Indexing and Statistical Arbitrage

Tracking error or cointegration?

Carol Alexander and Anca Dimitriu

Following the seminal work of Markowitz [1959], Sharpe [1964], Lintner [1965], and Black [1972], the fundamental statistical tool for traditional portfolio optimization is correlation analysis of asset returns. Optimization models for benchmark replication focus on minimizing the variance of the tracking error, with additional constraints concerning the correlation of the portfolio returns with the benchmark returns, or the transaction costs involved in rebalancing the portfolio.

There are a number of drawbacks to optimization models based on tracking error or on correlation measures, especially in application to a passive investment framework. First, minimizing tracking error with respect to an index that as a linear combination of stock prices entails a significant amount of noise may result in a portfolio that is very sample-specific and unstable under volatile market circumstances.

Additional limitations relate to the very nature of correlation as a measure of dependence; it is applicable only to stationary variables such as stock returns, so the method requires prior detrending of level variables (i.e., stock prices) and has the disadvantage of losing valuable information (i.e. the common trends in prices). Correlation is thus only a short-term statistic, and it lacks stability.

Finally, depending on the model used to estimate it, correlation can be very sensitive to the presence of outliers, non-stationarity, or volatility clustering, which limit the use of a long data history and can lead to erroneous conclusions about the nature of long-term dependencies.

Carol Alexander is a professor of risk management and director of research at the ISMA Centre of the University of Reading in the UK.
c.alexander@ismacentre.rdg.ac.uk

Anca Dimitriu is a Ph.D. student at the ISMA Centre of the University of Reading.
a.dimitriu@ismacentre.rdg.ac.uk

Given these limitations of correlation, it is not surprising that most applied financial econometric analyses employ a different tool for modeling dependencies between time series. Among economists, *cointegration* has gained far wider acceptance than correlation. Granger's [1966] influential work has engendered a vast academic literature in this area, and earned him the Nobel Prize in 2003.[1]

Cointegration is an extremely powerful statistical tool that, in a sense, generalizes correlation to non-stationary time series. Cointegration allows simple estimation methods such as least squares regression or maximum-likelihood to capture dependencies between non-stationary series such as stock prices, while still encompassing the dynamic correlation of the associated stationary series such as stock returns.

The fundamental characteristic that justifies the application of cointegration to a set of stock prices is that they can share a *common stochastic trend* (see Stock and Watson [1991]).[2] In this case, there is cointegration when there is at least one stationary linear combination of their prices, or, simply put, there is mean reversion in their price spreads. The finding that the spread in a system of prices is mean-reverting does not provide any information for forecasting the individual prices in the system, or the position of the system at some time in the future, but it does provide the valuable information that, irrespective of the position of the system, the prices will stay together on a long-run basis.

The cointegration approach to portfolio modeling, introduced by Alexander [1999], enables use of the entire information set in a set of stock prices. Since prices are long-memory processes, cointegration is able to explain their long-run behavior (Granger and Terasvirta [1993]).

The rationale for constructing portfolios based on a cointegration relationship with a benchmark rests on two main features. First, the price difference between the benchmark and the portfolio is, by construction, stationary, and this implies that the tracking portfolio will be tied to the benchmark in the long run. Second, stock weights, based on a long history of prices, will have more stability. These properties are the result of making full use of the information in stock prices prior to their detrending.

Correlation-based tracking strategies, however, are based on only partial information, and there is no mechanism ensuring the reversion of the portfolio to the benchmark over the longer term. If tracking error follows a random walk process, for example, the portfolio can diverge significantly from its benchmark, unless it is frequently rebalanced. Thus hedging strategies based on cointegration that focus on common trends only may be more effective in the long run.

Considering the important comparative advantages of using cointegration rather than correlation to optimize equity portfolios, it should be possible to exploit a long-run relationship between equity prices and the market index price, if found, to construct trading strategies. Such evidence is provided by Alexander, Giblin, and Weddington [2001], who investigate the performance of different long-short strategies developed in the S&P 100 stock universe. This application implies an extensive search for the best long-short combination over a large number of portfolios constructed on cointegration relationships and optimized on different model parameters such as training period, targeted tracking error, and number of assets in the portfolio. The results, even if based on a black box selection algorithm, indicate that cointegration-based optimization can ensure stable alpha, with low volatility and no correlation with market returns.

In another application of cointegration analysis to investment management that is particularly relevant to our line of research, Lucas [1997] considers the optimal asset allocation problem in the presence of cointegrated time series. Using an asset allocation problem with a risk-averse investment manager, Lucas shows that cointegrating combinations of time series have lower long-term volatility than their non-cointegrated counterparts. From a short-term or tactical asset allocation perspective, cointegration implies that the price series have an error-correcting behavior, allowing the anticipation of future developments. According to Lucas's results, the presence of cointegration relations also has important consequences for the short-term predictability of asset prices.

Clearly, the presence of cointegration relationships has a number of significant advantages for a trading strategy. To our knowledge, though, there are no rigorous academic studies that compare the theoretical and empirical properties of cointegration-based portfolios with traditional optimization models, such as tracking error variance minimization. We attempt to remedy this.

We find that cointegration-optimal tracking portfolios have a strong relationship to the benchmark, and the resulting statistical arbitrage portfolios clearly dominate their traditional equivalents.

COINTEGRATION MODEL FOR INDEX TRACKING AND STATISTICAL ARBITRAGE

Optimization models for active investments are generally more diverse and more sophisticated than those for passive investments, and these two categories often have little in common. Our approach is to extend index replication, a traditional passive investment strategy, into an active strategy. Assuming we can find an appropriate replication model, the passive strategy is extended by constructing portfolios to track artificial indexes, such as index-plus and index-minus, and trading on their spread. This is a standard statistical arbitrage strategy based on enhanced indexation.

Cointegration-Optimal Index Tracking

Cointegration-optimal portfolios are the tracking portfolios that are constructed on a cointegration relationship with a benchmark. The price difference between the tracking portfolio value and the benchmark, the price *spread*, is stationary only for a cointegration-optimal portfolio. Of the two stages in portfolio optimization, i.e., selecting the stocks to be included in the portfolio and then determining the optimal portfolio holdings in each stock, cointegration optimality is primarily a property of allocation rather than selection. That is, the optimization problem is: Given a set of assets $S_1, \ldots, S_n$, what is the allocation between these assets that gives the highest possible cointegration with the benchmark?

Having said this, the selection process can have a dramatic effect on the results of the optimization and consequently the tracking performance of the portfolio. It is easy to find strong and stable cointegrating relationships for some stock selections but more difficult for others. An important consideration is the number of stocks selected.

For instance, the portfolio of all stocks is trivially cointegrated with the reconstructed index (i.e., the index based on current weights).[3] As fewer stocks are included in the portfolio, cointegration relationships between the tracking portfolio and the benchmark become less stable, and indeed some portfolios may outperform or underperform the index significantly. Below some critical number of stocks, cointegration may be impossible to find.

Managers who seek to outperform an index will select proper subsets of the index stocks, but a simply mechanical stock selection procedure can be very time-consuming. To see why, suppose an index has N stocks and that at least n stocks are required for a stable cointegrating vector. Then a mechanical search would typically consider a very large number of portfolios—in fact, $N!/(N-n)!n!$ portfolios—as candidates for backtesting.

In any case, portfolio managers may prefer to select stocks according to a particular style of investing. Or stock selection can be the result of a proprietary selection model, technical analysis, or just the stock-picking skills of a portfolio manager. Some stock selection criteria will be more consistent with cointegration optimality than others. Some stock selection criteria will be more likely to produce portfolios with positive (or negative) alpha.

However critical, the selection process does not impose any special features on the cointegration optimization problem. We illustrate the difference between cointegration and tracking error variance allocation methods, given an identical set of stocks in the two portfolios, so here we apply only a naive approach to stock selection. Whenever possible, however, we do separate the effect of the stock selection criteria from the effect of the optimization method on the portfolio performance.

