Pairs Trading with General State Space Models[*]

Guang Zhang[†]

Department of Economics, Boston University, Boston, MA, 02215

November 24, 2020

Abstract

This study examines pairs trading using a general state space model framework. I model the spread between the prices of two assets as an unobservable state variable, and assume that it follows a mean-reverting process. This new model has two distinctive features: the (1) non-Gaussianity and heteroscedasticity of the innovations to the spread, and (2) nonlinearity of the mean reversion of the spread. I show how to use the filtered spread as the trading indicator to carry out statistical arbitrage. I also propose a new trading strategy and present a Monte Carlo-based approach to select the optimal trading rule. For empirical applications, I first apply the new model and the new trading strategy to two examples: PEP vs. KO and EWT vs. EWH. The results show that the new approach can achieve 21.86% (31.84%) annualized return for the PEP-KO (EWT-EWH) pair. Then, I consider all the possible pairs among the five largest and the five smallest U.S. banks listed on the NYSE. For these pairs, I compare the performance of the proposed approach with that of the existing popular approaches, both in-sample and out-of-sample. In almost all the cases considered, I find that our approach can significantly improve the return and the Sharpe ratio.

Keywords: pairs trading, state space models, mean-reverting, heteroscedasticity, Quasi Monte Carlo Kalman filter

JEL codes: C11, C32, C41, G11, G17.

1 Introduction

In the early 1980s, a group of physicists, mathematicians and computer scientists, led by quantitative analyst Nunzio Tartaglia, used a sophisticated statistical approach to find the opportunities for arbitrage trading (Gatev et al. 2006). Tartaglia's strategy, later coined pairs trading, is to find a pair of two stocks whose prices have moved similarly historically, and make profit by applying the simple contrarian principles. Since then, pairs trading has become a popular short-term arbitrage strategy used by hedge funds and is often considered as the "ancestor" of statistical arbitrage.

Pairs trading is a self-financing portfolio constructed with a long position in one security and a short position in the other. Given that the two securities have moved together historically, when a temporary anomaly occurs, one security would be overvalued, compared with the other, relative to the long-term equilibrium. An investor may be able to make money by selling the overvalued security, buying the undervalued security, and clearing the exposure when the two securities revert to their long-term equilibrium. Because the effect of the market's movement is hedged by this self-financing portfolio, pairs trading is market-neutral.

The methods for pairs trading can be broadly divided into nonparametric and parametric methods. In particular, Gatev et al. (2006) proposed a nonparametric distance-based approach to determine the securities for constructing the pairs. A pair is chosen by finding the securities that minimized the sum of squared deviations between two normalized prices. They argued that this approach "best approximates the description of how traders themselves choose pairs." They found that average annualized excess returns reach 11% for the top pair portfolios using CRSP daily data from 1962 to 2002. Other nonparametric methods on pairs trading can also be found in Bogomolov (2013), among others. Overall, the nonparametric distance-based approach provides a simple and general method for selecting "good" pairs; however, as Krauss (2016) and others pointed out, this selection metric is prone to choosing pairs with small variance of the spread, and thereby limiting the profitability of pairs trading.

By contrast, the parametric approach tries to capture the mean-reverting characteristic of the spread using a parametric model. For example, Elliott et al. (2005) proposed a mean-reverting Gaussian Markov chain model for the spread that is observed with Gaussian noise. See Vidyamurthy (2004), Cummins and Bucca (2012), Tourin and Yan (2013), Moura et al. (2016), Stübinger and Endres (2018), Clegg and Krauss (2018), Elliott and Bradrania (2018), and Bai and Wu (2018) for other parametric methods on pairs trading. Overall, the parametric approach provides tractable methods for the analysis of pairs trading. However, most of the existing parametric models are too simple to be capable of capturing the dynamics of asset prices, which substantially limits the returns from pairs trading.

Compared with the existing methods of pairs trading, the proposed approach has the following

features:

(1) It is based on a general state space model that includes nonlinear and non-Gaussian model. This model can capture several stylized features of financial asset prices, including heavy-tailedness, heteroscedasticity, volatility clustering and nonlinear dependence;

(2) The trading strategy is different from the existing ones. It uses the model's features such as heteroscedasticity and volatility clustering, and it can potentially achieve significantly higher returns and Sharpe ratios;

(3) The optimal trading rules are also different from the existing ones. Although this rule has no analytic solution, we show that it can be computed effectively using simulations; and finally,

(4) The optimal trading rule can adapt to various objectives, such as a high cumulative return, Sharpe ratio, or Calmar ratio.

I apply this approach to two pairs: PEP vs. KO and EWT vs. EWH. I find that our approach achieves an annualized return of 0.2186 (0.3184) and a Sharpe ratio of 2.9518 (3.8892) on the PEP-KO (EWT-EWH) pair. In comparison, a conventional approach applied to the same pairs can only achieve an annualized return of 0.1311 (0.1480) and a Sharpe ratio of 1.1003 (1.1277) for the PEP-KO (EWT-EWH) pair. Next, I test the approach using all the possible pairs among the largest five banks and the smallest five banks listed on the NYSE. I find significant improvements over the conventional approach for almost all the pairs. I also find that pairs of small banks produce higher returns than the pairs of large banks. This is likely because the spread between small banks is more volatile, providing more opportunities for active trading.

The remainder of this paper is organized as follows. In Section 2, I propose a new model for pairs trading. In Section 3, I propose a new trading strategy based on the mean-reverting property of spreads, and compare it with conventional trading strategies using simulations. In Section 4, I implement the proposed approach on actual data, and in Section 5 I present the conclusions.

2   A New Model for Pairs Trading

I propose the following general state space model with the linear observable process for pairs trading:

$$P_{A,t} = \lambda + \gamma P_{B,t} + x_t + \varepsilon_t \tag{1}$$

$$x_{t+1} = f(x_t; \theta) + g(x_t; \theta) * \eta_t \tag{2}$$

where P_A is the price of security A; P_B is the price of security B; γ is the hedge ratio between the two securities; and x is the unobservable true spread between P_A and P_B. I assume that x follows a mean-reverting process as in (2), $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$, and $\eta_t \sim p\left(\eta_t; \theta\right)$. Popular choices for f, g and p can be the following. Our framework applies to all of them.

- Linear mean-reverting (Ornstein–Uhlenbeck process): $f(x_t; \theta) = \theta_1 + \theta_2 x_t$

- Nonlinear mean-reverting model: $f(x_t; \theta) = \theta_1 + \theta_2 x_t + \theta_3 x_t^2$

- Ait-Sahalia's nonlinear mean-reverting model (Ait-Sahalia, 1996): $f(x_t; \theta) = \theta_1 + \theta_2 x_t^{-1} + \theta_3 x_t + \theta_4 x_t^2$

- Homoskedasticity model: $g(x_t; \theta) = \theta_0$

- Linear heteroscedastic model: $g(x_t; \theta) = \theta_0 + \sum_{i=1}^m \theta_i x_{t-i}$

- ARCH(m) nonlinear heteroscedastic model: $g(x_t; \theta) = \sqrt{\theta_0 + \sum_{i=1}^m \theta_i x_{t-i}^2}$

- APARCH(m, δ) nonlinear heteroscedastic model: $g(x_t; \theta) = \left(\theta_0 + \sum_{i=1}^m \theta_i \mid x_{t-i} \mid^\delta\right)^{\frac{1}{\delta}}$

- Gaussian distributed noise: $p(\eta; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mu-\eta)^2}{2\sigma^2}\right)$

- Student's t distributed noise: $p(\eta; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\eta^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

- Generalized error distributed noise: $p(\eta; \alpha, \beta, \mu) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} \exp\left(-(\mid \eta - \mu \mid /\alpha)^\beta\right)$

In models (1)-(2), since λ and θ_1 in the f function can not be identified simultaneously, I let $\lambda = 0$ and denote $\psi = (\gamma, \theta, \sigma_\varepsilon) \in \Psi$ as the identifiable parameter of the models (1)-(2). ψ is estimated based on the data set $\{P_{A,t}, P_{B,t}\}_{t=1}^T$.

It is easy to find that the general framework for pairs trading includes linear and Gaussian model as special case for the dynamics of the spread, and my approach can be applied to all these cases. The new model has three advantages compared with existing models for pairs trading, such as Elliott et al. (2005) and Moura et al. (2016). First, because η can be non-Gaussian, x can follow a non-Gaussian process. By allowing for this non-Gaussianity in η, the model can capture the distributional deviation from Gaussianity and reproduce heavy-tailed returns.

Second, the model can capture heteroscedasticity in financial data. A well-known feature of financial time-series is volatility clustering: "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes" (Mandelbrot, 1963). This feature was documented later in Ding et al. (1993), and Ding and Granger (1996) among others. In model (2), the volatility persistence can be represented by an ARCH-style modeling. Details about the application of an ARCH model in finance can be found in Bollerslev et al. (1992).

Third, to characterize the nonlinear dependence in financial data, I allow f to be nonlinear. Scheinkman and LeBaron (1989) found evidence indicating the presence of nonlinear dependence in the weekly returns of the CRSP value-weighted index. Ait-Sahalia (1996) found nonlinearity in the drift function of interest rate and concluded that "the principal source of rejection of existing (linear drift) models is the strong nonlinearity of the drift." I keep the functional form of f flexible and as a result, we can capture the nonlinear dependence in financial data.

3   A New Approach to Pairs Trading

In this section, I discuss the trading strategies and trading rules for pairs trading. In this study, a trading strategy is a method of buying and selling of assets in markets based on the estimation of the unobservable spread. A trading rule is the predefined value that generates the trading signal for a specific trading strategy with an investing objective. To implement a strategy and rule on pairs trading, we need the following quantities: (i) parameter estimates for the models (1)-(2), (ii) an estimate of the spread, and (iii) choice of a specific strategy and the optimal trading rule. I discuss these aspects in this section. More specifically, in Section 3.1, I present an algorithm for filtering the unobservable spread and parameter estimation. In Section 3.2, I discuss two benchmark trading strategies. In Section 3.3, I present and compare three popular trading rules associated with the benchmark trading strategies. In Section 3.4, I propose a new trading strategy, where I change the way we open or close a trade, and discuss the benefit of this new strategy compared with the benchmark strategies. Because the existing trading rule can not be simply applied to the models (1)-(2), I propose a new approach to calculate the optimal trading rule based on a simulation of the spread. The details of this simulation-based method are presented in Section 3.5. In Section 3.6, I summarize the procedures for pairs trading. These procedures can be applied to pairs trading for all the trading strategies and rules discussed in this paper.

3.1   Algorithm for Filtering and Parameter Estimation

For the specification of models (1)-(2), I run the following algorithm of Quasi Monte Carlo Kalman filter (QMCKF) for nonlinear and non-Gaussian state space models to estimate the unobservable spread and unknown parameters in the models, based on the observations $\{P_{A,t}, P_{B,t}\}_{t=1}^{T}$. This approach treats linear and Gaussian model as a special case, and its theoretical properties are studied in Zhang (2020). Suppose the initial spread x_0 follows $N(\mu, \Sigma)$ for any reasonable choices of μ and Σ.

- Step 1: For non-Gaussian density $p(\eta_t)$, we use Gaussian mixture density to approximate its pdf and denote the approximation as $\widetilde{p}(\eta_t) = \sum_{i=1}^{m} \alpha_i \phi(\eta_t - a_i, P_i), \sum_{i=1}^{m} \alpha_i = 1$ associated with parameters $\{\alpha_i, a_i, P_i\}_{i=1}^{m}$, and Gaussian density ϕ[1]. If η_t is Gaussian, then this step can be dropped.

