*ANZIAM J.* **63** (2021), 104–122
doi:10.1017/S1446181121000298

# ON OPTIMAL THRESHOLDS FOR PAIRS TRADING IN A ONE-DIMENSIONAL DIFFUSION MODEL

MASAAKI FUKASAWA[1], HITOMI MAEDA[1] and JUN SEKINE✉[1]

(Received 1 May, 2020; accepted 8 April, 2021; first published online 7 September, 2021)

**Abstract**

We study the static maximization of long-term averaged profit, when optimal preset thresholds are determined to describe a pairs trading strategy in a general one-dimensional ergodic diffusion model of a stochastic spread process. An explicit formula for the expected value of a certain first passage time is given, which is used to derive a simple equation for determining the optimal thresholds. Asymptotic arbitrage in the long run of the threshold strategy is observed.

2020 *Mathematics Subject Classification*: primary 60J70; secondary 91G15, 60J60.

*Keywords and phrases*: pairs trading, threshold rule, one-dimensional diffusion, first passage time, Pearson diffusion, long-time averaged profit, asymptotic arbitrage.

## 1. Introduction

In pairs trading, we focus on two securities whose spread of (log) prices is modelled as a mean-reverting stochastic process. When the spread widens, we make the higher security short and the cheaper one long. When the spread reverts to its mean, we clear the positions and make a profit. For a comprehensive review of studies on pairs trading, we can refer to Krauss [11] and the references therein, where various schemes of pairs trading are well categorized and explained. As for the formation period, that is, the period to find and identify the comovements of a pair of security price processes, the distance approach and the cointegration approach are introduced as two major ones. As for the trading period after a suitable pair of securities is selected in the formation period, several trading rules have been studied in the pairs trading literature. Among these, the so-called threshold rules are widely employed to trigger trading signals in constructing pairs trading strategies, and a one-dimensional Ornstein–Uhlenbeck (OU) process is popularly used as a simple tractable model for the mean-reverting

[1]Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan; e-mail: fukasawa@sigmath.es.osaka-u.ac.jp, u112299g@gmail.com, sekine@sigmath.es.osaka-u.ac.jp.

stochastic spread process. Zeng and Lee [14] studied a static optimization problem for determining "optimal" preset thresholds, where a series expansion formula for the expected value of a certain first passage time of an OU process is utilized (see Remark 2.7 in Section 2). A related study of the OU process model by Bertram [1] influenced Zeng and Lee's work. In the present paper, we present a unified approach to Zeng and Lee's problem for determining optimal thresholds for pairs trading with a *general* ergodic one-dimensional diffusion model for the stochastic spread process. Our results are summarized as follows.

(a) We present a general class of one-dimensional diffusion models (see Assumptions 2.1–4.2) which have symmetric stationary distributions. The class contains various tractable models: not only the OU process having a Gaussian stationary distribution, but also the Pearson diffusion process having a $t$-stationary distribution [7, 13], and the Jacobi diffusion process having a $\beta$-stationary distribution [7, 10].

(b) For the class of one-dimensional diffusion models stated in (a) above,

- an explicit, analytic formula for the expected value of a certain first passage time is derived (see Theorem 3.1 for details), and
- the static optimization problem for selecting the thresholds of pairs trading is solved, where the long-time averaged profit is used as the criterion function; a simple equation for the optimal thresholds, which involves a one-dimensional integral, is explicitly described (see Theorem 4.5 for details).

These analytically tractable results for general non-Gaussian models seem to be important, as the drawbacks of applying a Gaussian OU process model to non-Gaussian financial data have been pointed out by Bertram [1] and Krauss [11].

As evidence of the non-Gaussian nature of real financial time series data, we observe daily Nikkei 225 (Nikkei stock average) and TOPIX (Tokyo stock price index) data from the Japanese stock market from 1 June 2011 to 30 December 2020. Figure 1 shows price movements, Figure 2 shows log-price movements, Figure 3 shows the regression residual, log(Nikkei 225) − 1.144 × log(TOPIX) − 1.502 (the spread of log-prices), and Figure 4 shows the quantile–quantile plot (see Chambers et al. [3]) of the residual data, which seem to be non-Gaussian (with sample kurtosis 3.412061).

## 2. Model set-up

Consider two security price processes $(P_t)_{t\geq 0}$ and $(Q_t)_{t\geq 0}$ in a continuous-time economy. We may regard $P$ and $Q$ as discounted price processes for simplicity. Suppose that fir some constants $\beta, \lambda \in \mathbb{R}$,

$$X_t = \log P_t - \beta \log Q_t - \lambda, \quad t \geq 0, \tag{2.1}$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

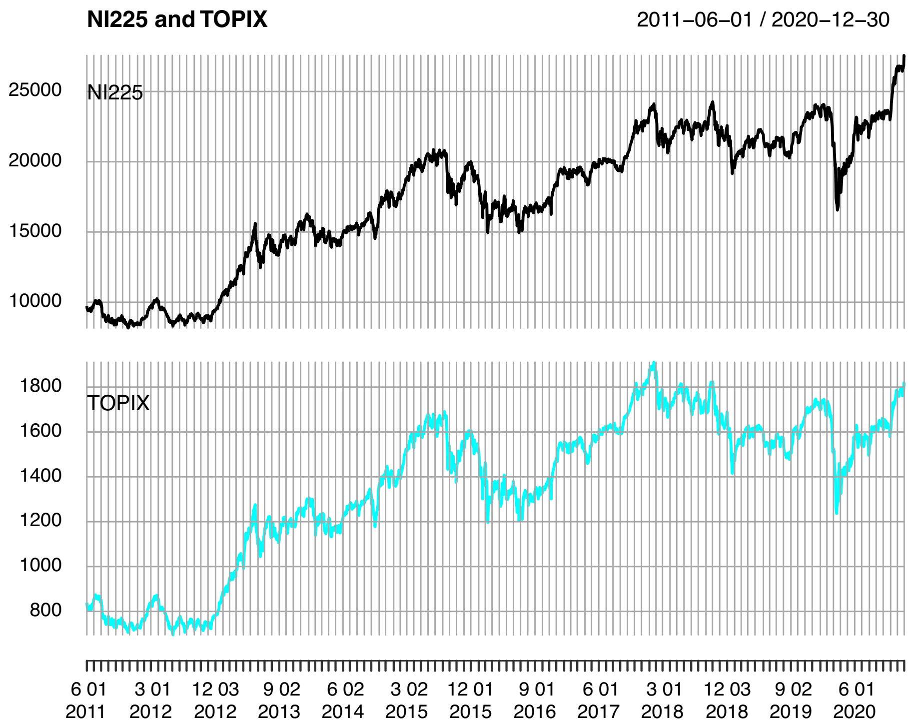


FIGURE 1. Nikkei 225 and TOPIX (2011–2020).

