

---

# Relative Implied-Volatility Arbitrage with Index Options

Manuel Ammann and Silvan Herriger

*In the study reported here, we investigated the efficiency of markets as to the relative pricing of similar risk by using implied volatilities of options on highly correlated indexes and a statistical arbitrage strategy to profit from potential mispricings. We first analyzed the interrelationships over time of the three most highly correlated and liquid pairs of U.S. stock indexes. Based on this analysis, we derived a relative relationship between implied volatilities for each pair. If this relationship was violated (i.e., if we detected a relative implied-volatility deviation), we suspected a relative mispricing. We used a simple no-arbitrage barrier to identify significant deviations and implemented a statistical arbitrage trade each time such a deviation was recorded. We found that, although many deviations can be observed, only some of them are large enough to be exploited profitably in the presence of bid-ask spreads and transaction costs.*

---

Arbitrage relationships in derivatives markets have been studied extensively. For example, option boundary conditions, as derived by Stoll (1969) and Merton (1973), have been the subject of numerous empirical studies; examples are Gould and Galai (1974), Klemkosky and Resnick (1979), and Ackert and Tian (1998, 1999). Index arbitrage has also been thoroughly investigated. Empirical studies include those of Figlewski (1984), Chung (1991), Sofianos (1993), and Neal (1996). Figlewski (1989) provided an example of option arbitrage in imperfect markets. Clearly, the testing of market efficiency in derivatives markets by using arbitrage relationships has drawn a great deal of interest.

Statistical arbitrage, however, be it in derivatives or other markets, has received surprisingly little attention in the literature, despite its high practical relevance. A possible reason may be the nature of the mispricings underlying statistical arbitrage. Statistical arbitrage is not based on theoretical, exact pricing relationships but, rather, on empirical, statistically established relationships. Consequently, statistical arbitrage involves risk. Omitting the study of such forms of pricing relationships from research agendas altogether, how-

ever, may lead to an incomplete understanding of market mechanisms and thus of market efficiency.

One study that used the statistical arbitrage approach to test market efficiency in equity markets is that of Gatev, Goetzmann, and Rouwenhorst (1999), who investigated the relative pricing mechanisms of securities that are close economic substitutes. In addition, motivated by the widespread intermarket hedging activities in commodities markets, a number of authors have analyzed various pricing relationships for commodity spreads. This interest explains the presence of several papers, such as Johnson, Zulauf, Irwin, and Gerlow (1991) and Poitras (1997), that applied statistical arbitrage to such markets.

A statistical arbitrage approach to test the efficiency of options markets has not yet been attempted. Thus, the aim of this study was to devise and implement a statistical arbitrage strategy for testing an aspect of market efficiency that the classical boundary conditions for options fail to reveal—namely, the efficiency of markets in pricing relative risk in highly correlated markets.

## Theory and Data

In what situations is a statistical arbitrage strategy possible and likely to lead to profitable trading? If two indexes are highly correlated (because of a securities overlap or other reasons), one should be able to calculate the relationship between their volatility levels. A similar relationship must also be

---

*Manuel Ammann is professor of finance at the University of St. Gallen, Switzerland. Silvan Herriger is an index derivatives trader at J.P. Morgan, London.*

valid for the implied-volatility levels of the respective index options. If the relationship between the implied volatilities is significantly different from the relationship observed between the two index volatilities, the option prices are misaligned, which should not occur in efficient markets. In such a case, a statistical arbitrage strategy can be implemented to take advantage of the relative implied-volatility deviation.

To test these ideas, we used significantly related U.S. equity indexes. Listed options are available for 11 stock indexes in the United States, and several of these indexes are closely related. This close relationship is often a result of the same stocks being included in several indexes; for example, every stock in the S&P 100 Index is also included in the S&P 500 Index. We studied the time period from January 1995 through February 2000.

Our statistical arbitrage methodology consisted of the following consecutive steps:

- First, after ensuring that the time-series returns were stationary, we calculated the correlations of the various index pairs. We then selected the pairs with the highest correlation coefficients and no longer considered the other indexes.
- Second, we studied the relationship of the daily returns of the index pairs by running OLS (ordinary least-squares) regressions to establish the past relationships between them. We also tested the robustness of the relationships. Because the linear relationship between two indexes is time varying, we estimated statistical boundaries for the OLS coefficients.
- Third, we established a conditional forecast of future variance based on the past relationship between the indexes' returns. The reason was to test, out of sample, the predictive powers of the boundaries estimated in the second step.
- Once the predictive capacity of the boundaries was confirmed for the historical volatilities, we applied the estimated relationship to implied volatilities, for which a similar relationship should prevail.
- Based on the implied volatilities and on the riskless rate recorded every trading day, we calculated the corresponding option prices. We incorporated bid-ask spreads in the process to ensure that, should we identify a deviation of a certain significant magnitude, we could implement and test an option strategy that took advantage of the suspected mispricing.
- Finally, we implemented a simple arbitrage trading strategy.<sup>1</sup>

The 11 stock indexes for which exchange-traded options are available in the United States are the S&P 500 Index (SPX), the S&P 100 Index (OEX),

the Nasdaq 100 Index (NDX), the NYSE Composite Index (NYA), the Philadelphia U.S. TOP 100 Index (PTPX), the Philadelphia Stock Exchange Utility Sector (UTY), the S&P Smallcap 600 Index (SML), the S&P Mid Cap 400 (MID), the Amex Major Market Index (XMI), the Russell 2000 Index (RTY), and the Dow Jones Industrial Average (INDU). These indexes formed the pool from which we chose index pairs for further study.

Following Harvey and Whaley (1991), we used for this study the implied volatility of at-the-money options with the shortest maturity. At-the-money options contain the most information about volatility. Also, we used the "front month" options because they are the most liquid. If fewer than 20 calendar days were left to expiration, we used the next available series. Thus, the implied volatilities were calculated from options ranging from 20 to 50 calendar days to expiration, or an average of 35 calendar days. Because 35 calendar days represent exactly 5 weeks and an average week has 5 trading days, an average of 25 trading days was used in the calculations. We based the calculation of the implied volatilities on the closing prices of the options and on the closing price levels of the indexes. We derived the volatility from an average of the implied volatilities of at-the-money options.

The riskless interest rate was used to calculate the daily option prices, and because the options had, on average, 35 calendar days left to expiration, we used the one-month Eurodollar LIBOR as the riskless rate.

## Choosing Indexes

The criteria for selection of the specific indexes to study were (1) stationarity of the index returns, (2) high correlation, and (3) liquidity of the market for index options.

To avoid spurious correlations and regressions, we first tested each index time series for stationarity. The standard stationarity tests revealed that all return time series (based on continuously compounded returns) can be considered stationary except the Philadelphia UTY. We thus dropped the UTY from further analysis.

We then calculated the correlation coefficients for the remaining 10 indexes. This criterion was motivated by the conjecture that index pairs with high correlations exhibit a strong linear relationship with each other. The correlation matrix is in **Table 1**.