---

[1]To get this approximation, we determine the values of $\{\alpha_i, a_i, P_i\}_{i=1}^{m}$ by minimizing the relative entropy between the true density $p(\eta_t)$ and its approximation $\widetilde{p}(\eta_t)$. The relative entropy is defined by

$$\mathcal{H}(p|\widetilde{p}) = \int \left(\log \frac{p(\eta)}{\widetilde{p}(\eta)} \right) \times p(\eta) \, d\eta.$$

The above relative entropy can be computed in a numerical manner if it does not admit a closed-form solution.

- Step 2: Generate a Box-Muller transformed Halton sequence $\{x_{t,s}^{(g)}\}_{g=1}^{G}$ with sequence size G from $\phi\left(x_t - b_{t,s}, P_{t,s}\right)$, where $b_{t,s}$ and $P_{t,s}$ are from Step 5 in the previous period. Compute and store

$$c_{t+1,i} = \frac{1}{G} \sum_{g=1}^{G} f\left(x_{t,s}^{(g)}\right) + g\left(x_{t,s}^{(g)}\right) * a_k.$$

and

$$Q_{t+1,i} = \frac{1}{G} \sum_{g=1}^{G} \left(f\left(x_{t,s}^{(g)}\right) - c_{t+1,i}\right)^2 + \left(g\left(x_{t,s}^{(g)}\right)\right)^2 * P_k,$$

where a_k and P_k, $k = 1, ..., m$, are from the approximation in Step 1. When $t = 0$, $\{x_{0,s}^{(g)}\}_{g=1}^{G}$ is sampled from $N(\mu, \Sigma)$.

- Step 3: Repeat Step 2 for $s = 1, 2, ..., J_{t+1}$, with $J_{t+1} = m^t$, and $k = 1, \ldots m$, and store $c_{t+1,i}$ and $Q_{t+1,i}$ for $i = 1, 2, ..., I_{t+1}$, with $I_{t+1} = J_{t+1} * m = m^{t+1}$.

- Step 4: Based on the results from Step 3, generate a Box-Muller transformed Halton sequences $\{x_{t+1,i}^{(g)}\}_{g=1}^{G}$ from $\phi\left(x_{t+1} - c_{t+1,i}, Q_{t+1,i}\right)$ for $i = 1, 2, ..., I_{t+1}$, with $I_{t+1} = m^{t+1}$. Then generate $P_{A,t+1,i}^{(g)} = x_{t+1,i}^{(g)} + \gamma * P_{B,t+1}$. Compute and store the following

$$\bar{P}_{A,t+1,i} = \frac{1}{G} \sum_{g=1}^{G} P_{A,t+1,i}^{(g)}, \quad V_{t+1,i} = \frac{1}{G} \sum_{g=1}^{G} \left(P_{A,t+1,i}^{(g)} - \bar{P}_{A,t+1,i}\right)^2 + \sigma_\varepsilon^2,$$

$$S_{t+1,i} = \frac{1}{G} \sum_{g=1}^{G} \left(x_{t+1,i}^{(g)} - c_{t+1,i}\right)\left(P_{A,t+1,i}^{(g)} - \bar{P}_{A,t+1,i}\right).$$

- Step 5: Compute $K_{t+1,i} = S_{t+1,i} V_{t+1,i}^{-1}$, $P_{t+1,i} = Q_{t+1,i} - K_{t+1,i}^2 V_{t+1,i}$, and $b_{t+1,i} = c_{t+1,i} + K_{t+1,i}\left(P_{A,t+1} - \bar{P}_{A,t+1,i}\right)$ where $Q_{t+1,i}$ and $c_{t+1,i}$ are from Step 3.

- Step 6: Repeat Steps 4-5 for $i = 1, 2, ..., I_{t+1}$, with $I_{t+1} = m^{t+1}$. Compute and store $\bar{x}_{t+1}$ and $\bar{P}_{t+1}$ where $\bar{x}_{t+1} = \sum_{i=1}^{I_{t+1}} \beta_{t+1,i} b_{t+1,i}$, and

$$\bar{P}_{t+1} = \sum_{i=1}^{I_{t+1}} \beta_{t+1,i}\left(P_{t+1,i} + b_{t+1,i}^2\right) - \left(\sum_{i=1}^{I_{t+1}} \beta_{t+1,i} b_{t+1,i}\right)^2,$$

$$\beta_{t+1i} = \frac{\phi\left(P_{A,t+1} - c_{t+1,i} - \gamma * P_{B,t+1}, V_{t+1,i}\right)}{\sum_{i=1}^{I_{t+1}} \phi\left(P_{A,t+1} - c_{t+1,i} - \gamma * P_{B,t+1}, V_{t+1,i}\right)}.$$

- Step 7: Repeat Steps 2-6 for $t = 0, 1, 2, ..., T$.

The output $\{\bar{x}_t\}_{t=1}^T$ from Step 6 is our estimation of the spread. To estimate the unknown parameter in the model, we first write the log-likelihood function as

$$L_T^G(\psi) \equiv \sum_{t=1}^T \log f^G(\psi; P_{A,t}, P_{B,t}) = \sum_{t=0}^{T-1} \log \left[\sum_i^{I_{t+1}} \frac{1}{\sqrt{2\pi V_{t+1,i}}} \exp\left(-\frac{\left(P_{A,t+1} - \bar{P}_{A,t+1,i}\right)^2}{2 * V_{t+1,i}} \right) \right]$$

and the MLE, $\hat{\psi}_{MLE}$, of the unknow parameter will be determined to maximize the above likelihood. That is,

$$\hat{\psi}_{MLE} = \operatorname*{argmax}_{\psi \in \Psi} L_T^G(\psi).$$

A straightforward grid search can be used to find $\hat{\psi}_{MLE}$ when the number of the unknown parameters is small. Otherwise, several optimization algorithms can be applied to find $\hat{\psi}_{MLE}$. See Hajivassiliou and McFadden (1998) and Cameron and Trivedi (2005) for details.

Also, a dual state-parameter estimation approach can be used to estimate the parameters. In this approach, we consider the unknown parameters as state variables and obtain the filtering estimation of the original state variables and unknown parameters simultaneously. See Wan, van der Merwe, and Nelson (1999) and Wan and Nelson (2000) for examples.

3.2 Benchmark Trading Strategies

As I discussed in Section 1, the basic idea for pairs trading is to open a trade (short one asset and long the other one) when the spread deviates from the equilibrium, and to close the trade when the spread returns to the equilibrium. The trading strategies for pairs trading are constructed based on this idea. I use Figures 1 and 2 to illustrate two benchmark trading strategies (hereinafter, Strategies A and B). In Figures 1 and 2, the same estimated spread is plotted as solid lines, while a preset upper boundary U and a preset lower boundary L are plotted as dashed lines. I discuss how to choose the optimal U and L in Section 3.3. The upper and lower boundaries act as thresholds to determine whether the spread has deviated enough from the long-term equilibrium, and we use these two criteria to open a trade. In addition, a preset value C acts as a threshold to determine whether the spread has returned to the long-term equilibrium, and we use this criterion to close a trade. In this study, I take C as the mean of the spread, and plot it as a solid green line in both Figures 1 and 2.

In Strategy A (illustrated in Figure 1), a trade is opened at t_1 when the spread is higher than or equal to U. In this case, we sell 1 share of stock A and buy γ shares of stock B. When the spread is less than or equal to the mean (i.e., C), we close the trade and clear the position. The return from this trade is thus $U - C$. At t_2, when the spread is less than or equal to L, we open a trade, buying 1 share of stock A and selling γ shares of stock B. We close this trade and clear the position at t_2', when the spread is higher than or equal to the mean. The return from this trade is $C - L$.
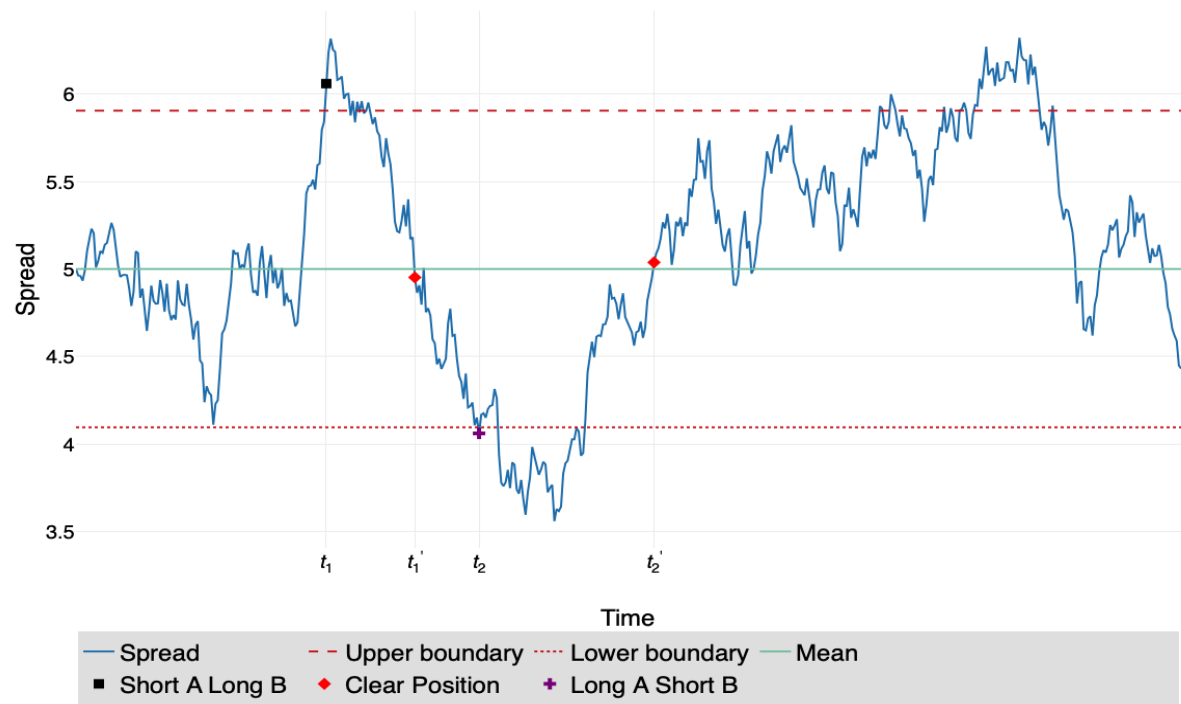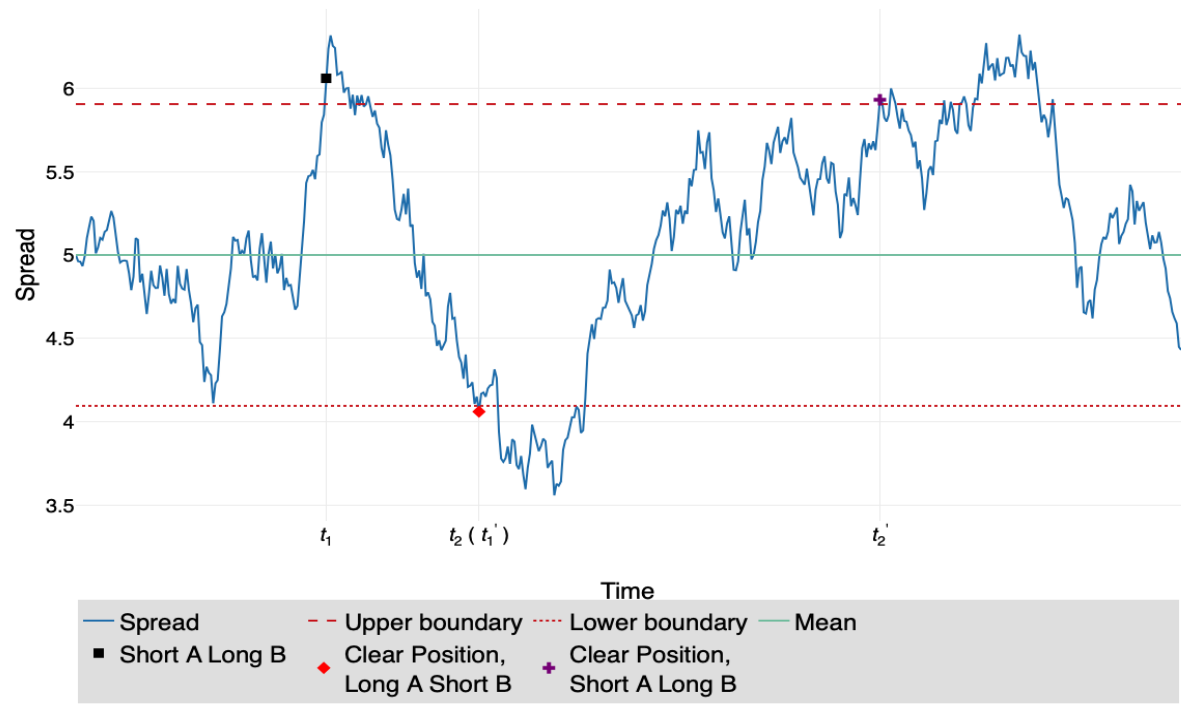
Figure 1: Trading Strategy A

Figure 2: Trading Strategy B

In Strategy B (illustrated in Figure 2), we open a trade when the spread crosses the upper boundary from below (e.g., at t_1) or crosses the lower boundary from above (e.g., at t_2). Unlike the Strategy A, we will hold the portfolio until we need to switch the position. Thus, in Strategy B, we clear the exposure at the same time when we open a new trade (i.e., t_2 and t'_1 coincide).