The second stage of constructing cointegration-optimal tracking portfolios is to determine the portfolio holdings in each of the selected stocks. This stage is based on a standard *cointegration regression* of the market index on a number of n stocks from its components:[4]

$$\ln(I_t) = c_1 + \sum_{k=1}^{n} c_{k+1} \ln(P_{k,t}) + \varepsilon_t \qquad (1)$$

where I_t is the index price, and $P_{k,t}$ is the k-th stock price at time t.

Note that the stocks must be selected so that the regression weights yield a price spread between the portfolio and the index ε_t that is stationary. In practice, as mentioned above, this entails choosing a high enough n. Since equity indexes are just linear combinations of stock prices, a stationary price spread can be easily obtained, provided that n is sufficiently large and, if the index is capitalization weighted, that the number of stocks in each issue is relatively stable.

The model specification in Equation (1) is not unique because the cointegration regression can also be estimated on level rather than (natural) log variables. The log variables specification has the advantage that when we take the first difference of (1) the expected return on the index will equal the expected return on the tracking portfolio. Moreover the estimated coefficients in Equation (1), further normalized to sum up to one, will provide the composition of the tracking portfolio.

The most general form of the model uses an ordinary least squares (OLS) criterion to estimate the stock coefficients. Note that the application of OLS to non-stationary dependent variables such as the log index price is valid only in the special case of a cointegration relationship. The residuals in (1) are stationary if, and only if, the index and the tracking portfolio are cointegrated. If the residuals from the regression are non-stationary, the OLS coefficient estimates will not be consistent, and further inference will not be valid. Therefore testing for cointegration is an essential step in constructing cointegration-optimal tracking portfolios.

We use the Engle and Granger [1987] methodology for cointegration testing, which is particularly appealing for portfolio optimization because of its intuitive and straightforward implementation.[5] The cointegration augmented Dickey-Fuller (ADF) [1979] regression estimated on the residuals of the cointegration regressions is:

$$\Delta\hat{\varepsilon}_t = \gamma\hat{\varepsilon}_{t-1} + \sum_{i=1}^{p}\alpha_i\Delta\hat{\varepsilon}_{t-i} + u_t \tag{2}$$

The null hypothesis tested is of no cointegration, i.e., $\gamma = 0$, against the alternative of $\gamma < 0$. The critical values for the t-statistic of γ are obtained using the response surfaces provided by MacKinnon [1991].

If the null hypothesis of no cointegration is rejected, the cointegration-optimal tracking portfolio based on OLS estimation of Equation (1) is expected to have very similar returns to the market index. That is, cointegration ensures that the price spread between the portfolio value and the benchmark is a mean-reverting process with minimum volatility.

Note that a highly dimensional system of stocks is likely to exhibit some cointegration relationships purely by chance. This parallels the findings of Plerou et al. [2002] and others that correlation matrices based on financial asset returns often have a high degree of randomness.

For instance, when a system of asset returns is *independently* shuffled in time, i.e., each series is shuffled separately from the others, the resulting random correlation matrix is very likely to have some significant eigenvalues. Similarly, when prices are compounded from this shuffled system of returns, some spurious cointegration relationships are likely to be found. This indicates that a single test, like the ADF statistic or one of the Johansen [1991] tests, may not by itself be capable of discriminating between genuine and spurious cointegration.

One might argue that since the benchmark is just a linear combination of those and other stock prices, selecting enough stocks from the benchmark will ensure that any cointegration relationship between the tracking portfolio and the benchmark represents a genuine equilibrium. A formal check of this hypothesis would be recommended. Standard cointegration tests can and should be supplemented by additional criteria that are capable of supporting cointegration as a stable property of multivariate time series.

Proposing a set of formal criteria to discriminate between genuine and spurious cointegration in highly dimensional systems is beyond the scope of this article. Rather, we simply point out several candidates for such criteria. A natural candidate is the stability of cointegration coefficients as the portfolio is rebalanced. A spurious cointegration vector is likely to fail a stability test because the coefficients of random equilibrium vectors will be highly variable on rolling sample estimation.

Second, the presence of many cointegration vectors should be an indication of a genuine relationship—it may be possible to find by chance one or even a few cointegrating vectors, but it is virtually impossible to find many cointegrating vectors simply by chance.

Third, the range of active weights (the difference between the tracking portfolio weights and the benchmark weights) is a common indication of spurious cointegration. A genuine cointegration relationship should not imply many extreme exposures to individual stocks, yet spurious cointegration relationships such as those found on shuffled price series will often violate normal position limits.

Cointegration tests applied to constrained portfolios are therefore a reliable indication of a genuine cointegration relationship, and spurious relationships will rarely show significant cointegration once concentration constraints have been imposed.[6]

Long-Short Cointegration-Optimal Portfolios

A natural extension of the simple cointegration tracking strategy presented above is to exploit the tracking potential of cointegration, attempting to replicate enhanced benchmarks constructed by adding to and subtracting from the index returns an annual excess return of α%, uniformly distributed over daily returns. Then, self-financing statistical arbitrage portfolios can be set up as a difference between two portfolios tracking a plus and a minus benchmark. This statistical arbitrage is expected to generate returns according to the plus/minus spread with low volatility. Moreover, if each plus and minus portfolio is tracking its benchmark accurately, and their tracking errors are not correlated with the market returns, statistical arbitrage will give a market-neutral portfolio (the effect of netting similar betas).

The new cointegration regressions can be written as:

$$\ln(I_t^+) = a_1 + \sum_{k=1}^{n} a_{k+1} \ln(P_{k,t}) + \varepsilon_t^+ \tag{3}$$

$$\ln(I_t^-) = b_1 + \sum_{k=1}^{n} b_{k+1} \ln(P_{k,t}) + \varepsilon_t^- \tag{4}$$

Naturally it will become more difficult to construct cointegrated portfolios, the more the benchmark diverges from the index. The cointegration relationship between the market index and its component stocks has a solid rationale, but this is not necessarily the case for portfolios tracking artificial benchmarks, which might conceivably be chosen to outperform the market index by 50%. In this case, the difficulty of finding an appropriate cointegration relationship would lead to an increased instability of stock weights, higher transaction costs, and higher volatility of returns. To avoid this, it is essential to ensure that all the portfolios tracking plus or minus benchmarks pass the cointegration test.

Note that stock weights need not be restricted to be positive in the tracking portfolios. In fact, it is likely we shall take some short positions in the portfolio tracking the minus benchmark. The stock holdings in the cointegration-optimal statistical arbitrage portfolio are obtained by netting their individual weights in the plus and minus portfolios.

Tracking Error Variance (TEV) Minimization Model

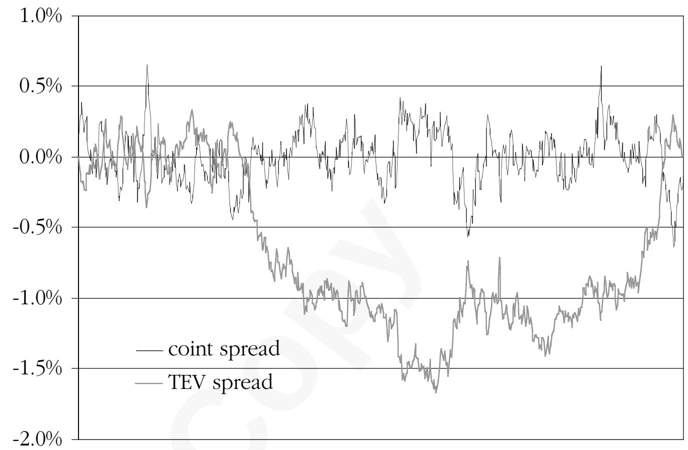
We compare the performance of cointegration-optimal portfolios with portfolios that are optimized using the standard tracking error variance (TEV) minimization model of Roll [1992].[7] Using a similar notation as in the cointegration-optimal tracking portfolio, the tracking error variance minimization model can be written as:

$$r_t = \sum_{k=1}^{n} c_k r_{k,t} + \eta_t \tag{5}$$

where r_t is the log return on the index at time t, $r_{k,t}$ is the log return on the stock k at time t, and η_t is the tracking error.