follows a one-dimensional diffusion process. Concretely, $X = (X_t)_{t\geq 0}$ satisfies the stochastic differential equation (SDE)

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geq 0, \quad X_0 \in E, \tag{2.2}$$

on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geq 0})$ endowed with the $\mathcal{F}_t$-Brownian motion $(W_t)_{t\geq 0}$, where $E = (l, r)$ $(-\infty \leq l < 0 < r \leq \infty)$ is the state space. Here, $\mu, \sigma : E \to \mathbb{R}$ are continuous functions and $\sigma(x)^2 > 0$ for all $x \in E$. Associated with $X$, we define the scale function

$$s(x) = \int_0^x \exp\Big\{-2\int_0^y \Big(\frac{\mu}{\sigma^2}\Big)(z)\,dz\Big\}\,dy, \quad x \in E,$$

and the speed measure on $E$ by

$$m(dx) = \frac{2}{\sigma(x)^2}\exp\Big\{2\int_0^x \Big(\frac{\mu}{\sigma^2}\Big)(y)\,dy\Big\}\,dx.$$

For the basic roles of the scale function and the speed measure in one-dimensional diffusion theory, we can refer to the works of Durret [4, Chapter 6] and Karatzas and Shreve [9, Section 5.5], for example. We make the following assumptions.

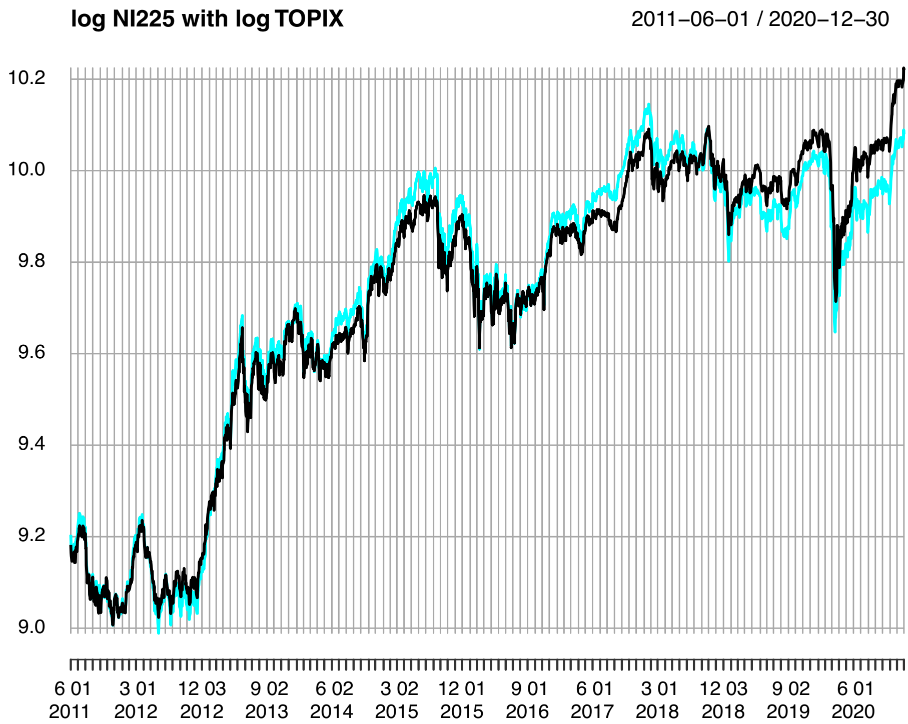


FIGURE 2. Log-prices: Nikkei 225 and TOPIX (2011–2020).

ASSUMPTION 2.1. The boundary values of the scale function are

$$\lim_{x\downarrow l} s(x) = -\infty, \qquad \lim_{x\uparrow r} s(x) = +\infty,$$

and $m(E) < \infty$.

ASSUMPTION 2.2. The boundaries of the state space are $-l = r > 0$. In the SDE (2.2), $\mu$ is an odd function and $\sigma^2$ is an even function, so, $\mu(x) = -\mu(-x)$ and $\sigma(x)^2 = \sigma(-x)^2$ for all $x \in E$.

REMARK 2.3. Assumption 2.1 implies that $X$ is recurrent and has the invariant probability measure, given by

$$\bar{m}(dx) = \frac{1}{m(E)} m(dx). \tag{2.3}$$

Hence, $X$ is ergodic, and its stationary distribution is given by (2.3). Moreover, we see that

$$\lim_{t\to\infty} \mathbb{P}_x(X_t < y) = \bar{m}((l, y)) \quad \text{for all } x, y \in E, \tag{2.4}$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

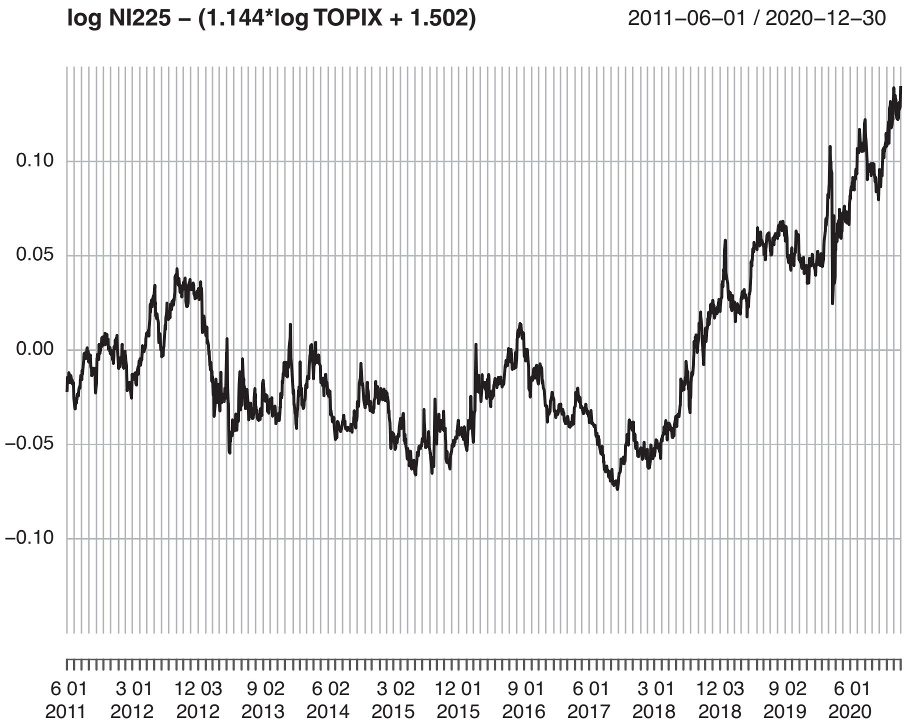


FIGURE 3. Residual of regression.

