

The Profitability of Pairs Trading Strategies Based on ETFs

JEL Classification Codes: G10, G11, G14

Keywords: Pairs trading, relative value arbitrage, statistical arbitrage, weak-form market efficiency, ETFs

Sept. 20, 2014

Susana Yu*

Montclair State University

Gwendolyn Webb

Baruch College/CUNY

*Contact Author

Susana Yu

Montclair State University

Department of Economics and Finance

1 Normal Avenue

Montclair, NJ 07043

917-834-5238

Susana Yu <yu.susana@gmail.com>

The Profitability of Pairs Trading Strategies Based on ETFs

Abstract

Our analysis will extend the body of empirical literature of pairs trading to the realm of ETFs in the U.S. We will test whether a pairs trading strategy on ETFs is profitable and whether the decline in pairs trading profitability on stocks also applies to ETFs. We will explore the robustness of our results to microstructure factors such as the bid-ask bounce, short-selling costs, and transaction costs. We will test two different approaches to normalizing prices on which the pairs are formed; compare the profitability of pairs trading strategies on equity versus commodity ETFs; and test long and short positions separately to determine whether it is only the long position that drives the results.

Introduction

Hedge funds follow a wide range of innovative and creative investment strategies, and employ them flexibly when they work, and abandon them quickly when they don't. One such strategy is relative value arbitrage, which involves taking long and short positions simultaneously in two securities that are in some way mispriced relative to each other. Because the positions are matched and offsetting, they are hedged to a certain extent. Because the short position generates a cash inflow and the long position a cash outflow, they are "self financing," in that they require relatively little net capital.¹ When a profit is guaranteed, this position becomes a true arbitrage. When a profit is not guaranteed, but expected most of the time, the position is not a true arbitrage, but may still be profitable on average. Since these positions are not riskless, profits may reflect compensation for risk bearing or for the provision of liquidity.

One form of relative value arbitrage that has been the subject of rigorous academic studies is pairs trading. The concept of pairs trading is straightforward and based solely on past price dynamics and contrarian principles. The trader begins by finding two stocks whose prices have been highly correlated in the past. The procedure is to observe the two stocks, and when the

¹ In many cases, the proceeds of a short sale are not made available to the short seller but must remain on account in the seller's brokerage account. In this way, the short/long combination uses capital, even though in theory the position can be considered self financing.

spread between them widens enough to set off a trigger, the trader shorts the winner and buys the loser. The position is maintained until the spread converges enough to set off a second trigger to close both sides of the position. If the deviation is temporary, prices will converge and the trader/arbitrageur will profit.

Weak-form market efficiency suggests that historical trading data, such as prices, cannot be used to predict future prices. A key consequence of weak-form market efficiency is that a trading rule based on historical data should not result in significant positive excess returns. Thus, if a pairs trading strategy is found to be profitable, it appears to violate weak-form market efficiency. However, if the trading strategy is not perfectly riskless, any observed profitability may be compensation for risk taking or liquidity provision.

It may be added that an interesting aspect of tests of pairs trading is that it provides a test of market efficiency that does not require a valuation model. Instead, it only requires that two securities are mispriced relative to each other. This makes tests of pairs trading strategies compelling because they offer the capability of providing significant insights into the strength or weakness of market efficiency.

The first empirical analysis pairs trading was that of Gatev, Goetzmann, and Rouwenhorst (1999). They find that pairs trading based on their initial sample period is profitable. In 2006 they updated the analysis and incorporated an out-of-sample test. They find that pairs trading was less profitable in the more recent years. Do and Faff (2010) extend their analysis to 2009 and find that the profitability of pairs trading strategies on individual stocks has declined in the still more recent years.

ETFs are investment funds traded on stock exchanges, much like stocks. They have several unique characteristics that may be useful in a pairs trading strategy. First, when based on

stock indexes like the S&P 500, ETF are designed to track a well-diversified portfolio of stocks. For this reason, ETFs should have much less idiosyncratic risk than individual common stocks. It has been shown that pairs trades are less profitable when the original price deviation is associated with a fundamental change in the value of the stock. Also, ETFs are not subject to risk of bankruptcy risk, as individual firms are. Second, ETFs can be sold short more easily than stocks because the uptick rule doesn't apply to them. For this reason, it may be easier for arbitrageurs to employ a pairs trading strategy on ETFs than on stocks.

