Self-Imposed Limits of Arbitrage

Philip Z. Maymin

A multi-billion-dollar, multi-year discrepancy between two identical share classes of Hong Kong and Shanghai Banking Corporation (HSBC) did not suffer from traditional external limits to arbitrage such as transactions costs and risk measures. One possible explanation is that self-imposed limits to arbitrage (SILTA) such as internal restrictions on position size allowed persistent mispricings. SILTA predicts a novel negative relation between relative volume and relative price. This prediction from SILTA holds not only for HSBC, but also other large mispriced pairs such as 3Com/Palm and Royal Dutch-Shell. Indeed, the implied overall maximum position size of arbitrageurs is roughly constant at one hundred days of trading volume for various mispriced pairs spanning different time periods and countries, suggesting SILTA as a possible explanation for all of them.

■ 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, as documented by de Jong, Rosenthal, and van Dijk (2009) and 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, and are documented in Lamont and Thaler (2003). 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.

HSBC is a serendipitous example combining the best of both twins and stubs: two identical but non-fungible share classes of HSBC co-existed from 1992 to 1999 with the mispricing reaching highs of eight percent in each direction.

On June 25, 1992, Hong Kong and Shanghai Bank (HSBC) acquired Midland Bank. Prior to the acquisition, HSBC had been listed primarily in Hong Kong, where its shares were denominated in Hong Kong dollars, with a secondary listing on the London Stock Exchange (LSE). These "Old" shares had a par value of HK$10, but traded in sterling on the LSE. To consummate the acquisition, a new sterling-denominated share class needed to be 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 the Old shares. 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.[1]

Philip Z. Maymin is an Assistant Professor of Finance and Risk Engineering at New York University - Polytechnic Institute in Brooklyn, NY.

I am grateful to Richard H. Thaler, Nicholas Barberis, Tobias Moskowitz, and Ľuboš Pástor for their detailed comments and continuous support. I also thank John Cochrane for comments on an earlier version of this paper, and seminar participants at the 9th Annual Columbia-JAFEE Conference on Quantitative Finance and two anonymous JAF referees for their feedback and comments.

[1] 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, yet it was the Old, HK-par shares which typically traded at a discount to the New, Great British Pound (GBP)-par shares.

The May 8, 1992 Listing Particulars of the New shares in connection with the acquisition of Midland Bank specifically states (HSBC Listing Particulars, 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 of HK$1.31.[2] Even on matters not explicitly spelled out, the Listing Particulars makes it clear that the two share classes were to be treated as equals (Paragraph 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.

HSBC is a serendipitous example combining the best of both twins and stubs: two identical but non-fungible share classes of HSBC co-existed from 1992 to 1999 with the mispricing reaching highs of eight percent in each direction.

The Listing Particulars then 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."

Even the taxation treatment of the two shares do not differ (Ellerton, 1997):

> 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 an equivalent exchange ratio: each share, Old or New, received three dollar-denominated shares.

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).

The two share classes were pure twins, but listed in the same country and on the same exchange. Thus, traditional, external limits to arbitrage such as home bias or short selling constraints did not apply, making it a rare example of a genuinely 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.

What could explain such a discrepancy? As a possibility, 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 among traders would not eliminate these self-imposed limits, and develop its two novel 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 pairs and stubs. 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, but 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, namely that the relative volume (the natural logarithm of the ratio of the daily trading volumes in shares) will be negatively related to the relative price (the natural logarithm of the ratio of the prices of the two share classes).

[2] 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.

Table I. Summary of the Two HSBC Share Classes

Standard errors where appropriate are in parentheses to FTSE-100 is calculated by Bloomberg using weekly data. The Fama-French rows represent the risk loadings each share class exhibited on 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 of 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)

Indeed, I find that the same negative correspondence between relative price and relative volume holds not only in HSBC, but also 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.

The 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, zero for Reed-Elsevier, which seems to have been ignored by arbitrageurs, and three percent for 3Com/Palm, which in fact was more likely to converge. 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.

The model also predicts roughly the same total position limits across different pairs. This prediction is largely borne out by the data as well.

I. Two Discrepancies

A. The 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 I 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 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.

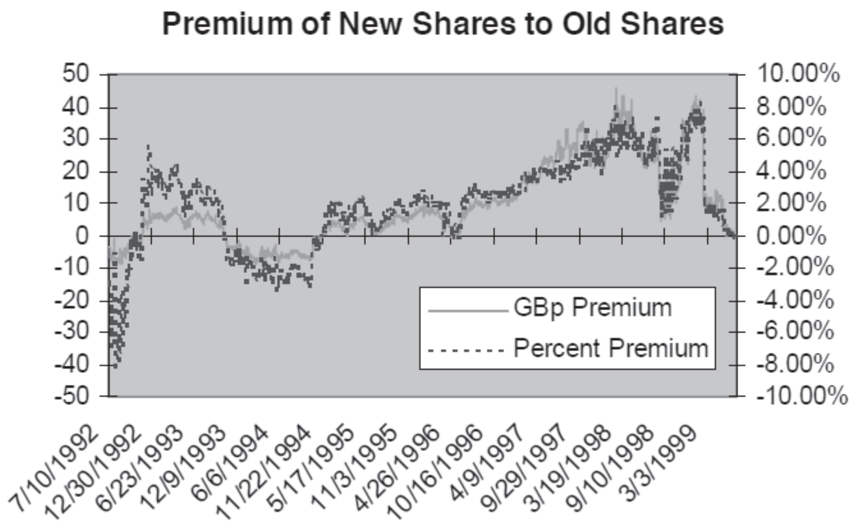
The average absolute value of the percentage premium of the New shares to the Old shares was 2.89%, peaking on February 2, 1999 at GBp 42 or 8.26% of the value of the Old shares. The average absolute value of the pence premium was GBp 11.6, peaking on February 25, 1998 at GBp 45.7 or 8.01% of the Old 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.