We set a minimum of 0.95 for the correlation coefficient as a criterion for inclusion in our index sample, which screened out all but five pairs of

**Table 1. Correlation Matrix, 31 March 1995 through 3 December 1999**

	SPX	OEX	NDX	NYA	PTPX	INDU	SML	MID	XMI	RTY
SPX	1	<b>0.988</b>	0.780	<b>0.988</b>	0.798	0.922	0.709	0.804	0.91	0.729
OEX		1	0.770	<b>0.968</b>	0.799	0.933	0.662	0.759	0.923	0.684
NDX			1	0.721	0.563	0.633	0.688	0.745	0.593	0.708
NYA				1	0.785	0.933	0.757	0.846	0.919	0.772
PTPX					1	0.896	0.653	0.748	0.892	0.676
INDU						1	0.637	0.750	<b>0.974</b>	0.651
SML							1	0.922	0.583	<b>0.983</b>
MID								1	0.700	0.922
XMI									1	0.595
RTY										1

Note: Correlation coefficients based on weekly returns.

indexes (those in boldface in Table 1) to be considered for further calculations.<sup>2</sup> Of these five pairs, we chose the three indexes with the most-liquid options markets to ensure that potential arbitrage trading strategies could be executed. The three chosen indexes are the SPX, OEX, and NYA.

That the S&P 500 is highly correlated with the S&P 100 is not surprising because the OEX is an integral part of the SPX. We expected the OEX to be more volatile than the SPX, however, because of their comparative respective constituents (100 stocks versus 500 stocks). We also expected the overlap of the SPX with the NYA to be large because many of the stocks in the SPX are listed on the NYSE. The NYA has more stocks than the SPX, so we expected the SPX to be slightly more volatile. The relationship between the OEX and the NYA is more surprising than the other relationships. Apparently, the OEX tracks the larger SPX so well that it also manages to track the NYA, which is even larger. We expected the OEX definitely to be more volatile than the NYA because of the greater diversification of the NYA.

## Relationships between Index Pairs

For every pair of indexes, we used OLS regression to regress the daily returns of one index onto the daily returns of the other:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i, \quad (1)$$

where

$Y_i$  = daily return of Index Y at time  $t$

$X_i$  = daily return of Index X at time  $t$

$\hat{u}_i$  = disturbance term

The sample used in every case was half a year of daily returns, or 125 trading days, from the period January 1995 through February 2000.<sup>3</sup> Consequently, the first regression was made 125 days

into the data, with the use of the past 125 daily returns, and for every day after this regression, the regression was run again using the previous 125 days. The result was a rolling 125-day regression in which the oldest data point was dropped every time a new one was added.

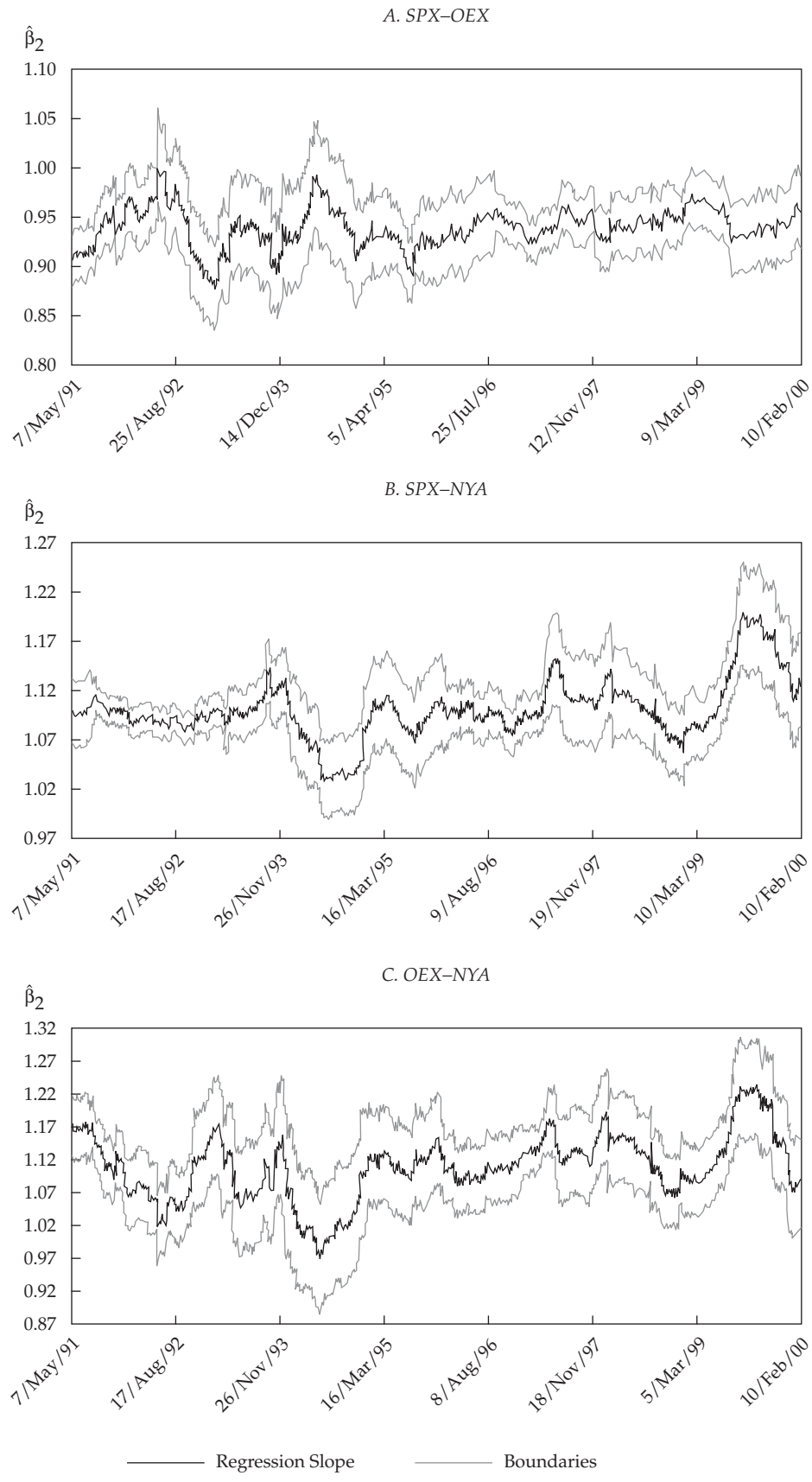
Because the regressions were rolled day by day for a long period for every index pair, a large number of regressions had to be calculated. The panels in **Figure 1** present the resulting graphs of regression coefficient  $\hat{\beta}_2$  plus the calculated lower and upper boundaries (to be described).

We used  $t$ -tests to test the significance of the regression coefficients and found each estimated slope coefficient to be significantly different from zero at the 99 percent confidence level. Because the results regarding the significance of the estimated slope coefficients were very similar for the several thousand regressions, we do not present these results in detail.

The situation is different for the estimated  $\hat{\beta}_1$  intercepts. For the large majority, the hypothesis that they were not different from zero could not be rejected at a 95 percent level of confidence. In the rare cases in which the estimated intercepts were found to be statistically significant, the recorded value of the intercept was very low (for practical purposes, zero).