where $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$. This relation is seen in Karatzas and Shreve [9], which is an extension of a result obtained by Pollack and Siegmund [12]. On the other hand, Assumption 2.2 implies the symmetry of the stationary distribution $\bar{m}$ on $E$. Further, it implies that $s$ and $s''$ are odd functions and $s'$ is an even function.

Here are some examples which satisfy Assumptions 2.1 and 2.2.

EXAMPLE 2.4 (Ornstein–Uhlenbeck process). Let $\sigma \in \mathbb{R}_{++}$ be constant, let $\mu(x) = -\kappa\sigma^2 x$ with $\kappa \in \mathbb{R}_{++} := (0, \infty)$, and let $E = \mathbb{R}$. The associated process is written as

$$dX_t = -\kappa\sigma^2 X_t\, dt + \sigma\, dW_t. \tag{2.5}$$

In this case, we see that

$$s'(x) = e^{\kappa x^2}, \quad m(dx) = \frac{2}{\sigma^2} e^{-\kappa x^2}\, dx$$

and the stationary distribution is a centred Gaussian distribution.

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

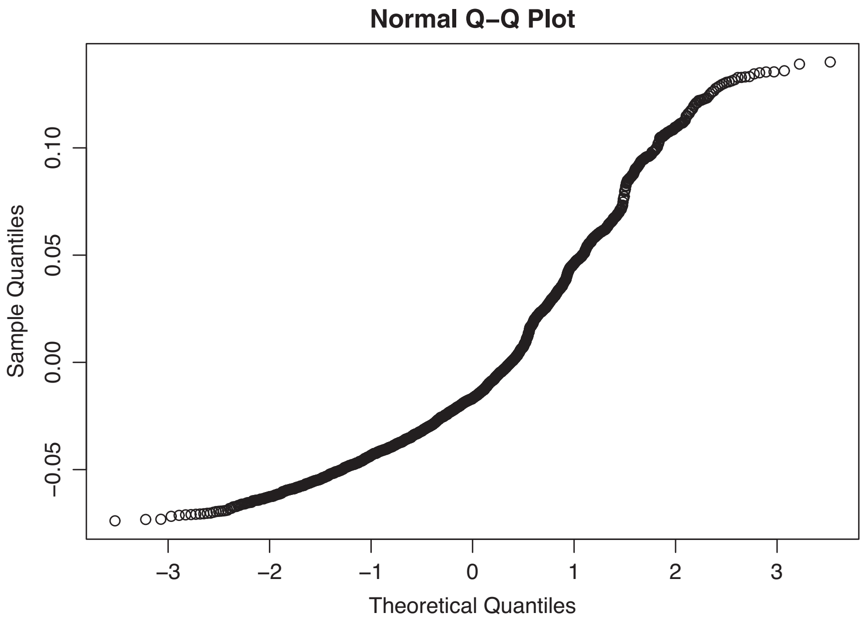


FIGURE 4. Quantile–quantile plot of residuals.

EXAMPLE 2.5 (Pearson diffusion). Let $\mu(x) = -\kappa\gamma^2 x$ and $\sigma(x) = \gamma\sqrt{\delta + x^2}$ with $\kappa, \gamma, \delta \in \mathbb{R}_{++}$, and let $E = \mathbb{R}$. The associated process is written as

$$dX_t = -\kappa\gamma^2 X_t\,dt + \gamma\sqrt{\delta + X_t^2}\,dW_t. \tag{2.6}$$

In this case, we see that

$$s'(x) = (\delta + x^2)^{\kappa}, \quad m(dx) = \frac{2}{\gamma^2}(\delta + x^2)^{-(1+\kappa)}\,dx,$$

and the stationary distribution is a scaled $t$-distribution.

EXAMPLE 2.6 (Jacobi diffusion). Let $\mu(x) = -\kappa\gamma^2 x$ and $\sigma(x) = \gamma\sqrt{\delta^2 - x^2}$ with $\kappa, \gamma, \delta \in \mathbb{R}_{++}$, and let $E = (-\delta, \delta)$. The associated process is written as

$$dX_t = -\kappa\gamma^2 X_t\,dt + \gamma\sqrt{\delta^2 - X_t^2}\,dW_t.$$

In this case, we see that

$$s'(x) = (\delta^2 - x^2)^{-\kappa}, \quad m(dx) = \frac{2}{\gamma^2}(\delta + x)^{\kappa-1}(\delta - x)^{\kappa-1}\,dx,$$

and the stationary distribution is a centred and scaled beta distribution.

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

REMARK 2.7. Forman and Sørensen [7] studied the parametrized one-dimensional diffusion process

$$dX_t = -\theta(X_t - \mu)\,dt + \sqrt{2\theta(aX_t^2 + bX_t + c)}\,dW_t,$$

with $\theta > 0$, $\mu, a, b, c \in \mathbb{R}$, which contains the above three examples, and shows the feasibility of explicit statistical inference of parameters.

## 3. Expected value formula for first passage time

For $y \in E$, denote the first hitting time by

$$\tau_y = \inf\{t \geq 0 \mid X_t = y\},$$

where we make the interpretation $\inf \emptyset = +\infty$, with $\emptyset$ being the empty set. Using the simplified notation $\mathbb{E}_x[(\cdot)] = \mathbb{E}[(\cdot) \mid X_0 = x]$, we obtain the following theorem.