Each shareholder who purchased New shares at a premium could have received a higher dividend yield without the loss of fundamental exposure to the overall performance of the

Figure 1. HSBC Historical Premium of New Shares to Old Shares

This 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.



company by purchasing Old shares instead.

Analysts at banks began to issue reports in 1997 when the premium widened again after a three-year lull and continued to do so until convergence in 1999. Chris Ellerton of SBC Warburg wrote that institutions buying the New shares in preference to the Old shares were "in effect sacrificing 1.8p of net dividend, equivalent to 15bp of yield" and that those institutions "appear to be wasting the money" and should buy the New shares instead. BZW Securities similarly wrote that the two listings, a "relic" of the merger, are "identical in dividends, rights, voting, and all other material matters" but are not fungible because of the different in nominal "par" value. They also conclude that "this difference in entirely nominal."

B. The Relative Volume Discrepancy

Table I also 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 relative volume discrepancy is the relationship between the relative volume and the relative price. Specifically, the volume differential was often low when the price discrepancy was widest. 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.

Figure 2 shows how well the relative price and the opposite relative shares volume move together. That figure plots the opposite relative volume in the time series to make it is easier to spot comovement. The relative price, drawn in gray, moves 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 3 shows the same data smoothed with monthly (21-day) moving averages where the effect is even more pronounced.

The scatterplot and regression in Figure 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).

Replacing the relative shares volume with the relative shares turnover 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.

Replacing the relative shares volume with the relative notional volume

$$\log \frac{P_{New} V_{New}}{P_{Old} V_{Old}} = \log \frac{P_{New}}{P_{Old}} + \log \frac{V_{New}}{V_{Old}}, \tag{1}$$

where P_i is the price time series of share class i, makes

Figure 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.

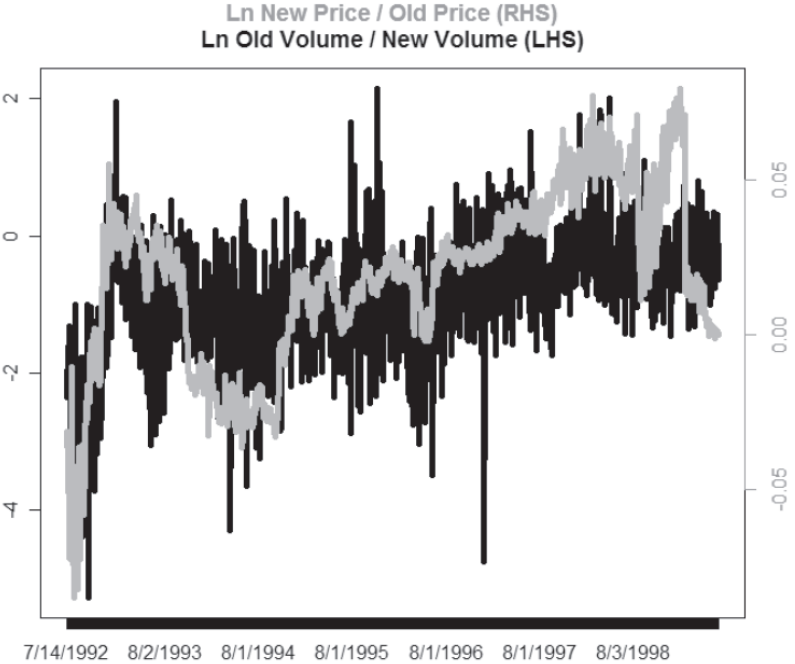


Figure 3. HSBC Time Series of Monthly Averages of Log Price Ratio and Opposite Log Volume Ratio

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.