The boundaries for Figure 1 were set as follows. The lowest and highest recorded  $R^2$  values for the coefficients of determination for all regressions are in **Table 2**. The coefficients confirm a strong linear relationship between the selected indexes. The relationships of the examined index pairs are not constant, however, over time. This simple observation had a profound impact on this study because the relative volatility forecasts had to be based on these relationships. Time-invariant relationships between index pairs with consistently

**Figure 1. Regression Coefficient and Boundaries, 7 May 1991 through 10 February 2000**



**Table 2. Minimum and Maximum  $R^2$ s**

Index Pair	Minimum $R^2$	Maximum $R^2$
SPX-OEX	0.891	0.988
SPX-NYA	0.935	0.996
OEX-NYA	0.831	0.976

high  $R^2$  values would have been ideal. Then, establishing their interrelationships once would have sufficed to predict the relative volatilities for any future time interval. With both varying interrelationships and varying goodness of fit of the models, however, an alternative method had to be found to account for these instabilities.

The slope coefficients that were estimated from the previous 125 trading days were not sufficient as a prediction tool. Such point estimates are subject to estimation error because coefficients have been found to vary over time. Thus, an upper and a lower boundary are needed to render the estimated slope coefficients more robust as a prediction tool. Applying the method of interval estimation would be inappropriate because ordinary interval estimation makes a statement about the confidence level with which the calculated interval will contain the true slope coefficient; the true slope coefficient is thus assumed to be constant. In our study, because we were dealing with rolling regressions, the degree of variance of the slope coefficients over time also had to be considered when establishing boundaries for the estimators. Consequently, we chose, instead, an empirical boundary calculation based on a simple minimum-maximum approach.

To construct the boundaries, we recorded the largest variations of the relevant parameters during a time span that matched the forecasting horizon. This method reflects the market situation: In a volatile situation, when the index relationships vary greatly, the boundaries are wider; when the rela-

tionships are relatively constant, the boundaries are narrower.

Because 25 trading days formed the horizon of the various forecasting calculations, we recorded the largest percentage change of the estimated slope coefficients measured in any preceding 25-day interval during the preceding 250 trading days.<sup>4</sup> These extreme changes of the beta coefficient were then used for establishing minimum and maximum boundaries at each point in time to ensure robust beta forecasts. The boundary estimation methodology is illustrated in **Figure 2**.

The following equations were used to calculate the boundaries of the estimated slope coefficients at time  $t$ :

Lower boundary:

$$\hat{\beta}_{2\ low(t)} = \hat{\beta}_{2(t)} - \max(\Delta\hat{\beta}_2)\hat{\beta}_{2(t)}, \quad (2)$$

Higher boundary:

$$\hat{\beta}_{2\ high(t)} = \hat{\beta}_{2(t)} + \max(\Delta\hat{\beta}_2)\hat{\beta}_{2(t)}, \quad (3)$$

where  $\hat{\beta}_{2(t)}$  is the estimated slope coefficient at time  $t$  based on the sample  $t_{-124}$  to  $t$  and  $\max(\Delta\hat{\beta}_2)$  is the largest percentage change of the estimated slope coefficients measured over any 25-trading-day period during the subsample (250 trading days) preceding the time of estimation.<sup>5</sup>

### Volatility Relationships

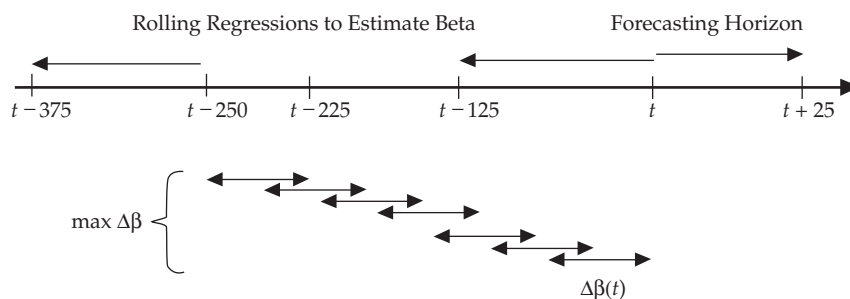
For the purpose of this study, the relationships that were established between the daily returns of index pairs needed to be transformed into relationships between the respective volatilities. For the relationship between random variables  $X$  and  $Y$  such that

$$Y = a + bX + u, \quad (4)$$

we obtained

$$\text{var}(Y) = \text{var}(a + bX + u). \quad (5)$$

**Figure 2. Beta Boundary Estimation**



Note: Because betas were estimated using 125 trading days of data, the first boundary estimation available was 375 days into the data.

By basic statistics, if  $a$  and  $b$  are constants and  $X$  and  $u$  are independent random variables,

$$\text{var}(Y) = b^2 \text{var}(X) + \text{var}(u). \quad (6)$$

Applied to the regression outputs, this expression can be written as

$$\text{var}(R_{Yi}) = \hat{\beta}_2^2 \text{var}(R_{Xi}) + \text{var}(\hat{u}_i). \quad (7)$$

Based on this relationship, we attempted a forecast of the future interrelationships of the index pairs. Equation 7 indicates that two inputs need to be estimated to forecast the variance of  $R_{Yi}$ —the slope coefficients,  $\hat{\beta}_2$ , and the variance of the sample disturbance term,  $\hat{u}_i$ . Estimates of the slope coefficients were described previously. To estimate  $\hat{u}_i$ , we chose an approach similar to the estimation of  $\hat{\beta}_2$ : Both the smallest and the largest 25-day variance of the residuals during the 250 trading days preceding a point of estimation were recorded. Formally,

Lower boundary:

$$\text{var}(\hat{u}_i)_{low} = \min[\text{var}_{25 \text{ days}}(\hat{u}_i)], \quad (8)$$

Higher boundary:

$$\text{var}(\hat{u}_i)_{high} = \max[\text{var}_{25 \text{ days}}(\hat{u}_i)]. \quad (9)$$

Once all necessary parameters had been estimated, we proceeded to test their forecasting power.

**Out-of-Sample Tests of Volatility Boundaries.** To test the volatility boundaries around  $\hat{\beta}_2$  calculated previously, we used the boundaries to make daily rolling forecasts of the variable to be explained. In other words, we used the established relationships between the indexes to forecast their relative future volatilities over a 25-trading-day period and then compared this forecast with the actual recorded volatility levels. We applied the statistical properties introduced in Equations 6 and 7. For example, to test the established SPX–OEX relationship at time  $t$ , we applied the following equation:

$$\begin{aligned} & \left[ \hat{\beta}_2_{low(t)} \right]^2 \text{var}(OEX) + \text{var}(\hat{u}_t)_{low} \\ & \leq \text{var}(SPX) \\ & \leq \left[ \hat{\beta}_2_{high(t)} \right]^2 \text{var}(OEX) + \text{var}(\hat{u}_t)_{high} \end{aligned} \quad (10a)$$

where  $\text{var}(OEX)$  and  $\text{var}(SPX)$  are the realized future 25-day variances.

If the realized SPX variance was within the forecasted boundaries, we deemed the forecasting successful. This test was rolled daily, similarly to

the regressions. We took the percentage of successful forecasts as an indication of the forecasting power of the established boundaries.