3.3 Conventional Trading Rules

When implementing pairs trading, the trading rule for a specific trading strategy is the computation of optimal thresholds, U and L, for fulfilling an investing objective based on that strategy. There are three popular approaches for computing the optimal thresholds, when the model (2) is linear mean-reverting, homoscedastic, and Gaussian (i.e., f is linear, g is a constant, and η is a Gaussian noise). The optimal trading rule for a general specification of model (2) is given in Section 3.5.

- Rule I: Ad hoc boundaries

Rule I takes U to be one (1-σ rule) or two (2-σ rule) standard deviations above the mean, L to be one or two standard deviations below the mean, and C is the mean of the spread. This rule is simple and popular in practice. In particular, the 2-σ rule was first applied by Gatev et al. (2006) and later checked by Moura et al. (2016), Zeng and Lee (2014), and Cummins and Bucca (2012). The 1-σ rule was discussed in Zeng and Lee (2014), and the performance of the 1-σ rule and 2-σ rule was compared in the same paper.

- Rule II : Boundaries based on the first-passage-time

This rule was first adopted by Elliott et al. (2005) and later by Moura et al. (2016). Suppose x_t, the spread, follows a standardized Ornstein–Uhlenbeck process:

$$dx_t = -x_t dt + \sqrt{2} dW_t$$

Let T_{0,x_0} be the first passage time of x_t, and we know T_{0,x_0} has a pdf known explicitly:

$$f_{0,x_0}(t) = \sqrt{\frac{2}{\pi}} \frac{|x_0| e^{-t}}{(1 - e^{-2t})^{3/2}} \exp\left(-\frac{x_0^2 e^{-2t}}{2(1 - e^{-2t})}\right)$$

Here, given the value of current spread, t^*, the optimizer of the above pdf, is the likeliest time that the spread will revert to the mean. In model (2), if the spread follows the Ornstein–Uhlenbeck process, then we can first standardize x, and use t^* to construct the optimal C. A similar idea can be applied to compute U and L.

- Rule III: Boundaries based on the renewal theorem

This rule was first proposed by Bertram (2010), and then extended by Zeng and Lee (2014). Under this rule, each trading cycle is separated into two parts, τ_1 and τ_2, defined by

$$\tau_1 = \inf \{t; x_t = C | x_0 = U\}, \quad \text{and} \quad \tau_2 = \inf \{t; x_t = U | x_0 = C\},$$

where x is the spread. Suppose T is the time taken to complete each trade cycle, i.e., $T = \tau_1 + \tau_2$, and N_τ is the number of transactions we can have in the period $[0, \tau]$. Then, by the renewal theorem, the return per unit time is given by:

$$(U - C) \lim_{\tau \to \infty} \frac{E(N_\tau)}{\tau} = \frac{U - C}{E(T)} = \frac{U - C}{E(\tau_1) + E(\tau_2)}.$$

where $E(\tau_1)$ and $E(\tau_2)$ can be computed based on the density of the first passage time, as mentioned in Rule II.

The problem of this rule, as Zeng and Lee (2014) have pointed out, is that when there is no transaction cost, this strategy implies that U (and L) will be arbitrarily close to C. This suggests that the trader values the trading frequency more than the profit per trade. Consequently, this could increase the risk of the portfolio significantly.

3.4 The New Trading Strategy

I summarize the new trading strategy (hereinafter, Strategy C) in Figure 3. The basic idea of Strategy C is similar to both Strategies A and B: opening a trade when the spread is far away from the equilibrium and closing the trade when the spread returns to the equilibrium. Unlike Strategies A and B, in Strategy C, we open a trade when the spread crosses the upper boundary from above (or crosses the lower boundary from below), and we clear the position when the spread crosses the mean, or crosses the boundaries (U and L) after a trade has been opened (i.e., the spread crosses the upper boundary from below or the lower boundary from above). For example, in Figure 3a for a homoscedastic model, at t_1, t_2, t_3, and t_4, we open a trade; and at t'_1, t'_2, t'_3, and t'_4, we clear the exposure. In Figure 3b for a heteroscedastic model, we open a trade at t_1 and t_2, and we close the trade at t'_1, and t'_2.

Now I discuss the properties of this trading strategy when the model (2) is homoscedastic (i.e., the g function is constant) and when it is heteroscedastic (i.e., g is non-constant). In the first situation, the main benefit of Strategy C is that we can avoid holding the portfolio when the spread is larger than the upper boundary (or smaller than the lower boundary). This would significantly decrease the portfolio risk and drawdown. The main drawback of Strategy C is that the return can be lower because we open a trade when the spread is closer to the mean of the spread than in the case of Strategy A. Therefore, there is a trade-off between the risk and the return. When model (2) is heteroscedastic, this strategy can not only reduce the risk but also improve the return

9

because opening a trade now depends on the level of volatility, and consequently, the boundaries are no longer constant over time. The logic of this new strategy is illustrated in Figures 3a and 3b, for homoscedastic and heteroscedastic cases, respectively.

3.5 Simulation Based Method for Optimal Trading Rule

For a general specification of models (1)-(2), the conventional trading rules in Section 3.3 are difficult to apply. For example, the 1-σ or 2-σ rules can not be applied when model (2) is heteroscedastic. For a complicated specification of model (2), it is impossible to derive the density of the first passage time explicitly; thus, rules II and III are unavailable in this case.

To compute the optimal trading rule under model (2) for all of the trading strategies, I propose to select the optimal boundaries (i.e., for U and L; I set C as the mean of the spread by default[2]) based on the Monte Carlo simulation of the spread (equation (2) given the estimation of the unknown parameters). Different criteria or investing objectives, such as expected return, Sharpe ratio or Calmar ratio, can be used to determine the optimal boundaries for a given trading strategy.

Now, I use the following five specifications of model (2) to describe the details of the computation of the new trading rules:

- Model 1: $x_{t+1} = 0.9590 * x_t + 0.0049 * \eta_t$, $\eta_t \sim N(0,1)$

- Model 2: $x_{t+1} = 0.9 * x_t + 0.2590 * x_t^2 + 0.0049 * \eta_t$, $\eta_t \sim N(0,1)$

- Model 3: $x_{t+1} = 0.9590 * x_t + \sqrt{(0.00089 + 0.08 * x_t^2)} * \eta_t$, $\eta_t \sim N(0,1)$

- Model 4: $x_{t+1} = 0.9590 * x_t + \frac{0.0049}{\sqrt{3}} * \eta_t$, $\eta_t \sim$ t-distribution with 3 degrees of freedom

- Model 5: $x_{t+1} = 0.9 * x_t + 0.2590 * x_t^2 + \frac{0.0049}{\sqrt{3}} * \eta_t$, $\eta_t \sim$ t-distribution with 3 degrees of freedom

Model 1 is a linear, Gaussian, and homoscedastic model. This is the most popular model used for pairs trading. See Elliott et al. (2005) and Moura et al. (2016) for examples of this model. Model 2 is a nonlinear mean-reverting and Gaussian model; Model 3 is a linear mean-reverting, Gaussian, and nonlinear heteroscedastic model; Model 4 is a linear mean-reverting and non-Gaussian model; and Model 5 is a nonlinear mean-reverting and non-Gaussian model. The last four models are different extensions of Model 1 and have never been discussed in the literature on pairs trading. These five models can be considered as the benchmark models for pairs trading. Further extensions are available based on the combination of these models, and our simulation-based method for the optimal trading rule can also be applied to them.

---

[2]I assume C as the mean of the spread for simplicity, and this assumption can be relaxed to construct more flexible strategies. For example, we can define $C \equiv ($ mean $\pm \Delta)$ for a parameter Δ and we close the trade when the spread enters into this region. My simulation-based method can also be applied to compute the optimal Δ. See an example of this strategy in Tie et al. (2018). A more detailed discussion of this strategy is provided in the online appendix.