These distinguishing characteristics of ETFs motivate our research. Our analysis will extend the body of empirical literature of pairs trading to the realm of ETFs in the U.S. Our first tasks are to test whether a pairs trading strategy on ETFs is profitable and whether the decline in pairs trading profitability on stocks also applies to ETFs. To do this we will examine the profitability, risk, and return characteristics of pairs trading on exchange-traded funds (ETFs) in the U.S. In addition, we will explore the robustness of our results to microstructure factors such as the bid-ask bounce, short-selling costs, and transaction costs. Since pairs trading strategies are a form of technical analysis, there are several challenges in evaluating the test results, especially problems associated with data-snooping, transaction costs, and liquidity.

We plan several specific tests of the effectiveness of the pairs trading strategy. We will test two different approaches to normalizing prices on which the pairs are formed. We further plan to compare the profitability of pairs trading strategies on equity versus commodity ETFs. Prior research indicates that most of the profit of pairs trading strategy on common stocks comes from the long end. We will test long and short positions separately to determine whether it is only the long position that drives the results.

The remainder of the article is organized as follows. Section 1 reviews prior studies on pairs trading strategy and ETFs. The next section describes our data and methodology of constructing pairs and calculating returns. The empirical results are then described, and conclusions will follow.

I. Review of Literature on Pairs Trading

Gatev, Goetzmann, and Rouwenhorst (1999) were the first to test the profitability of pairs trading strategies based on common stocks. They find that the strategy is significantly profitable. They updated their analysis in 2006 and extended the time period of the data by several years. They were careful to retain the design and parameters of their original strategy, so their extension serves as a true out-of-sample test. This qualification is very important because it reduces the chance that their results are due to data-snooping. Their algorithm matches pairs by minimizing the historical price spreads over 12 months and trades them in the following 6 months by using two historical standard deviations as the opening trigger. The extension of their analysis to a more recent sample (GGR 2006) confirms the existence of the strategy's profitability.

Engelberg, Gao, and Jagannathan (2009) focus on determining the sources of profitability of the pairs trading strategy. They find that the strategy is less profitable in cases when the cause of the price divergence is related to news that affects the value of one of the stocks. This suggests that trades which are triggered by stock specific, or idiosyncratic, news events are less profitable. The strategy is more profitable when the cause of the price divergence is related to temporary liquidity factors, or to news that affects both stocks, but one reacts more quickly to the news. They extend the analysis to test for the optimal time to exit from a pairs trade, and find that it is

more profitable to close out positions after 10 days if the prices of the two stocks don't converge on their own within that time period.

Another line of inquiry into the nature of the profitability is Papadakis and Wysocki (2008), who examine pairs formed around accounting events such as earnings announcements. They find that pairs are often triggered by these events, and these pairs are often less profitable than those not triggered by them. This is consistent with Engelberg, Gao, and Jagannathan's findings that pairs are less profitable when the triggering event is associated with a change in value of the underlying stock. That is, when the trigger is tied to some form of idiosyncratic risk.

Do and Faff (2010) extend GGR's (1999) test period by another six years and find a downward trend in profitability of the pairs trading strategy. They further distinguish between different possible sources of profitability, focusing on improvements in market efficiency and higher arbitrage risks. There are three sources of arbitrage risk: fundamental risk, noise trader risk, and synchronization risk. They conclude that the general decline in profitability is due to worsening arbitrage risks rather than to an increase in market efficiency. An interesting implication of this is that the lower profitability of the strategy is not due to large numbers of arbitrageurs employing it. However, when they break their sample down into four subperiods, they find that the balance of more efficiency vs. higher arbitrage risks varies from one time period to another. Finally, they test whether the strategy is more profitable when the pairs are formed from stocks in the same industry, and verify that selecting pairs of similar or related stocks improves the results of the strategy.

Chen, Chen, and Li (2012) test whether the profitability of the pairs strategy is due to different rates of information diffusion between stocks or whether it represents compensation for provision of liquidity. Their results provide more support for the delay in diffusion hypothesis

than to the liquidity provision hypothesis. For example, the pairs trades are more profitable when the stocks are less well known and followed by fewer analysts, which are cases with slower informational diffusion. The strategy has become less profitable over time, possibly as the result of more efficient markets, and consistent with faster diffusion associated with more arbitrageurs following the strategy. Overall their evidence provides little support to the liquidity provision hypothesis.

Several studies have examined pairs trading in other countries. Perlin (2009) applies various pairs-trading strategies to the Brazilian financial market in the 2000-2006 period. His empirical results suggest that pairs trading strategies are potentially profitable in the Brazilian financial market, especially for strategies using daily data. Bogomolov (2010) tests the strategy on Australian data, and finds little profitability in that market. These comparative results are consistent with the hypothesis that the Australian financial markets are more efficient in the weak-form sense than those in Brazil.