THEOREM 3.1. *Under Assumptions 2.1 and 2.2, and for* $l < -\alpha < \beta < \alpha < r$,

$$\mathbb{E}_\alpha[\tau_\beta] + \mathbb{E}_\beta[\tau_{-\alpha} \wedge \tau_\alpha] = \frac{m(E)}{2}\{s(\alpha) - s(\beta)\}.$$

REMARK 3.2. For the OU process given in Example 2.4, Zeng and Lee [14] derived the series expansion formula

$$\mathbb{E}_\alpha[\tau_\beta] + \mathbb{E}_\beta[\tau_{-\alpha} \wedge \tau_\alpha] = \frac{1}{2\kappa\sigma^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)!} \times \{(2\sqrt{\kappa}\alpha)^{2n+1} - (2\sqrt{\kappa}\beta)^{2n+1}\}\Gamma\Big(\frac{2n+1}{2}\Big),$$

where we use the notation

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx \quad (z \in \mathbb{C},\ \Re(z) > 0)$$

for the gamma function. Using Theorem 3.1 for Example 2.4, we have the following different analytic representation:

$$\mathbb{E}_\alpha[\tau_\beta] + \mathbb{E}_\beta[\tau_{-\alpha} \wedge \tau_\alpha] = \frac{\sqrt{2\pi}}{2\kappa\sigma^2}\int_\beta^\alpha e^{\kappa x^2}\,dx.$$

PROOF. Let $l < a < x < b < r$. Recall that

$$\mathbb{E}_x[\tau_a \wedge \tau_b] = \int_E G_{a,b}(x, y)m(dy),$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

where we define the Green function by

$$G_{a,b}(x,y) = \begin{cases} \dfrac{(s(x)-s(a))(s(b)-s(y))}{s(b)-s(a)} & \text{if } a \le x \le y \le b, \\ \dfrac{(s(y)-s(a))(s(b)-s(x))}{s(b)-s(a)} & \text{if } a \le y \le x \le b, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

(see, for example, Durret [4, Chapter 6] or Karatzas and Shreve [9, Section 5.5]). Note that

$$\begin{aligned} \mathbb{E}_x[\tau_a \wedge \tau_b] = & \frac{s(x)-s(a)}{s(b)-s(a)} \int_x^b (s(b)-s(y))m(dy) \\ & + \frac{s(b)-s(x)}{s(b)-s(a)} \int_a^x (s(y)-s(a))m(dy). \end{aligned} \tag{3.2}$$

Letting $b \uparrow r$ and using Assumption 2.1 yields

$$\mathbb{E}_x[\tau_a] = (s(x)-s(a)) \int_x^r m(dy) + \int_a^x (s(y)-s(a))m(dy). \tag{3.3}$$

Then, combining (3.2) with $x = \beta$, $a = -\alpha$, $b = \alpha$ and (3.3) with $x = \alpha$, $a = \beta$, we have

$$\begin{aligned} &\mathbb{E}_\beta[\tau_{-\alpha} \wedge \tau_\alpha] + \mathbb{E}_\alpha[\tau_\beta] \\ &= \frac{s(\beta)-s(-\alpha)}{s(\alpha)-s(-\alpha)} \int_\beta^\alpha (s(\alpha)-s(y))m(dy) + \frac{s(\alpha)-s(\beta)}{s(\alpha)-s(-\alpha)} \int_{-\alpha}^\beta (s(y)-s(-\alpha))m(dy) \\ &\quad + (s(\alpha)-s(\beta)) \int_\alpha^r m(dy) + \int_\beta^\alpha (s(y)-s(\beta))m(dy) \\ &= \frac{s(\beta)+s(\alpha)}{2s(\alpha)} \int_\beta^\alpha (s(\alpha)-s(y))m(dy) + \frac{s(\alpha)-s(\beta)}{2s(\alpha)} \int_{-\alpha}^\beta (s(y)+s(\alpha))m(dy) \\ &\quad + (s(\alpha)-s(\beta)) \int_\alpha^r m(dy) + \int_\beta^\alpha (s(y)-s(\beta))m(dy) \\ &= \frac{s(\alpha)-s(\beta)}{2}\Big( \int_\beta^\alpha + 2\int_\alpha^r + \int_{-\alpha}^\beta \Big) m(dy) + \frac{s(\alpha)-s(\beta)}{2s(\alpha)}\Big( \int_{-\alpha}^\beta + \int_\beta^\alpha \Big) s(y)m(dy) \\ &= \frac{s(\alpha)-s(\beta)}{2} m(E), \end{aligned}$$

where we use the property that $s$ is an odd function, that is, $s(x) = -s(-x)$ for $x \in E$, and the symmetry of $m$, that is, $m(A) = m(-A)$ for $A \in \mathcal{B}(E)$, which follow from Assumption 2.2. This completes the proof. □

REMARK 3.3. Under Assumption 2.1 (without imposing Assumption 2.2), we have for $l < \beta < \alpha < r$,

$$\mathbb{E}_\alpha[\tau_\beta] + \mathbb{E}_\beta[\tau_\alpha] = m(E)\{s(\alpha) - s(\beta)\}. \tag{3.4}$$

This assertion follows by combining equation (3.2) with $x = \alpha$, $a = \beta$, and

$$\mathbb{E}_x[\tau_b] = \int_x^b (s(b) - s(y))m(dy) + (s(b) - s(x)) \int_l^x m(dy) \tag{3.5}$$

with $x = \beta$, $b = \alpha$. Equation (3.5) follows from (3.2), by letting $a \downarrow l$ and using Assumption 2.1. Note that formula (3.4) appeared in Karatzas and Shreve [9]. Also, Bertram [1] has derived this formula for the OU process given in Example 2.4.

## 4. Optimal thresholds for pairs trading

In this section, we apply Theorem 3.1 to compute optimal thresholds for pairs trading. Consider the situation where the stochastic spread process $X = (X_t)_{t\geq 0}$ (2.1) is governed by the one-dimensional SDE (2.2), and that Assumptions 2.1 and 2.2 are satisfied. First, we note that the SDE (2.2) has a weak solution, which is unique in the sense of probability law (for the definition of the uniqueness of the solution of SDE in the sense of probability law, see [9, Chapter 5]). For the proof, we refer to Karatzas and Shreve [9], where the case $E = \mathbb{R}$ was treated, which can be straightforwardly modified to apply to our situation with $E \subset \mathbb{R}$. Next, we note that

$$X^{(x)} \underset{\text{law}}{\equiv} -X^{(-x)},$$

where the superscript $(x)$ denotes the starting position of the process $X$, that is, $x = X_0^{(x)} \in E$, which follows from the uniqueness of the solution of SDE (2.2) and Assumption 2.2.