Figure 3: Trading Strategy C

(a) Trading Strategy C in Homoscedastic Model
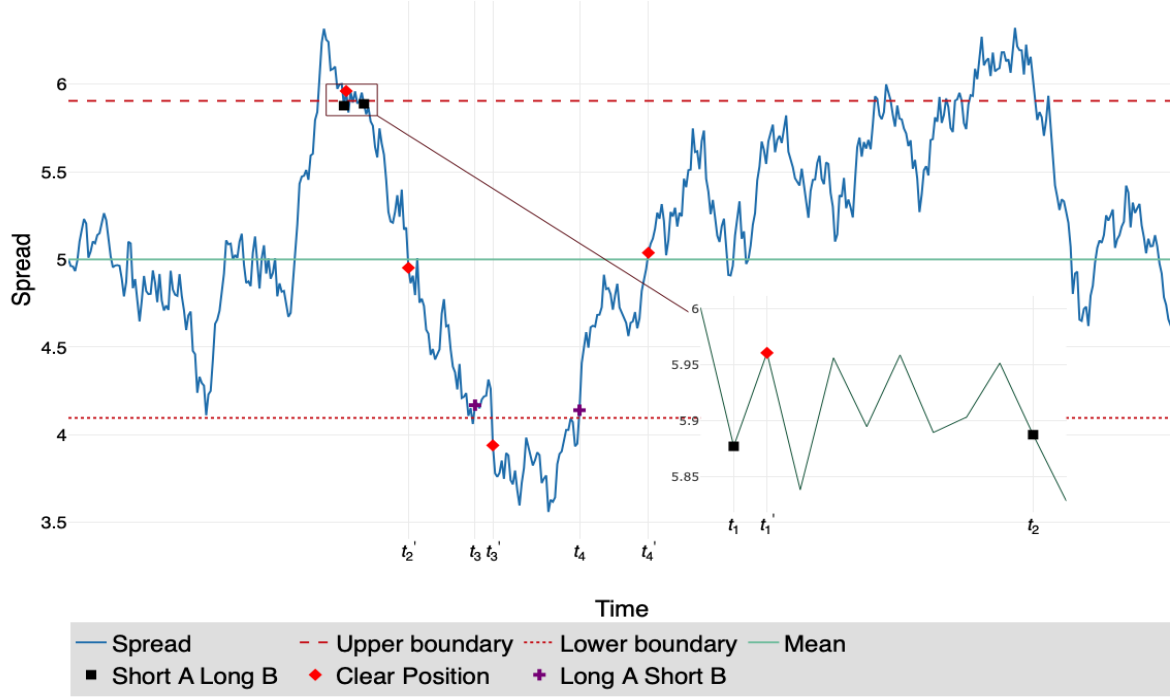
(b) Trading Strategy C in Heteroscedastic Model
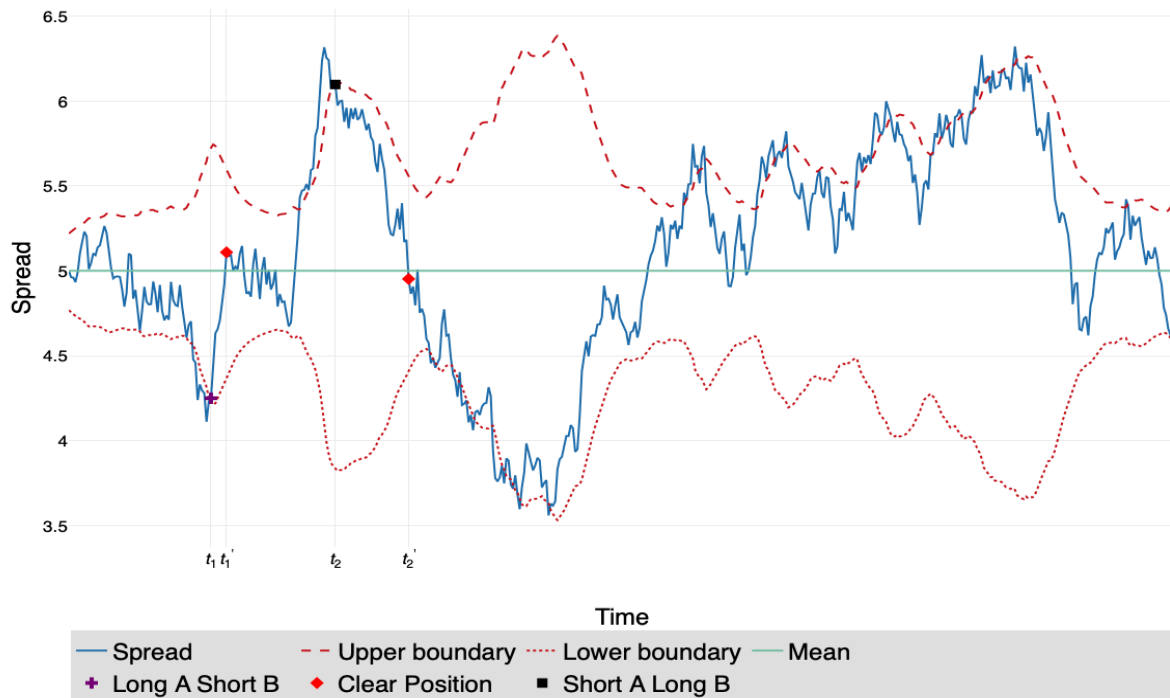
For every specification of Models 1-5, I calculate the optimal trading rules through the N simulations of the spread for Strategies A, B and C and compare the resulting performances of the three strategies based on the expected return and Sharpe ratio. More specifically, across all of the examples, I represent the optimal trading rule (upper boundary U and lower boundary L) as the ratio of its difference with the mean of the spread to one standard deviation of the spread, and I consider the upper boundary U between $[0.1, 2.5]$ and lower boundary L between $[-2.5, -0.1]$ for a grid size of 0.1. For every specification of Models 1-5 and every realization of the process of the spread $\{x_t^{(m,n)}\}_{t=1}^T$, where m is for different models ($m = 1, 2, 3, 4, 5$), and n is for different realizations of the spread in the simulation ($n = 1, \ldots N$), I choose U_i from $[0.1, 2.5]$ and L_j from $[-2.5, -0.1]$, where $i, j = 1, \ldots, 25$, and compute the resulting cumulative return and Sharpe ratio for difference strategies. More specifically, I denote the cumulative return and Sharpe ratio as $CR_{i,j}^{m,k,n}$ and, $SR_{i,j}^{m,k,n}$, respectively, where k is for difference strategies ($k = 1, 2, 3$). For model m and strategy k, the resulting expected cumulative return $CR_{i,j}^{m,k}$ and Sharpe ratio $SR_{i,j}^{m,k}$ are computed as

$$CR_{i,j}^{m,k} = \frac{1}{N} \sum_{n=1}^{N} CR_{i,j}^{m,k,n}, \quad \text{and} \quad SR_{i,j}^{m,k} = \frac{1}{N} \sum_{n=1}^{N} SR_{i,j}^{m,k,n}.$$

Then, the optimal trading rule $(U_{m,k}^*, L_{m,k}^*)$ is selected to maximize $CR_{i,j}^{m,k}$ or $SR_{i,j}^{m,k}$, that is,

$$\left[U_{m,k}^*, L_{m,k}^* \right] \quad = \quad \arg\max_{U_i, L_j} z_{i,j}^{m,k}$$

where $z = CR$ or SR. Across all of the examples, I set the total trading period to be 1,000 trading days (or approximately four years), and I set the simulation size to be $N = 10,000$. For simplicity, I assume that the transaction cost is 20 bp (0.2%) [3], and the annualized risk-free rate is set at 0.

In Table 1, I report the optimal trading rule for every combination of the five models and three strategies, and the resulting expected cumulative return and Sharpe ratio[4]. As we can find from this table, when the model is heteroscedastic, Strategy C outperforms the other two strategies in terms of both the cumulative return and the Sharpe ratio. Moreover, for the other homoscedastic models (Models 1, 2, 4, and 5), the Sharpe ratio of Strategy C is competitive, although the cumulative return is not. This supports our discussion of this new strategy in Section 3.4.

I leave the detailed results of the simulation method in the appendix. More precisely, the expected cumulative returns and Sharpe ratio as functions of various choices of U and L are given in Figures A1-A5 for every possible combination of the three strategies and four models.

---

[3]This transaction cost is on one asset of the pair. Since a complete trading includes transactions on two assets, the total transaction cost of one complete trading is 40 bp.

[4]If the spread and the strategy are symmetric around the mean, then the optimal upper boundary and lower boundary should also be symmetric around zero, i.e, $U^* = -L^*$. However, due to the approximation error in gridding, the absolute values of U^* and L^* may not be exactly the same in Table 1.

Table 1: Optimal Selection of Trading Rule for Cumulative Return and Sharpe Ratio

Model	Strategy	U^*	L^*	CR	U^*	L^*	SR
Model 1	A	0.7	-0.7	0.3868	1.1	-1.1	0.0882
	B	0.5	-0.5	0.4245	0.5	-0.5	0.0807
	C	1	-1	0.2990	0.9	-0.9	0.1044
Model 2	A	0.8	-0.8	0.5562	1.2	-1.3	0.1308
	B	0.6	-0.6	0.6085	0.6	-0.6	0.1203
	C	1.2	-1.3	0.3300	1.2	-1.3	0.1163
Model 3	A	0.3	-0.2	3.9413	0.4	-0.4	0.0751
	B	0.1	-0.1	4.0139	0.1	-0.1	0.0743
	C	0.8	-0.8	6.6763	0.1	-0.1	0.2499
Model 4	A	0.6	-0.6	0.3792	1	-1	0.0881
	B	0.4	-0.5	0.4071	0.5	-0.5	0.0782
	C	1	-1	0.2243	1	-1	0.0829
Model 5	A	0.7	-0.7	0.5359	1.2	-1.2	0.1293
	B	0.5	-0.5	0.5760	0.5	-0.5	0.1145
	C	1.2	-1.2	0.2423	1.4	-1.4	0.0961

Note: The third and forth columns are the optimal boundaries based on maximizing the cumulative return, and the fifth column is the resulting cumulative return. The sixth and seventh columns are the optimal boundaries based on maximizing the Sharpe ratio, and the eighth column is the resulting Sharpe ratio. The cumulative return is displayed in decimal.

3.6 Summary

Now I summarize the procedures for pairs trading based on models (1)-(2) and conclude this section.

- Step 1: Choose a specific model for (1)-(2). Given this model and observations $\{P_{A,t}, P_{B,t}\}_{t=1}^{T}$, we run the Quasi Monte Carlo Kalman filter and obtain the filtered estimation of the spread $\{\bar{x}_t\}_{t=1}^{T}$ and the estimation of the unknown parameters in the model. The details of running QMCKF are discussed in Section 3.1.

- Step 2: Choose a trading strategy, and determine the optimal trading rule (i.e., the optimal U and L) for a specific criterion using the Monte Carlo simulation based on the data until time T. The details of this step can be found in Sections 3.2-3.5.

- Step 3: For $t > T$, we run QMCKF and estimate $\bar{x}_t$ with $\psi = \hat{\psi}$, the estimate of the parameters obtained in Step 1. We use this $\{\bar{x}_t\}_{t>T}$ and follow the preset trading strategy and optimal trading rule from Step 2 to generate the signal for trading.

13

4 Applications

In this section, I test the performance of pairs trading through general state space modeling for different trading strategies. Across all of the applications in this section, I assume that the transaction cost is 20 bp and the annualized risk-free rate is 2%, and I test the performance of Strategies A, B, and C for two specifications of model (2):

- Model I: $x_{t+1} = \theta_0 + \theta_1 x_t + \theta_2 * \eta_t$, $\eta_t \sim N(0, 1)$

- Model II: $x_{t+1} = \theta_0 + \theta_1 x_t + \sqrt{\theta_2 + \theta_3 x_t^2} * \eta_t$, $\eta_t \sim N(0, 1)$

As I explained in Section 3.4, the new trading strategy (Strategy C) can significantly improve the performance of pairs trading when the model (2) is heteroscedastic. See Model 3 of Table 1 for a comparison based on simulation. In the following, I use actual data to compare the performance of various strategies and show the improvement of the heteroscedastic modeling combined with the proposed strategy.

4.1 Pepsi vs. Coca

In this example, I examine the performance of pairs trading for PEP (Pepsi) and KO (Coca). The data consist of daily observations of adjusted closing prices of PEP and KO from 01/03/2012 to 06/28/2019. Panel A of Table 2 reports the parameter estimation of both Models I and II for this pair. The annualized performance metrics are provided in Panel A of Table 3. It is clear that in Model II, the annualized return of Strategy C is almost 50% higher than those of Strategies A and B, while Strategy C maintains almost half of the risk (measured by annualized standard deviation) of Strategy A or B. By comparing these matrices, we find that this improvement is significant. However, the difference in the performances of Strategies A and B across the two models is limited. This implies that the effect of heteroscedasticity modelling on the performances of Strategies A and B is not significant. This is because in Strategies A and B, the hedging portfolio is held until the spread is around the mean; therefore, the frequency of changing positions is lower in Strategies A or B than that in Strategy C.