Andrade, di Pietro and Seasholes (2005) hypothesize that purchases or sales by uninformed buyers are the causes of triggers. To do this, they employ data on stocks in Taiwan. There is just a single stock market, and they are better able to identify uninformed buyers there than in the CRSP NYSE data, that reflect trades in regional markets as well as the NYSE. They first find that the strategy is profitable, and then that there is a clear association between profitability and uninformed trading. This leads back to the hypothesis that returns to arbitrageurs are associated with compensation for liquidity provision.

Recent literature has extended to tests using ETFs. Schizas, Thomakos, and Wang (2011) examine the pairs trading strategy on a set of international ETFs. They find that the strategies are profitable, that there is a difference between the profitability of the long and short sides of the

positions, and that profitability can be explained by fundamental factors, such as EPS, dividend yield, and unemployment.

II. Data and Methodology

Our data sample covers all securities with share code of 73 in the CRSP daily database. Daily total return and price data are from the CRSP daily database. As shown in Table 1, since the introduction of first ETF in January 1993, it took nearly 10 years for the number of ETFs in the U.S. to cross the first 100 milestone (2001) and the another 10 years to cross the 1,000 mark (2011). In order to have meaningful number of pairs in our sample space, we limit our research to the years 2002 through 2013.

Following GGR, we implement the pairs trading strategy in two stages: the formation period and the trading period. We first form ETF pairs by using the 12 months of daily return data and then trade them in the subsequent 6- month period. For a given pair, we initiate a position each time the absolute distance between the prices is above a preset threshold value. If the ETFs, after the divergence, return to their historical relation, then we expect that the one with higher price will decline in value and the one with the lower price will increase in value. All long and short positions are taken according to this procedure. The strategy is implemented every month, effectively allowing up to six portfolios simultaneously.

For each ETF, we form a pair by finding another ETF that minimizes the sum of squared differences (SSD) in the normalized prices of the two ETFs. We use two normalization procedures because ETFs in different categories may have distinctive risk characteristics.

Normalization based on total return index. Following GGR (2006), Do and Faff (2010) normalize the stock prices with the daily total returns, dividends included, scaling both

price series to start at \$1 in the formation period. Hence, our first method of normalizing ETF prices is the total return index. Once we have identified all ETF pairs, we also scale the price series to \$1 at the beginning of the trading period.

When the new paired normalized price spread widens by more than two standard deviations from the historical spread calculated over the formation period, we long (and short) the ETF with lower (higher) normalized prices. Both positions are unwound at the next crossing of the normalized prices, on the day a stock was delisted, or on the last day of the trading period, whichever occurs first. Positions are allowed be opened and closed multiple times during the 6-month trading period.

Normalization based on risk and expected return. The second method of price normalization, as used by Perlin (2009), adjusts ETF prices in the formation period with the mean and standard deviation computed over the same period. This method takes into consideration an EFT's risk characteristics. The transformation employed is the normalization of the price series based on its mean and standard deviation, equation (1).

$$P_{it}^* = \frac{P_{it} - E(P_{it})}{\sigma_i} \quad (1)$$

The value P_{it}^* is the normalized price of ETF i at time t , $E(P_{it})$ is the expected value of the price, P_{it} , and is the average price in our case, and σ_i is the standard deviation of the respective ETF price. Once we have identified all ETF pairs, we also normalize the ETF prices with their respective means and standard deviations from the formation period. When the new normalized price spread widens by more than two standard deviations, we long (and short) the ETF with lower (higher) normalized price.

Analysis of the trading strategy's outcome. Once we have determined our ETF pairs in the formation period, we study the top 20 pairs and the next 20 pairs with the smallest historical

distance measures. On the day following the last day of the formation period, we begin to trade according to a pre-specified rule. Once we have identified pairs, we take a \$1 long–short position in a pair. Daily returns are compounded to form monthly returns. This has the straightforward interpretation of a buy-and-hold strategy.

Our methodology involves two key design decisions: (1) How to trigger a long/short position based on pairs-trading strategy and (2) How to evaluate the performance of the trading signals.

We define monthly excess returns to the portfolio as the equally weighted average of returns by the six “managers.”