The results of this out-of-sample test for the boundaries with four tolerance levels are in **Table 3**. We used tolerance levels to widen the boundaries and thus make the forecasts more robust. For a given tolerance  $\psi$ , the forecast was successful if

$$\begin{aligned} & \left[ \hat{\beta}_2_{low(t)} \right]^2 \text{var}(OEX) + \text{var}(\hat{u}_t)_{low} - \psi \\ & \leq \text{var}(SPX) \\ & \leq \left[ \hat{\beta}_2_{high(t)} \right]^2 \text{var}(OEX) + \text{var}(\hat{u}_t)_{high} + \psi. \end{aligned} \quad (10b)$$

Clearly, the wider the boundaries are, the higher the probability is that the future volatility will fall between those boundaries. This fact is important when evaluating the power of such a forecasting test: The ideal would be a narrow boundary that includes all future volatility readings, but the narrower the boundary, the smaller the probability, all else being equal, of including the future volatility. This opposition in the goals is why we included tolerance levels in Table 3; because we are dealing with statistical arbitrage, the forecasting need not be perfect, but it must be sufficiently close to allow for arbitrage trades to be triggered when a deviation has been detected and thus a mispricing is suspected.

Keep in mind the interpretation of Figure 1. A result of 100 percent with no tolerance and a narrow boundary band would indicate, in effect, the existence of a quasi-deterministic relationship between the considered indexes; that is, knowing the volatility of one would be enough to determine precisely and repeatedly the volatility of the other. That such a result is not shown in Table 3 is not surprising, but the table does indicate that, with a 1 percent tolerance level, the percentage of successful forecasts is more than 90 percent for all three index pairs. Considering the relatively narrow boundaries depicted in Figure 1, this degree of precision is satisfactory.

The next step is to apply the inferences made so far regarding relative future volatility of index pairs to their relative implied-volatility levels.

**Arbitrage Boundaries for Relative Implied Volatilities.** In the previous sections, we developed a method that, based on the past relationship of daily returns, provides a robust forecast of the future relative volatility of three index pairs. In this section, the lower and higher boundaries of these relationships are used to identify relative implied-volatility deviations in the option markets.

**Table 3. Successful Out-of-Sample Relative-Volatility Forecasts from Established Relationship Boundaries, 7 May 1991 through 10 February 2000**

Index Pair	Tolerance Level			
	Zero	0.25	0.50	1.00
SPX–OEX	68.1%	83.6%	90.8%	97.7%
SPX–NYA	60.6	82.6	92.4	97.3
OEX–NYA	68.0	76.7	84.5	91.8

We want to stress that we are not postulating any relationship between the magnitude of the historical volatility and the implied volatility.<sup>6</sup> We are addressing, rather, the relationship between the *relative* volatilities—that is, the relative difference in volatility levels, whether historical or implied. In other words, we are establishing a relationship not between implied and realized volatility but between the implied volatility of options on two distinct indexes. Therefore, once we have established a precise relationship between the relative future volatility of index pairs, this relationship must also hold for the relative implied volatility of options on the two indexes. The boundaries calculated for the relative future volatility must also hold for relative implied volatility, and Equation 10a must also apply to implied volatilities.

Using the previous example of OEX and SPX options, the relationship we were examining is given by

$$\begin{aligned}
 & \left[ \hat{\beta}_{2 \text{ low}(t)} \right]^2 \text{var}_{\text{impl}}(\text{OEX}) + \text{var}(\hat{u}_t)_{\text{low}} \\
 & \leq \text{var}_{\text{impl}}(\text{SPX}) \\
 & \leq \left[ \hat{\beta}_{2 \text{ high}(t)} \right]^2 \text{var}_{\text{impl}}(\text{OEX}) + \text{var}(\hat{u}_t)_{\text{high}},
 \end{aligned} \tag{11}$$

where  $\text{var}_{\text{impl}}(\text{OEX})$  and  $\text{var}_{\text{impl}}(\text{SPX})$  are the observed implied variances (squared implied volatilities) of at-the-money options with 25 trading days left to expiration. If such a boundary was violated, a theoretical mispricing was suspected.<sup>7</sup> The nature of this mispricing is not absolute; it is a relative mispricing (i.e., the relative difference of the implied volatilities of a particular index pair at a specific time differs significantly from the calculated relative future volatility difference between the two indexes).

The size of the theoretical deviation (suspected mispricing) was defined as the difference between the violated boundary and the observed implied volatility. The deviations for call and put options for the three combinations and for every trading day are presented in **Figure 3**, **Figure 4**, and **Figure 5**.

In testing the use of deviations in possible arbitrage trading strategies, because of transaction

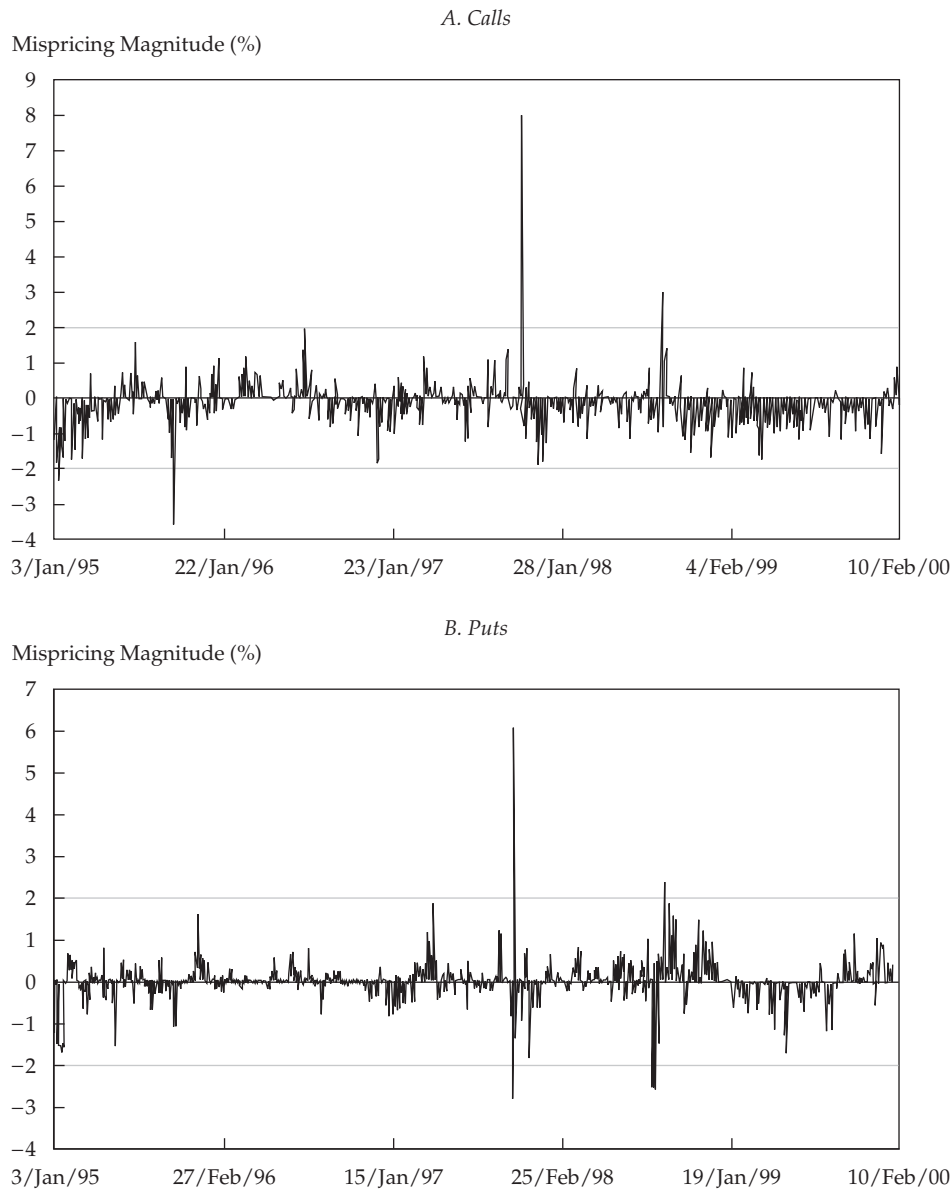
costs, we introduced a security margin similar to the tolerance level in Table 3 to identify significant deviations before an arbitrage trade was initiated. This security margin can be interpreted as a form of no-arbitrage barrier. Its magnitude was fixed at two times the bid–ask spread of at-the-money options. Whenever we observed such a significant deviation, we simulated a statistical arbitrage trade. In other words, the relative implied volatility falling outside the bounds implied by Equation 11 was not sufficient to signal a trade. The deviation also had to be of a certain minimum size before it was considered a statistical arbitrage opportunity. This rule underscores the conservative approach we took to identifying potential trading opportunities. The no-arbitrage security margins are identified by the shaded lines in Figures 3, 4, and 5.