4.2 EWT vs. EWH

In this example, I examine the performance of pairs trading for EWT and EWH. The data consist of daily adjusted closing prices of EWT and EWH from 01/01/2012 to 05/01/2019. EWT is the iShares MSCI Taiwan ETF managed by BlackRock. It seeks to track the investment results of an index composed of Taiwanese equities, while EWH similarly corresponds to Hong Kong equities. Following the example of PEP vs. KO, I test the performance of Strategies A, B, and C for Models

Table 2: Parameter Estimation of Model I and Model II on PEP vs. KO and EWT vs. EWH

	Panel A: PEP vs. KO		Panel B: EWT vs. EWH	
	Model I	Model II	Model I	Model II
γ	1.98	2.03	1.40	1.42
σ_{ε}^2	0.012	0.011	0.0007	0.0006
θ_0	-0.0001	-0.001	-0.0004	-0.0015
θ_1	0.9572	0.9330	0.9898	0.9589
θ_2	0.029	0.0003	0.0337	0.0016
θ_3	-	0.1283	-	0.1136

Table 3: Annualized Performance of Pairs Trading on PEP vs. KO and EWT vs. EWH

Strategy and Model	Return	Std Dev	Sharpe	Calmar	Pain index
Panel A: PEP vs. KO					
Strategy A, Model I	0.1311	0.0988	1.1003	1.3742	0.0195
Strategy B, Model I	0.1385	0.1153	1.0052	1.2204	0.0334
Strategy C, Model I	0.0618	0.0534	0.7649	0.8243	0.0087
Strategy A, Model II	0.1340	0.1038	1.0751	1.4040	0.0200
Strategy B, Model II	0.1407	0.1139	1.0366	1.2398	0.0258
Strategy C, Model II	0.2186	0.0659	2.9518	8.2384	0.0030
Panel B: EWT vs. EWH					
Strategy A, Model I	0.1480	0.1111	1.1277	1.3042	0.0156
Strategy B, Model I	0.1109	0.1362	0.6531	0.7836	0.0328
Strategy C, Model I	0.1294	0.0740	1.4458	3.0926	0.0080
Strategy A, Model II	0.1402	0.1223	0.9622	1.2354	0.0196
Strategy B, Model II	0.1093	0.1349	0.6473	0.7717	0.0306
Strategy C, Model II	0.3184	0.0752	3.8892	10.3005	0.0032

I and II. I report the parameter estimation in Panel B of Table 2, and the annualized performance in Panel B of Table 3. Comparing the annualized performance, we find that the heteroscedasticity modeling can improve the performance of Strategy C significantly, while it has no effect on Strategy A or B.

4.3 Pairs Trading on U.S. Banks Listed on the NYSE

I use this example to illustrate the improvement of our new modeling and strategy by implementing pairs trading on U.S. banks listed on the NYSE from 01/01/2013 to 01/10/2019. To avoid data snooping and make our results more concrete, I use a simple method in choosing assets and constructing pairs. Specifically, based on market capacity, I select the five largest banks to construct a group of large banks and the five smallest banks to construct a group of small banks. The large bank group includes JPM, BAC, WFC, C, and USB[5], and the small bank group includes CPF, BANC, CUBI, NBHC, and FCF[6]. I compare the performance of Model I combined with Strategy A against Model II combined with Strategy C. Model I combined with Strategy A is a popular approach in the existing literature on pairs trading, making it a good benchmark for comparison.

In Panel A of Table A1, I report the performance of these two approaches on 10 pairs of the large banks. The performance on 10 pairs among the small banks is given in Panel B of Table A1. It is apparent that Model II combined with Strategy C outperforms Model I combined with Strategy A for almost all the pairs, either in terms of annualized returns or annualized Sharpe ratios. The improvement in the Sharpe ratio of Model II combined with Strategy C is much more significant than that in returns. For example, when trading is implemented on pairs among large banks, the improvement in return is 41.29%, and the improvement in the Sharpe ratio is 89.23%. If trading is implemented on pairs of small banks, the improvement in returns is 74.41%, and that in the Sharpe ratio is 151.8%.

In addition, by comparing the results in both panels of Table A1, we find that the performance of pairs of small banks is better than that of large banks, either for Model I combined with Strategy A or Model II combined with Strategy C is applied for trading. For example, applying Model I combined with Strategy A, the mean of the returns of all pairs of large banks is 0.0703, and that of small banks can be improved to 0.1524. Moreover if Model II combined with Strategy C is applied, we can get an improvement of 0.1664 (from 0.0994 to 0.2658) by switching from trading on large banks to trading on small banks. In Table A2, I report the performance of the two approaches of pairs trading for all possible pairs between large banks and small banks, that is, I pair one large bank with one small bank. For some pairs, such as JPM-CUBI and BAC-CUBI, the spread is far from being mean-reverting, thus resulting in poor pairs trading performance.

To further investigate the performance of pairs trading, I check the out-of-sample performance of the two approaches on the 10 bank stocks. More precisely, I separate 01/10/2012 to 01/12/2019 into two periods: 01/10/2012 to 01/01/2018 (in-sample period) and 01/01/2018 to 01/12/2019

---

[5]JPM is for J P Morgan Chase & Co; BAC is for Bank of America Corporation; WFC is for Wells Fargo & Company; C is for Citigroup Inc.; USB is for U.S. Bancorp.

[6]CPF is for CPB Inc.; BANC is for Banc of California, Inc.; CUBI is for Customers Bancorp, Inc.; NBHC is for National Bank Holdings Corporation; FCF is for First Commonwealth Financial Corporation.

(out-of-sample period). I use the in-sample data to train the model, estimate the parameter of the model, and determine the optimal trading rules. In the out-of-sample period, I use the parameters and optimal trading rules based on in-sample data to generate the trading signal. The results are given in Tables A3-A6. From these tables, we can also confirm our earlier findings that Model II combined with Strategy C outperforms Model I combined with Strategy A in terms of both the returns and Sharpe ratio, and the improvement is more significant with regard to the Sharpe ratio.

5 Conclusion

Pairs trading is a statistical arbitrage involving the long/short position of overpriced and underpriced assets. The results in this study show that digging into the modeling and trading strategy can improve the performance of pairs trading significantly and harness its greater potential of pairs trading in the financial market. This finding can help the empirical research on the general profitability of pairs trading and discussions on the tests of market efficiency, which I leave for future research to address.

References

Ait-Sahalia, Y. (1996): *Testing Continuous-Time Models of the Spot Interest Rate*, Review of Financial Studies, 9, 385-426

Avellaneda, M., and J.H. Lee (2010): *Statistical arbitrage in the US equities market.* Quantitative Finance 10:761–782.

Bai, Y., and L. Wu (2018): *Analytic value function for optimal regime-switching pairs trading rules*, Quantitative Finance, 2018, Vol.18(4)

Barndorff-Nielsen, O.E., and N. Shephard (2001): *Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics*, Journal of Royal Statistical Society. B 63, Part 2, pp. 167-241

Bertram, W.K. (2010): *Analytic solutions for optimal statistical arbitrage trading*, Physica A: Statistical Mechanics and its Applications, 389(11):2234–2243.

Bogomolov, T. (2013): *Pairs trading based on statistical variability of the spread process*, Quantitative Finance 13:1411–1430.

Bollerslev, T., R.Y. Chou, and K.F. Kroner (1992): *ARCH modeling in finance: A review of the theory and empirical evidence*, Journal of Econometrics 52 5-59.

Cameron, A.C., and P.K. Trivedi (2005): *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, May 2005

Clegg, M., and C. Krauss (2018): *Pairs trading with partial cointegration*, Quantitative Finance 18 (1), 121–138.

Cont, R. (2001): *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance VOL 1 223–236

Cummins, M., and A. Bucca (2012): *Quantitative spread trading on crude oil and refined products markets*, Quantitative Finance, Vol. 12 Issue 12, p1857-1875.

Ding, Z., and C.W.J. Granger (1996): *Modeling volatility persistence of speculative returns: A new approach*, Journal of Econometrics, vol. 73, issue 1, 185-215

Ding, Z., C.W.J. Granger, and R.F. Engle (1993): *A long memory property of stock market returns and a new model*, Journal of Empirical Finance, Volume 1, Issue 1, Pages 83-106

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

Elliott, R.J., and R. Bradrania (2018): *Estimating a regime switching pairs trading model*, Quantitative Finance, Vol.18(5), pp.877-883

Fama, E.F., and J.D. MacBeth (1971): *Risk, return, and equilibrium*, The Journal of Political Economy Vol. 81, No. 3, pp. 30–55

Fama, E.F. (1970): *Efficient Capital Markets: A Review of Theory and Empirical Work*, The Journal of Finance, Vol. 25, No. 2, Papers and Proceedings of the Twenty Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, pp. 383-417

Focardi, S.M., F.J. Fabozzic, and I.K. Mitov (2016): *A new approach to statistical arbitrage: Strategies based on dynamic factor models of prices and their performance*, Journal of Banking and & Finance, 65 134-155

Gatev, E.G., W.N. Goetzmann, and K.G. Rouwenhorst (2006): *Pairs Trading: Performance of a Relative Value Arbitrage Rule*, The Review of Financial Studies, 19, 797-827.

Hajivassiliou, V.A., and D.L. McFadden (1998): *The Method of Simulated Scores for the Estimation of LDV Models*, Econometrica, Jul., 1998, Vol. 66, No. 4, pp. 863-896

Hsieh, D.A. (1989): *Testing for Nonlinear Dependence in Daily Foreign Exchange Rates*, The Journal of Business, Vol. 62, No. 3, pp. 339-368

Kon, S. (1984): *Models of stock returns: a comparison*, The Journal of Finance XXXIX 147–65

Lei, Y., and J. Xu (2015): *Costly arbitrage through pairs trading*, Journal of Economic Dynamics & Control, 56 1-19

Mandelbrot B.B. (1963): *The variation of certain speculative prices*, Journal of Business XXXVI 392–417

Mandelbrot, B.B. (1971): *When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models*, The Review of Economics and Statistics, Vol. 53, No. 3 pp. 225-236

Moura, C.E., A. Pizzinga, and J. Zubelli (2016): *A pairs trading strategy based on linear state space models and the Kalman filter*, Quantitative Finance

Rad, H., R.K.Y. Low, and R. Faff (2016): *The profitability of pairs trading strategies: distance, cointegration and copula methods*, Quantitative Finance 16:1541–1558.

Scheinkman, J.A., and B. LeBaron (1989): *Nonlinear dynamics and stock returns, The Journal of Business*, Vol. 62, No. 3, pp. 311-337

Stübinger, J., and S. Endres (2018): *Pairs trading with a mean-reverting jump-diffusion model on high-frequency data*, Quantitative Finance Volume 18, Issue 10

Suzuki, K. (2018): *Optimal pair-trading strategy over long/short/square positions—empirical study*, Quantitative Finance, Volume 18, Issue 1

Tie, J., H. Zhang, and Q. Zhang (2018): *An optimal strategy for pairs trading under geometric brownian motions*, Journal of Optimization Theory and Applications, volume 179, pages 654–675

Tourin, A., and R. Yan (2013): *Dynamic pairs trading using the stochastic control approach*, Journal of Economic Dynamics and Control, 37 1972-1981.

Vidyamurthy, G. (2004): *Pairs trading: Quantitative methods and analysis*, J. Wiley, Hoboken, N.J.

Wan, E.A., R. van der Merwe, and A.T. Nelson (1999): *Dual estimation and the unscented transformation*, NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems, November 1999, Pages 666–672

Wan, E.A., and A.T. Nelson (2000): *Dual Kalman Filtering Methods for Nonlinear Prediction, Smoothing, and Estimation*, Advances in neural information processing systems, 2000

Yang, S.Y., Q. Qiao, P.A. Beling, W.T. Scherer, and A.A. Kirilenko (2015): *Gaussian process-based algorithmic trading strategy identification*, Quantitative Finance, 1–21.

Zeng, Z., and C.G. Lee (2014): *Pairs trading: optimal thresholds and profitability*, Quantitative Finance 14 (11), 1881–1893.

Zhang, G. (2020): *Quasi Monte Carlo Kalman filter for nonlinear and non-Gaussian state space models*, Unpublished manuscript, Department of Economics, Boston Univeristy.