Now, inspired by Zeng and Lee [14], we consider the following trading strategy. Let $a > 0$ and $b \in [-a, a]$ be two preset thresholds, which determine the trading signals. We set $T_0 = 0$, and, for $n \in \mathbb{N}$, determine the starting time $S_n$ and the completion time $T_n$ of the $n$th trading cycle by

$$\begin{aligned} S_n &= \inf\{t > T_{n-1} \| X_t| = a\}, \\ T_n &= \inf\{t > S_n | X_t = \operatorname{sgn}(X_{S_n})b\}, \end{aligned}$$

respectively. The $n$th trading cycle consists of the following two steps.

(A) At time $S_n$, if $X_{S_n} = a$ (respectively, $X_{S_n} = -a$), set a one dollar short (respectively, long) position of security $P$ and a $\beta$ dollar long (respectively, short) position of security $Q$. Keep these positions until time $T_n$.

(B) At time $T_n$, clear the positions and make a profit. Wait until the next starting time $S_{n+1}$.

The profit of the $n$th trading cycle is (approximately) given by

$$\begin{aligned} -\operatorname{sgn}(X_{S_n})\Big(\frac{P_{T_n} - P_{S_n}}{P_{S_n}} - \beta\frac{Q_{T_n} - Q_{S_n}}{Q_{S_n}}\Big) - c &\approx -\operatorname{sgn}(X_{S_n})\Big(\log\frac{P_{T_n}}{P_{S_n}} - \beta\log\frac{Q_{T_n}}{Q_{S_n}}\Big) - c \\ &= -\operatorname{sgn}(X_{S_n})(X_{T_n} - X_{S_n}) - c = a - b - c, \end{aligned}$$

where $\operatorname{sgn}(x)$ denotes the sign of $x \in \mathbb{R}$, and $c > 0$ is the total transaction cost in the trading cycle. Here, recall that

$$U_n = T_n - T_{n-1}, \quad n \in \mathbb{N},$$

are independent, identically distributed random variables, which are deduced from the time-homogeneity of $X$ and the strong Markov property (the strong Markov property of the solution of the Markovian SDE can be found in Karatzas and Shreve [9, Chapter 5]). We then define

$$N_t = \sup\{n \in \mathbb{N} \mid T_n \le t\}, \quad t \ge 0,$$

which is a renewal process. The cumulative profit obtained from the completed trading until time $t$ is

$$(a - b - c)N_t,$$

and the long-time averaged profit is

$$\lim_{t\to\infty} \frac{1}{t}(a - b - c)N_t = \frac{a - b - c}{\mathbb{E}[U_1]} = L(a, b), \quad \text{almost surely,} \tag{4.1}$$

where we use the strong law of large numbers result in renewal theory [2, Chapter 10]. To determine the optimal threshold levels, we are interested in the maximization problem

$$\max_{(a,b)\in\mathcal{A}} L(a, b), \tag{4.2}$$

where

$$\mathcal{A} = \{(x, y) \in E^2 \mid -x \le y \le x - c\}, \tag{4.3}$$

and the objective function can be expressed as

$$L(a, b) = \frac{a - b - c}{\mathbb{E}_b[\tau_a \wedge \tau_{-a}] + \mathbb{E}_a[\tau_b]} = \frac{2(a - b - c)}{m(E)\{s(a) - s(b)\}} \tag{4.4}$$

by Theorem 3.1. To solve the problem expressed by (4.2)–(4.4), we impose the following conditions.

ASSUMPTION 4.1. For all $x \in E \setminus \{0\}$, $x\mu(x) < 0$.

ASSUMPTION 4.2. In the case where $r = \infty$, $\lim_{x\uparrow r} s'(x) = \infty$.

REMARK 4.3. Assumption 4.1 implies that $X$ is mean-reverting: the drift $\mu$ of $X$ is always directed "inward". It also implies that for all $x \in (0, r)$,

$$s''(x) = -2\Big(\frac{\mu}{\sigma^2}\Big)(x)s'(x) = -s''(-x) > 0,$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

where we recall that for all $x \in E$,

$$s'(x) = \exp\left\{ -2 \int_0^x \left(\frac{\mu}{\sigma^2}\right)(y)\, dy \right\} = s'(-x) > 0,$$

which follow from Assumption 2.2 and Remark 2.3. Assumption 4.2 ensures that $L(a, b) \to 0$ as $a \uparrow r(= \infty)$; the long-time averaged profit becomes small if the threshold $a$ is too large. The details can be found in the proof of Theorem 4.5.

REMARK 4.4. It is easy to check that Examples 2.4–2.6 satisfy Assumptions 2.1–4.2.

We obtain the following theorem concerning the solution of the maximization problem (4.2).

THEOREM 4.5. *Suppose that Assumptions 2.1–4.2 are satisfied, and let*

$$h(a) = s(a) - \left(a - \frac{c}{2}\right)s'(a), \quad a \geq 0.$$

*Then there exists a unique constant $a^* > 0$ such that*

$$h(a^*) = 0 \quad \textit{and} \quad a^* \in \left(\frac{c}{2}, r\right),$$

*and it defines the maximizer for (4.2) with (4.3) and (4.4) as follows:*

$$\max_{(a,b)\in\mathcal{A}} L(a, b) = L(a^*, -a^*) = \frac{2(a^* - c/2)}{m(E)s(a^*)} = \frac{2}{m(E)s'(a^*)}.$$

PROOF. Let

$$f(a, b) = \frac{a - b - c}{s(a) - s(b)}.$$

We see that

$$\partial_a f(a, b) = \frac{\{s(a) - s(b)\} - (a - b - c)s'(a)}{\{s(a) - s(b)\}^2},$$

$$\partial_b f(a, b) = \frac{-\{s(a) - s(b)\} + (a - b - c)s'(b)}{\{s(a) - s(b)\}^2},$$

and $\partial_b f(a, b) \leq 0$ for any $(a, b) \in \mathcal{A}_+ = \{(a, b) \in \mathcal{A} \mid b \geq 0\}$. Indeed, the inequality

$$-\{s(a) - s(b)\} + (a - b - c)s'(b) \leq -cs'(b) \leq 0$$

holds for $(a, b) \in \mathcal{A}_+$ as $s$ is convex on $(0, r)$. Hence, we deduce that

$$\max_{(a,b)\in\mathcal{A}} f(a, b) = \max_{(a,b)\in\mathcal{A}_-} f(a, b), \tag{4.5}$$

where $\mathcal{A}_- = \{(a, b) \in \mathcal{A} \mid b \leq 0\}$. Then we check the following.

