Journal of Accounting, Finance, Economics, and Social Sciences
Vol 8, No 1, January - June 2023, pp.1-12
Received: Apr 2023 | Revised: May 2023 | Accepted: Jun 2023
DOI: https://doi.org/10.62458/jafess.160224.8(1)1-12

An Empirical Study of Pairs Trading in Cambodia Securities Exchange

Siphat Lim*

ABSTRACT

A pair trading strategy of two shares works when there is a co-integration between the two equities. This paper uses the two-step Engle and Granger method to establish whether or not the two stock price series had a long-run relationship. Since the residual term predicted from the sample regression function between the dependent variable, PPSP, and the independent variable, PWSA, is stationary, as indicated by the ADF test, all listed stocks in CSX, a pair of equities, PPSP and PWSA, are co-integrated. During the research period, the price spread increased three times to two standard deviations. At any point, one unit of outperformance stock is short, and one unit of co-integrated ratio is long. Pair trading has an average investment return of 8.0264 percent, outperforming the weighted average return of 1.1651 percent.

Keywords: Pair trading; CSX; Co-integration; ADF test

INTRODUCTION

Investors are seeking new methods, instruments, and solutions in the face of growing market volatility and upheaval to support their capital. Many structured products have been developed and are too complex for average investors to approach, such as Synthetic Options and Knock-In Knock-Out options or the growth of derivatives products. A strategy for investment instead remained on newspaper pages despite the decade-long fall from one-pair trading, one of the most widely used investment techniques.

Pair trading is a market-neutral strategy that enables investors to take advantage of various market movements. Gerry Bamberger pioneered the pair trading strategy, which Nunzio Tartaglia's Investment Bank Morgan Stanley team built in the 1980s (Wilmontt, 2004). If their connection is less in the near term owing to the news and income, one stock is rising, and one stock is decreasing, it is a way to cut a prolonged and excessive stock, risking the spread of prices between two stocks to average. The shift in demand and supply, large purchases or sales volumes, and market reaction to the news or revenue reporting of the firm, among other factors, are believed to generate this short-term variation.

Since April 2012, the Cambodia Securities Exchange (CSX) has been in operation. There are now seven listed stocks, such as Pestech (Cambodia) Plc. (PEPC), ACLEDA Bank Plc. (ABC), Sihanoukville Autonomous Port (PAS), Phnom Penh Special Economic Zone (SEZ) Plc. (PPSP), Phnom Penh Autonomous Port (PPAP), Grand Twins International (Cambodia) Plc. (GTI), and Phnom Penh Water Supply Authority (PWSA). The listed firms have been divided into five distinct industries, which are as follows: financial, port services, SEZ developer, apparel clothing, and power. Even though the market capitalization of the listed stocks in CSX was USD 2.4 billion in June 2021 and has been classified as one of the world's smallest capital markets, it is worthwhile to investigate if a pair trading method exists in this market. The primary goal of this research is to provide a system for constructing pair trading that can be used for any pair of stocks listed on CSX.

LITERATURE REVIEW

According to Gatev et al. (1999), each stock is paired with a matching partner with the minimum normalized historical price deviation. The pair-trading introduced by Gatev et al. (1999) yields profitable results after an allowance for trading costs, and these profits are inherently different from a pure mean-reversion strategy. Patanapol (2001) further confirmed that the approach could generate an abnormal return, exceeding the average market return in the Stock Exchange of Thailand (SET) during 2001-2002. Nath (2003) is more concerned with risk control and prevents huge losses by employing a stop-loss trigger to close the position once the distance widens further

* Siphat Lim, Ph.D., Professor, CamEd Business School
Email: lsiphat@cam-ed.com

to hit the 5th percentile. The additional measure can improve the portfolio performance relative to various benchmarks. The distance approach mainly employs the statistical relationship between two pairing stocks. While the method is not prone to model misspecification and misestimating, being nonparametric statistics suggests that the strategy cannot forecast the portfolio's expected holding period or convergence time.

Co-integration test on long-term co-movement of stock prices is considered a critical criterion for pairs-trading. Lin et al. (2006) employed Co-integration-based analysis using daily closing price data of two shares from the Australian Stock Exchange between January 2, 2001, and August 30, 2002. The findings indicate that the trading strategy can generate a substantial return and is unrestricted by implementing a reasonable minimum profit condition to secure against losses. Hong & Susmel (2003) undertake pairs-trading analyses on 64 Asian shares listed in their local market and the U.S. as ADRs (American Depositary Receipts). The outcomes of this study demonstrate that pair trading in this market could yield significant profits. The results are robust to various profit measures of the strategies and holding periods. Further, research on Co-integration-based pairs-trading was done by Vidyamurthy (2004) to develop the forecasting framework, though empirical results still needed to be done. The approach presented by Vidyamurthy (2004) may be subject to errors arising from the econometric models, particularly misspecification and misestimating. The Engle-Granger method's two-step Co-integration procedure raises a question about selecting the independent and dependent variables in the equation.