Table A1: Performance of Pairs Trading on Pairs of Large Banks and Pairs of Small Banks

Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
Panel A: Pairs of Large Banks								
1	JPM	BAC	0.1185	1.0030	0.0961	1.1126	-18.90	10.93
2	JPM	WFC	0.0229	0.2268	0.0581	0.7434	153.7	227.8
3	JPM	C	0.0567	0.5359	0.1049	1.3486	85.01	151.7
4	JPM	USB	0.0412	0.3971	0.0663	0.7832	60.92	97.23
5	BAC	WFC	0.0451	0.3455	0.0695	0.6046	54.10	74.99
6	BAC	C	0.0874	0.8158	0.1369	1.7516	56.64	114.7
7	BAC	USB	0.0554	0.3786	0.0923	1.0077	66.61	166.2
8	WFC	C	0.1031	0.8041	0.1014	0.9731	-1.649	21.02
9	WFC	USB	0.0591	0.5631	0.0674	0.8934	14.04	58.66
10	C	USB	0.1140	0.9040	0.2009	2.0862	76.23	130.8
Mean			0.0703	0.5974	0.0994	1.1304	41.29	89.23
Min			0.0229	0.2268	0.0581	0.6046	153.7	166.6
Max			0.1185	1.0030	0.2009	2.0862	69.54	108.0
Median			0.0579	0.5495	0.0942	0.9904	62.69	80.24
Panel B: Pairs of Small Banks								
1	CPF	BANC	0.1832	0.6745	0.2158	1.3428	17.79	99.08
2	CPF	CUBI	0.1092	0.4736	0.2374	1.3563	117.4	186.4
3	CPF	NBHC	0.1436	0.7694	0.1912	1.2573	33.15	63.41
4	CPF	FCF	0.1162	0.7127	0.2175	1.7210	87.18	141.5
5	BANC	CUBI	0.1583	0.5199	0.4820	1.9742	204.5	279.7
6	BANC	NBHC	0.2105	0.8353	0.1807	1.1435	-14.16	36.90
7	BANC	FCF	0.1669	0.5830	0.3094	2.1898	85.38	275.6
8	CUBI	NBHC	0.1575	0.6049	0.2392	1.4485	51.87	139.5
9	CUBI	FCF	0.1362	0.5593	0.2718	1.5292	99.56	173.4
10	NBHC	FCF	0.1425	0.8161	0.3132	2.5273	119.8	209.7
Mean			0.1524	0.6549	0.2658	1.6490	74.41	151.8
Min			0.1092	0.4736	0.1807	1.1435	65.48	141.4
Max			0.2105	0.8353	0.4820	2.5273	129.0	202.6
Median			0.1506	0.6397	0.2383	1.4889	58.29	132.7

Note: Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Table A2: Performance of Pairs Trading on Pairs Between Large Banks and Small Banks.

Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
1	JPM	CPF	0.0670	0.3965	0.1833	1.4799	173.6	273.2
2	JPM	BANC	0.0587	0.2396	0.0935	0.8334	59.28	247.8
3	JPM	CUBI	-0.0604	-0.2669	0.0423	0.3536	170.0	232.5
4	JPM	NBHC	0.1860	0.9750	0.2683	2.1385	44.25	119.3
5	JPM	FCF	0.1151	0.7230	0.2594	2.3479	125.4	224.7
6	BAC	CPF	0.0778	0.3770	0.2486	1.5596	219.5	313.7
7	BAC	BANC	0.0565	0.2124	0.1383	0.7916	144.8	272.7
8	BAC	CUBI	-0.0959	-0.3612	0.0473	0.5852	149.4	262.0
9	BAC	NBHC	0.1942	0.9496	0.3420	2.4948	76.11	162.7
10	BAC	FCF	0.1729	0.9061	0.2541	2.1954	46.96	142.3
11	WFC	CPF	0.0420	0.2149	0.1138	1.2746	171.0	493.1
12	WFC	BANC	0.1671	0.6058	0.2071	1.0214	23.94	68.60
13	WFC	CUBI	0.0606	0.2572	0.2053	1.3002	238.8	405.5
14	WFC	NBHC	0.1410	0.7844	0.1237	0.9464	-12.27	20.65
15	WFC	FCF	0.1058	0.5948	0.1366	1.3104	29.11	120.3
16	C	CPF	0.1421	0.7000	0.2214	2.1513	55.81	207.3
17	C	BANC	0.0244	0.0961	0.1999	1.1101	719.3	1055
18	C	CUBI	-0.0031	-0.0138	0.0617	0.4357	2090	3257
19	C	NBHC	0.2164	1.0536	0.2927	2.3896	35.26	126.8
20	C	FCF	0.1520	0.7687	0.2246	1.8611	47.76	142.1
21	USB	CPF	0.0782	0.4494	0.2408	2.0902	207.9	365.1
22	USB	BANC	0.1435	0.5450	0.2361	1.7444	64.53	220.1
23	USB	CUBI	-0.0678	-0.2938	0.0700	0.3497	203.2	219.0
24	USB	NBHC	0.1911	1.2574	0.2384	2.1422	24.74	70.37
25	USB	FCF	0.0789	0.5077	0.1206	1.1142	52.85	119.5
Mean			0.0898	0.4671	0.1828	1.4409	103.6	208.4
Min			-0.0959	-0.3612	0.0423	0.3497	144.1	196.8
Max			0.2164	1.2574	0.3420	2.4948	58.04	98.41
Median			0.0789	0.5077	0.2053	1.3104	160.2	158.1

Note: Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Table A3: In Sample and Out of Sample Performance of Pairs Trading on Pairs of Large Banks

Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
Panel A: In Sample Performance								
1	JPM	BAC	0.1145	0.8864	0.1501	1.8003	31.09	103.1
2	JPM	WFC	0.0160	0.1461	0.0795	0.9451	396.9	546.9
3	JPM	C	0.0664	0.5686	0.1013	1.5193	52.56	167.2
4	JPM	USB	0.0186	0.2172	0.0629	1.4293	238.2	558.1
5	BAC	WFC	0.0027	0.0179	0.0568	0.4748	2004	2553
6	BAC	C	0.0920	0.7252	0.1193	1.5417	29.67	112.6
7	BAC	USB	0.0603	0.3936	0.1535	1.5144	154.6	284.8
8	WFC	C	0.0827	0.5918	0.1219	1.2283	47.40	107.6
9	WFC	USB	0.0600	0.6432	0.0739	0.9603	23.17	49.30
10	C	USB	0.1146	0.8553	0.1695	1.7648	47.91	106.3
Mean			0.0628	0.5045	0.1089	1.3178	73.42	161.2
Min			0.0027	0.0179	0.0568	0.4748	2004	2553
Max			0.1146	0.8864	0.1695	1.8003	47.91	103.1
Median			0.0634	0.5802	0.1103	1.4719	74.11	153.7
Panel B: Out of Sample Performance								
1	JPM	BAC	-0.0503	-0.4730	-0.0500	-0.4760	0.5964	-0.6342
2	JPM	WFC	-0.0809	-0.5693	-0.0361	-0.3281	55.38	42.37
3	JPM	C	-0.0841	-0.6845	0.0299	0.3228	135.6	147.2
4	JPM	USB	0.0867	0.9267	0.1297	1.6816	49.60	81.46
5	BAC	WFC	0.0364	0.4593	0.0464	0.4636	27.47	0.9362
6	BAC	C	-0.0512	-0.3766	0.0149	0.2612	129.1	169.4
7	BAC	USB	-0.0037	-0.0252	0.0587	0.5169	1686	2151
8	WFC	C	-0.0586	-0.3472	0.0698	0.7619	219.1	319.5
9	WFC	USB	-0.1029	-0.6961	0.0269	0.3591	126.4	151.6
10	C	USB	-0.0486	-0.2948	0.0942	0.7796	293.8	364.5
Mean			-0.0357	-0.2081	0.0384	0.4343	207.6	308.7
Min			-0.1029	-0.6961	0.0500	-0.4760	51.41	31.62
Max			0.0867	0.9267	0.1297	1.6816	49.60	81.46
Median			-0.0508	-0.3619	0.0382	0.4114	175 2	213.7

Note: The data is from 01/10/2012 to 01/01/2018. Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Table A4: In Sample and Out of Sample Performance of Pairs Trading on Pairs of Small Banks

Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
Panel A: In Sample Performance								
1	CPF	BANC	0.2713	0.9758	0.3513	2.0574	29.49	110.8
2	CPF	CUBI	0.1226	0.4404	0.4457	1.9114	263.5	334.0
3	CPF	NBHC	0.1905	0.9823	0.2559	1.7188	34.33	74.98
4	CPF	FCF	0.1855	1.2385	0.2453	2.5505	32.24	105.9
5	BANC	CUBI	0.2500	0.6928	0.4076	1.9505	63.04	181.5
6	BANC	NBHC	0.2406	0.8926	0.1699	1.4127	-29.38	58.27
7	BANC	FCF	0.2056	0.7819	0.3308	1.8279	60.89	133.8
8	CUBI	NBHC	0.1130	0.3808	0.2164	1.8059	91.50	374.2
9	CUBI	FCF	0.1125	0.4133	0.1886	1.1579	67.64	180.2
10	NBHC	FCF	0.1026	0.5723	0.2523	1.8035	145.9	215.1
Mean			0.1794	0.7371	0.2864	1.8197	59.64	146.9
Min			0.1026	0.3808	0.1699	1.1579	65.59	204.1
Max			0.2713	1.2385	0.4457	2.5505	64.28	105.9
Median			0.1880	0.7374	0.2541	1.8169	35.16	146.4
Panel B: Out of Sample Performance								
1	CPF	BANC	0.1856	0.7541	0.1649	0.8297	-11.15	10.03
2	CPF	CUBI	-0.0924	-0.3528	0.2424	1.8467	362.3	623.4
3	CPF	NBHC	-0.0769	-0.3944	0.1621	1.0216	310.8	359.0
4	CPF	FCF	-0.0373	-0.1906	0.2094	1.4249	661.4	847.6
5	BANC	CUBI	0.1266	0.7454	0.4109	2.5902	224.6	247.5
6	BANC	NBHC	-0.1577	-0.6720	-0.0797	-0.3926	49.46	41.58
7	BANC	FCF	0.0107	0.0821	0.1601	1.3930	1396	1596
8	CUBI	NBHC	-0.1475	-0.5514	0	-	100	-
9	CUBI	FCF	-0.1137	-0.4079	0	-	100	-
10	NBHC	FCF	-0.0578	-0.3088	0.1520	1.0421	363.0	437.4
Mean			-0.0360	-0.1296	0.1422	0.9756	494.6	852.6
Min			-0.1577	-0.6720	-0.0797	-0.3926	49.46	41.58
Max			0.1856	0.7541	0.4109	2.5902	121.4	243.5
Median			-0.0674	-0.3308	0.1611	1.0319	339.2	411.9