EXHIBIT 1

In-Sample Spread (cumulative return difference) Between Tracking Portfolios and Benchmark



The analytic solution of Roll [1992] is not applicable for portfolios with only a subset of the stocks in the benchmark. Instead, the allocations are estimated using a numerical optimization to minimize the variance of the tracking error subject to the constraints of zero expected tracking error and a unit sum on the coefficients.

The main difference between Equations (1) and (5) concerns the objective of the index tracking strategy. In the first case, it is the (squared) price spread between the replica portfolio and benchmark that is minimized, while in the second it is the (squared) change in this spread (i.e., the return). This seemingly innocuous difference actually results in quite different tracking error characteristics. With cointegration, the values of the index and the tracking portfolio are tied together; with TEV optimization, the value difference can diverge significantly before returning to zero.

To illustrate this, we use a three-year daily data sample from the Dow Jones Industrial Average (DJIA) stocks. Exhibit 1 compares the in-sample price spread between the index and the tracking portfolio optimized according to Equations (1) and (5). By construction, the spread between the cointegration tracking portfolio and the index must be both mean-reverting and low-variance. The TEV criterion, however, targets the return spread, not the price spread. The price spread shown in Exhibit 1 (which is typical for TEV portfolios) moves quite far from the benchmark before returning to zero. In-sample, the price spread must return to zero by definition of OLS, but out-of-sample, the price spread of the TEV optimal portfolio need *not* return to zero after a period of time. This may happen because there is nothing in the TEV model to ensure that the price spread is mean-reverting.

This simple example shows that, in theory, cointegration-based index tracking is more appropriate than standard tracking error variance minimization—or indeed any other traditional strategy that is based on *return* optimization rather than *price* optimization.

DYNAMIC PERFORMANCE TESTING METHODOLOGY

We investigate the empirical properties of cointegration-optimal portfolios using a database of daily closing prices over the period January 1990–December 2003 for the stocks in the Dow Jones Industrial Average index as of December 31, 2003. As a benchmark we use a DJIA historical series reconstructed to match the last available membership of the index. By doing this, we compare the performance of portfolios comprising the stocks currently included in the DJIA with a market index constructed from the same stocks. The use of the reconstructed index ensures consistency in the treatment of dividends and stock splits (both index and stock prices are adjusted for dividends and stock splits) and eliminates a potential survivorship bias.[8]

The 30 stock price series were downloaded from yahoo-financial.com and any missing observation was replaced by the last closing price available for that particular stock. As expected, according to standard unit root tests such as (2), all stock price series proved to have significant stochastic trends, but the associated returns series were clearly stationary, thus satisfying the conditions for cointegration analysis.

To test the performance of different cointegration portfolios, we generate optimal portfolios based on different parameters: *number of stocks in the tracking portfolios* (20, 25, and 30 stocks, selected according to their price ranking, starting with the highest-priced stocks); *calibration period* (up to five years of daily data prior to the moment of portfolio construction); *model specification*—with or without constraints on the portfolio weights; *rebalancing period* (every two weeks, monthly, every three months, and every six months); and the *spread* between the benchmarks tracked (up to 30% per year).

The first cointegration-optimal tracking portfolios are constructed in January 1993. All portfolios are rebalanced at the given frequency, with stock selection based on the new stock rankings and optimal weights based on the new coefficients of the cointegration regression. At each rebalancing, the cointegration regression is reestimated over the rolling fixed-length calibration period preceding the rebalancing time. The number of shares held in each stock is determined by the previous portfolio value, the current stock price, and the stock weight in the cointegration regression. In between rebalancings, the portfolios are left unmanaged, i.e., the number of stocks is kept constant. We follow exactly the same process for the TEV optimal portfolios.

To account for the impact of the bid-ask spread and brokerage fees on the portfolio returns, we assume a fixed amount of 20 basis points in transaction costs on each trade value. This is in line with other research on transaction costs incurred on the New York Stock Exchange (Chalmers, Edelen, and Kadlec [1999]; "Comparing Bid-Ask Spreads on the NYSE and Nasdaq" [2001]).

The repo costs are normally low so these are computed at 0.25% of the increase in the short position in a bull market for a particular stock (defined as an increase in price over the last ten trading days) and at 0.35% on the same amount in a bear market for that particular stock (defined as a decline in price over the last ten trading days). These cost levels are conservative, given the fact that DJIA stocks are known to be very liquid, and their trading has market impact. In any case, we are interested primarily in the *comparative* effect of little transaction costs on different strategies, and on a relative basis the particular cost rate we choose is less important.

INDEX TRACKING

The first target of our empirical analysis is the out-of-sample performance of the two approaches to index tracking, using the cointegration model specification in Equation (1) and the TEV model specification in Equation (5). We start the analysis with a standard two-week rebalancing frequency and no constraints on the portfolio weights, and subsequently extend it to other rebalancing frequencies and introduce constraints.

Number of Stocks and Calibration Period

We first note that the degree of cointegration between the cointegration-optimal tracking portfolio and the benchmark increases with the number of stocks in the portfolio and marginally with the sample size. A number of portfolios with either too few stocks (fewer than 20 stocks) or weights based on a too-short calibration period (one or two years) proved not to be sufficiently cointegrated with the market index, so we exclude them from the analysis following. Second, note that *none* of the TEV

Exhibit 2

Tracking Performance of Cointegration-Optimal and TEV Portfolios

A. Unconstrained Weights

	2-week rebalancing			monthly rebalancing			3-month rebalancing			6-month rebalancing		
Cointegration	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks
Average annual TE	-2.76%	-1.36%	0.68%	-2.32%	-0.61%	0.71%	-2.29%	-0.92%	0.38%	-2.10%	-1.09%	-0.02%
Tracking error annual	4.63%	3.59%	2.77%	4.46%	3.48%	2.83%	4.33%	3.32%	2.72%	4.35%	3.32%	2.75%
Correlation TP/DJIA	0.97	0.98	0.99	0.97	0.98	0.99	0.97	0.98	0.99	0.97	0.98	0.99
Correlation TE/DJIA	0.19	0.20	0.12	0.09	0.11	0.07	0.03	0.04	0.02	-0.02	-0.01	-0.03
Tracking error skew	0.08	0.10	0.01	-0.09	-0.04	-0.17	-0.14	-0.10	-0.08	-0.20	-0.29	-0.24
Tracking error xs kurt	3.57	3.02	2.68	2.40	2.63	4.09	1.81	1.39	1.94	2.16	1.86	2.45
ADF statistic	-6.41	-6.72	-7.01	-6.40	-6.75	-7.00	-6.39	-6.71	-7.00	-6.51	-6.74	-6.93
Average trans costs	0.97%	0.69%	0.28%	0.58%	0.42%	0.21%	0.31%	0.24%	0.14%	0.18%	0.14%	0.10%
Sharpe ratio TP	0.18	0.26	0.38	0.20	0.29	0.37	0.18	0.26	0.34	0.21	0.27	0.33
Beta TP	1.05	1.04	1.02	1.02	1.02	1.01	1.01	1.01	1.00	1.00	1.00	1.00
Prob(TE<0)	52.37%	51.72%	49.69%	51.47%	51.36%	49.27%	52.21%	51.22%	49.05%	51.20%	50.61%	48.80%
TEV	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks
Average annual TE	-2.37%	-1.02%	0.90%	-1.90%	-0.65%	0.88%	-2.28%	-0.75%	0.64%	-2.31%	-0.84%	0.37%
Tracking error annual	3.51%	2.74%	2.21%	3.46%	2.70%	2.24%	3.26%	2.52%	2.06%	3.30%	2.53%	2.04%
Correlation TP/DJIA	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99
Correlation TE/DJIA	0.12	0.14	0.13	0.03	0.05	0.06	-0.03	-0.01	-0.01	-0.08	-0.07	-0.05
Tracking error skew	-0.02	0.24	0.44	-0.14	-0.01	-0.06	-0.08	0.05	0.12	-0.24	-0.03	-0.01
Tracking error xs kurt	4.71	4.32	5.27	4.64	5.53	7.75	3.66	2.94	2.51	3.42	4.39	3.72
ADF statistic	-2.00	-1.85	-1.58	-1.92	-1.85	-1.53	-1.86	-1.90	-1.56	-1.82	-1.81	-1.55
Average trans costs	0.55%	0.41%	0.19%	0.33%	0.24%	0.12%	0.18%	0.12%	0.07%	0.11%	0.07%	0.05%
Sharpe ratio TP	0.21	0.29	0.40	0.22	0.30	0.38	0.19	0.27	0.35	0.20	0.29	0.36
Beta TP	1.02	1.02	1.02	1.01	1.01	1.01	1.00	1.00	1.00	0.98	0.99	0.99
Prob(TE<0)	52.84%	51.43%	49.47%	51.69%	50.70%	49.38%	52.40%	50.19%	49.66%	51.72%	50.06%	49.76%