Somewhat surprisingly, the number of theoretical deviations (thus, suspected mispricings) shown in Figures 3–5 is rather large. Most deviations failed to surpass the no-arbitrage barrier, however, and were thus not considered substantial enough to represent arbitrage opportunities.

Note that the deviations seem to have occurred in clusters. In certain periods, one index was persistently over-/undervalued relative to the other as to its implied volatility.<sup>8</sup> A particularly clear example is the OEX–NYA relationship depicted in Figure 5. In such periods, the market seems to be persistently mistaken by failing to recognize the correct relationship between the future volatilities of the indexes. Under our conservative rules, however, this persistent deviation could not be eliminated by arbitrage as long as it stayed inside the no-arbitrage barrier formed by the security margins.

We have not been able to identify factors that could cause the concentration of volatility deviations in phases. The explanation that simply links an increase of deviations to an increase in market volatility, although intuitively appealing, is unsatisfactory for two reasons. First, although deviations (significant or not) appear to be slightly more frequent in highly volatile markets, they can also be observed in periods of low volatility. Second, linking the concentration of deviations to volatility

**Figure 3. SPX–OEX Relative Implied-Volatility Deviations, 3 January 1995 through 10 February 2000**



levels fails to account for the fact that the observed deviations are persistent overvaluations of one index vis-à-vis the other in some phases and undervaluations in other phases. An unknown factor seems to be influencing the subjective relative risk perception of market participants in phases.

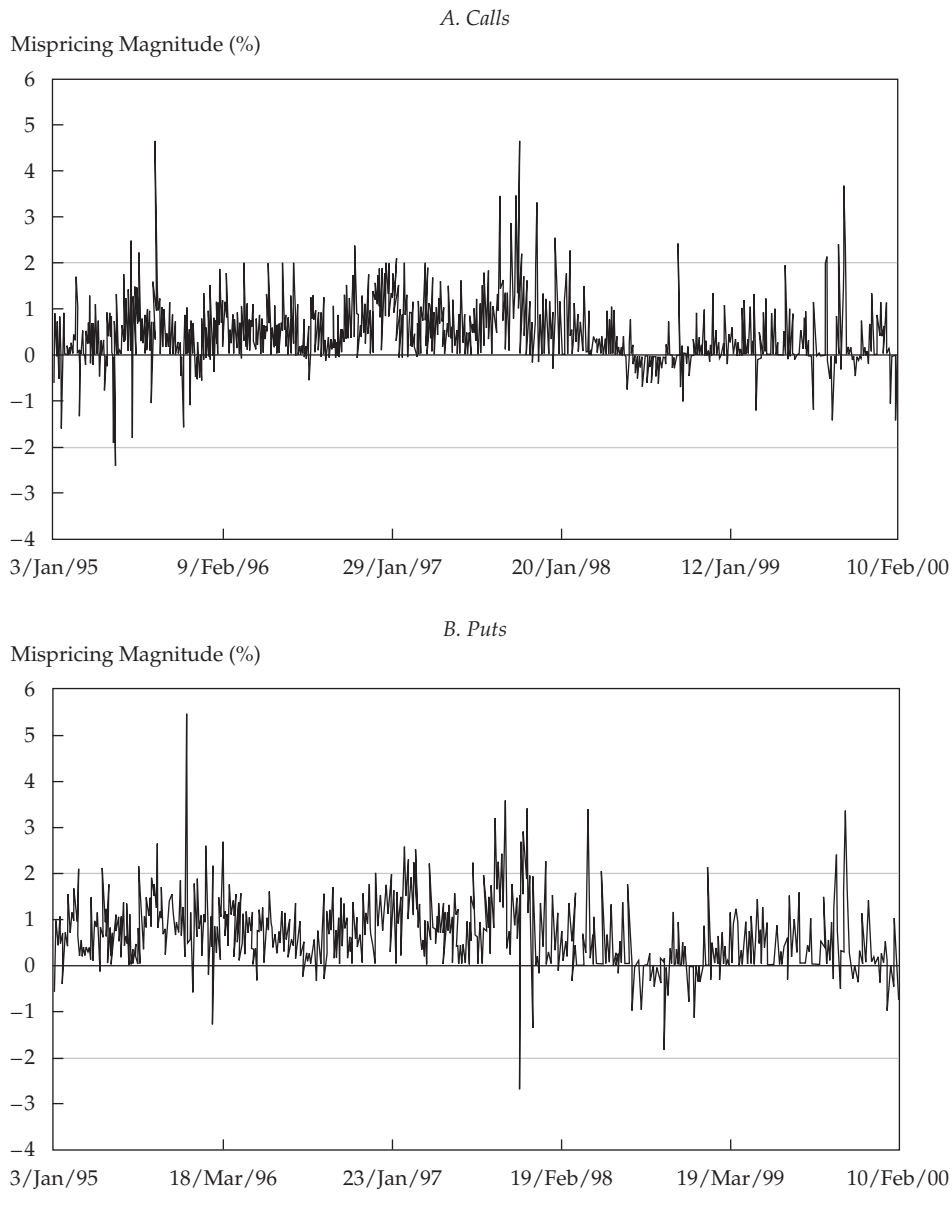
### Arbitrage Trading Strategy

To test whether the suspected mispricings can be used for profitable arbitrage trading, we implemented an arbitrage strategy using at-the-money options. Option bid–ask prices were calculated from implied-volatility data by using a bid–ask spread (in

terms of implied-volatility percentage) around the volatility implied by the closing price of the options. The bid–ask spread we used was 1 percent, which for near-month at-the-money options in liquid options markets is a conservative assumption.<sup>9</sup> We used 35 calendar days to expiration and the relevant one-month Eurodollar LIBOR. Because the SPX options are European-style options and thus not exercisable prior to maturity, we used the Black–Scholes option-pricing model for them. For the American-style OEX and NYA options, we used a binomial tree that incorporated the early exercise feature.