Zapart (2004) introduces an approach in which stocks from the same industry are paired following their wavelet correlation measure. Correlation analysis for the two chosen stocks based on time series would be performed using a highly optimized wavelet correlation measure with artificial neural networks and genetic algorithms. Based on twenty-five NYSE stocks, trading simulations suggest the model can exploit consistent profits in rising and falling markets even when a typically established long-short arbitrage strategy remains flat. Comparing the two measures by the statistical arbitrage fund managers, the static wavelet correlation measure outperforms a trading system based on a dynamic risk model during the studied period between 2000 and 2004.

Elliott et al. (2005) estimate a parametric spread model using a Kalman filter. The method can provide a consistent and robust model to predict the spread for making an investment decision, with several advantages from the empirical perspectives. First, it utilizes the mean reversion strategy of pairs-trading. Second, the continuous-time feature of the model makes it capable of forecasting the expected holding period. However, the fundamental issue of this approach is that the model restricts the long-term relationship between the two stocks to one of return parity. In the long run, the two chosen stocks must offer the same return, and any deviation from it will be expected to be corrected in the future. In practice, finding such two stocks with identical returns is challenging.

According to the research of Jurek and Yang (2007) on dynamic portfolio selection in arbitrage, if two equities are co-integrated, and one share is overpriced while the other is underpriced at some point, the prices of the mispriced pair should coincide. Thus, a position should be opened by shorting the overpriced share and longing for the underpriced share. This technique is known as the "Long-Short (LS) Strategy." In terms of portfolio weight or number of shares, the position should be taken at the same magnitude but with the opposite sign.

Profit was made by employing the pair trading strategy on the New York Mercantile Exchange. The trading concept arose from the mean-reversion process of the price gap between two separate assets under consideration in the same or different sectors. The spread's strong mean-reversion and high volatility contribute to the expected return from trading. Kanamura et al. (2008) discovered that this approach works effectively with natural gas, heating oil, and WTI crude oil futures (2008).

Krauss (2017) identified five types of pair trading approaches: distance trading, co-integration trading, time series trading, stochastic control trading, and other trading strategies. Distance and co-integration techniques are the most popular, with several literature evaluations done by several researchers. Several distance measures based on the distance method are utilized throughout the formation stage to discover co-moving shares. During the trading session, simple nonparametric threshold criteria activate trade alerts. The significant advantages of this method are its simplicity and transparency, which enable large-scale empirical applications. The findings establish distance pairs trading as a pro table over a wide range of markets, asset classes, and time periods. For the co-integration approach, in a formation phase, Engle and Granger co-integration or other co-integration tests are employed to detect co-moving stocks. Simple algorithms are used to produce trade signals during the trading period, with the bulk of them based on the Gatev et al. (1999) threshold criterion mentioned earlier. Of course, this research study is likewise conducted by these criteria.

To evaluate the five pair trading techniques: correlation, distance, stochastic, stochastic differential residual, and co-integration, Blázquez and Román (2018) used empirical pair trading approaches to determine how pair trading is selected. The research focuses on the banking industry in the United States, represented by the S&P500 index. This study analyzed daily stock prices from January 1, 2008, until December 31, 2013. The empirical results of this study show that co-integration and distance procedures of a pair of stocks meet the property of residual series better than other approaches. The results of the two techniques also show that the larger the co-integrated ratio of a pair of stocks, the better the residual series fulfills the essential properties. The outcome of the stochastic approach used in this study is also consistent with the findings of Elliott et al. (2005), Jurek and Yang (2007), and Kanamura et al. (2008), which revealed to be dependent on the premise that the strategy's residual series is represented by an Ornstein-Uhlenbeck stochastic model with constant k, θ and σ parameters, which is extremely difficult to solve.

Namwong et al. (2019) applied a Markov-switching GARCH model in two different regimes, which have different variances in each regime, and the sum of the two variances is used as a pair-trading signal in the Stock Exchange of Thailand (SET). The empirical results demonstrate that the Markov-switching GARCH model trading signals provide positive returns for all selected pairings and give the greatest return of up to 14.27 percent. Furthermore, pair trading yields a larger return than solo stock trading.

Due to the need for an efficient approach to determining the long or short ratio of investment in the equity of a particular pair of companies, Ramos et al. (2020) created a novel strategy for defining the investment ratio. Six distinct techniques have been discovered in determining weighted components indicated as b: equal weight, stock price standard deviation, minimal distance of the log-prices, correlation of return, co-integration of prices, and lowest Hurst exponent of the pair. This research paper employed four pair trading methods: correlation, co-integration, distance method, and Hurst exponent on the components of the Nasdaq 100 index technological sector in the United States of America. Daily stock price sample sizes were separated into two sub-periods: January 1999 to December 2003 and January 2007 to December 2012. The first sub-sample experienced the "dot.com boom and burst," whereas the second experienced the "subprime mortgage crisis." According to the outcomes of this study, a novel approach to computing weighted factors caused the return to surpass the current methods employed.

It is essential to check whether the data series are stationary or non-stationary in time series econometric techniques. The time series data are usually non-stationary, and modeling with non-stationary data series may increase the possibility of a spurious regression problem from which no valid statistical inference can be made. Moreover, the variance of a non-stationary series changes with time, and the critical assumption of OLS estimation breaks down.