(i) For $(a, b) \in \mathcal{A}_-$, we see that

$$f(a, b) \leq \frac{2a - c}{s(a)} = \bar{f}(a),$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

as $s$ is increasing. Further, we see that $\lim_{a\uparrow r}\bar{f}(a) = 0$ by Assumption 2.1 (if $r < \infty$) and l'Hôpital's rule combined with Assumption 4.2 (if $r = \infty$).

(ii) We compute the boundary values of $f$ on $\mathcal{A}_-$ as

(a) $f(a,b) = f(a, a-c) = 0$ on $\{(a,b) \in \mathcal{A} \mid b = a - c, a \geq c/2\}$.
(b) $f(a,b) = f(a,0) = (a-c)/s(a)$ on $\{(a,b) \in \mathcal{A} \mid b = 0, a \geq c\}$.
(c) $f(a,b) = f(a,-a) = (a-c/2)/s(a)$ on $\{(a,b) \in E^2 \mid b = -a, a \geq c/2\}$.

Hence, we deduce that the maximum of $f(a,b)$ on $\mathcal{A}_-$ exists, and that the maximizer(s) exist(s) in $\mathcal{A}_0 \cup \partial_1 A$, where

$$\mathcal{A}_0 = \{(a,b) \in \mathcal{A}_- \mid \partial_a f(a,b) = \partial_b f(a,b) = 0\},$$
$$\partial_1 \mathcal{A} = \left\{(a,b) \in E^2 \;\middle|\; b = -a, a \geq \frac{c}{2}\right\}.$$

Here, we see that $\mathcal{A}_0 \subset \partial_1 A$. Indeed, $\partial_a f(a,b) = \partial_b f(a,b) = 0$ for $(a,b) \in \mathcal{A}_-$ implies

$$0 = \{s(a) - s(b)\} - (a-b-c)s'(a) = -\{s(a) - s(b)\} + (a-b-c)s'(b),$$

from which we deduce the relations $s'(a) - s'(b) = 0$ and $b = -a$. Now, to compute the maximizer of

$$g(a) = f(a,-a) = \frac{1}{s(a)}\left(a - \frac{c}{2}\right)$$

on $\partial_1\mathcal{A} (= \mathcal{A}_0 \cup \partial_1 A)$, we first note the results

$$g\left(\frac{c}{2}\right) = 0 \quad \text{and} \quad \lim_{a\uparrow r} g(a) = 0, \tag{4.6}$$

where the latter follows from $g(a) \leq \bar{f}(a) \to 0$ as $a \uparrow r$ as we have seen. We next compute the derivative as

$$g'(a) = \frac{h(a)}{s(a)^2}.$$

Here, note that $g'(c/2) = 1/s(c/2) > 0$ for $c > 0$ and that $h'(a) = -(a - c/2)s''(a) < 0$ for $a > c/2$. So, combining these observations with (4.6), we deduce that there exists a unique $a^* \in (c/2, r)$ such that $h(a^*) = 0 = g'(a^*)$, and we have that $g'(a) > 0$ for $a \in (c/2, a^*)$ and $g'(a) < 0$ for $a \in (a^*, r)$. Therefore, recalling (4.5), we conclude that

$$\max_{a \geq c/2} g(a) = g(a^*) = f(a^*, -a^*) = \max_{(a,b)\in\mathcal{A}} f(a,b). \qquad \square$$

REMARK 4.6. The resulting trading strategy with the optimal thresholds $(a^*, -a^*)$ is called the new optimal rule (NOR) by Zeng and Lee [14]. Step (B) of this strategy has no waiting time, which is different from that of the conventional trading rule with thresholds $0 = b < a$. Thus, one trading cycle of the NOR is described as follows.

(A) When $X_{t_1} = a^*$ (respectively, $X_{t_1} = -a^*$), set a one dollar short (respectively, long) position of security $P$ and a $\beta$ dollar long (respectively, short) position of security $Q$. Keep these positions as long as $X_t > -a^*$ (respectively, $X_t < a^*$).

(B) When $X_{t_2} = -a^*$ (respectively, $X_{t_2} = a^*$) ( $t_2 > t_1$), clear the positions and make a profit. Restart immediately with step (A).

## 5. Asymptotic arbitrage

Pairs trading is sometimes explored in the context of so-called "statistical arbitrage", as it is sometimes executed algorithmically at high frequency. In this section, we consider the asymptotic arbitrage property of the threshold strategy, which is a weak form of arbitrage in the long run. Let us consider the threshold strategy described in Section 4. The cumulative gain of the (self-financing) threshold strategy until time $t$ is given by

$$\sum_{j=1}^{N_t} -\mathrm{sgn}(X_{S_j})\left(\frac{P_{T_j} - P_{S_j}}{P_{S_j}} - \beta\frac{Q_{T_j} - Q_{S_j}}{Q_{S_j}}\right) - cN_t$$
$$+ \left[-\mathrm{sgn}(X_{S_{N_t+1}})\left(\frac{P_t - P_{S_{N_t+1}}}{P_{S_{N_t+1}}} - \beta\frac{Q_t - Q_{S_{N_t+1}}}{Q_{S_{N_t+1}}}\right) - c\right]1_{\{S_{N_t+1}\le t<T_{N_t+1}\}},$$

where we regard $P$ and $Q$ as discounted price processes to cancel the interest rate effect. Here, we make the interpretation that

$$S_{N_t+1} = \sum_{j=0}^{\infty} S_{j+1}1_{\{N_t=j\}} \quad \text{and} \quad X_{S_{N_t+1}} = \sum_{j=0}^{\infty} X_{S_{j+1}}1_{\{N_t=j\}},$$

for example. As we have seen in Section 4, we can approximate the cumulative gain

$$\begin{aligned} G_t &= \sum_{j=1}^{N_t} -\mathrm{sgn}(X_{S_j})\left(\log\frac{P_{T_j}}{P_{S_j}} - \beta\log\frac{Q_{T_j}}{Q_{S_j}}\right) - cN_t \\ &\quad + \left[-\mathrm{sgn}(X_{S_{N_t+1}})\left(\log\frac{P_t}{P_{S_{N_t+1}}} - \beta\log\frac{Q_t}{Q_{S_{N_t+1}}}\right) - c\right]1_{\{S_{N_t+1}\le t<T_{N_t+1}\}} \\ &= (a-b-c)N_t + [-\mathrm{sgn}(X_{S_{N_t+1}})(X_t - a) - c]1_{\{S_{N_t+1}\le t<T_{N_t+1}\}}. \end{aligned} \tag{5.1}$$