**Figure 4. SPX–NYA Relative Implied-Volatility Deviations, 3 January 1995 through 10 February 2000**



When a volatility deviation was identified as significant, we sold one at-the-money option of the overvalued index and bought the point-equivalent,  $\hat{\beta}_2$ -adjusted amount of at-the-money options of the undervalued index.<sup>10</sup>

Analytically, for the SPX–OEX example, if either

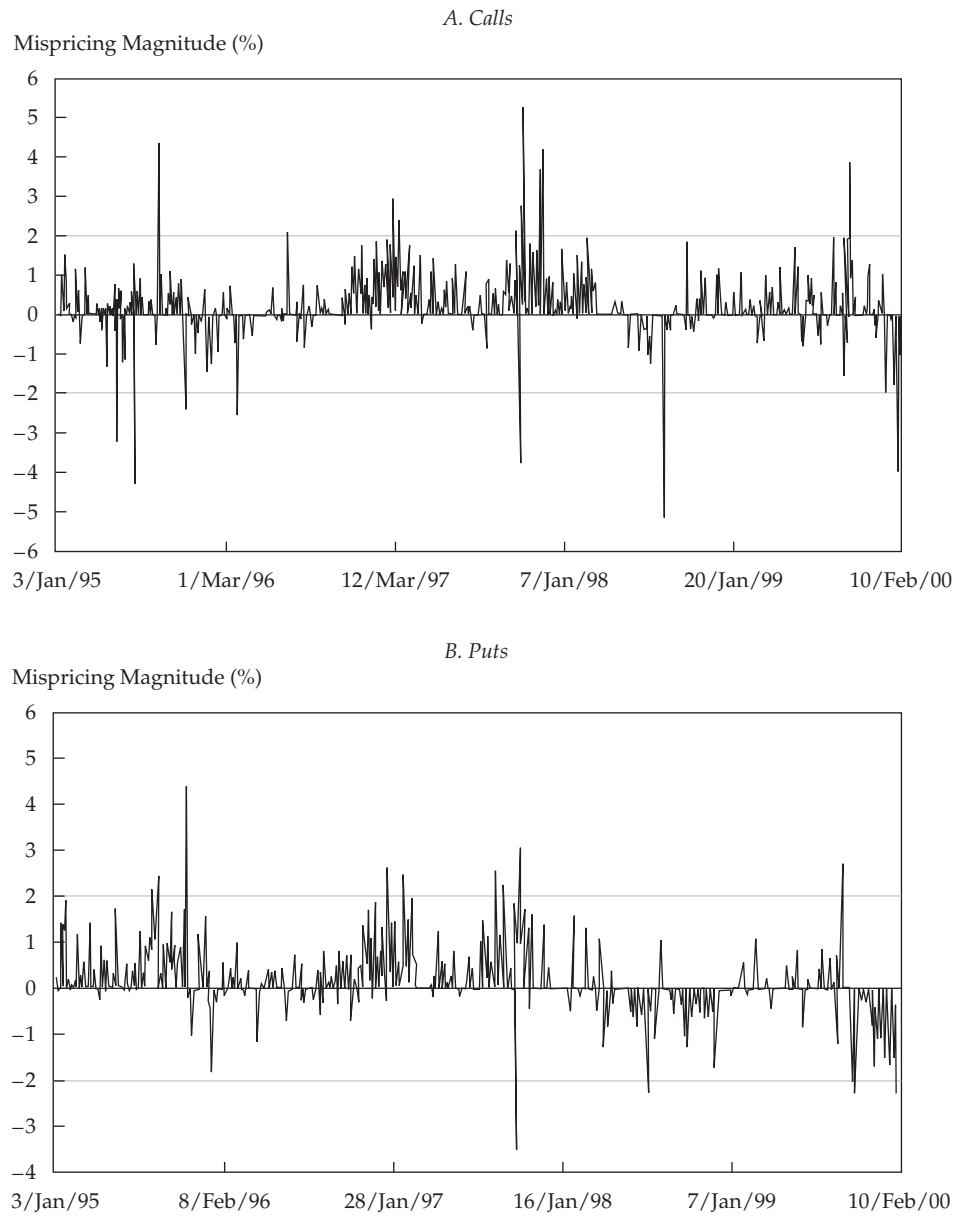
$$\left\{ \left[ \hat{\beta}_{2\ low(t)} \right]^2 \text{var}_{impl}(OEX)_t + \text{var}(\hat{u}_t)_{low} \right\} - \text{var}_{impl}(SPX)_t \geq \psi\% \quad (12a)$$

or

$$\text{var}_{impl}(SPX)_t - \left\{ \left[ \hat{\beta}_{2\ high(t)} \right]^2 \text{var}_{impl}(OEX)_t + \text{var}(\hat{u}_t)_{high} \right\} \geq \psi\%, \quad (12b)$$

we opened an arbitrage trade. If the first (second) condition was satisfied, we entered into a short (long) position of  $\hat{\beta}_2(SPX/OEX)$  at-the-money options on the OEX and a long (short) position of one at-the-money option on the SPX.<sup>11</sup> The terms  $SPX_t$  and  $OEX_t$  represent the index levels, and  $\psi$  is the security margin (in volatility percentage). A

**Figure 5. OEX–NYA Relative Implied-Volatility Deviations, 3 January 1995 through 10 February 2000**



simple example in **Exhibit 1** explains this procedure.

Translated from the equations, whenever we observed a significant deviation, we sold one at-the-money option (put or call) on the relatively overvalued index at the bid price (short position) and bought the point-equivalent,  $\hat{\beta}_2$ -adjusted amount of at-the-money options on the undervalued index at the ask price (long position).<sup>12</sup> We held the two positions without rebalancing for delta neutrality until either (1) the deviation disappeared (i.e., the observed implied volatility of the indexes went back inside the calculated bound-

aries) or (2) the options expired before the deviation disappeared.

If the volatility deviation disappeared before expiration, we unwound the position by buying back the previously overvalued option at the ask price and selling the previously undervalued amount of options at the bid price. If the deviation persisted until expiration of the options, we calculated and summed up the intrinsic values of the options. The total cash flows upon opening and closing the positions (interest on cash flows was neglected) could then be calculated.

If the method used to define the interrelationship of the indexes and to forecast their relative

**Exhibit 1. Example Offsetting Hedge Strategy**

The procedure was as follows: Assume that

- Index A has 100 points,
- Index B has 1,000 points,
- one at-the-money call on Index A is worth \$8 (35 days to expiration),
- one at-the-money call on Index B is worth \$170 (35 days to expiration),
- the relative-volatility differential (the estimated slope coefficient,  $\hat{\beta}_2$ ) of Index A to Index B is 2 (for every percentage A moves, B will move by 2 percent), and
- Index A and Index B are perfectly correlated.

If the implied volatility on Index B is overvalued relative to the implied volatility on Index A (meaning that the relative implied-volatility differential is higher than 2), the following position is taken: One sells 1 Index B call (+\$170). Then, to render the positions point equivalent, one must buy 10 (i.e., 1,000/100) calls on Index A (because A is relatively undervalued); the result is  $10 \times -\$8 = -\$80$ . This point-equivalent position then has to be adjusted by the relative-volatility differential (the  $\hat{\beta}_2$  adjustment) to keep the offsetting hedge delta and gamma neutral. This adjustment is achieved by multiplying the long position by the relative-volatility differential (2 in this example). A net long position of -\$160 and a net short position of +\$170 would be the result of this trade.

volatility relationship is correct, the sum of the cash flows should, on average, be positive. This statistical arbitrage strategy may not be profitable for every single trade, however, for two reasons. First, the option positions were constructed to be delta and gamma neutral under the assumption that the underlying indexes vary according to the statistical inferences made about their relationships. When the index pairs failed to behave as predicted, the option position gained in delta exposure, which can result in a losing trade even though the observed implied-volatility deviation may have disappeared. Second, an inherent risk in statistical arbitrage is that an observed deviation, however large, may persist or even increase. The statistical arbitrageur must always be aware of this risk when deciding on the size of an arbitrage trade.