optimal portfolios is cointegrated with the benchmark, whatever number of stocks included. They all have non-stationary in-sample price spreads, such as that illustrated in Exhibit 1.

Once the minimum calibration period (of about three years) for ensuring cointegration is used, increasing it has little impact on the tracking performance of the cointegration-optimal portfolios. For reasons of space, we focus on a calibration period of three years. To allow a *ceteris paribus* comparison, we must use the same calibration period for the tracking error variance model.

In an independent application of TEV to portfolios with relatively few stocks, the calibration period could be shorter than three years, so that the non-stationary price spread (such as that illustrated in Exhibit 1) becomes less of a problem. Apart from a better in-sample model specification, though, shorter calibration periods result in TEV portfolios that are more sample-specific, have a less stable structure, and require more frequent rebalancing.

Exhibit 2 reports summary statistics for the cointegration- and TEV-optimal tracking portfolios of different numbers of stocks based on a calibration period of three years, over the 1993-2003 period. For each tracking portfolio, we report the average annual tracking error (after transaction costs), its volatility, the correlation of the tracking portfolio returns with the benchmark returns, the correlation of the tracking error with the benchmark returns, the skewness and excess kurtosis of the tracking error, the in-sample ADF statistic for the cointegration regression, the average annual transaction costs, the Sharpe ratio of the tracking portfolios after transaction costs, the tracking portfolio beta, and the empirical probability of observing a negative tracking error.

Optimization Alpha

The first observation is that the tracking portfolios of 20 and 25 stocks underperform the benchmark, even before transaction costs, while the 30-stock portfolios consistently outperform it. This is true for *both* models.

E X H I B I T 2 (continued)

Tracking Performance of Cointegration-Optimal and TEV Portfolios

B. Constrained Weights

	2-week rebalancing			monthly rebalancing			3-month rebalancing			6-month rebalancing		
Cointegration	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks
Average annual TE	-1.96%	-0.72%	1.46%	-1.40%	0.13%	1.48%	-1.61%	-0.18%	1.21%	-1.67%	-0.36%	0.90%
Tracking error annual	4.72%	3.86%	3.09%	4.64%	3.80%	3.17%	4.62%	3.64%	3.11%	4.66%	3.68%	3.15%
Correlation TP/DJIA	0.97	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98	0.97	0.98	0.98
Correlation TE/DJIA	0.14	0.13	0.06	0.02	0.04	0.01	-0.04	-0.02	-0.04	-0.07	-0.06	-0.07
Tracking error skew	0.12	0.15	0.05	0.07	0.04	-0.10	0.08	-0.04	-0.08	-0.07	-0.15	-0.13
Tracking error xs	2.88	2.07	1.83	2.90	1.98	2.64	2.67	0.86	1.54	1.53	1.32	1.83
ADF statistic	-5.64	-6.25	-6.48	-5.62	-6.28	-6.46	-5.61	-6.25	-6.40	-5.58	-6.26	-6.29
Average trans costs	0.88%	0.63%	0.26%	0.52%	0.38%	0.20%	0.31%	0.21%	0.14%	0.17%	0.13%	0.09%
Sharpe ratio TP	0.23	0.30	0.43	0.25	0.34	0.42	0.22	0.30	0.38	0.24	0.31	0.39
Beta TP	1.04	1.03	1.01	1.01	1.01	1.00	0.99	1.00	0.99	0.98	0.99	0.99
Prob(TE<0)	52.73%	51.65%	48.79%	51.58%	51.10%	48.13%	53.05%	51.29%	48.17%	52.28%	51.02%	49.35%
TEV	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks	20 stocks	25 stocks	30 stocks
Average annual TE	-1.81%	-0.43%	1.54%	-1.31%	-0.07%	1.52%	-1.74%	-0.21%	1.18%	-1.78%	-0.33%	0.91%
Tracking error annual	3.72%	2.95%	2.48%	3.69%	2.94%	2.52%	3.57%	2.79%	2.37%	3.65%	2.83%	2.38%
Correlation TP/DJIA	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99
Correlation TE/DJIA	0.07	0.07	0.06	-0.02	-0.01	-0.01	-0.08	-0.07	-0.08	-0.13	-0.12	-0.12
Tracking error skew	-0.02	0.17	0.30	-0.13	-0.03	-0.06	-0.13	-0.03	0.03	-0.23	-0.10	-0.05
Tracking error xs	3.57	2.70	3.23	3.33	3.69	4.82	2.91	1.72	1.83	2.44	2.58	2.47
ADF statistic	-2.05	-1.92	-1.70	-2.00	-1.91	-1.65	-1.92	-1.93	-1.68	-1.90	-1.90	-1.66
Average trans costs	0.61%	0.43%	0.20%	0.37%	0.25%	0.13%	0.20%	0.13%	0.08%	0.12%	0.08%	0.05%
Sharpe ratio TP	0.24	0.32	0.44	0.26	0.33	0.42	0.22	0.31	0.39	0.24	0.32	0.39
Beta TP	1.02	1.01	1.01	1.00	1.00	1.00	0.98	0.99	0.99	0.97	0.98	0.98
Prob(TE<0)	51.87%	49.73%	48.68%	50.81%	49.63%	48.64%	52.13%	49.09%	48.59%	52.20%	48.98%	49.24%

Calibration period: 3 years.

Exhibit 3 plots the cumulative returns of the tracking portfolios of 30 stocks. While the outperformance of the 30-stock portfolios is clearly generated by the optimization models, the underperformance of the 20- and 25-stock portfolios can be caused either by the stock selection method or by the optimization algorithm.

To distinguish between these possibilities, we construct a simple price-weighted *index20* portfolio of the same 20 (or 30) stocks that are selected for the tracking portfolios. Then the difference between the index20 (or index 25) portfolio and the corresponding tracking portfolio is due only to the optimization of allocations.

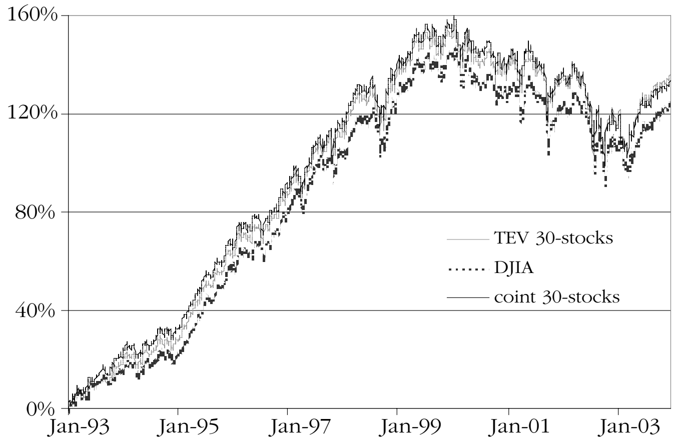
During the calibration period, the index20 portfolio outperforms the index30 portfolio on average by 1% in annual terms, over the period 1993-2003.[9] Also, the index25 portfolio performs the index30 portfolio by 0.7% per year on average. Out-of-sample, over a ten-day no-trading horizon, the index20 and index25 portfolios significantly underperform the index30 by 3.3% (respectively, 2.0%) on average in annual terms.