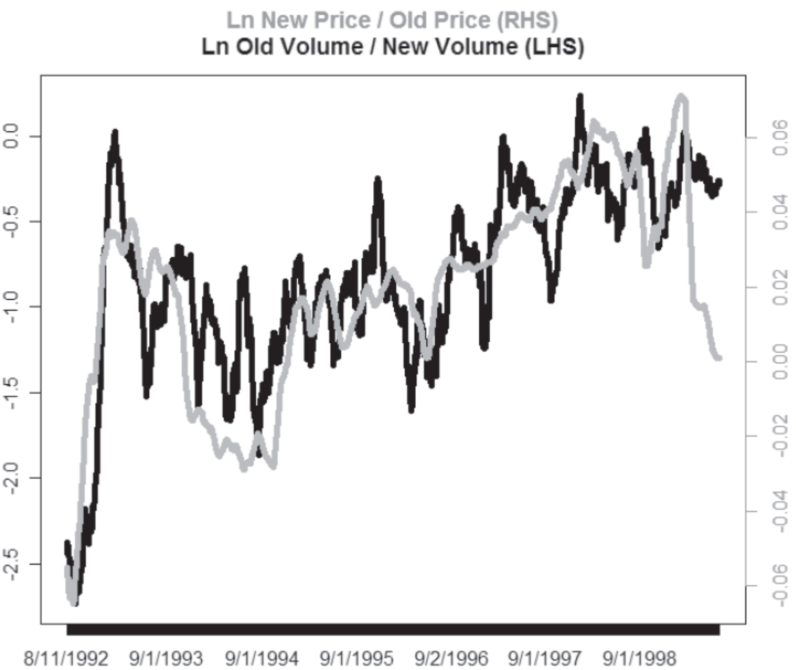
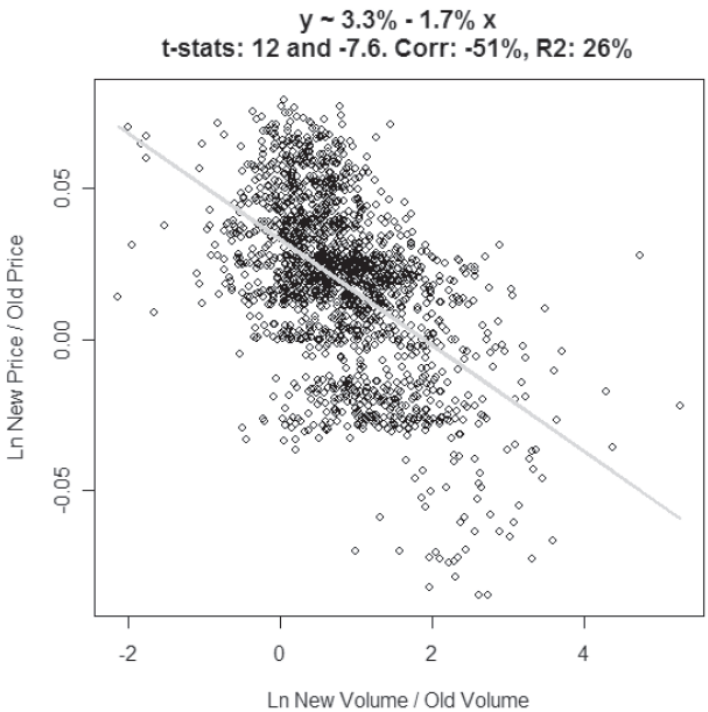


Figure 4. HSBC Log Price Ratio and Log Shares Volume Ratio

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.



essentially no difference in the case of HSBC because it simply rewrites the regression equation.[3]

Replacing both the log price ratio and the log shares volume ratio with their standardized counterparts *does* make a difference, halving the correlation. Table II 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.

II. External Limits to Arbitrage Did Not Bind

"Limits to arbitrage" theories state that pricing discrepancies persist because of external restrictions on arbitrageurs such as fundamental risk, the 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 such as home bias (de Jong, Rosenthal, and van Dijk, 2009). Both fundamentally and in terms of liquidity, we have seen in Table I that the two share classes are virtually identical.[4] If the discrepancy existed because of an inability by arbitrageurs to raise additional capital in times of need, then we would expect the discrepancy to widen during August 1998 during the time of Long-Term Capital Management's collapse, but the discrepancy actually collapsed four-fifths of the way down to zero, from 5.5% at the end of July 1998 to 1.1% at the end of August 1998. Further, home bias was not an issue for HSBC.

A. Transactions Costs

Could transactions costs have explained the discrepancy? 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 swaps with counterparties to avoid the stamp tax of 50 basis points

[3] If, $\log(P_{New}/P_{Old}) = \alpha + \beta \log(V_{New}/V_{Old})$, where $\alpha = 3.3\%$ and $\beta = -1.7\%$ as reported above, then by algebra we can write $\log(P_{New}/P_{Old}) = \alpha/(1+\beta) + \beta/(1+\beta)\log(P_{New}/P_{Old} . V_{New}/V_{Old})$, and because $1+\beta = 0.983$ is so close to one, the regression results will hardly differ.

[4] Could the discrepancy be linked to insiders? To be sure, 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. Nevertheless, ownership type could not be the driving force behind the discrepancies because it did not change substantially over time.

Table II. 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 $y = \log(P_{New} / P_{Old})$ is the relative price and $x = \log(V_{New} / V_{Old})$ is the relative shares volume or $x = \log(V_{New} / V_{Old} . P_{New} / P_{Old})$ 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 the Newey-West corrected t-statistics are in parentheses.

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%

charged on all stock purchases on the LSE. 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 (which requires retail customers post initial margins of at least 50%).[5]

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, Rosenthal, and van Dijk (2009) cite commissions of 25 basis points per side and bid-offer spreads of 40 basis points in their study of 12 dual-listed companies but their estimates are based on "conversations with a large number of investment firms" and likely reflect averages for arbitrageurs, whereas top-tier arbitrageurs could negotiate lower commissions than the average. In addition, they follow Mitchell, Pulvino, and Stafford (2002) in imposing special borrow rates for the shorts despite the fact that shares in twins are often available at general collateral rates. Mitchell, Pulvino, and Stafford (2002) confirm that the average borrow fee for general collateral stocks, albeit in the US 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.

The round-turn costs, meaning the costs to both enter and ultimately exit the strategy, assuming price convergence within one year are 180 basis points, as summarized in Table III. (The round-turn costs assuming structural convergence within one year are 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 at this time. A position long £100 and short £100 required collateral of £10, and if the arbitrageur allocated an additional £10 of risk capital, the return on capital for a one-year price convergence from a 5% initial discrepancy would be:[6]

$$\frac{100}{20} \cdot \left(5.0\% - 1.8\%\right) = 16\%. \quad (2)$$

The absolute value of the log ratio of the New share to the Old share exceeded 180 basis points more than 70% of the

[5] De Jong, Rosenthal, and van Dijk (2009) and Mitchell, Pulvino, and Stafford (2002) use far more conservative leverage numbers by assuming Regulation T applies, under which initial margins allow only two times leverage. However, arbitrageurs actually use swaps and contracts-for-differences to exempt themselves from Regulation T, Regulation X, and similar restrictions on leverage.