So stationary or lack of it is an important property of time series processes. In general, a collection of N-dimensional random vectors $y_{t-1}; y_t; \dots, y_{t+1}$ is called a stochastic *stationary* process if,

i. All the random vectors have the same mean vector $E[y_t] = \mu$ for all t, so $E[y_t] = E[y_{t+1}]$ for any *t* and *k*.

ii. The variance of all involved random variables is a finite constant σ_y^2 $var(y_t) = \sigma_y^2$ for all *t*, so that $var(y_t) = var(y_{t+k})$ for any *t* and *k*.

iii. The covariance matrices of vector y_t and y_{t+k} that are *k* periods apart do not depend on *t* but only on *k*.

$$cov(y_t; y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)'] = \Gamma \text{ for all } t$$

so that $cov(y_t; y_{t+k}) = cov(y_{t+n}; y_{t+n+k})$ for any *t k* or *n*

Estimating the abovementioned conditions implies that the time series under consideration must not have trends, fixed seasonal patterns, or time-varying variances. If time series variables do not possess the properties (i), (ii), and (iii), the variables are said to be generated by a non-stationary process. The significant difference between stationary and non-stationary time series is that shocks to a stationary time series are necessarily short-lived.

The “carry over” effect of an old shock on the current value of the series would be insignificant if the shock happened long enough ago. Over time, the results of the shock will dissipate, and the series will return to its long-run mean level. In other words, a stationary series will have a well-determined mean which will not vary significantly with the sampling period. On the other hand, in the case of a non-stationary series, an old shock will still have a noticeable impact on the current value of the series. A non-stationary series’s mean and/or variance are time-dependent, and we cannot generally properly use the term mean without referring to some particular time period.

The simplest example of a non-stationary process is the random walk, which is well represented by the following equation:

$$x_t = x_{t-1} + \varepsilon_t \qquad (1)$$

where $\varepsilon_t \sim iid(0,\sigma^2)$, so that if $x_0=0$,

$$x_t = \sum_{t=1}^{T} \varepsilon_t$$

The previous equation implies that old shocks have equal weight to new shocks in determining the current value of x_t. The variance of x_t is $t\sigma^2$, which becomes indefinitely large as $t \rightarrow \infty$. It is also clear that the concept of a mean value for x_t has no meaning. While modeling time series data, it is essential to know whether or not the underlying stochastic process that generated the series can be assumed to be invariant concerning time.

If the characteristics of the stochastic process change over time, i.e., if the process is non-stationary, it will be challenging to represent the time series over past and future time intervals in a simple algebraic model. Moreover, the statistical properties of regression analysis and estimators using non-stationary time series are dubious, as evidenced by the substantial literature on “spurious regression.” For example, the Gauss-Markov theorem will not hold if a random walk has no finite variance. Hence, OLS would not yield a consistent parameter estimator. If the series are non-stationary, the series is likely to end up with a model showing promising diagnostic test statistics even when there is no sense in the regression analysis.

To overcome the problem of non-stationary data and “spurious regression,” the usual and common practice is to differentiate the time series to achieve stationary. A non-stationary series is said to contain an integrated component, and it should be differenced before the estimation process to achieve stationary. Following a formal definition by Engel and Granger (1987) who introduced the concept of integrated series into econometrics, a series x_t is said to be integrated of order d (denoted $I(d)$) if it is a series which has a stationary, invertible, non-deterministic *ARMA* representation after differencing d times.

Based on this definition, a stationary series is said to be integrated of order zero, $x_t \sim I(0)$. For a linear combination of two series, each integrated at different levels, the resulting series will be integrated at the higher of the two orders of integration. For example, suppose:

$$Z_t = bx_t + cy_t \text{ where, } x_t \sim I(d_x),\ x_t \sim I(d_y) \qquad (2)$$

Then, in general, $Z_t = I\ (max(d_x, d_y))$.

However, this need is always the case. The exception to this rule gives rise to the new concept of co-integration. An important exception to this rule occurs when the common integrating factor of two or more variables exactly offsets each other to give a stationary Z series $Z \sim I(0)$. The basic idea of a set of co-integrating variables is that if two or more series move closely together in the long run, even though the series themselves are trended, the difference between them is constant. A detailed discussion of the concept of co-integration will follow (Hamilton, 1994; Hayashi, 2000).

The following discussion outlines the basic features of unit root tests. By necessity, the discussion will be brief.

Consider a simple *AR(1)* process:

$$y_t = \rho y_{t-1} + x_t^{'} \delta + \epsilon_t \qquad (3)$$

Where x_t are optional exogenous regressors consisting a constant or a constant and trend, ρ and δ are parameters to be estimated, and the ϵ_t are assumed to be white noise. If $|\rho| \geq 1$, y is a nonstationary series, and the variance increases with time and approaches infinity. If $|\rho| < 1$, y is a (trend-) stationary series. Thus, the hypothesis of (trend-) stationarity can be evaluated by testing whether the absolute value of ρ is strictly less than one.

The unit root tests in some programs provide a general test the null hypothesis H_0: $\rho=1$ against the one-sided alternative H_1: $\rho<1$. In some cases, the null was tested against a point alternative. In contrast, the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS)

Lagrange Multiplier test evaluates the null of H_0: $\rho<1$ against the alternative H_1: $\rho=1$ (KPSS, 1992).