Table 4 shows the results of the statistical arbitrage strategy. Only 10 SPX–OEX statistical arbitrage trades were triggered over the observed time period; trades involving NYA options occurred more often. The reason could be that the relationship between the SPX and the OEX indexes is more obvious and thus more widely monitored by market participants.

For every index pair, the average profit/loss per trade was positive. Although some trades generated losses, as expected for a statistical arbitrage

strategy, most were profitable. Because we used only daily closing prices and implicitly assumed an ability to trade at the signal price, the profits of an actual trading strategy could be different from our results.

One trade needs to be singled out because of its magnitude: As Figure 3 (volatility deviations of the SPX–OEX pair) shows, a significant deviation was identified on 27 October 1997, when the SPX and the OEX lost 7.11 percent and 7.09 percent, respectively, followed by a bounce on the following day of 4.99 percent and 5.61 percent, respectively. Clearly, in such times, the probability of a relative deviation increases. Bid–ask spreads are obviously wider in such times, but even if one were to double the spread on these days, a significant deviation would have been identified and the ensuing trade would have remained profitable. If we assumed that no trading in options was possible at all during those two days, the average profit/loss for the SPX–OEX index pair shown in Table 4 would fall greatly. The average would be \$0.61 achieved from 8 instead of 10 trades (on 27 October 1997, the strategy called for entering into both call and put positions). Nevertheless, the value from the strategy would still be significantly positive and of a magnitude comparable to that achieved by the other two index pairs.

**Table 4. Profitability of Arbitrage Trades, 3 January 1995 through 10 February 2000**

Index Pair	Number of Trades	Number of Winning Trades	Average Profit	Number of Losing Trades	Average Loss	Average Time Held	Total Average Profit/Loss
SPX–OEX	10	9	\$3.62	1	-\$0.12	2.9 days	\$3.24
SPX–NYA	37	31	2.00	6	-1.57	9.6	1.42
OEX–NYA	28	22	1.07	6	-0.84	3.6	0.66

Notes: Trades effectuated every time a mispricing was larger than 2 percent. Profit per profitable position and average profit per arbitrage trade were normalized to a position of one option of the first-named index of the pair.

Of the two potential risks involved with this strategy (persistence of volatility deviation and varying relationship of indexes), persistence, surprisingly, never materialized. The theoretical deviations seem to have been strongly mean reverting, with the mean defined as the historical relationship between relative volatility levels. No significant relative deviation lasted longer than 35 days.

The experience in this study of the other potential risk, however, illustrates the difficulty of making exact statistical inferences about the future relationships of stock indexes. All the losing positions were caused by indexes varying in a manner different from what was predicted by a linear relationship. Those positions ended up in the loss column even though the implied-volatility deviation disappeared. The reason is that we took the option positions in such a way as to ensure they remained delta neutral as long as the indexes varied relative to one another according to their past relationship. If they ceased to vary this way during the time the option position was open, a delta exposure arose, introducing a new risk. Because the chosen indexes are strongly correlated, however, and because of the robust statistical inferences (rolling regressions with min-max boundaries), the impact of this risk was kept to a minimum.

Although the amount and magnitude of volatility deviations tended to increase when the underlying stock indexes were volatile, the majority of significant deviations recorded were not linked to such extraordinary market conditions.

The results of Table 4 do not account for transaction costs. Even retail transaction costs would be low, however, compared with the arbitrage profits displayed in Table 4. Commissions are usually charged per traded contract, with one contract typically representing 100 options. Assuming rates of \$1.25 per contract (\$0.0125 per option),<sup>13</sup> the result is transaction costs of approximately \$0.05 per roundtrip arbitrage trade ( $4 \times \$0.0125$ ).<sup>14</sup> Such low costs would have enabled even the retail investor to take advantage of the observed deviations. Professional market participants, who benefit from lower transaction costs, could have profited to an even greater extent.

Some might argue that our simulated arbitrage strategy might not be fully replicating real market conditions because the option prices calculated from the implied-volatility data are not exact. For example, because the option prices used for simulated arbitrage trades were theoretical prices, not real market prices, the simulated arbitrage opportunities might not have been, in fact, real arbitrage opportunities. But our analysis was based on implied volatilities computed from real market

option prices. Therefore, by computing option prices from implied volatilities again, we were performing a simple reverse transformation without distorting actual arbitrage opportunities.

Nevertheless, the statistical arbitrage opportunities we detected might not have been implementable in some cases for other reasons. First, the option prices were calculated from the closing prices of the indexes, which as Harvey and Whaley pointed out, is an imprecise method. Because option markets close 15 minutes after stock markets, calculated option prices may be different from real option prices. Second, we corrected the implied volatilities used to calculate the bid and ask option prices by a bid-ask implied-volatility spread. Although we chose this spread to be conservatively large, compared with normal market circumstances, a more exact approach would be to use the actual intraday option bid and ask prices for the simulated trades. For example, bid-ask spreads may occasionally be much larger than usual, in which case, although an arbitrage opportunity would be detected, the arbitrage could not be implemented (or would at least be less profitable than we found). Third, when an arbitrage opportunity is detected, an arbitrage trade taking advantage of this opportunity has to go through at the next price. That price may be different from the initial price and may no longer allow arbitrage. Furthermore, an arbitrage trade consists of two legs. Often, these legs cannot be executed at the same time. We assumed that trading at the signal price is possible. Thus, our results regarding the profitability of the arbitrage strategy should be interpreted with caution.

Under the assumption that the results in Table 4 could have been attained in practice, can we conclude that option markets are inefficient? Not really. Statistical arbitrage is not riskless. It returns a profit during normal market situations but may result in losses during unusual circumstances. Perhaps the unusual market circumstances never materialized in our sample of only a few dozen arbitrage transactions. To determine conclusively whether the profits of a statistical arbitrage strategy arise from market inefficiency, a much larger sample of transactions would be needed.

## Conclusion

We illustrated and tested an aspect of option market efficiency that cannot be tested by using conventional exact arbitrage pricing relationships. We used options on several pairs of highly correlated stock indexes to establish the efficiency with which market participants value the relative risk (implied

volatility) of these indexes. We calculated boundaries on the basis of the historical covariance of the index-pair volatilities. Whenever the relative implied volatilities violated such a boundary, a relative implied-volatility deviation was identified and a mispricing suspected.

Each of the three investigated index pairs was found to have had a large number of such deviations, which seemed to be concentrated in phases during the time of the study. After we included bid-ask spreads in the calculations, however, only a small fraction of the deviations could be identified as an arbitrage opportunity.

We thus found some indication that deviations occasionally occur that can be exploited profitably by statistical arbitrage. Their number is small, however, and the possibility exists that not all of the arbitrage opportunities we detected could have been executed because of various arbitrage barriers in certain market situations and our assumption that the arbitrage trade could go through at the

signal price. To test the presence and persistence of potential relative mispricings and to determine conclusively whether a statistical arbitrage strategy based on relative implied volatility could have been implemented profitably in the period we studied would require further research using intraday tick data.

