5%	1501	-115bps	261	-384bps
6%	1637	-139bps	125	-423bps

threshold and is -216 basis points when it is above it. The slope for the entire sample is -175 basis points.

We can run a regression to confirm that the "below" and "above" numbers of the critical threshold in row 3 of table 5.2 are significantly different. Table 5.3 regresses the relative price on the relative volume of HSBC as well as an indicator variable that is one if and only if the absolute value of the relative price exceeds 1.8%, the threshold transactions costs for arbitrageurs. The indicator variable appears twice in the regression: once to adjust the constant term and once to adjust the slope on relative volume. The difference in slopes between the "below" and "above" subsamples is statistically significant: the coefficient on the indicator scaling the slope of the relative volume has a t-statistic of -6.3.

The liquidity risk model does not explain why the relation between relative price and relative volume should be stronger for greater average relative price. One pos-

Table 5.3: Price-Volume Regressions for HSBC With Threshold Indicator

The first row of this table shows the regression measuring the correspondence between the relative price and the relative volume with an indicator variable measuring the presence or absence of arbitrageurs depending on whether the relative price exceeds transactions costs: $y = \alpha + \beta x + \delta_1 I + \delta_2 I x + \epsilon$, where $y = \log \frac{P_{New}}{P_{Old}}$ is the relative price, $x = \log \frac{V_{New}}{V_{Old}}$ is the relative shares volume, and $I = 1$ if $|y| > 1.8\%$ and $I = 0$ otherwise. The second row repeats the earlier reported regression results without the indicator variable to ease comparison. The R^2 of the regression and Newey-West corrected t-statistics (in brackets) are also reported.

Regression	α	β	δ_1	δ_2	R^2
With Indicator	0.7% [+4.3]	−0.3% [−3.5]	3.3% [+10.7]	−1.8% [−6.3]	39%
Without Indicator	3.3% [+12.0]	−1.7% [−7.6]			26%

sibility would be that when relative price is higher, the range for relative volume is tighter and lower, thus causing a stronger regression result, but this is not the case. For a threshold of 1.8%, the range on the "below" subsample is essentially the same as the range on the "above" subsample.

Explanations based on liquidity are ambitious in that they attempt to explain both the price and the volume puzzles without resorting to noise traders and arbitrageurs. They imply that the relative prices differ because of inequality in some measure of liquidity, and that this difference is also reflected in the opposite direction by the relative volume. Such sharp theories leave no room for noise traders at all and are contradicted by the data, most notably the non-zero constant term in the regression of relative price on relative volume. If the theories are relaxed to allow some noise traders, they essentially revert to square one as they must also allow arbitrageurs, and they must explain why arbitrageurs with no external limits on them do not eliminate the excess price discrepancy caused by the noise traders over and above that caused by liquidity factors.

It is impossible to completely rule out any effect of any liquidity or liquidity-related measure on the two puzzles. However, the first two rows of table 5.2 suggest that liquidity-related effects explain at most ten percent of the volume puzzle. Whatever the actual number, some theory like self-imposed limits to arbitrage would still be needed to explain the remainder.

CHAPTER 6

CONCLUSION

Can the market compute equality? The case of HSBC requires no multiplication or even addition. The two share classes were designed to be as equal as possible on a one-for-one basis, and in virtually all respects were identical, but not fungible. The absence of fungibility created room for mispricing, which swung from an 8% discount to an 8% premium. One notable difference was in their trading volume. The New shares had about four times greater turnover than the Old shares.

A large part of the discrepancy is explained by the contemporaneous movement of the trading volume of the two share classes. The share class that had a higher volume relative to the other one tended to have a relatively lower price. The effect is approximately 120 basis points for a doubling in relative volume, and the correlation is 50%.

Often pricing discrepancies persist because of limits to arbitrage. Traditionally these limits are external restrictions imposed on arbitrageurs, such as an inability to raise capital in times of need, transactions costs, noise trader risk such as home bias, or short selling constraints. The unique case of HSBC combines the benefits of stubs (side-by-side trading in the same currency, in the same country, and on the same exchange) with the benefits of large twins (easily borrowable shares) and eliminates the disadvantages of each. The HSBC pair was not subject to home bias or short sale constraints.

I propose self-imposed limits to arbitrage to explain both why the HSBC discrepancy was able to persist and why the relative volume was negatively related to the relative price. Arbitrageurs may be subject not just to the traditional external limits to arbitrage, but also to internal restrictions they voluntarily place on themselves.

The specific restriction I explore is a willful limitation on position size.