The standard *DF* test was carried out by estimating equation (3) after subtracting y_{t-1} from both sides of the equation:

$$\Delta y_t = \alpha y_{t-1} + x_t' \delta + \epsilon_t \qquad (4)$$

where $\alpha= \rho-1$. The null and alternative hypothesis may be written as,

$$H_0: \alpha = 0$$

$$H_1: \alpha < 0$$

and evaluated using the conventional *t-ratio* for α:

$$t_\alpha = \frac{\hat{\alpha}}{(se(\hat{\alpha}))} \qquad (5)$$

where $\hat{\alpha}$ is the estimate of α, and $se(\hat{\alpha})$ is the coefficient standard error.

Dickey and Fuller (1979) show that under the null hypothesis of a unit root, this statistic does not follow the conventional Student's t-distribution, and they derive asymptotic results and simulate critical values for various tests and sample sizes. More recently, MacKinnon (1996) implemented a much larger set of simulations than those tabulated by Dickey and Fuller. In addition, MacKinnon estimates response surfaces for the simulation results, permitting the calculation of Dickey-Fuller critical values and p-values for arbitrary sample sizes. Some programs constructing test output use the more recent MacKinnon critical value calculations. The simple Dickey-Fuller unit root test described above is valid only if the series is an *AR(1)* process. If the series is correlated at higher order lags, the assumption of white noise disturbances ϵ_t is violated. The Augmented Dickey-Fuller *(ADF)* test constructs a parametric correction for higher-order correlation by assuming that the series follows an *AR(p)* process and adding *p* lagged difference terms of the dependent variable to the right-hand side of the test regression:

$$\Delta y_t = \alpha y_{t-1} + x_t' \delta + \beta_1 \Delta y_{t-1} + \beta_2 \Delta y_{t-2} + \ldots + \beta_p \Delta y_{t-p} + \epsilon_t \quad (6)$$

This augmented specification is then used to test (4) using the -ratio (5). A significant result obtained by Fuller is that the asymptotic distribution of the *t-ratio* for α is independent of the number of lagged first differences included in the *ADF* regression. Moreover, the assumption that *y* follows an autoregressive *(AR)* process may seem restrictive.

A frequent question raised in time series analysis is whether one economic variable can forecast another economic variable. According to Granger (1969), testing causality involves using *F-tests* to identify whether lagged information on a variable *x* provides any statistically significant information about a variable y in the presence of lagged *y*. If not, then "*x* does not Granger-cause *y*."

A bivariate vector autoregression is specified and applied to implement the Granger causality case. Assume a particular autoregressive lag length *p*, and estimate the following equation by ordinary least squares *(OLS)*:

$$y_t = \theta + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \varepsilon_t \qquad (7)$$

The null and alternative hypotheses are set as follows,

H_0: $\beta_i = 0$ for all *i*

H_1: $\beta_i \neq 0$

Then, an *F-test* is conducted of the null hypothesis by estimating the following equation by OLS:

$$y_t = \delta + \sum_{i=1}^{p} \varphi_i y_{t-i} + \mu_t \qquad (8)$$

Compare their respective sum of squared residuals.

$$RSS_\varepsilon = \textstyle\sum_{t=1}^{T} \varepsilon_t^2 \qquad , \qquad RSS_\mu = \sum_{t=1}^{T} \varepsilon_t^2$$

If the statistic test

$$S = \frac{(RSS_\varepsilon / RSS_\mu)/p}{RSS_\varepsilon/(T-2p-1)}$$

is greater than the specified critical value, then reject the null hypothesis that *x* does not Granger-cause *y* (Greene, 2003).

Co-integration test is an economical technique for measuring the correlation between non-stationary time series variables. If two or more time series are non-stationary, but a linear combination of them is stationary, then the series is said to be co-integrated.

In regression model

$$y_t = \alpha + \beta x_t + \varepsilon_t \qquad (9)$$

where y_t and x_t are both non-stationary and *I*(1) processes. The difference between them is increasing, not stable as time passes. Unless there is a relationship between y_t and x_t, then there should be α and β such that

$$\varepsilon_t = y_t - \alpha - \beta x_t \qquad (10)$$

is $I(0)$ or stationary. Intuitively, if two series are both $I(1)$, then this partial difference might be stable around a fixed mean. Two series that satisfy this requirement are said to be co-integrated. A long-run relationship between y_t and x_t can be estimated in such a case.

The most well-known test, suggested by Engle and Granger (1987), is to run a static regression (after having verified that y_t and x_t are both I(1)). In the single equation approach, the co-integration test can be conducted in the following manner. First, y_t and x_t are regressed by ordinary least squares method and obtain the OLS residuals, ε_t.

$$\varepsilon_t = y_t - \alpha - \beta x_t \qquad (11)$$

where α and β denote the OLS estimates.

$$H_0: \beta=0$$

$$H_1: \beta<0$$

Accepting the null hypothesis, β=0, implied that the residual contains unit root and is non-stationary. Rejection of the null hypothesis, β<0, means that the residual converges to the long-term mean or being a stationary time series.