REMARK 5.1. On $\{T_{N_t} \le t < S_{N_t+1}\}$,

$$G_t = (a-b-c)N_t \ge 0,$$

which implies that the cumulative gain is always nonnegative and constant, until the new $(N_t+1)$th trading cycle starts at time $S_{N_t+1}$ after the previous $N_t$th trading cycle completes. On the other hand, note that, on $\{S_{N_t+1} \le t < T_{N_t+1}\}$,

$$G_t = (a-b-c)N_t - \mathrm{sgn}(X_{S_{N_t+1}})(X_t - a) - c,$$

which implies the cumulative gain can be negative, and hence the threshold strategy becomes risky until the new $(N_t + 1)$th trading cycle completes after starting at time $S_{N_t+1}$.

We recall the notion of asymptotic arbitrage in the long run, which was introduced by Föllmer and Schachermayer [6].

DEFINITION 5.2 (Asymptotic arbitrage). The cumulative gain process $(G_t)_{t\geq 0}$ realizes a strong asymptotic arbitrage (SAA) in the long run if it satisfies the following conditions:

**(AA1)** $G_0 = 0$;

**(AA2)** for any $\epsilon > 0$ sufficiently small, there exists $T > 0$ such that

$$G_T > -\epsilon \text{ almost surely} \quad \text{and} \quad \mathbb{P}(G_T > \epsilon^{-1}) \geq 1 - \epsilon.$$

Let us impose the following condition.

ASSUMPTION 5.3. $\int_E |x| m(dx) < \infty.$

We then obtain the following result.

PROPOSITION 5.4 *Suppose Assumptions 2.1, 2.2 and 5.3 hold and let $a - b - c > 0$. Then the cumulative gain process $(G_t)_{t\geq 0}$ given by (5.1) satisfies*

$$\lim_{t\to\infty} \frac{G_t}{t} = L(a, b) \quad \text{in } L^1(\mathbb{P}),$$

*where $L(a, b)$ is given by (4.1), and the existence of an SAA follows.*

PROOF. First, recall the relation

$$|G_t - (a - b - c)N_t| \leq |X_t| + a + c \quad \text{for almost every } (t, \omega) \in \mathbb{R}_+ \times \Omega. \tag{5.2}$$

Further, we see that

$$\lim_{t\to\infty} \mathbb{E}_x[|X_t|] = \int_E |y| \bar{m}(dy) < \infty \tag{5.3}$$

for any $x \in E$, where we use Assumption 5.3 and relation (2.4). Combining (5.2), (5.3) and the law of large numbers for the renewal process $(N_t)_{t\geq 0}$, we deduce that

$$\lim_{t\to\infty} \mathbb{E}\Big[\Big|\frac{G_t}{t} - L(a, b)\Big|\Big] = 0.$$

So, for any $\delta > 0$,

$$\lim_{t\to\infty} \mathbb{P}\Big(\Big|\frac{G_t}{t} - L(a, b)\Big| \geq \delta\Big) = 0$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

follows from Markov's inequality. Further, for any $\delta, \epsilon > 0$ sufficiently small, there exists $T_* > 0$ such that

$$\mathbb{P}(G_T \le \{L(a,b) - \delta\}T) < \epsilon \quad \text{for all } T \ge T_*.$$

Hence, for $T \ge \max(T_*, (\{L(a,b) - \delta\}\epsilon)^{-1})$,

$$\mathbb{P}(G_T > \epsilon^{-1}) = 1 - \mathbb{P}(G_T \le \epsilon^{-1}) \ge 1 - \mathbb{P}(G_T \le \{L(a,b) - \delta\}T) \ge 1 - \epsilon,$$

and an SAA is realized. □

REMARK 5.5 (Statistical arbitrage). Hogan et al. [8] used the following definition. If the cumulative gain process $(G_t)_{t\ge0}$ satisfies the following four conditions, then we say that a statistical arbitrage (SA) opportunity exists:

- **(SA1)** $G_0 = 0$;
- **(SA2)** $\lim_{t\to\infty} \mathbb{E}[G_t] > 0$;
- **(SA3)** $\lim_{t\to\infty} \mathbb{P}(G_t < 0) = 0$;
- **(SA4)** $\lim_{t\to\infty}(1/t)\mathbb{V}[G_t] = 0 \quad \text{if } \mathbb{P}(G_t < 0) > 0, \quad \text{for all } t \ge 0.$

We conjecture that an SA does not exist in the cumulative gain process (5.1) of the threshold strategy. Indeed, (SA4) seems to be violated, if we recall the central limit theorem for the renewal process $(N_t)_{t\ge0}$ discussed in Section 7.

## 6. Numerical experiment

In this section, using two examples of the stochastic spread processes, that is, Examples 2.4 and 2.5, we show some numerical experiments. Recall that the OU process (2.5) has the stationary distribution

$$\bar{m}(dx) = \frac{m(dx)}{m(\mathbb{R})} = \sqrt{\frac{\kappa}{\pi}} e^{-\kappa x^2}\, dx \ \sim N\Big(0, \frac{1}{2\kappa}\Big),$$

the centered normal distribution with variance $1/2\kappa$, and that Pearson diffusion process (2.6) has the stationary distribution,

$$\bar{m}(dx) = \frac{m(dx)}{m(\mathbb{R})} = \frac{\Gamma(\kappa+1)}{\sqrt{\delta\pi}\,\Gamma(\kappa+1/2)}\Big(1 + \frac{x^2}{\delta}\Big)^{-(\kappa+1)} dx,$$

which is a scaled $t$-distribution; concretely, we have

$$\lim_{t\to\infty} \mathbb{P}\Big(\sqrt{\frac{2\kappa+1}{\delta}} X_t \in dx\Big) \sim f_{2\kappa+1}(x)\, dx$$

with

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\Big(1 + \frac{x^2}{\nu}\Big)^{-(\nu+1)/2}.$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

Hence, the limiting variance and kurtosis are given by

$$\begin{aligned} V^{\infty} &= \lim_{t\to\infty} \mathbb{V}[X_t] = \frac{\delta}{2\kappa - 1}, \\ K^{\infty} &= \lim_{t\to\infty} \mathbb{K}[X_t] = \frac{6}{2\kappa - 3}, \end{aligned} \tag{6.1}$$

respectively, where $\mathbb{V}[(\cdot)]$ denotes the variance and $\mathbb{K}[(\cdot)]$ denotes the kurtosis.