[6] 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. The dividend yield averaged 3.5% so the arbitrageur would receive an additional 17.5 basis points per year. Also, the transactions costs described above do not include mitigation 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.

time. Therefore, transactions costs and short borrow fees were not limits on arbitrageurs either.

III. Self-Imposed Limits to Arbitrage

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 their maximal arbitrage positions.[7] Some examples common to various hedge funds are restricting position size to a portion of fund capital, e.g., five percent, restricting trading activity to a percentage of average daily trading volume, e.g. fifteen percent, and restricting total position size to a multiple 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, Kempf, and Wagner, 2006).

In principle such restrictions of oneself 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, setting a fixed limit of five percent of fund capital for any and all stocks ignores the differences in risks among stocks. Therefore such a limit is more likely to be self-imposed, rather than be the result of transactions costs or risk measures.

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).

These self-imposed restrictions could also be *rational* and *structural*. The justification 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 at 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%, 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% held back until the year-end audit is completed (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.

[7] Counterparties can and do restrict position size through credit limits, but counterparties can only restrict the size of positions they see. In practice, an arbitrageur's position size is not materially restricted by the credit limits of counterparties so long as the underlying products are broadly available. In the cases of multiple share classes and equity carve-outs that we consider here, virtually any counterparty would be able to provide the necessary swaps. Therefore external restrictions from counterparties do not bind in these examples.

Different arbitrageurs may have different motivations but as long as they set a maximum position size, they must determine how to optimally trade a price discrepancy. In particular, arbitrageurs will seek to minimize market impact on entry while maximizing the eventual position size and, ultimately, profit.

So long as the price impact of trading quickly substantially exceeds the cumulative price impact of trading slowly, and the likelihood of fundamental convergence is relatively small, arbitrageurs will choose to impose this trading restriction on themselves, and the pricing discrepancy will persist. This argument is similar in principle 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. However that model 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, arbitrageurs 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 (the effect on the market price as a result of trading) 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 , then an arbitrageur will choose t^* to maximize expected profit:

$$\max_t \sum_{s=1}^{t-1} \frac{s}{t} Np(1-p)^{s-1}\left(E_0(D_t)-f\left(\frac{s}{t}N,s\right)\right) + N(1-p)^t\left(E_0(D_t)-f(N,t)\right). \qquad (3)$$

In other words, the arbitrageur decides over how many days he wants to spread out his trading, taking into account both the market impact of his large trades and the possibility that the mispricing disappears before he is able to trade his entire position.[8]

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, $\frac{\partial}{\partial t} f(N,t)$, the arbitrageur will choose to trade over t^* days such that trades of size N/t^* experience minimal market impact. In practice this is typically on the order of 1-5% of daily volume. In other words, if convergence is impossible, the arbitrageur will spread out his trading as long as necessary to minimize his market impact.

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:

$$\max_t \frac{N}{t}\left(D_0 - f\left(\frac{N}{t},1\right)\right). \tag{4}$$

In other words, the arbitrageur chooses 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.[9]

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. Thus, when aggressive arbitrageurs enter the market, we should observe 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 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.

How would self-imposed limits to arbitrage impact relative volume? Let V_i be the daily shares volume of class $i = 1,2$, composed of a "standard" volume S_i, i.e., the volume that share class i would have experienced had there been no arbitrageur activity, and k_i, the number of shares traded by arbitrageurs. Arbitrageurs trade a portion, $\theta = N/t^*$, of the standard volume of the less traded class:

$$k \equiv k_1 = k_2 = \theta \min(S_1, S_2). \tag{5}$$

Note that k, the number of shares traded by arbitrageurs, is the same for each share class, because arbitrageurs establish positions on an equal share basis. Note also that arbitrageurs limit themselves to a portion of average or standard volume, not the actual volume of the day; this allows θ to exceed one if necessary.

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

$$\lim_{k\to\infty} \frac{V_1 + k}{V_2 + k} = 1. \tag{6}$$

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) 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.

If the expensive share class also has higher standard volume, then $S_1 > S_2$. This is the case with HSBC and is 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 $V_1 = S_1 + \theta S_2 > S_2 + \theta S_1 = V_2$. Then:

$$\frac{V_1}{V_2} = \frac{S_1 + \theta S_2}{S_2 + \theta S_2} = \frac{S_1/S_2 + \theta}{1+\theta} = 1 + \frac{\chi}{1+\theta}, \tag{7}$$

or $\ln\frac{V_1}{V_2} = \ln\left(1 + \frac{\chi}{1+\theta}\right) \approx \frac{\chi}{1+\theta}$. So the relative volume $\ln V_1/V_2$ decreases in the amount of arbitrage activity θ and increases in the excess normal volume χ.

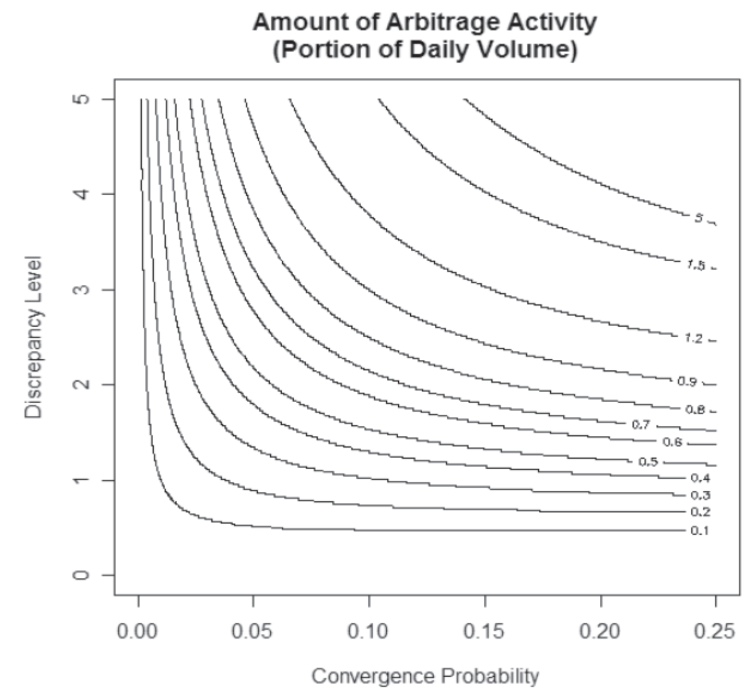
How does the amount of arbitrage activity θ vary with the