Stock prices change over time, following a random walk. Testing the hypothesis that there is a statistically significant correlation between two series of stock price can be performed based on a co-integration procedure. Thus, the formula is presented as follows:

$$\bar{\alpha} = y_t - \bar{\beta}\, x_t \qquad (12)$$

Where,

$\bar{\alpha}$: Estimated mean spread of stock *x* and *y*

$\bar{\beta}$: Estimated Co-integration coefficient

y_t : Price of stock *y* at time *t*

x_t : Price of stock *x* at time *t*

This formula will produce a buying or selling signal for the trading program. In the event of a buying signal, a long position will be entered on one share of stock *y* and simultaneously going short on $\bar{\alpha}$ share of stock *x*, and vice versa for a selling signal.

METHODOLOGY

To check whether a series of stock price has a unit root, non-stationary, or stationary, Augmented Dickey-Fuller (ADF) test is carried out. If each data series is non-stationary, the ADF test is applied again on the first difference of the series. The series is said to be integrated of order one or $I(1)$, when stationary result is found on the first difference of the data series, but non-stationary is found on the level of data series. If two stock prices are integrated of order one, the Granger causality test is applied to determine which stock price is classified as the dependent and independent variable. The linear relationship between two stock prices is defined by using two steps Engle and Granger co-integration test. Firstly, the logarithms of stock Y are regressed again, and the logarithm of stock X as a co-integration equation is

$$log(p_t^Y) = \mu + \vartheta log(p_t^X) + \varepsilon_t \qquad (12)$$

Where,

p_t^Y : Price of stock Y,

p_t^X : Price of stock X,

μ : Intercept,

θ : Co-integration coefficient,

ε_t : Residual term,

The estimated method is Fully Modified Least Squares (FMOLS). The predicted residual from equation (12) could be generated as follows:

$$\hat{\varepsilon}_t = log(p_t^Y) - \hat{\mu} - \hat{\vartheta} log(p_t^X) \qquad (13)$$

Secondly, unit root test is performed on the predicted residual value from equation (13). If the residual term is stationary or has no unit root, it could be concluded that a linear relationship between the two stock prices exists.

After estimation of the co-integration coefficient, a trading position is opened when price spread between two stocks climbs up to 2 standard deviations (SD) and the position is closed out when the price spread reverts back to the mean or climbs up to 4 standard deviations or any price spread during the last trading day. A position is opened by a long one underperforming stock and a short $\hat{\theta}$ over-performing stock

Ratio of share number = 1:$\hat{\vartheta}$

Amount of invested capital = KHR100,000

$$Ratio\ of\ short\ unit = \frac{Invested\ capital}{Price\ of\ short\ unit\ at\ time\ t_t}$$

Ratio of long unit = Ratio of short unit*$\hat{\vartheta}$

$$Investment\ return = \frac{Cash\ inflow\ at\ time\ t_{t+1} - Cash\ outflow\ at\ time\ t_t}{Invested\ capital} \times 100$$

If the trade rule is complied with, the abovementioned trading is repeatedly applied. The average return on investment is determined for the study period. In order to assess the performance of Pair Trading, two techniques compare the weighted average portfolio return and the average return of the CSX Index.

Weighted average return during the period=

$$(C_Y \times R_Y) + (C_X \times R_X)$$

Where, C_Y is co-integration of stock Y, C_X is co-integration of stock X, R_Y is the return of stock Y and R_X is the return of stock X. The return on stock *i* or the CSX index is computed as follows, $R_{i,t}$=Log($S_{i,t}$/$S_{i,t}$-1) ×100. Where, $S_{i,t}$ is the stock i price at time t, $S_{i,t}$-1 is the stock i price at time t-1 and Log is logarithm. The research will run from June 7, 2017 through July 7, 2021. Daily stock prices are obtained from the Bloomberg terminal.

EMPIRICAL RESULT

Of course, there are seven listed stocks on the Cambodia Securities Exchange; however, during the study period from June 7, 2017 to July 7, 2021, only four stocks have enough data set: Sihanoukville Autonomous Port (PAS), Phnom Penh Special Economic Zone (SEZ) Plc. (PPSP), Grand Twins International (Cambodia) Plc. (GTI), and Phnom Penh Water Supply Authority (PWSA). The equities under consideration have been divided into four industries: port services, SEZ developer, apparel garments, and water utility. Table 1 shows the descriptive statistics of the daily returns of the selected equities and market index.

To begin, each stock price or stock index series is given a natural logarithm to make the series continuous data. Second, the series is converted to the first difference and then multiplied by 100, resulting in each stock's daily growth rate or daily return. The average return on stock PAS and PWSA is 0.1062 percent and 0.0629 percent, respectively, which is higher than the average return on the CSX index. Furthermore, the coefficient of variation (CV) of the two stocks, which shows the ratio of standard deviation over mean, is higher than the market one, at 4.8997 percent for PAS and 3.8708 percent for PWSA, compared to 3.2338 percent for the CSX index. Based on the CV comparison, it is possible to conclude that the returns of PAS and PWSA beat the market. In contrast, GTA's daily average return is 0.0000 percent, while PPSP's daily average return is -0.0243 percent, lower than the CSX index's daily return. Nonetheless, the CV of the two stocks is smaller than that of the market. As a result, the returns on the two stocks underperform the market return.