Note: The data is from 01/10/2012 to 01/01/2018. Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Table A5: In Sample Performance of Pairs Trading on Pairs Between Large Banks and Small Banks

| Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
1	JPM	CPF	0.1668	0.9415	0.2866	3.0567	71.82	224.7
2	JPM	BANC	0.2067	0.7134	0.2581	1.5501	24.87	117.3
3	JPM	CUBI	0.0649	0.9832	0.2576	1.6633	296.9	69.17
4	JPM	NBHC	0.1505	0.8387	0.2735	2.2745	81.73	171.2
5	JPM	FCF	0.2083	1.3273	0.3281	2.9235	57.51	120.3
6	BAC	CPF	0.1572	0.7484	0.2099	1.7310	33.52	131.3
7	BAC	BANC	0.2361	0.7452	0.1708	1.0044	-27.66	34.78
8	BAC	CUBI	0.0789	0.2755	0.1669	1.4519	111.5	427.0
9	BAC	NBHC	0.2608	1.2323	0.3354	2.5663	28.60	108.3
10	BAC	FCF	0.1918	1.0401	0.2653	2.3337	38.32	124.4
11	WFC	CPF	0.0376	0.1924	0.0988	0.6388	162.8	232.0
12	WFC	BANC	0.2371	0.8323	0.2165	1.0599	-8.690	27.53
13	WFC	CUBI	0.0729	0.2682	0.2307	1.9597	216.5	630.7
14	WFC	NBHC	0.0974	0.5548	0.0917	0.6167	-5.850	11.16
15	WFC	FCF	0.0656	0.3971	0.1413	1.1406	115.4	187.2
16	C	CPF	0.0571	0.2873	0.1766	1.4015	206.3	387.8
17	C	BANC	0.2454	0.8899	0.2154	1.9512	-12.22	119.3
18	C	CUBI	0.0715	0.2696	0.1589	1.0954	122.2	306.3
19	C	NBHC	0.1279	0.6511	0.2125	1.5321	66.15	135.3
20	C	FCF	0.1160	0.6154	0.1790	1.3736	54.31	123.2
21	USB	CPF	0.0654	0.4915	0.2126	1.9990	225.1	306.7
22	USB	BANC	0.2164	0.7529	0.3389	1.9118	56.61	153.9
23	USB	CUBI	0.0565	0.2443	0.2826	1.9450	400.2	696.2
24	USB	NBHC	0.1340	0.9289	0.1947	1.5321	45.30	64.94
25	USB	FCF	0.0922	0.6221	0.2167	2.1579	135.0	246.9
Mean			0.1366	0.6737	0.2208	1.7148	61.61	154.5
Min			0.0376	0.1924	0.0917	0.6167	143.9	220.5
Max			0.2608	1.3273	0.3389	3.0567	29.95	130.3
Median			0.1279	0.7134	0.2154	1.6633	68.41	133.2

Note: The data is from 01/10/2012 to 01/01/2018. Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Table A6: Out of Sample Performance of Pairs Trading on Pairs Between Large Banks and Small Banks

Pair	Stock #1	Stock #2	Model I + Strategy A		Model II + Strategy C		Improvement (in %)	
			Return	Sharpe	Return	Sharpe	Return	Sharpe
1	JPM	CPF	0.1514	0.8997	0.2731	2.3058	80.38	156.3
2	JPM	BANC	0.2190	0.9752	0.2023	1.1630	-7.626	19.26
3	JPM	CUBI	0.0965	1.1227	0.1610	1.0135	66.84	-9.727
4	JPM	NBHC	0.0303	0.1492	0.1799	1.8165	493.7	1117
5	JPM	FCF	0.0878	0.4209	0.1682	1.0338	91.57	145.6
6	BAC	CPF	0.0379	0.1702	0.1592	1.3579	320.1	697.8
7	BAC	BANC	0.1763	0.6913	0.1693	0.8830	-3.971	27.73
8	BAC	CUBI	0.0926	0.3435	0.1014	0.4298	9.503	25.12
9	BAC	NBHC	-0.0212	-0.0999	0.0144	0.7148	167.9	815.5
10	BAC	FCF	0.0196	0.0899	0.1117	0.8152	469.9	8.6.8
11	WFC	CPF	-0.0625	-0.2981	-0.0061	0.6388	90.24	314.3
12	WFC	BANC	0.0583	0.2249	0.1282	0.6058	119.9	169.4
13	WFC	CUBI	-0.0181	-0.0652	0.2826	1.5870	1661	2534
14	WFC	NBHC	-0.1181	-0.5631	0.0447	0.2594	137.8	146.1
15	WFC	FCF	-0.0821	-0.3725	0.1225	0.8413	249.2	325.9
16	C	CPF	-0.0072	-0.0314	0.1433	1.1894	2090	3888
17	C	BANC	0.1238	0.4691	0.0839	0.6480	-32.23	38.13
18	C	CUBI	0.0459	0.1692	0.2568	1.2778	459.5	655.2
19	C	NBHC	-0.0648	-0.2911	0.2108	2.1138	425.3	826.1
20	C	FCF	-0.0265	-0.1143	0.2174	1.2651	920.4	1207
21	USB	CPF	0.2108	2.2429	0.2652	2.4946	25.81	11.22
22	USB	BANC	0.1951	0.8939	0.1909	1.3332	-2.153	49.14
23	USB	CUBI	0.1516	0.7685	0.2356	1.5712	55.41	104.5
24	USB	NBHC	-0.0242	-0.1258	0.1514	0.9637	725.6	866.1
25	USB	FCF	0.0037	0.0192	0.1979	1.2151	5249	6229
Mean			0.0510	0.3076	0.1626	1.1815	218.6	284.2
Min			-0.1181	-0.5631	-0.0061	0.2594	94.84	146.4
Max			0.2190	2.2429	0.2826	2.4946	29.04	11.22
Median			0.0379	0.1692	0.1682	1.1630	343.8	587.4

Note: The data is from 01/01/2018 to 01/12/2019. Return is the annualized return, displayed in decimal. Sharpe is the annualized Sharpe ratio.