[8] Different models could also be developed for arbitrageurs choosing other self-imposed limits, for example by restricting trading activity. The overall position limit is chosen here for concreteness and because of its widespread prevalence in practice.

[9] 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 treat it as a sequence of mini-days, e.g. minutes, with a probability of convergence as per Equation (1).

Figure 5. 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.



relative price? Figure 5 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 $f(N,t)=\sqrt{\theta}$, a market impact function commonly used both by academics and practitioners. This same shape appears for other values of N: for any given probability p, the arbitrage activity always increases in the discrepancy level.

The discrepancy level D and the market impact function $f(N,t)$ were defined in terms of dollars, but they can be in any units so long as they are consistent. For example, they could both 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, to obviate 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 5 shows that θ increases for higher D for all probability levels, and we saw above that relative volume decreases for higher D when $\chi>0$. Further, standardizing $\chi/(1+\theta)$ 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 relation between standardized relative price and standardized relative volume:

$$Z\left[\ln\frac{P_1}{P_2}\right]=a+bZ\left[\ln\frac{V_1}{V_2}\right], \tag{8}$$

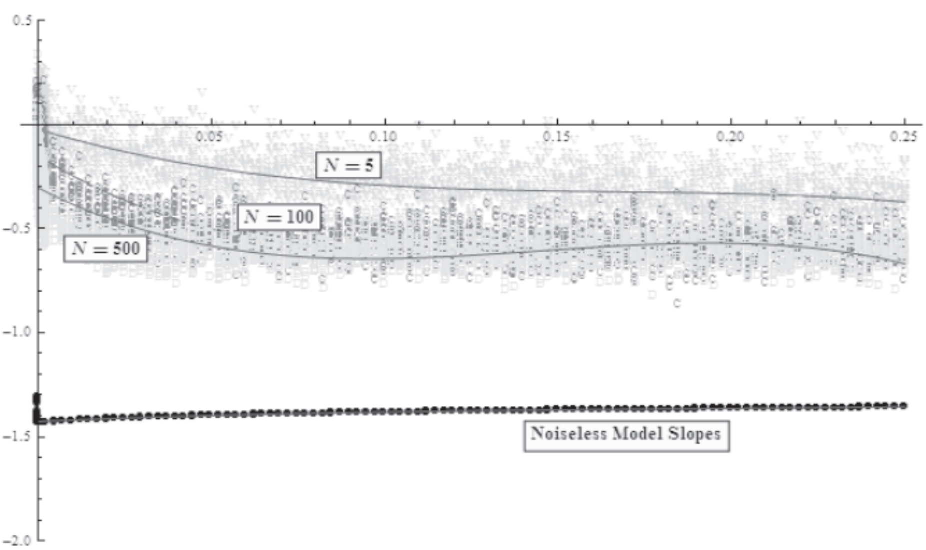
where $Z[\cdot]$ is the standardization function that demeans, descales, and detrends its input. This negative volume-price link holds only when the two share classes have substantially different volume and the more expensive share class tends to have more volume, so that $\chi>0$, as is the case with HSBC.

At this point, we have seen that the model of self-imposed limits to arbitrage, specifically restrictions on overall position size and trading activity, does two things: first, it allows a pricing discrepancy to persist and, second, it predicts 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:

Figure 6. 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: $D = a + bZ\left[\ln\left(1+\chi\left(1+\theta\right)^{-1}\right)\right] + \epsilon$ where χ is distributed lognormally with mean μ = 1.01 and standard deviation σ = 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 (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 are essentially the same for any N.



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

$$a + bZ\left[\ln\left(1+\frac{\chi}{1+\theta}\right)\right] \approx a + bZ\left[\ln\left(1+\frac{1}{1+\theta}\right)\right], \qquad (9)$$

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

The bottom of Figure 6 shows the ordinary least squares (OLS) slopes of the regressions of D on the standardized log of $1+(1+\theta)^{-1}$ for 50 values of p from 0.00005 through 0.00250 in steps of 0.00005 and for the 99 values of p from 0.00500 through 0.25000 in steps of 0.0025, each having 200 values of D ranging from 0.025 to 5 in steps of 0.025. In other words, each point represents the coefficient b in the regression above.

The noiseless model slope lies in a range between -145% and -139%. Yet, from Table II, 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 typically 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, since presumably arbitrageurs do not trade there.

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 II.A. The top part of Figure 6 shows the OLS slope estimate of b computed in the regression between the standardized relative price D and the standardized log of $1+\chi\left(1+\theta\right)^{-1}$ for the same values of p and D as for the noiseless model slopes described above, where N = 5 and 1+ χ is simulated by being drawn from a lognormal distribution with log mean μ = 1.01 and log standard deviation σ = 0.81. The figure also plots a fitted cubic. The top part of Figure 6 also plots the results for N = 100 and N = 500.