Table 1: Descriptive Statistics

	DLOGCSX	DLOGGTI	DLOGPAS	DLOGPPSP	DLOGPWSA
Mean	0.0485%	0.0000%	0.1062%	-0.0243%	0.0629%
Median	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
Maximum	7.9529%	9.5310%	9.5003%	9.3010%	9.5310%
Minimum	-8.9566%	-12.9356%	-10.5361%	-10.5361%	-10.4106%
Std. Dev.	1.4998%	3.0856%	2.1675%	1.9860%	1.6250%
Skewness	-0.094325	-0.015820	0.304799	0.181084	0.273979
Kurtosis	12.996490	6.569771	10.516650	11.380160	15.350000
CV	3.2338%	0.0000%	4.8997%	-1.2236%	3.8708%

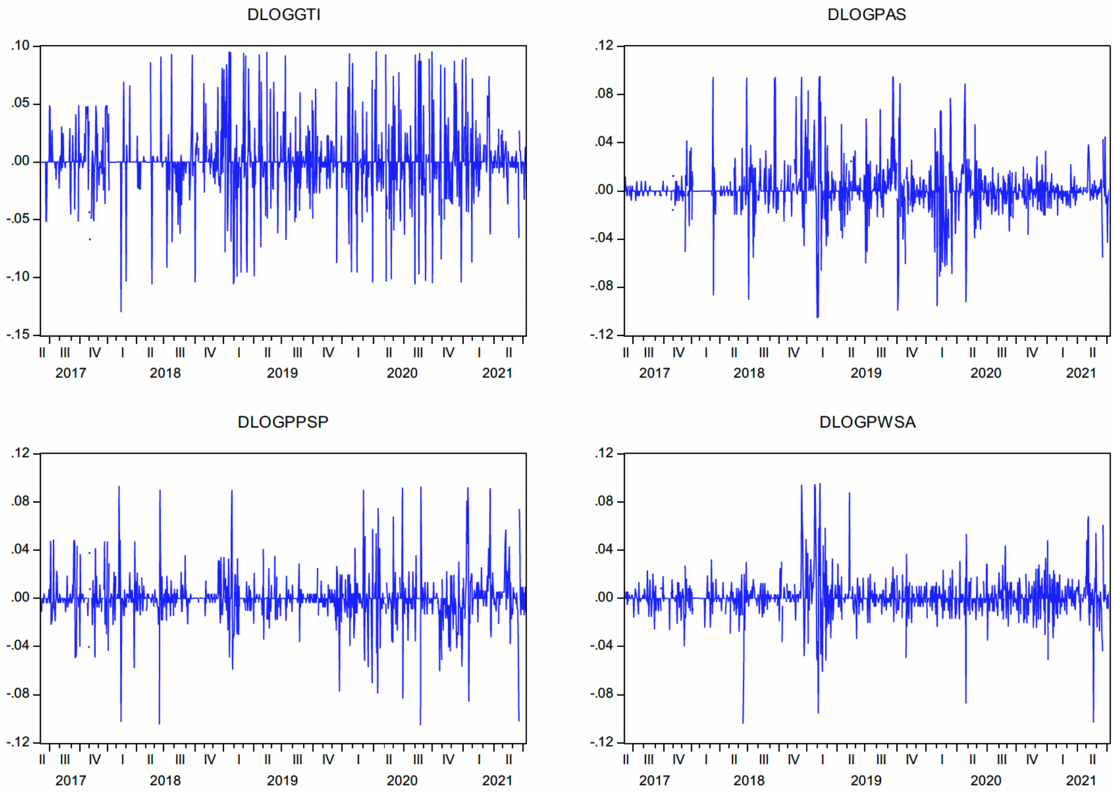

Figure 1: Daily Return of Stock

The Augmented Dickey-Fuller (ADF) unit root test employs three regression models: model with constant, model with constant and trend, and model without constant and trend. First, the test is performed on the logarithm of a stock price or level series. The null hypothesis: The variable has a unit root and failed to reject at the 5 percent significant level in all models for PAS, PPSP, and PWSA, as shown in Table 2. GTI series, on the other hand, is shown to be stationary at the level using models with constant and model with constant and trend since the null hypothesis is rejected at 5 percent and 10 percent, respectively. However, the series becomes non-stationary when a model without a constant and a trend is used.

Table 2: ADF Unit Root Test

Null Hypothesis: The variable has a unit root					
	At Level				
		LOGGTI	LOGPAS	LOGPPSP	LOGPWSA
With Constant	t-Statistic	-3.0070	-1.2205	-1.3646	-0.7542
	Prob.	**0.0346**	**0.6674**	**0.6008**	**0.8307**
		**	n0	n0	n0
With Constant & Trend	t-Statistic	-3.1919	-0.7453	-2.1312	-1.7989
	Prob.	**0.0866**	**0.9687**	**0.5271**	**0.7048**
		*	n0	n0	n0
Without Constant & Trend	t-Statistic	-0.0515	1.4499	-0.4330	1.2053
	Prob.	**0.6654**	**0.9639**	**0.5268**	**0.9420**
		no	no	no	no
	At First Difference				
		d(LOGGTI)	d(LOGPAS)	d(LOGPPSP)	d(LOGPWSA)
With Constant	t-Statistic	-34.6566	-29.3565	-33.9637	-10.8126
	Prob.	**0.0000**	**0.0000**	**0.0000**	**0.0000**
		***	***	***	***
With Constant & Trend	t-Statistic	-34.6386	-29.3669	-33.9470	-10.8102
	Prob.	**0.0000**	**0.0000**	**0.0000**	**0.0000**
		***	***	***	***
Without Constant & Trend	t-Statistic	-34.6747	-29.3066	-33.9763	-10.7398
	Prob.	**0.0000**	**0.0000**	**0.0000**	**0.0000**
		***	***	***	***
Notes:					
a: (*)Significant at the 10%; (**)Significant at the 5%; (***) Significant at the 1% and (no) Not Significant					
b: Lag Length based on SIC					
c: Probability based on MacKinnon (1996) one-sided p-values.					