On the one hand, this is evidence of mean reversion in stock returns.[10] On the other hand, for the purpose of our analysis we can conclude that the underperformance of 20- (or 25-) stock tracking portfolios is *not* due to the optimization models, but is a result of the stock selection criterion used. After transaction costs, the optimized portfolios have less tracking error, in absolute terms, than the underperformance of the index20 and index25 portfolios. Thus, both cointegration and TEV optimization models enhance the returns of tracking portfolios over the price-weighted portfolio of the identical stocks.

We show in Exhibit 4 that this outperformance is not uniform over the sample. The cointegration- and the TEV-optimal portfolios of all 30 stocks outperform the reconstructed DJIA index in our out-of-sample tests over 1993-2003, but with the exception of the year 1994, outperformance occurred during the main market crisis periods: the Asian crisis, the Russian crisis, and

Exhibit 3

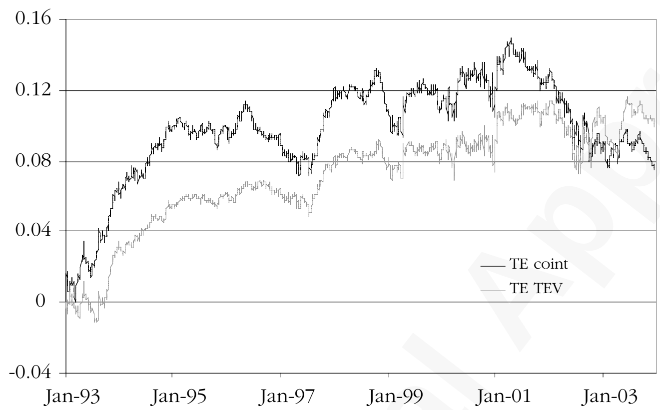
Cumulative Out-of-Sample Returns on DJIA, Cointegration-Optimal, and TEV Tracking Portfolios



Calibrated on 3-year sample period and rebalanced every 10 trading days.

Exhibit 4

Cumulative Out-of-Sample Tracking Error of 30-Stock Cointegration-Optimal and TEV Tracking Portfolios



the technology market crash. Why should this be so?

The answer is that the tracking portfolio weights are constructed on a relatively long calibration period so they tend to ignore short-term movements in stock prices, although these are immediately accounted for in the price-weighted index structure. This results in tracking error, which nevertheless mean-reverts when prices mean-revert. If in addition to mean reversion in prices there is a marked asymmetry (in that prices fall faster than they rise), the gains in the price spread will outweigh the losses.

If we take the example of a stable trending market where the prices of certain stocks increase well above their historical average, then the tracking portfolio weights lag behind the index weights, and consequently generate relative losses for the tracking portfolios for as long as the trend continues. Then when the prices of these stocks revert to their historical equilibrium levels, the tracking portfolios, as still underweighted on them, realize relative gains.

In some cases (such as during market crash periods), the gains are more significant than the losses. This is due to the asymmetry of stock markets; it is well documented that stock prices usually fall more rapidly than they rise.[11] When prices increase *slowly*, they allow the tracking portfolios to adjust to the new information, and to track the benchmark reasonably well. When prices fall *suddenly*, the tracking portfolios, with weights that are lagging the market weights, realize large gains.

It is notable that the cointegration-optimal tracking portfolio consistently underperformed the benchmark during the years 2001–2003. We discuss this later when we examine the statistical arbitrage performance of the two strategies.

Comparison of Cointegration and TEV Tracking Portfolios

We have seen that the out-of-sample performance of the cointegration-optimal and tracking error variance portfolios is very similar. Even though in-sample the cointegration portfolio is tied to the benchmark while the TEV is not, both models produce returns that are very highly correlated with the benchmark out-of-sample. The tracking errors are uncorrelated with the market, and the tracking portfolio betas are very close to one.

Both portfolios generate a positive alpha when the effect of stock selection is neutralized (illustrated in Exhibit 4), and throughout most of the sample period the cumulative tracking error from the cointegration-optimal model is well above that of the TEV model, although the opposite happens during the last three years of the sample. As one would expect, since the TEV portfolios are specifically constructed to minimize the variance of the tracking error, the TEV portfolios have lower tracking error volatility.

The TEV portfolios generate marginally lower transaction costs, and they also have slightly better Sharpe ratios. The tracking error, however, has a distribution closer to normality, with less excess kurtosis, in the case of the cointegration-optimal portfolios. Also, the probability of underperforming the benchmark is in general marginally lower for the cointegration-optimal portfolios.

Clearly, the extra feature of cointegration with the benchmark can be achieved at no significant cost for the tracking portfolio. In this simple index tracking exercise, with no weight constraints, neither advantages nor limitations of a cointegration relationship with the benchmark are empirically evident.

Rebalancing Period and Weight Constraints

The investigation of transaction costs in Exhibit 2, Panel A, is essential for understanding the characteristics of optimized tracking portfolios. For both models, the transaction costs decline significantly as the number of stocks in the tracking portfolios rises. In fact, this is simply the stock selection effect. As at each rebalancing we select the first *k* stocks according to their price, and then optimize the portfolio to replicate the index, as the composition of the first *k* stocks group changes, the portfolio weights will also change significantly.

One way to reduce the transaction costs is to lengthen the rebalancing schedule. Given the length of the calibration period used to optimize the tracking portfolios, rebalancing less often should not hurt tracking performance dramatically. Indeed, the results in Exhibit 2, Panel A, show that, as rebalancing is reduced to as few as two times per year, the tracking portfolios constructed with both models continue to perform well. The correlation of the tracking portfolios with the benchmark continues to be very high, while the Sharpe ratios of the tracking portfolios are maintained. In fact, the average tracking error gross of transaction costs declines slightly when we rebalance less often, but the transaction costs are significantly lowered, as is the volatility of the tracking error.

It is important to note that robustness to curtailing rebalancing arises for different reasons. In the cointegration portfolio, it is driven by the stability of the cointegration relationship; for the TEV model, it is the result of using a long calibration period. With a shorter calibration period, as one could use in an independent application of TEV, we have found that the portfolio structure is much less stable, and overall performance deteriorates markedly with longer rebalancing intervals.

Finally, what is the effect of imposing no short sales and concentration constraints on the tracking portfolio weights? Such constraints are widely used in practice by mutual funds and other institutional investors.

Our first observation is that in the presence of such constraints, we still find cointegration relationships between the tracking portfolios and the benchmark, and the price spread for the tracking error variance portfolios remains non-stationary. Indeed, given the results in Exhibit 2, Panel B, the alpha tends to rise in the presence of constraints for both models, at the expense of slightly higher tracking error volatility. The correlation of the tracking portfolios with the benchmark remains very high, and overall the Sharpe ratios of the tracking portfolios are slightly improved.

The main benefit of imposing constraints turns out to be in terms of excess kurtosis, which is significantly reduced. This is consistent with the finding of Jagannathan and Ma [2003] that imposing constraints can lessen model errors arising from outliers in the sample.

To conclude, the empirical performance of the cointegration index tracking model is very similar to that of the traditional TEV-minimizing model. When the impact of the selection criteria is neutralized, both models enhance the performance of the benchmark. When we measure performance relative to a reconstructed index of the same stocks and thus exclude the possibility that survivorship bias or dividend effects have influenced our results, we still find that both models produce a *significant positive* tracking error out-of-sample and better Sharpe ratios than the benchmark, even after transaction costs.