Unlike the noiseless model slopes of the bottom of Figure 6, the noisy model slopes of the top of Figure 6 vary with the convergence probability. From Table II, 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

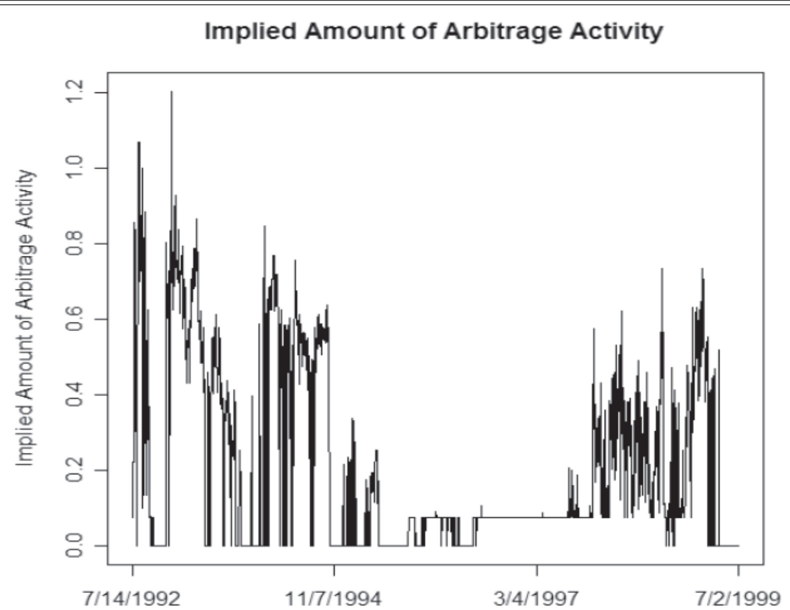
Table IV. Length of HSBC Mispricings

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

Figure 7. 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 (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%.



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:

$$(1-p)^{126} = 0.50 \Rightarrow p = 0.55\%. \tag{10}$$

Ex post, the discrepancy crossed parity after exceeding transactions costs seven times during the seven year period. Table IV 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 the implied daily convergence probability is 30 basis points. If we use only the average of the first six periods (146 business days), then the implied probability is 47 basis points. However, arbitrageurs would not establish their maximum position on the first day of each period, so perhaps the better estimate 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

Figure 8. 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.

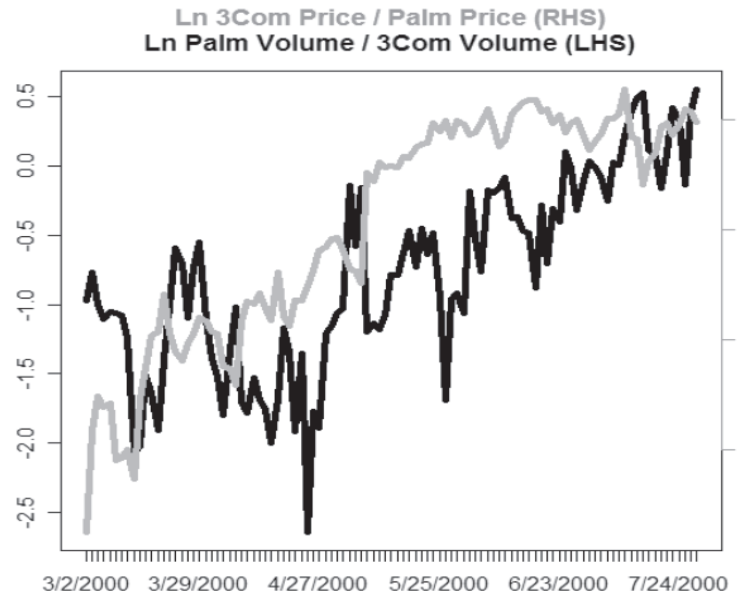


Figure 9. 3Com/Palm Log Relative Price vs. Log Relative Volume

The scatterplot shows 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.

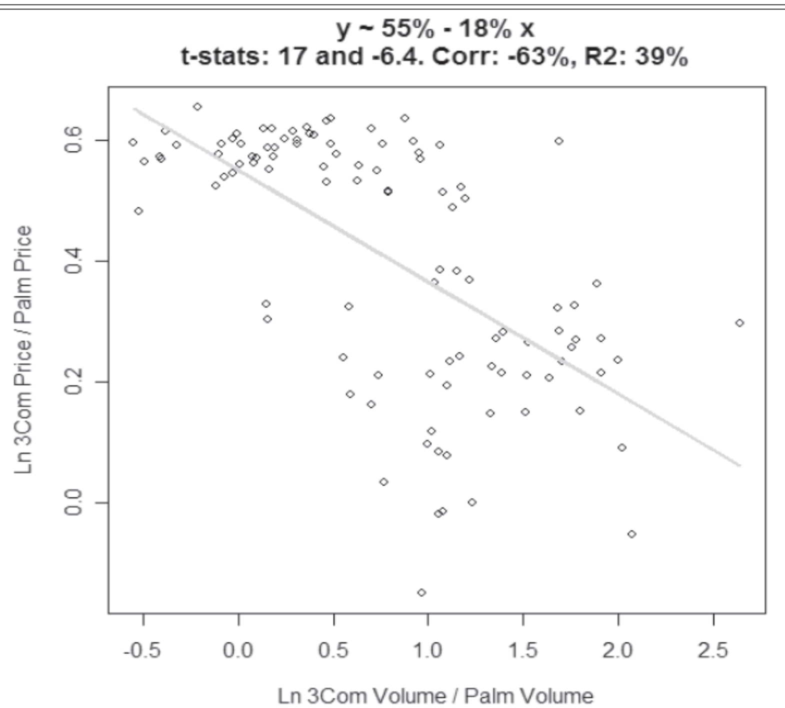


Table V. Price-Volume Regressions for 3Com/Palm

This table shows four regressions measuring the correspondence between the relative price and the relative volume: $y = \alpha + \beta x + \epsilon$, where $y = \log(P_{3Com} / P_{Palm})$ is the relative price, $x = \log(V_{3Com} / V_{Palm})$ is the relative shares volume or $x = \log(V_{3Com} / V_{Palm} \cdot P_{3Com} / P_{Palm})$ 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 the Newey-West corrected t-statistics are reported in parentheses.