After each data series has been modified to be first different, the ADF result has changed substantially. For all models, the null hypothesis that the variable has a unit root is strongly rejected at the 1% significance level. This finding may be understood as the logarithm of each stock price series being integrated into order one, I(1). A pair trading relationship has been created between PPSP and PWSA. The Granger causality test examines the causal link between PPSP and PWSA, whether run from PPSP to PWSA or vice versa, or to identify which variable is dependent and independent.

Table 3: Granger Causality Test, PPSP and PWSA

Null Hypothesis:	Obs	F-Statistic	Prob.
LOGPPSP does not Granger Cause LOGPWSA	961	3.02854	0.0489
LOGPWSA does not Granger Cause LOGPPSP		3.44856	0.0322

Table 3 shows that PPSP Granger causes PWSA and vice versa since the p-value is less than 5% significant. As demonstrated by this finding, it makes no difference whether PPSP is considered a dependent or independent variable; hence, PPSP is categorized as a dependent variable, and PWSA is classed as an independent variable. A Fully Modified Least Squares (FMOLS) method is used to identify sample parameters.

Table 4: Co-integration Equation between PPSP and PWSA

Dependent Variable: LOGPPSP				
Method: Fully Modified Least Squares (FMOLS)				
Sample (adjusted): 6/08/2017 7/08/2021				
Included observations: 962 after adjustments				
Cointegrating equation deterministics: C				
Long-run covariance estimate (Bartlett kernel, Newey-West fixed				
bandwidth = 7.0000)				
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
LOGPWSA	-0.366341	0.057243	-6.399729	0.0000
C	10.93578	0.487396	22.43715	0.0000
R-squared	0.222608	Mean dependent var		7.818329
Adjusted R-squared	0.221799	S.D. dependent var		0.197561
S.E. of regression	0.174280	Sum squared resid		29.15846
Long-run variance	0.209194			

A sample regression function between PPSP and PWSA is written as below:

$$\widehat{LogPPSP}_t = -0.366341 LogPWSA_t + 10.93578$$

Based on this estimated result, PWSA is highly significant in explaining PPSP at 1 percent level. This sample regression function is used to predict the residual term, and the unit root test is applied to check whether the term is stationary or non-stationary.

Table 5: Unit Test, Residual Term between PPSP and PWSA

Null Hypothesis: RESID01_PPSP&PWSA has a unit root			
Exogenous: None			
Lag Length: 0 (Automatic- based on SIC, maxlag=21)			
		t-Statistic	**Prob.***
Augmented Dickey-Fuller test statistic		-1.918696	0.0526
Test critical values:	1% level	-2.567371	
	5% level	-1.941153	
	10% level	-1.616478	
*MacKinnon (1996) one-sided p-values.			

According to the ADF test results in Table 5, the residual term is stationary at a 10% significant level, indicating that the two data series are co-integrated or have a long-run connection. The stock price spread between PWSA and PPSP is computed next, and summary statistics are shown in Table 6.

Table 6: Summary Statistics, Stock Price Spread between PWSA and PPSA

	SPREAD_PWSA_PPSP
Mean	2596.376
Median	3220
Maximum	6270
Minimum	180
Std. Dev.	1499.936
Observations	963

The mean spread is KHR 2,596.376, with a standard deviation (SD) of KHR 1,499.936. Furthermore, the spread's 2SD and 4SD are KHR 2,999.872 and KHR 5,999.744, respectively. As previously stated, if the spread clams up to 2SD, a position will be opened by shorting one unit of the outperformance stock and longing for some quantity of share equal to the co-integration ratio between the two stocks obtained in the co-integration calculation above. The position will be closed when the spread reverts to the mean, and the investment return is determined. If the trading rule is met, this procedure will be performed several times during the research. KHR 100,000 is set to be the starting investment. PWSA and PPSP have co-integrated ratios of 0.36634 and 0.62669, respectively.