We have explained why the periods responsible for the greatest part of the positive tracking error coincide with the main market crises during our data sample, the Asian and Russian crises and the burst of the technology bubble. Tracking performance is shown to be robust to imposing weight constraints and reducing rebalancing.

These results are thus of great relevance for institutional investors such as mutual index funds.

STATISTICAL ARBITRAGE STRATEGIES

To examine the performance of the cointegration-optimal statistical arbitrage model, for reasons of space, we do not consider the effect of the stock selection criteria and use all 30 stocks in each portfolio. We also impose no constraints on the portfolio weights. Statistical arbitrage implies very dynamic portfolio management, so we assume the highest rebalancing frequency in the tracking simulations, i.e., every ten trading days. Also, we consider only portfolios based on a three-year estimation period (although fine-tuning these model parameters can improve the strategy performance significantly).

As a basis for tracking, we create six plus/minus benchmarks by adding and subtracting annual returns of 5%, 10%, and 15% to and from the reconstructed DJIA

EXHIBIT 5

Statistical Arbitrage Performance of Cointegration-Optimal and TEV Portfolios with 30 Stocks

A. Entire Sample

		coint				TEV			
Jan 1993 – Dec 2003		**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**	**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**
minus 0%	Annual return	N/A	1.03%	0.93%	0.42%	N/A	0.57%	0.12%	-0.59%
	Annual vol	N/A	2.82%	5.15%	7.94%	N/A	1.93%	3.88%	6.24%
	Skewness	N/A	-0.14	-0.13	-0.16	N/A	0.09	-0.01	-0.06
	Xs kurtosis	N/A	1.98	1.34	1.58	N/A	3.88	1.42	1.39
	Correlation	N/A	0.20	0.20	0.19	N/A	0.32	0.33	0.30
	Sharpe ratio	N/A	0.37	0.18	0.05		0.29	0.03	-0.09
minus 5%	Annual return	-0.92%	0.11%	0.01%	-0.50%	-1.64%	-1.07%	-1.52%	-2.23%
	Annual vol	4.77%	5.84%	8.81%	11.75%	4.00%	4.73%	7.19%	9.70%
	Skewness	-0.17	-0.12	-0.14	-0.18	-0.34	-0.15	-0.17	-0.17
	Xs kurtosis	1.65	1.56	1.63	1.75	2.28	1.32	1.34	1.45
	Correlation	0.16	0.23	0.20	0.20	0.19	0.29	0.29	0.27
	Sharpe ratio	-0.19	0.02	0.00	-0.04	-0.41	-0.23	-0.21	-0.23
minus 10%	Annual return	-0.62%	0.41%	0.30%	-0.20%	-2.07%	-1.50%	-1.95%	-2.66%
	Annual vol	7.27%	8.55%	11.52%	14.44%	6.10%	6.97%	9.41%	11.92%
	Skewness	-0.15	-0.11	-0.13	-0.16	-0.26	-0.14	-0.16	-0.16
	Xs kurtosis	1.57	1.51	1.57	1.68	1.72	1.27	1.30	1.38
	Correlation	0.18	0.21	0.20	0.20	0.22	0.28	0.28	0.27
	Sharpe ratio	-0.09	0.05	0.03	-0.01	-0.34	-0.21	-0.21	-0.22
minus 15%	Annual return	-0.06%	0.98%	0.87%	0.37%	-2.41%	-1.84%	-2.29%	-3.00%
	Annual vol	9.72%	11.10%	14.04%	16.97%	8.22%	9.15%	11.59%	14.08%
	Skewness	-0.10	-0.08	-0.11	-0.13	-0.21	-0.13	-0.15	-0.15
	Xs kurtosis	1.63	1.52	1.56	1.65	1.51	1.23	1.26	1.33
	Correlation	0.18	0.21	0.20	0.20	0.23	0.28	0.28	0.27
	Sharpe ratio	-0.01	0.09	0.06	0.02	-0.29	-0.20	-0.20	-0.21

returns, uniformly distributed. Then, using both the cointegration and the tracking error variance models, we construct portfolios to track these artificial indexes. Even though the artificial benchmarks diverge significantly from the actual index values (up to plus or minus 15% annually), we are still able to find portfolios that are cointegrated with them. The price spreads between the cointegration-optimal tracking portfolios and their artificial benchmarks prove to be stationary. Finally, we set up statistical arbitrage strategies that are long on the plus tracking portfolios and short on the minus tracking portfolios.

The summary out-of-sample performance results for the two statistical arbitrage strategies over the 1993-2003 period are presented in Exhibit 5, Panel A. For both models, the best performance is produced by strategies tracking narrow spreads, such as plus 5% hedged with the portfolio tracking the actual benchmark. As the spread between the benchmarks tracked widens, the returns are significantly more volatile, without the compensation of additional returns.

Panel A actually shows a negative relationship between the long-short portfolio returns and the spread between the benchmarks tracked. This is the combined effect of higher transaction costs and increased volatility. Portfolios tracking wide spreads tend to assume more aggressive positions. Before transaction and repo costs, the arbitrage returns are substantially higher when tracking large plus or minus spreads. As the degree of cointegration lessens, the stock weights become more unstable, which results in higher transaction and repo costs.

Another notable feature is that the minus portfolios tend to be more volatile than their plus equivalents.

Both cointegration and TEV models exhibit low unconditional correlation with the market returns and close-to-normal return distributions, with negative but not significant skewness and excess kurtosis in the range of 1 to 2, much lower than that of the market index.

The most important finding in Exhibit 5, Panel A, is that the cointegration-optimal portfolios clearly dominate the TEV statistical arbitrage portfolios. As expected, the TEV portfolios generally are less volatile, but they also have much lower returns, and this results in lower Sharpe ratios. In fact, the great majority of TEV statistical arbitrage portfolios actually generate negative average returns, while most of the cointegration-based statistical arbitrage portfolios have positive average returns. Also, the TEV portfolios are slightly more correlated with the market returns.

EXHIBIT 5 (continued)

Statistical Arbitrage Performance of Cointegrational-Optimal and TEV Portfolios with 30 Stocks

B. Subsamples

		coint				TEV			
Jan 1993 – Dec 1999		**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**	**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**
minus 0%	Annual return	N/A	2.63%	3.56%	4.28%	N/A	1.58%	2.01%	2.30%
	Annual vol	N/A	2.35%	4.78%	7.51%	N/A	1.63%	3.69%	6.00%
	Skewness	N/A	-0.01	0.03	0.01	N/A	0.03	-0.07	-0.06
	Xs kurtosis	N/A	0.46	0.69	0.75	N/A	1.07	1.00	1.47
	Correlation	N/A	0.09	0.15	0.16	N/A	0.16	0.24	0.24
	Sharpe ratio	N/A	1.12	0.74	0.57	N/A	0.97	0.54	0.38
minus 5%	Annual return	-0.88%	1.75%	2.68%	3.41%	-0.99%	0.59%	1.01%	1.30%
	Annual vol	4.63%	5.80%	8.70%	11.48%	3.63%	4.62%	6.96%	9.32%
	Skewness	-0.03	0.05	0.03	0.01	-0.16	-0.09	-0.12	-0.11
	Xs kurtosis	0.48	0.54	0.54	0.59	1.40	1.23	1.25	1.40
	Correlation	0.21	0.21	0.20	0.19	0.30	0.29	0.28	0.27
	Sharpe ratio	-0.19	0.30	0.31	0.30	-0.27	0.13	0.15	0.14
minus 10%	Annual return	0.40%	3.03%	3.96%	4.68%	-0.58%	1.00%	1.43%	1.72%
	Annual vol	7.25%	8.56%	11.45%	14.23%	5.79%	6.85%	9.17%	11.53%
	Skewness	0.01	0.05	0.04	0.02	-0.14	-0.09	-0.12	-0.11
	Xs kurtosis	0.45	0.56	0.55	0.59	1.25	1.17	1.19	1.31
	Correlation	0.22	0.21	0.20	0.20	0.30	0.29	0.29	0.28
	Sharpe ratio	0.05	0.35	0.35	0.33	-0.10	0.15	0.16	0.15
minus 15%	Annual return	2.19%	4.82%	5.75%	6.47%	-0.08%	1.50%	1.93%	2.22%
	Annual vol	9.79%	11.15%	14.01%	16.79%	7.97%	9.06%	11.38%	13.72%
	Skewness	0.08	0.09	0.07	0.05	-0.12	-0.09	-0.11	-0.11
	Xs kurtosis	0.65	0.69	0.64	0.66	1.17	1.12	1.15	1.25
	Correlation	0.22	0.21	0.20	0.20	0.30	0.29	0.29	0.28
	Sharpe ratio	0.22	0.43	0.41	0.39	-0.01	0.17	0.17	0.16