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 Descaled	0% (0.0)	-82% (-9.7)	-82%	67%
Shares	Demeaned, Descaled, and Detrended	0% (0.0)	-43% (-2.8)	-43%	19%

average of all seven periods suggests a daily convergence probability of 60 basis points, but using the average of just the first six periods, perhaps the better estimate, suggests a daily convergence probability of 95 basis points.

A convergence probability of 95 basis points is close to the 70 basis points implied by the model for $N = 100$. It is the model implied probability for $N = 77$.

How does $N = 77$ or $N = 100$ relate to the informal statement that arbitrageurs impose something like as their position limit?

Suppose there are 20 arbitrageurs trading the pair. Because transactions costs and funding costs have to be minimized, this would likely 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 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 $N = 5$.

HSBC is unique because it has all the advantages of twins and stubs. Yet, do self-imposed limits to arbitrage help explain other pairs too?

Using $p = 0.95$ as the ex ante daily convergence probability and $N = 77$ as the total position size by arbitrageurs facing square-root market impact costs, Figure 7 plots the time series of the implied amount of arbitrage activity θ by solving Equation (3) 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 arbitrageur activity was 20% of daily volume on average during the seven years the HSBC discrepancy existed, with typical peaks in the range of 60% to 80% and some as high as 100% to 120%.

IV. Other Mispriced Pairs

HSBC is unique because it has all the advantages of twins and stubs. Yet, do self-imposed limits to arbitrage help explain other pairs too?

A. 3Com/Palm

As Lamont & Thaler (2003) describe, 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 spun-off its remaining shares of Palm on July 27.

Figure 8 shows how the relative price of 3Com to Palm seems to comove with the opposite relative volume. Each 3Com share is entitled to 1.5 Palm shares, so 3Com is exactly worth its holdings of Palm when the log price ratio is ln(1.5) = 0.4, which occurs in May, 2000. I use Center for Research in Security Prices (CRSP) data for prices and volumes on all stubs in this paper to match Lamont & Thaler (2003). The scatterplot in Figure 9 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.

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 Palm's price and its volume, consistent with a theory of convenience yield. However, that correlation is between the absolute levels of Palm's price and volume. The negative correlation we see here and in HSBC are between the relative price and the

Table VI. Market Regressions

This table 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^2.

Pair	$\beta = p$	Std. Err.	*t*-Stat	R^2
HSBC New-Old	-22%	5%	-4.2	5%
3Com/Palm	-43%	15%	-2.9	19%
RD-Shell 2002-2007	-25%	7%	-3.4	6%
Unilever 2002-2007	-23%	6%	-3.8	5%
Reed-Elsevier 2002-2007	-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%

relative volume.[10] Table V summarizes the four regression results to ease comparison with the similar Table II for HSBC.

B. Other Twins and Stubs

I also run the analysis for the five years of daily data from April 25, 2002 through April 24, 2007 for three other dual-listed companies, namely, Royal Dutch-Shell, Unilever, and Reed-Elsevier.[11] All had one share class trading in London and one in the Netherlands, and long-standing equalization agreements. These last two pairs also highlight that the discrepancy depends on more than just the country of residence because 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 & 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, 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 \$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 VI 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 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 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

[10] Why the seeming discrepancy? Because 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. We can see why this can occur if we let $p_i = a_i + b_i v_i + \epsilon_i$ for i=1,2 denote the regression results of log prices $p_i = \log P_i$ on log volumes $v_i = \log V_i$. Then the regression of log relative price on log relative volume is: $p_1 - p_2 = (a_1 - a_2) + \frac{b_1 v_1 - b_2 v_2}{v_1 - v_2}(v_1 - v_2) + (\epsilon_1 - \epsilon_2)$, and the regression coefficient can be anything, as it depends among other things on the correlation between the log relative volume and the relative residuals ϵ_1-ϵ_2. In the case of 3Com/Palm, that correlation is -0.85.

[11] Volume data prior to this time does not always agree between Datastream and Bloomberg so I focus only on the most recent data. I also 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.

Table VII. Implied Convergence Probabilities

This table shows the market implied convergence probability 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 securities in the given pair.

Pair	μ	σ	$\hat{p}(100)$	Traded Value
HSBC New-Old	1.01	0.81	70bps	£24 million
3Com/Palm	0.78	0.70	340bps	$230 million
RD-Shell 2002-2007	0.27	0.38	120bps	€300 million
Unilever 2002-2007	0.40	0.44	60bps	€130 million
Reed-Elsevier 2002-2007	0.57	0.35	0bps	€34 million

using the entire history, because it is unclear if there was ever a time that the discrepancy was within the transactions costs bounds.

Table VII 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 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 VII, Reed-Elsevier's trading volume was far lower than that of the two other London-Dutch multiple share classes, Royal Dutch-Shell and Unilever.[12]

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, 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 Internal Revenue Service (IRS). If the probability of convergence within six months was 99%, then the implied daily probability p would be 359 basis points, close to the 340 basis points implied for $N = 100$.