Figure A1: Performance of Strategy A, B and C, based on Model 1

(a) Return of Strategy A, Model 1
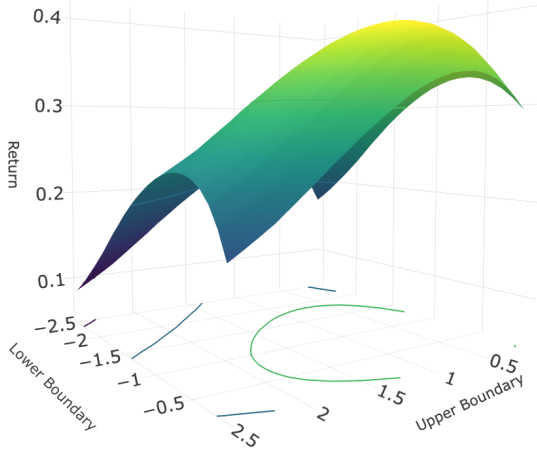
(b) Sharpe Ratio of Strategy A, Model 1
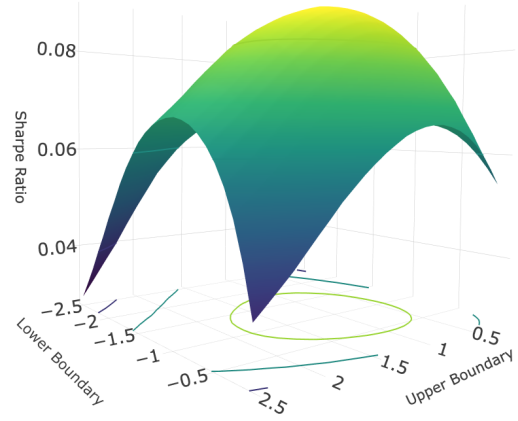
(c) Return of Strategy B, Model 1
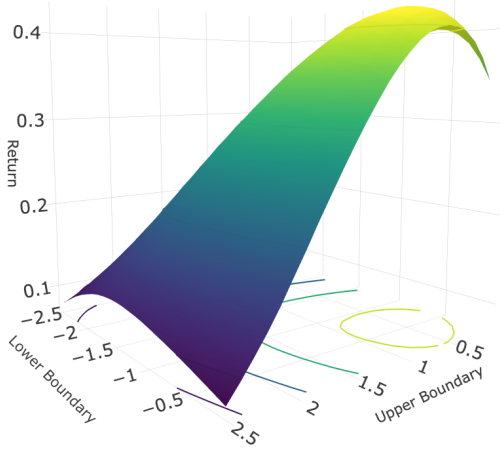
(d) Sharpe Ratio of Strategy B, Model 1
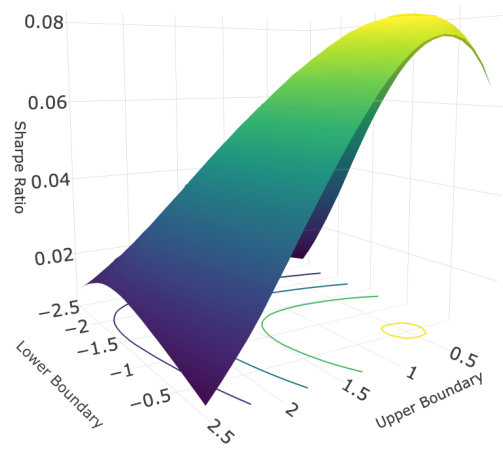
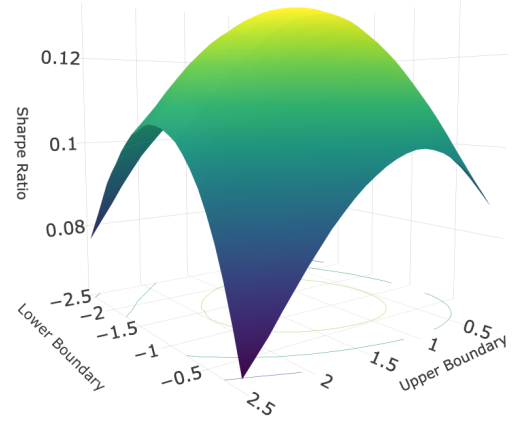(e) Return of Strategy C, Model 1
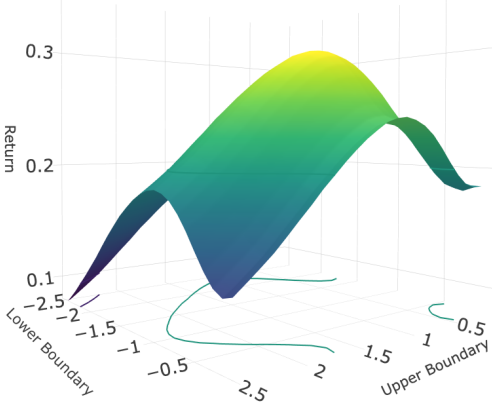
(f) Sharpe Ratio of Strategy C, Model 1
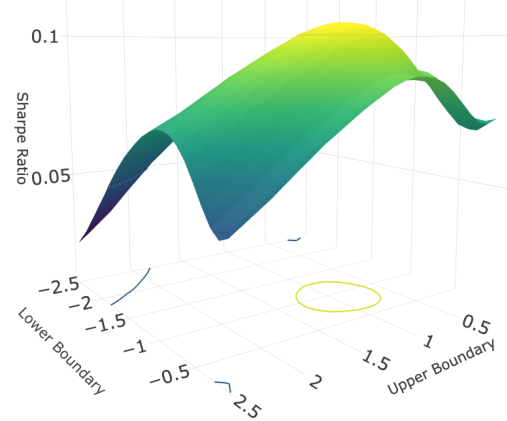
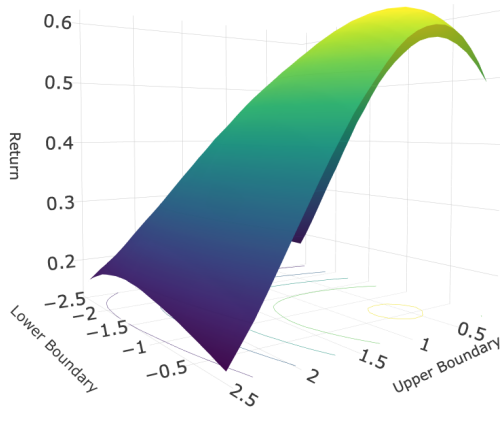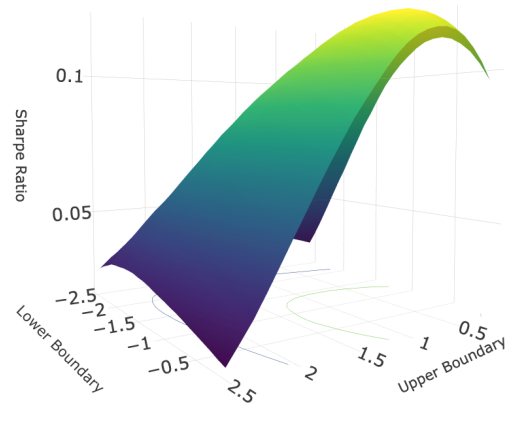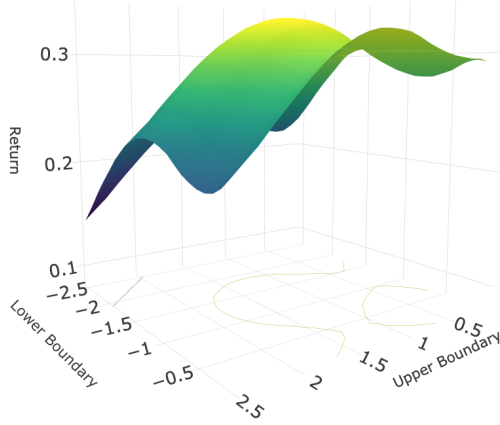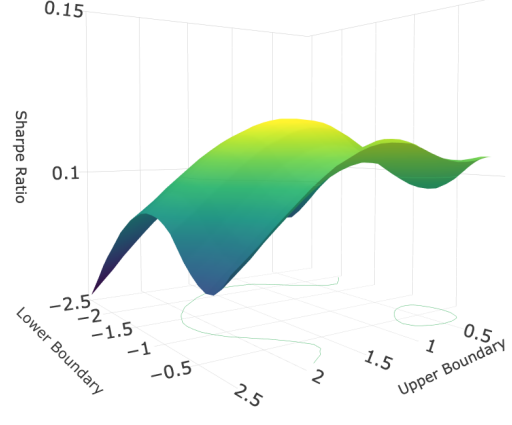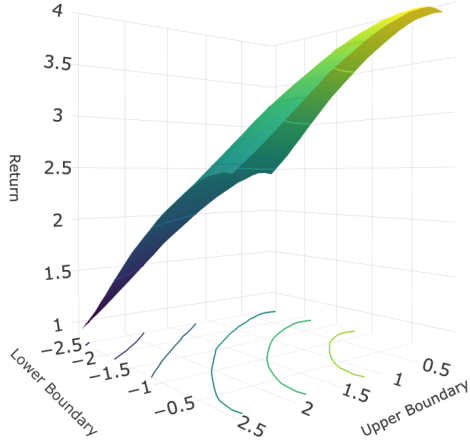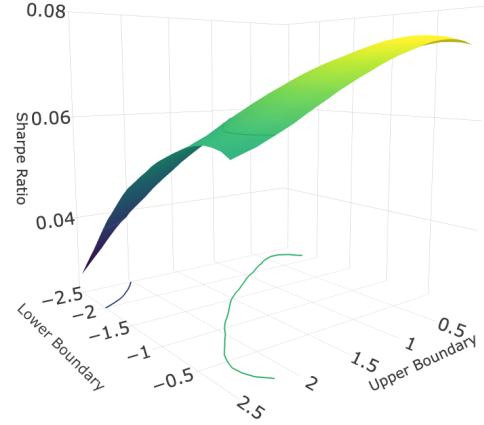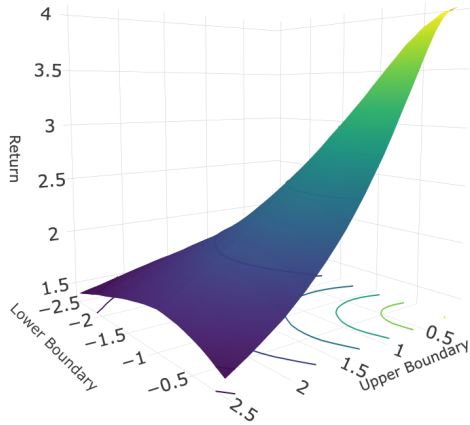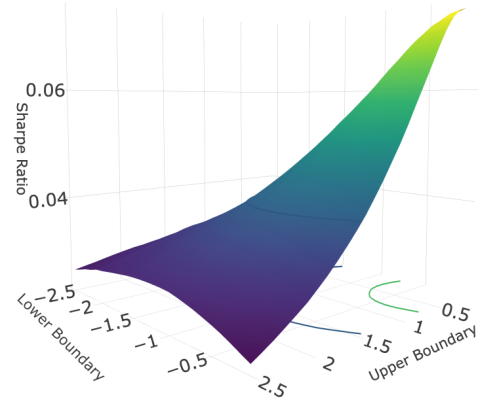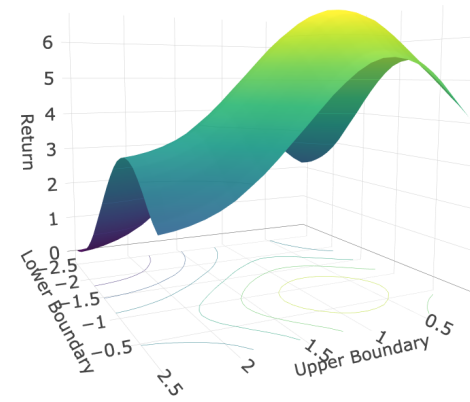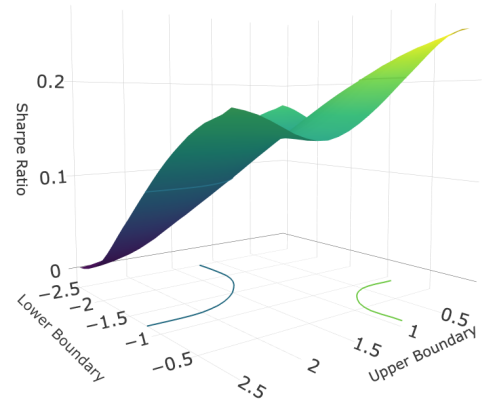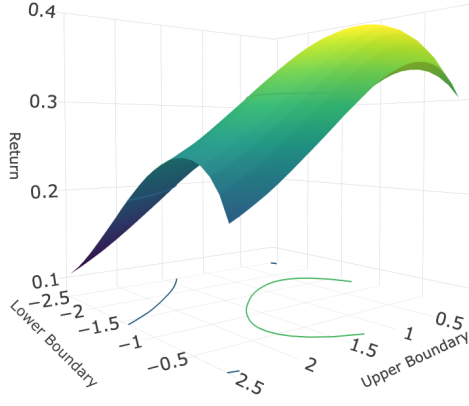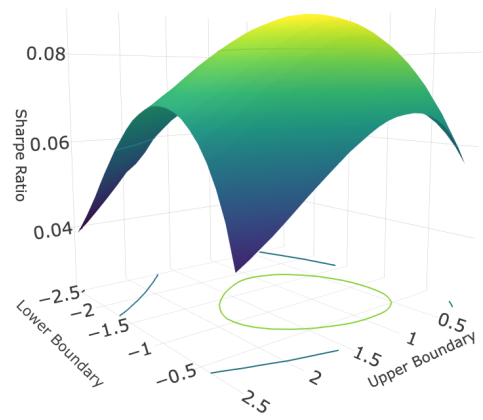Figure A2: Performance of Strategy A, B and C, based on Model 2

(a) Return of Strategy A, Model 2
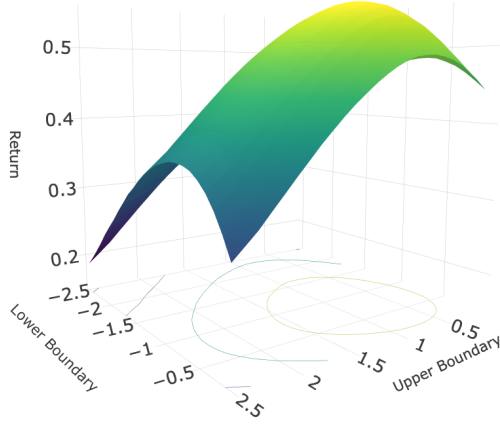
(b) Sharpe Ratio of Strategy A, Model 2

(c) Return of Strategy B, Model 2

(d) Sharpe Ratio of Strategy B, Model 2

(e) Return of Strategy C, Model 2

(f) Sharpe Ratio of Strategy C, Model 2

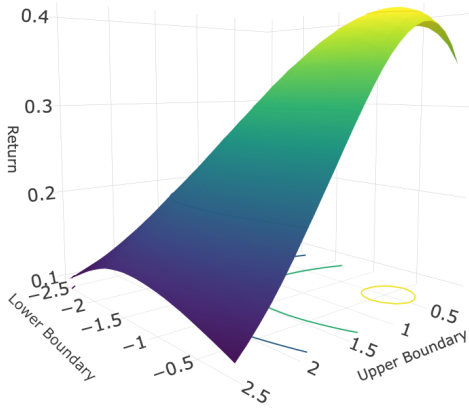Figure A3: Performance of Strategy A, B and C, based on Model 3

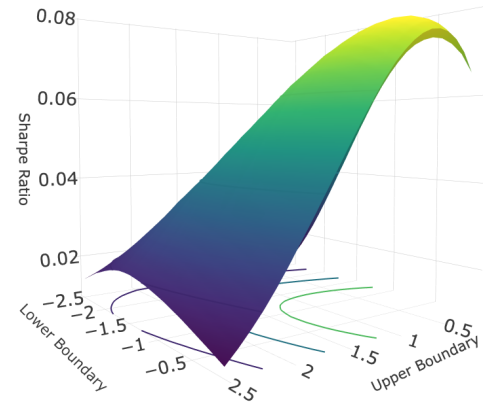(a) Return of Strategy A, Model 3

(b) Sharpe Ratio of Strategy A, Model 3

(c) Return of Strategy B, Model 3

(d) Sharpe Ratio of Strategy B, Model 3

(e) Return of Strategy C, Model 3

(f) Sharpe Ratio of Strategy C, Model 3

Figure A4: Performance of Strategy A, B and C, based on Model 4

(a) Return of Strategy A, Model 4

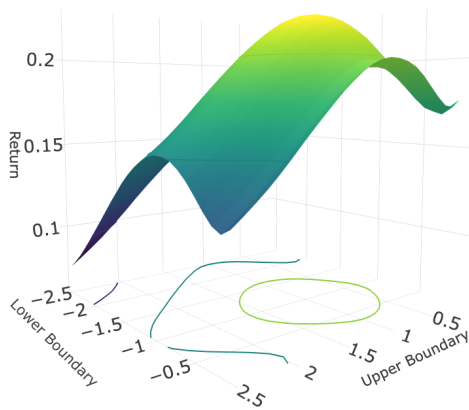(b) Sharpe Ratio of Strategy A, Model 4

(c) Return of Strategy B, Model 4

(d) Sharpe Ratio of Strategy B, Model 4

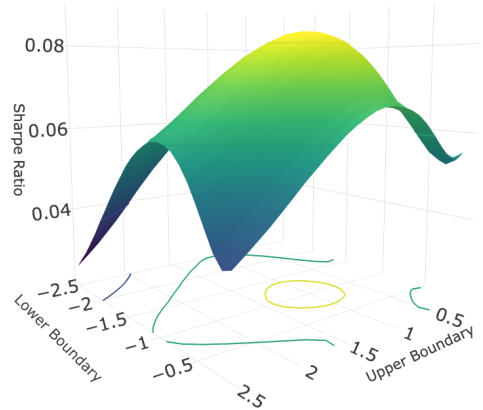(e) Return of Strategy C, Model 4

(f) Sharpe Ratio of Strategy C, Model 4

Figure A5: Performance of Strategy A, B and C, based on Model 5

(a) Return of Strategy A, Model 5

(b) Sharpe Ratio of Strategy A, Model 5

(c) Return of Strategy B, Model 5

(d) Sharpe Ratio of Strategy B, Model 5

(e) Return of Strategy C, Model 5

(f) Sharpe Ratio of Strategy C, Model 5

31