		coint				TEV			
Jan 2000 – Dec 2003		**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**	**plus 0%**	**plus 5%**	**plus 10%**	**plus 15%**
minus 0%	Annual return	N/A	-1.80%	-3.76%	-6.44%	N/A	-1.23%	-3.23%	-5.73%
	Annual vol	N/A	3.49%	5.73%	8.64%	N/A	2.38%	4.19%	6.64%
	Skewness	N/A	-0.12	-0.24	-0.33	N/A	0.20	0.08	-0.04
	Xs kurtosis	N/A	1.59	1.61	2.19	N/A	3.84	1.77	1.23
	Correlation	N/A	0.28	0.25	0.23	N/A	0.45	0.44	0.37
	Sharpe ratio	N/A	-0.52	-0.66	-0.75	N/A	-0.52	-0.77	-0.86
minus 5%	Annual return	-1.00%	-2.80%	-4.76%	-7.44%	-2.78%	-4.01%	-6.02%	-8.51%
	Annual vol	5.00%	5.90%	9.00%	12.21%	4.58%	4.91%	7.57%	10.33%
	Skewness	-0.38	-0.40	-0.42	-0.44	-0.46	-0.22	-0.21	-0.23
	Xs kurtosis	3.14	3.22	3.27	3.28	2.43	1.40	1.40	1.40
	Correlation	0.10	0.25	0.22	0.20	0.09	0.30	0.29	0.28
	Sharpe ratio	-0.20	-0.48	-0.53	-0.61	-0.61	-0.82	-0.79	-0.82
minus 10%	Annual return	-2.44%	-4.24%	-6.19%	-8.88%	-4.71%	-5.94%	-7.94%	-10.44%
	Annual vol	7.30%	8.54%	11.62%	14.81%	6.62%	7.16%	9.82%	12.57%
	Skewness	-0.43	-0.40	-0.42	-0.44	-0.37	-0.21	-0.20	-0.22
	Xs kurtosis	3.51	3.18	3.24	3.27	2.02	1.40	1.40	1.39
	Correlation	0.13	0.23	0.21	0.20	0.14	0.28	0.28	0.27
	Sharpe ratio	-0.33	-0.50	-0.53	-0.60	-0.71	-0.83	-0.81	-0.83
minus 15%	Annual return	-4.05%	-5.85%	-7.81%	-10.49%	-6.56%	-7.79%	-9.79%	-12.29%
	Annual vol	9.59%	11.01%	14.08%	17.26%	8.64%	9.31%	11.96%	14.70%
	Skewness	-0.44	-0.39	-0.41	-0.43	-0.32	-0.19	-0.19	-0.20
	Xs kurtosis	3.49	3.02	3.11	3.17	1.86	1.40	1.40	1.39
	Correlation	0.15	0.22	0.21	0.20	0.16	0.27	0.27	0.26
	Sharpe ratio	-0.42	-0.53	-0.55	-0.61	-0.76	-0.84	-0.82	-0.84

30 Stocks. Calibration period 3 years. Rebalanced every 2 weeks.

Why does the cointegration-based statistical arbitrage perform better than the TEV statistical arbitrage? Given the similarity of the empirical performance of the two optimization strategies for index tracking that we have documented, this is a surprising result.

To answer the question, we emphasize the practical difference between tracking an index and tracking an enhanced index. In the first case, one aims to identify the portfolio that stays closest to a real index. Most portfolios with a sufficient number of stocks are likely to stay close to their market index, irrespective of the presence of cointegration. In these circumstances, one cannot really observe the advantage of a cointegration relationship between the tracking portfolio and the benchmark.

The replication task becomes significantly more difficult when one aims to identify portfolios tracking artificial indexes that are designed to underperform or outperform the actual index. In this case, ensuring a stationary spread between the portfolio value and the index starts to pay off and enhances the out-of-sample performance of the statistical arbitrage.

In fact, the mean reversion of returns prevents the TEV-based statistical arbitrage from generating consistent out-of-sample returns. Simply identifying the outperformers or underperformers in-sample does not guarantee that relationships will continue out-of-sample. The fact that the cointegration-optimal tracking portfolio is tied to the artificial benchmark gives a more reliable basis for statistical arbitrage.

Despite the more attractive features of the cointegration-optimal statistical arbitrage, the average performance of both models over the entire 11-year period from 1993 through 2003 is not very encouraging. In terms of Sharpe ratios, the best statistical arbitrage strategy outperformed the market index (a Sharpe ratio of 0.37, as compared to 0.30), but most strategies had lower ratios. So, given the time variability identified in the simple tracking performance, and the fact that this seemed to deteriorate after year 2000, we split the sample into two, 1993-1999 and 2000-2003, and report the results in Exhibit 5, Panel B.

Indeed, there is a great difference in the performance of both statistical arbitrage models in 1993-1999 as compared to 2000-2003. Still, the consistent result over both subsamples is the dominance of the cointegration-optimal statistical arbitrage over the TEV strategy. In the first subsample, both models produce positive average returns with relatively low volatility and distributions close to normality. The average Sharpe ratio produced by the best cointegration statistical arbitrage is 1.12, while the highest average Sharpe ratio produced by the TEV model is 0.97. The arbitrage returns tend to increase with the spread between the benchmarks tracked, but so does their volatility. Thus the best performance continues to be achieved with portfolios tracking narrow spreads around the benchmark.

In the second subsample, 2000-2003, all statistical arbitrage strategies generate negative returns and higher volatility. We can put these losses into perspective if we consider that the market index over the same period lost an annual average of 4.8%, with an average volatility of 21%. The statistical arbitrage returns remained close to market-neutrality during this period, so for the cause of their ineffectiveness one has to look beyond the general market decline in the period analyzed.

In fact, the reason for the poor performance lies in the features of the calibration period. The three years preceding 2000 and the years 2000-2002 were marked by several market crises in a general background of increased volatility. The long-run equilibrium relationships between sectors and industries were affected. The calibration of statistical arbitrage portfolios on such eventful samples is very difficult—it is the quality of the calibration data that is responsible for their poor out-of-sample performance.

These results illustrate only the raw performance of the statistical arbitrage strategies. There is considerable scope for enhancement, by using stock selection, appropriate calibration periods and rebalancing frequencies, or by imposing portfolio constraints.

To summarize the results in this section, both strategies yield returns according to the spread between the benchmarks tracked, and have less volatility than the market, low market correlations, and near-to-normal return distributions. We find these results depend on the quality of the calibration data, and that optimizing portfolios on stressed samples is risky. Targeting wider spreads is penalized in terms of volatility and transaction costs, so the net returns are not linearly related to the arbitrage spread. We have shown that the benefit of ensuring a cointegration relationship with the benchmark pays off for the statistical arbitrage strategies, which clearly dominate their TEV equivalents.

CONCLUDING REMARKS

The theoretical benefits of a cointegration relationship between a tracking portfolio and its benchmark are clear. The two are tied together in the long run; their price spread has minimum volatility; and the model makes full use of the information in stock prices, including that in their common trends.

We conduct a thorough investigation of trading strategies based on cointegration in a realistic out-of-sample framework. Tests of the empirical performance of simple index tracking portfolio optimization models show that the out-of-sample performance of cointegration-based strategies is similar to that of the traditional tracking error variance-minimizing model.