We can extend the implied probability results of Table VII 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 10 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 VIII lists the results.[13]

We can now address the *constancy* implication of self-imposed limits to arbitrage described in Section III. 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

[12] Reed-Elsevier's average traded value was similar to that of HSBC, but the HSBC discrepancy occurred nearly ten years earlier, and volume had 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 a unique multiple share class opportunity where both share classes traded on the same exchange and in the same country, suggesting that it is not the best comparison for Reed-Elsevier in terms of critical volume.

[13] 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. Also note that the final two rows of Table VIII 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 discussed towards the end of Section III.

Figure 10. 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 VII. This figure shows the plot of N vs. p.

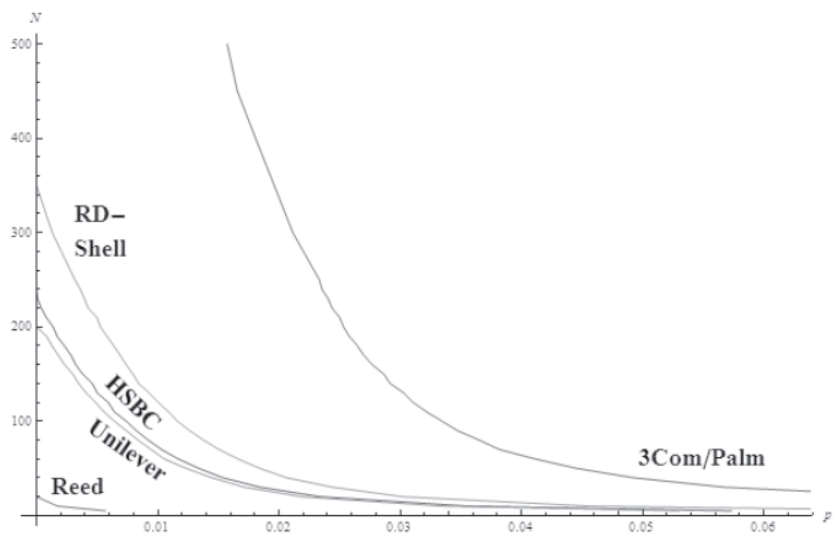


Table VIII. Implied Maximum Positions

This table 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

identical suggests that it is indeed self-imposed limits to arbitrage driving these discrepancies across the different pairs.

V. 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, but not fungible. The absence of fungibility created room for mispricing, which swung from an 8% discount to an 8% premium.

Often pricing discrepancies persist because of external limits to arbitrage, 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.

Much of the discrepancy is explained by the relative trading volume, a novel puzzle. The share class that had a higher volume relative to the other one tended to have a relatively lower price. 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. One possible explanation of the HSBC discrepancy is self-imposed limits to arbitrage, explaining both why the pricing discrepancy was able to persist and why the relative volume was negatively related to the relative price.

A model of self-imposed limits to arbitrage generates plausible values of the market implied amount of arbitrage activity in HSBC, averaging around 20% of daily volume. The model further 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. Also, 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.

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.■

References

Benartzi, S. and R.H. Thaler, 1995, "Myopic Loss Aversion and the Equity Premium Puzzle," *Quarterly Journal of Economics* 110 (No. 1), 73-92.

Cochrane, J.H., 2003, "Stock as Money: Convenience Yield and the Tech-stock Bubble," in W. C. Hunter, G. G. Kaufman & M. Pomerleano, Eds., *Asset Price Bubbles,* MIT Press, Cambridge, MA, 175-203.

de Jong, A., L. Rosenthal, and M.A.V. Dijk, 2009, "The Risk and Return of Arbitrage in Dual-Listed Companies," *Review of Finance* 13 (No. 3), 495-520.

Dow, J. and G. Gorton, 1994, "Arbitrage Chains," *Journal of Finance* 49 (No. 3), 819-849.

Drachter, K., A. Kempf, and M. Wagner, 2007, "Decision Processes in German Mutual Fund Companies: Evidence from a Telephone Survey," *International Journal of Managerial Finance* 3 (No. 1), 49-69.

Ellerton, C., 1997, "HSBC: Mind the Gap," *Analyst report, SBC Warburg*.

Froot, K.A. and E.M. Dabora, 1999, "How are Stock Prices Affected by the Location of Trade?" *Journal of Financial Economics* 53 (No. 2), 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."

Nicholas, Joseph G., 2004, *Hedge Fund of Funds Investing: An Investor's Guide,* Bloomberg Press, New York, NY.

Kondor, P., 2009, "Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading," *Journal of Finance* 64 (No. 2), 631-655.

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 (No. 2), 227-268.

Mitchell, M., T. Pulvino, and E. Stafford, 2002, "Limited Arbitrage in Equity Markets," *Journal of Finance* 57 (No. 2), 551-584.

Newey, W.K. and K.D. West, 1994, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies* 61 (No. 4), 631-653.

Shleifer, A. and R.W. Vishny, 1997, "The Limits of Arbitrage," *Journal of Finance* 52 (No. 1), 35-55.

Thaler, R.H. and H.M. Shefrin, 1981, "An Economic Theory of Self-Control," *Journal of Political Economy* 89 (No. 2), 392-406.

World Bank, 2005, "Project Information Document Report AB1634, Appraisal Stage."

Zeileis, A., 2004, "Econometric Computing with HC and HAC Covariance Matrix Estimators," *Journal of Statistical Software* 11 (No. 10), 1-17.