An interesting area for future research would be an analysis of why relative-volatility deviations tend to appear in clusters. Common factors causing such deviations could possibly be identified.

Additionally, other highly correlated markets, such as certain currency or commodity markets, could be investigated with the methodology presented here.

---

*All data used in this study were provided by Bloomberg Professional information service through Wegelin and Company. We thank David Rey and Maud Capelle for helpful comments.*

---

## Notes

1. Unless indicated otherwise, the term "arbitrage" here refers to statistical arbitrage.
2. Although our choice of the minimum limit for the correlation coefficients is somewhat arbitrary, it was guided by two factors. First, we wanted the coefficient to be as high as possible, but second, we needed several indexes to be left in the sample for further examination.
3. Because the regressions were used to establish the relationships between the index pairs to allow forecasts of relative volatility, an optimal sample size had to be found. Sample sizes of 3, 6, and 12 months were tested. Sample sizes of 3 months were unsatisfactory because they resulted in comparatively unstable relationships between index pairs. The cause could be "seasonal" relationships, which would prove to be unreliable in forecasting future relative developments in a different "season." Relationships between index pairs do change structurally over time, however, and the longer the sampling period, the slower the reaction of the forecast to this change. The 12-month sampling period proved to be too slow in adapting to such changes. Thus, we used a 6-month sample, which was long enough not to be seasonally biased but short enough to remain somewhat flexible as to structural changes.
4. Deciding how many trading days to use as a subsample to establish such maximum variations is a delicate matter. The longer the sampling period, the more conservative the result, because a large variation will widen the boundaries for a long period of time. The boundaries are designed to reflect current variation, however, not the variation long past that is no longer reflected in the market situation. So, we chose 250 trading days, approximately one trading year, as an informal compromise between conservatism and the relevance of information.
5. For example, assume that during the 250 trading days preceding time  $t$  of the ABC-XYZ index-pair sample, the largest change of the estimated slope coefficient over any 25-trading-day period was  $x$  percent. At time  $t$ , the estimated slope coefficient is  $\beta_{2t}$ ; thus, at that time, the lower boundary around this estimated slope coefficient would be  $\beta_{2t} - (\beta_{2t} x\%)$  and the higher boundary would be  $\beta_{2t} + (\beta_{2t} x\%)$ .
6. For an analysis of such a relationship, see, for example, Ncube (1994).
7. In the context of statistical arbitrage, a mispricing can never be identified with certainty. Consequently, a statistical arbitrage trade attempting to take advantage of a suspected mispricing is not guaranteed to be profitable.
8. The terms "undervalued" and "overvalued" are used here in reference to the relative implied volatility valuation of the indexes, not to their values measured in index points.
9. Bid-ask spreads were recorded on a sample of days when realized volatility was relatively high and a sample of days when realized volatility was relatively low so as to give a realistic reflection of such spreads in different market conditions. In relatively volatile markets, bid-ask spreads of approximately 1 percent were recorded for near-month at-the-money options for all three indexes, with slightly lower readings for the OEX. In less volatile markets, the spreads tended to narrow, so the decision to use 1 percent as an overall spread in the calculations can be regarded as a conservative assumption.
10. Point equivalency refers to the index points of the underlying indexes.
11. Because the idea behind this offsetting hedge strategy was an approximate immunization of the entire position against all parameter changes, except for the normalization of the observed implied volatility mispricing, taking positions with equivalent "weight" and adjusting them to reflect the relative volatility differential was crucial.
12. In this study, we implicitly assumed that trading at the signal prices was possible (i.e., that the mispricing persisted long enough to allow a trade). Because only end-of-day volatility data were available for the study, however, whether the persistence in the mispricings was sufficient to allow for the implementation of a trade remains unclear. A two-tick persistence (the first tick for the signal and the

- second for the trade) may be sufficient to implement such a trade.
13. Rates charged by the U.S. discount broker Ameritrade as of March 2000.
14. The opening and closing of the offsetting trades involves the buying and selling of approximately two options (point equivalency and  $\beta_2$  adjustments would increase or decrease this number somewhat).

---

## References

- Ackert, L., and Y. Tian. 1998. "The Introduction of Toronto Index Participation Units and Arbitrage Opportunities in the Toronto 35 Index Option Market." *Journal of Derivatives*, vol. 5, no. 4 (Summer):44–54.
- . 1999. "Efficiency in Index Options Markets and Trading in Stock Baskets." Working paper series, Federal Reserve Bank of Atlanta, vol. 99, no. 5.
- Chung, P. 1991. "A Transactions Data Test of Stock Index Futures Market Efficiency and Index Arbitrage Profitability." *Journal of Finance*, vol. 46, no. 5 (November):1791–1809.
- Fama, E. 1965. "The Behavior of Stock-Market Prices." *Journal of Business*, vol. 38, no. 1 (January):34–105.
- Figlewski, S. 1984. "Hedging Performance and the Basis Risk in Stock Index Futures." *Journal of Finance*, vol. 39, no. 3 (July):657–669.
- . 1989. "Options Arbitrage in Imperfect Markets." *Journal of Finance*, vol. 44, no. 5 (December):1289–1311.
- Gatev, E., W. Goetzmann, and K. Rouwenhorst. 1999. "Pairs Trading: Performance of a Relative Value Arbitrage Rule." Working Paper 7032, National Bureau of Economic Research (March).
- Gould, J., and D. Galai. 1974. "Transactions Costs and the Relationship between Put and Call Prices." *Journal of Financial Economics*, vol. 1, no. 2 (July):107–129.
- Harvey, C., and R. Whaley. 1991. "S&P 100 Index Option Volatility." *Journal of Finance*, no. 46, no. 4 (September):1551–61.
- Johnson, R., C. Zulauf, S. Irwin, and M. Gerlow. 1991. "The Soybean Complex Spread: An Examination of Market Efficiency from the Viewpoint of the Production Process." *Journal of Futures Markets*, vol. 11, no. 1 (February):25–37.
- Klemkosky, R., and B. Resnick. 1979. "Put-Call Parity and Market Efficiency." *Journal of Finance*, vol. 34, no. 5 (December):1141–55.
- Merton, R. 1973. "Theory of Rational Option Pricing." *Bell Journal of Economics and Management Science*, vol. 4, no. 1 (Spring):141–183.
- Ncube, M. 1994. "Modeling Implied Volatility with OLS and Panel Data Models." Discussion Paper No. 194, LSE Financial Markets Group Discussion Paper Series, London.
- Neal, R. 1996. "Direct Tests of Index Arbitrage Models." *Journal of Financial and Quantitative Analysis*, vol. 31, no. 4 (December):541–562.
- Poitras, G. 1997. "Turtles, Tails and Stereos: Arbitrage and the Design of Futures Spread Trading Strategies." *Journal of Derivatives*, vol. 5, no. 2 (Winter):71–88.
- Sofianos, G. 1993. "Index Arbitrage Profitability." *Journal of Derivatives*, vol. 1, no. 1 (Fall):6–20.
- Stoll, H. 1969. "The Relationship between Put and Call Options Prices." *Journal of Finance*, vol. 24, no. 5 (December):801–822.