My model of self-imposed limits to arbitrage generates plausible values of the market implied amount of arbitrage activity in HSBC, averaging around 20 percent of daily volume. The model also predicts that relatively small pairs will not experience the negative relation between relative volume and relative price, and indeed, for five stubs much smaller than 3Com/Palm the relation is absent. Additionally, for the one cross-country twin much smaller than Royal Dutch-Shell and Unilever, the market implied convergence probability is zero. Finally, the model predicts relatively constant overall position limits for a variety of pairs, and the evidence indeed suggests the implied position limits are roughly 100 days of trading volume. This supported prediction distinguishes self-imposed limits to arbitrage from models based solely on market transactions costs or risk measures.

The intuition behind the negative association between relative volume and relative price implied by the model of self-imposed limits to arbitrage is that wider discrepancies lead to more trading by arbitrageurs and, because arbitrageurs trade both securities in essentially equal volume, more arbitrageur activity tends to bring the relative volume closer to one. To the extent the more expensive share class tends to have higher volume, as is typically the case, the relative volume would fall towards one as the relative price increases.

Two rules of thumb emerge for mispriced pairs. One is that an increase of one standard deviation in the relative volume coincides with a decrease of about one quarter of a standard deviation in its relative price. The other is that the implied overall maximum arbitrageur position limits tends to be roughly 100 days of trading volume.

Future research could further explore the model to extend its predictions from being based solely on which share class on average trades more volume to being based on the distribution of that relative volume. For example, model predictions could be calibrated depending on the probability of the more expensive share class having greater volume.

BIBLIOGRAPHY

Almgren, R., C. Thum, E. Hauptmann, and H. Li (July 2005). Equity market impact. *Risk*, 57–62.

Barberis, N. and M. Huang (2001). Mental accounting, loss aversion, and individual stock returns. *Journal of Finance 56:4*, 1247–1292.

Barberis, N. and M. Huang (2004). Stocks as lotteries: The implications of probability weighting for security prices. *Working Paper*.

Barberis, N. and A. Shleifer (2003). Style investing. *Journal of Financial Economics 68*, 161–199.

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*, 307–343.

Benartzi, S. and R. H. Thaler (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics 110:1*, 73–92.

Brunnermeier, M. K. and S. Nagel (October 2004). Hedge funds and the technology bubble. *Journal of Finance 59*, 2013–2040.

Cochrane, J. H. (2003). Stock as money: Convenience yield and the tech-stock bubble. In W. C. Hunter, G. G. Kaufman, and M. Pomerleano (Eds.), *Asset Price Bubbles.* Cambridge, MA: MIT Press.

Coval, J. D. and T. Shumway (2005). Do behavioral biases affect prices? *Journal of Finance 60:1*, 1–34.

de Jong, A., L. Rosenthal, and M. A. V. Dyck (2005). The limits of arbitrage: Evidence from dual-listed companies. *EFA 2004 MAASTRICHT Available at SSRN Number 525282 November 2005.*

Dow, J. and G. Gorton (July 1994). Arbitrage chains. *Journal of Finance 49*, 819–849.

Drachter, K., A. Kempf, and M. Wagner (2006). Decision processes in German mutual fund companies: Evidence from a telephone survey. *Center for Financial Research Working Paper 06-07.*

Ellerton, C. (February 14, 1997). HSBC: Mind the gap. *Analyst report, SBC Warburg.*

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics 49*, 283–306.

Froot, K. A. and E. M. Dabora (1999). How are stock prices affected by the location of trade? *Journal of Financial Economics 53*, 189–216.

HSBC (1992). Listing particulars relating to the issue of new HSBC Holdings shares and new HSBC Holdings bonds in connection with the offer for Midland Bank Plc.

Johnson, T. (2007). Volume liquidity and liquidity risk. *Journal of Financial Economics Forthcoming*.

Joseph G. Nicholas, J. (2004). *Hedge Fund of Funds Investing: An Investor's Guide*. Bloomberg Press.

Kondor, P. (2006). Risk in dynamic arbitrage: Price effects of convergence trading. *Working Paper*.

Lamont, O. A. and R. H. Thaler (2003). Can the market add and subtract? Mispricing in tech stock carve-outs. *Journal of Political Economy 111:2*, 227–268.

Mitchell, M., T. Pulvino, and E. Stafford (2002). Limited arbitrage in equity markets. *Journal of Finance 57*, 551–584.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*, 631–653.

Pastor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Sheinkman, J. and W. Xiong (2003). Overconfidence and speculative bubbles. *Journal of Political Economy 111*, 1183–1219.

Shleifer, A. and R. W. Vishny (March 1997). The limits of arbitrage. *Journal of Finance 52*, 35–55.

Strahilevitz, M. A. and G. Loewenstein (1998). The effect of ownership history on the valuation of objects. *The Journal of Consumer Research 25:3*, 276–289.

Thaler, R. H. and H. M. Shefrin (April, 1981). An economic theory of self-control. *Journal of Political Economy 89*, 392–406.

World Bank (2005, May). Project information document report AB1634, appraisal stage.

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software 11*, 1–17.