Ensuring a cointegration relationship between a tracking portfolio and a benchmark does not seem to bring any obvious advantage or cost. This is merely because, if enough stocks are included in the portfolio, most models will return a reasonably good index tracking performance, whatever any cointegration relationship.

Yet our comparison of the cointegration-optimal and TEV statistical arbitrage strategies has revealed some interesting results. Depending on the characteristics of the calibration data for statistical arbitrage portfolios, both strategies yield returns according to the spread between the benchmarks tracked, and have less volatility than the market, low market correlations, and near-to-normal return distributions, but the cointegration-optimal statistical arbitrage strategies clearly dominate their TEV equivalents over a very long performance analysis.

Thus the benefit of cointegration relationships appears to be that they are more robust, out-of-sample, than relationships that are identified on returns. This ensures a reliable foundation for statistical arbitrage, reducing the risk of overhedging and the associated trading costs.

ENDNOTES

[1]See Alexander [1999] for a survey of this literature.

[2]According to Beveridge and Nelson [1981], a variable has a stochastic trend and is integrated of order one, if its first difference has a stationary invertible ARMA(p, q) representation plus a deterministic component.

[3]Optimization methods are normally based on current-weighted, i.e., reconstructed benchmarks.

[4]The use of regression to determine optimal stock holdings for tracking a constant weighted index dates to Hersom, Sutti, and Szego [1973]. The novel idea in the cointegration regression is to determine the holdings using a prices on prices (or log prices on log prices) regression.

[5]Its well-known limitations (small-sample problems, asymmetry in treating the variables, at most one cointegration vector) are not effective in this case. Our estimation sample is typically large; there is a strong economic background to treat the market index as the dependent variable; and identifying only one cointegration vector, i.e., one cointegration-optimal portfolio, is sufficient for our purposes. Moreover, from all cointegrated vectors that can be identified through the maximum-likelihood method of Johansen [1991], the OLS estimated coefficients ensure the least volatile spread between the portfolio value and the benchmark value.

[6]Instead of a two-stage process, i.e., normalizing the coefficients after estimation, the unit-sum constraint is normally implemented directly using a constrained least squares method. Indeed this approach must be taken when additional constraints such as no short sales or maximum exposures are imposed on stock weights.

[7]The return difference between the tracking portfolio and the benchmark is called *tracking error*. This terminology differs from the practitioner's use of this term. Practitioners define the tracking error itself as the standard deviation of the returns difference between the portfolio and the benchmark. We call the (squared) standard deviation of the return difference the *tracking error variance*.

[8]To ensure that the survivorship does not impact the relative performance statistics, we test an alternative database, comprising the stocks included in the DJIA at the beginning of 1990; the results are very similar to results obtained for the database of stocks in the DJIA at the end of 2001. This confirms that the relative performance is not affected by a potential survivorship bias.

[9]*Index30* is the price-weighted portfolio of all stocks; that is, the reconstructed DJIA.

[10]If we assume that the highest-priced stocks had above-average performance prior to the stock selection moment, their below-average performance over the next two weeks following the portfolio construction moment indicates mean reversion in stock returns. This phenomenon has been extensively studied (e.g., De Bondt and Thaler [1985]; Lo and MacKinlay [1988]; Poterba and Summers [1988]; and Jegadeesh and Titman [1993]), and behavioral explanations have been provided for it (e.g., Odean [1999]; De Long et al. [1990]; Lakonishok, Shleifer, and Vishny [1994]; and Shleifer and Vishny [1997]).

[11]This is the result of the leverage effect (Black [1976]; Christie [1982]; French, Schwert, and Stambaugh [1987]) and the presence of positive feedback. An initial sell reaction to some bad news will be followed by more selling, driving prices faster below their fundamental levels (De Long et al. [1990]).

REFERENCES

Alexander, C.O. "Optimal Hedging Using Cointegration." *Philosophical Transactions of the Royal Society* Series A, 357, (1999), pp. 2039-2058.

Alexander, C., I. Giblin, and W. Weddington. "Cointegration and Asset Allocation: A New Active Hedge Fund Strategy." *Research in International Business and Finance,* Vol. 16 (2001), pp. 65-90.

Beveridge, S., and C.R. Nelson. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle." *Journal of Monetary Economics*, 7 (1981), pp. 151-174.

Black, F. "Capital Market Equilibrium with Restricted Borrowing." *Journal of Business,* 45 (1972), pp. 444-454.

Chalmers, J.M.R., R.M. Edelen, and G.B. Kadlec. "An Analysis of Mutual Fund Trading Costs." Social Sciences Research Network, 1999.

"Comparing Bid-Ask Spreads on the New York Stock Exchange and Nasdaq Immediately following Nasdaq Decimalization." Research paper, New York Stock Exchange, 2001. Available at www.nyse.com.

De Long, J.B., A. Shleifer, L.H. Summers, and R.J. Waldmann. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4) (1990), pp. 703-738.

De Bondt, W., and R.H. Thaler. "Does the Stock Market Overreact?" *Journal of Finance*, 40 (1985), pp. 793-808.

Dickey, D.A., and W.A. Fuller. "Likelihood Ratio Statistics-Distributions of the Estimates for Autoregressive Time Sseries with a Unit Root." *Journal of the American Statistical Association*, 74 (1979), pp. 427-429.

Granger, C.W.J. "The Typical Spectral Shape of an Economic Variable." *Econometrica*, 34 (1966), pp. 150-161.

Granger, C.W.J., and T. Terasvirta. *Modelling Nonlinear Economic Relationships*. Chapter 5. Oxford: Oxford University Press, 1993.

Hersom, S.E., C. Sutti, and G.P Szego. "On Best Linear Fit Index." Technical Report 46, Hatfield Polytechnic Centre for Numerical Optimization, 1973.

Jagannathan, R., and T. Ma. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58 (2002), pp. 1651-1684.

Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (1993), pp. 65-91.

Johansen, S. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, 59 (1991), pp. 1551-1581.

Lakonishok, J., A. Shleifer, and R. Vishny. "Contrarian Investment, Extrapolation and Risk." *Journal of Finance*, 49 (1994), pp. 1541-1578.

Lintner, J. "The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, 4 (1965), pp. 13-37.

Lo, A., and C. MacKinlay. "Stock Market Prices Do Not Follow Random Walks—Evidence from a Simple Specification Test." *Review of Financial Studies,* 1 (1988), pp. 41-66

Lucas, A. "Strategic and Tactical Asset Allocation and the Effect of Long-Run Equilibrium Relations." Research Memorandum 1997-42, Free University of Amsterdam, 1997.

MacKinnon, J. "Critical Values for Cointegration Tests." In R.F. Engle and C.W.J. Granger, eds., *Long-Run Economic Relationships*. Oxford: Oxford University Press, 1991, pp. 267-276.

Markowitz, H. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons, 1959.

Odean, T. "Do Investors Trade Too Much?" *American Economic Review*, 89 (1999), pp. 1279-1298.

Plerou, V., P. Gopikrishnan, B. Rosenow, N. Amaral, T. Guhr, and E. Stanley. "Random Matrix Approach to Cross Correlations in Financial Data." *Physical Review E*, Vol. 65 (2002), pp. 1-18.

Poterba, J., and L. Summers. "Mean Reversion in Stock Returns: Evidence and Implications." *Journal of Financial Economics*, 22 (1988), pp. 27-59.

Sharpe, W. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance,* 19 (1964), pp. 425-442.

Shleifer, A., and R. Vishny. "Limits of Arbitrage." *Journal of Finance*, 52 (1997), pp. 35-55.

Stock, J.H., and M.W. Watson. "Testing for Common Trends." *Journal of American Statistical Association*, 83 (1988), pp. 1097-1107.

To order reprints of this article, please contact Ajani Malik at amalik@iijournals.com or 212-224-3205.