Figure 2: Stock Price Spread between PWSA and PPSP

Table 7: Open and Close Position with respect to Trading Rule

			1st Price Spread Climb up to 2SD			
	Units	**Position**	**Date**		**PWSA**	**PPSP**
			t_0	1/28/2019	KHR 6,200	KHR 3,500
Short (PWSA)	15.87	Open	t_1	1/29/2019	KHR 6,300	KHR 3,300
Long (PPSP)	5.81	Close	t_2	1/31/2019	KHR 5,700	KHR 3,090
			2nd Price Spread Climb up to 2SD			
			Date		**PWSA**	**PPSP**
			t_0	2/7/2019	KHR 5,940	KHR 3,100
Short (PWSA)	16.13	Open	t_1	2/8/2019	KHR 6,200	KHR 3,090
Long (PPSP)	5.91	Close	t_2	2/18/2019	KHR 5,580	KHR 2,960
			2nd Price Spread Climb up to 2SD			
			Date		**PWSA**	**PPSP**
			t_0	2/22/2019	KHR 5,660	KHR 2,960
Short (PWSA)	16.67	Open	t_1	2/25/2019	KHR 6,000	KHR 2,960
Long (PPSP)	6.11	Close	t_2	3/12/2019	KHR 5,600	KHR 2,940

During the research period, the stock price gap between PWSA and PPSP rose to 2SD three times before reverting to mean, on January 29, 2019, February 8, 2019, and February 25, 2019. The position is opened at time t_1 when the spread reaches 2SD and terminated at time t_2 when the spread returns to mean. Table 7 shows that t_0 indicates the day before the spread reached 2SD. When comparing t_1 and t_0, the PWSA's price rises while the PPSP's price falls all three times; the spread reaches 2SD and is converted back to the mean.

Table 8: Weighted Average and Pair-trading Return

Number of pair-trading	Weighted average return	Pair-trading return
1	-0.1318%	8.3027%
2	1.4339%	9.2319%
3	2.1933%	6.5446%
Average	1.1651%	8.0264%

At each point, the return on pair trading always surpasses the weighted average return between PWSA and PPSP, as illustrated in Table 8. Only two stocks, PWSA and PPSP, demonstrated a long-term relationship or co-integration over the study period.

CONCLUSION

Identifying an investing strategy in the stock market that would provide the best possible return is difficult. Investors may use numerous trading strategies to maximize their return on investment, with pair trading being one of the most prominent. Pairs trading is a sophisticated trading technique that entails opening one long of underperformance stock and one short position of over-performance stock. When two stocks have a long-run relationship (co-integration), the strategy may effectively create outperformance profit. Even though the Cambodia Securities Exchange only has seven listed stocks from six distinct industries: financial, port services, SEZ developer, apparel garments, power, and water utility, co-integration between two listed stocks, PPSP—Special Economic Zone developer and PWSA—Water utility, has been identified. The empirical value for the co-integrated ratio is 0.366341. The stock price spread between PWSA and PPSP during the study period reached 2SD three times. Shorting one unit of outperformance stock and longing 0.366341 units of underperformance stock is used to establish a trade each time. In contrast, the trade is closed when the spread is converted to mean. The findings of this study show that the pair trading method works in a tiny securities market like Cambodia since the return on investment from the technique outperforms the weighted average return between the two equities.

REFERENCES

Blázquez, M. C., & Román, C. P. (2018). Pairs trading techniques: An empirical contrast. *European Research on Management and Business Economics, 24*(3), 160-167.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association, 74*, 427–431.

Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica, 55*, 251-276.

Elliott, R. J., Van Der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance, 5*(3), 271–276.

Gatev, E. G., Goetzmann, W. N., & Rouwenhorst, K. G. (1999). *Pairs Trading: Performance of a Relative Value Arbitrage Rule.* National Bureau of Economic Research.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica: *Journal of the Econometric Society,* 424–438.

Greene, W. (2003). *Econometric analysis.* Prentice Hall.

Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press.

Hayashi, F. (2000). *Econometrics.* Princeton University Press.

Hong, G., & Susmel, R. (2003). *Pairs-trading in the Asian ADR market.* University of Houston, Unpublished Manuscript.

Jurek, J., & Yang, H. (2007). *Dynamic portfolio selection in arbitrage.* Harvard University working paper.

Kanamura, T., Rachev, S.T., & Fabozzi, F.J. (2008). *The application of pairs trading to energy future markets.* Yale School of Management.

Krauss, C. (2017). Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys, 31*(2), 513-545.

Kwiatkowski, D., Peter C. B. P., Peter, S., & Yongcheol, S. (1992). Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics, 54*, 159-178.

Lin, Yan-Xia, Mccrae, M., & Gulati, C. (n.a.). Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Journal of Applied Mathematics and Decision Sciences, 1*-16.

MacKinnon, J. G. (1996). Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics, 11*, 601-618.

Namwong, N., Yamaka, W., & Tansuchat, R. (2019, January). Trading signal analysis with pairs trading strategy in the Stock Exchange of Thailand. In *International Conference of the Thailand Econometrics Society* (378-388). Springer, Cham.

Nath, P. (2003). High frequency pairs trading with US treasury securities: Risks and rewards for hedge funds. SSRN 565441.

Patanapol, T. (2001). *Pairs trading: Performance of a relative value arbitrage rule in the stock exchange of Thailand.* Unpublished Paper, Chulalongkorn University.

Ramos-Requena, J. P., Trinidad-Segovia, J. E., & Sánchez-Granero, M. Á. (2020). Some notes on the formation of a pair in pairs trading. *Mathematics, 8*(3), 348.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative methods and analysis.* John Wiley & Sons.

Wilmontt, P. (2004). *The best of Wilmott: Incorporating the quantitative finance review.* John Wiley & Sons.

Zapart, C. A. (2004). Long-short equity pairs trading with optimum wavelet correlation measures. Financial Engineering and Applications. November 8-10, 2004. MIT, Cambridge.