We monitor pairs trading under each of the alternative trading rules: (1) the positions are opened and closed on the same day of the trigger; and (2) positions are opened or closed after a delay of one-day. Using the latter rule, as suggested by GGR, we attempt to alleviate concerns regarding the potential upward bias in the reported returns induced by bid-ask bounce.

Following prior researchers, we examine two excess return measures: the return on committed capital (the total mark-to-market payoff for all pairs divided by the number of pairs in the portfolio) and the return on employed capital (the total payoff divided by the number of pairs that are actually traded). The former scales the portfolio payoffs by the number of pairs that are selected for trading, the latter divides the payoffs by the number of pairs that open during the trading period. The former measure of excess return is clearly more conservative: if a pair does not trade for the whole of the trading period, we still include a dollar.

In addition to "unrestricted" pairs, we will also present results by ETF categories, where we restrict both ETFs to belong to the same broad categories: equities, bonds, and commodities. The minimum-distance criterion is then used to match ETFs within each of the groups.

References

- Abreu, D., and M. Brunnermeier. 2002. Synchronization risk and delayed arbitrage. *Journal of Financial Economics*, 66(2-3), 341–360.
- Andrade, S., V. di Pietro, and M. Seasholes. 2005. Understanding the profitability of pairs trading. Working paper, University of California, Berkeley, and Northwestern University.
- Bogomolov, T., Pairs trading in the land down under (November 30, 2010). Finance and Corporate Governance Conference 2011 Paper. Available at SSRN: <http://ssrn.com/abstract=1717295> or <http://dx.doi.org/10.2139/ssrn.1717295>
- Chen, H., S. Chen, and F. Li, Empirical investigation of an equity pairs trading strategy (September 27, 2012). Available at SSRN: <http://ssrn.com/abstract=1361293> or <http://dx.doi.org/10.2139/ssrn.1361293>
- De Long, J., A. Shleifer, L. Summers, and R. Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4), 703–738.
- Do, B., and R. Faff, 2010, Does simple pairs trading still work? *Financial Analysts Journal*, 66(4), 83-95.
- Engelberg, J., P. Gao, and R. Jagannathan. 2009. An anatomy of pairs trading: the role of idiosyncratic news, common information and liquidity. Working paper, University of North Carolina at Chapel Hill, University of Notre Dame, and Northwestern University.
- Fama, E., and K. French. 1997. Industry costs of equity. *Journal of Financial Economics*, 43(2), 153–193.
- Gatev, E., W. Goetzmann, and K. Rouwenhorst. 1999. Pairs trading: performance of a relative value arbitrage rule. Working paper, Yale School of Management.
- Gatev, E., W. Goetzmann, and K. Rouwenhorst, 2006, Pairs trading: performance of a relative-value arbitrage rule, *Review of Financial Studies*, 19(3), 797-827.
- Hogan, S., R. Jarrow, M. Teo, and M. Warachka. 2004. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics*, 73(3), 525–565.
- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1), 65–91.
- Khandani, A., and A. Lo. 2007. What happened to the quants in August 2007? *Journal of Investment Management*, 5(4), 5–54.
- Kondor, P. 2009. Risk in dynamic arbitrage: the price effects of convergence trading. *Journal of Finance*, 64(2), 631–655.

Mitchell, M., T. Pulvino, and E. Stafford. 2002. Limited arbitrage in equity markets. *Journal of Finance*, 57(2), 551–584.

Papadakis, G. and P. Wysocki, 2007, Pairs-trading and accounting information, Working paper, Boston University School of Management and MIT Sloan School of Management.

Perlin, M., 2009, Evaluation of pairs-trading strategy at the Brazilian financial market, *Journal of Derivatives & Hedge Funds*, 15(2), 122-136.

Rudy, J., C. Dunis, and J. Laws, Profitable pair trading: a comparison using the S&P 100 constituent stocks and the 100 most liquid ETFs (November 10, 2010). Available at SSRN: <http://ssrn.com/abstract=2272791> or <http://dx.doi.org/10.2139/ssrn.2272791>

Schizas, P., D. Thomakos, and T. Wang, Pairs trading on international ETFs (November 12, 2011). Available at SSRN: <http://ssrn.com/abstract=1958546> or <http://dx.doi.org/10.2139/ssrn.1958546>

Table 1
Population of ETF Funds by Year

Year	Number of Funds
1993	1
1994	1
1995	2
1996	19
1997	19
1998	29
1999	32
2000	93
2001	118
2002	130
2003	134
2004	169
2005	221
2006	375
2007	643
2008	723
2009	830
2010	957
2011	1,144
2012	1,192
2013	1,282