For the optimization problem studied in Section 4, we consider the following numerical experiment for Pearson diffusion model (2.6).

(i) Set $\delta = (0.2)^2$, $\gamma = 1$, and choose several values of $\kappa$, which control the limiting variance and kurtosis (6.1) of the $t$-stationary distribution.

(ii) For each parameter set, numerically compute the optimal threshold value $a_{\mathrm{P}}^*$ and the optimal long-time averaged profit

$$L_{\mathrm{P}}^* = L_{\mathrm{P}}(a_{\mathrm{P}}^*, -a_{\mathrm{P}}^*),$$

given in Theorem 4.5, where $L_{\mathrm{P}}(a, b)$ is given by (4.4) for the Pearson diffusion model with $c = 0.01$.

Further, as a comparison, we consider the following numerical experiment for the OU process model (2.5).

(iii) Set the limiting variance $1/2\kappa_{\mathrm{OU}}$ of the OU process equal to the limiting variance $V^{\infty} = \delta/(2\kappa - 1)$ of the Pearson diffusion process, to solve for the parameter value $\kappa_{\mathrm{OU}}$:

$$\frac{1}{2\kappa_{\mathrm{OU}}} = \frac{\delta}{2\kappa - 1} \quad \Leftrightarrow \quad \kappa_{\mathrm{OU}} = \frac{2\kappa - 1}{2\delta}.$$

Set $\sigma = 0.2$.

(iv) For each corresponding parameter set, numerically compute the optimal threshold value $a_{\mathrm{OU}}^*$ and the optimal long-time averaged profit

$$L_{\mathrm{OU}}^* = L_{\mathrm{OU}}(a_{\mathrm{OU}}^*, -a_{\mathrm{OU}}^*),$$

given in Theorem 4.5, where $L_{\mathrm{OU}}(a, b)$ is given by (4.4) for the OU model with $c = 0.01$.

(v) Compute

$$L_{\mathrm{mis}} = L_{\mathrm{P}}(a_{\mathrm{OU}}^*, -a_{\mathrm{OU}}^*),$$

which is interpreted as the long-time expected profit for a “misspecified” agent who observes the limiting variance and “misapplies” OU model (2.5) instead of Pearson model (2.6). Also, compute the loss rate,

$$\mathrm{Loss} = \frac{L_{\mathrm{P}}^* - L_{\mathrm{mis}}}{L_{\mathrm{P}}^*}.$$

https://doi.org/10.1017/S1446181121000298 Published online by Cambridge University Press

TABLE 1. Pearson model result and comparison with OU model.

| $\kappa$ | $\sqrt{V^\infty}$ | $K^\infty$ | $a^*_{\mathrm{P}}$ | $L^*_{\mathrm{P}}$ | $a^*_{\mathrm{OU}}$ | $L^*_{\mathrm{OU}}$ | $L_{\mathrm{mis}}$ | Loss |
|---|---|---|---|---|---|---|---|---|
| 1.6 | 0.1348 | 30.00 | 0.059 25 | 0.1347 | 0.065 88 | 0.1050 | 0.1345 | 1.634e-03 |
| 1.7 | 0.1291 | 15.00 | 0.058 09 | 0.1379 | 0.064 03 | 0.1093 | 0.1377 | 1.427e-03 |
| 1.8 | 0.1240 | 10.00 | 0.056 97 | 0.1409 | 0.062 37 | 0.1134 | 0.1407 | 1.259e-03 |
| 1.9 | 0.1195 | 7.500 | 0.055 93 | 0.1439 | 0.060 87 | 0.1173 | 0.1437 | 1.120e-03 |
| 2.0 | 0.1155 | 6.000 | 0.054 96 | 0.1468 | 0.059 51 | 0.1210 | 0.1466 | 1.003e-03 |
| 2.1 | 0.1118 | 5.000 | 0.054 07 | 0.1496 | 0.058 27 | 0.1246 | 0.1494 | 9.039e-04 |
| 2.2 | 0.1085 | 4.286 | 0.053 23 | 0.1523 | 0.057 13 | 0.1281 | 0.1522 | 8.194e-04 |
| 2.3 | 0.1054 | 3.750 | 0.052 58 | 0.1550 | 0.056 07 | 0.1314 | 0.1549 | 7.454e-04 |
| 2.4 | 0.1026 | 3.333 | 0.051 76 | 0.1576 | 0.055 08 | 0.1347 | 0.1575 | 6.829e-04 |
| 2.5 | 0.1000 | 3.000 | 0.051 02 | 0.1602 | 0.054 17 | 0.1378 | 0.1601 | 6.278e-04 |
| 2.6 | 0.097 57 | 2.727 | 0.050 33 | 0.1627 | 0.053 31 | 0.1409 | 0.1626 | 5.791e-04 |
| 2.7 | 0.095 35 | 2.500 | 0.049 70 | 0.1651 | 0.052 51 | 0.1438 | 0.1651 | 5.360e-04 |
| 2.8 | 0.093 25 | 2.308 | 0.049 12 | 0.1676 | 0.051 75 | 0.1467 | 0.1675 | 4.976e-04 |
| 2.9 | 0.091 29 | 2.143 | 0.048 58 | 0.1699 | 0.051 04 | 0.1495 | 0.1698 | 4.631e-04 |
| 3.0 | 0.089 44 | 2.000 | 0.048 08 | 0.1722 | 0.050 36 | 0.1523 | 0.1722 | 4.319e-04 |

From the result (Table 1) we see that $L^*_{\mathrm{P}} > L^*_{\mathrm{OU}}$ always holds though a model misspecification in the profit is rather small (as we see in the loss rate, Loss). Also, we see that higher mean-reversion (with larger $\kappa$ and $\kappa_{\mathrm{OU}}$) yields higher optimized profits.

## 7. Concluding remarks

The long-time maximization of profit for pairs trading with thresholds, discussed in Section 4, is an idealized problem, based on the law of large numbers in the long run:

$$\lim_{t\to\infty} \frac{N_t}{t} = \lim_{t\to\infty} \frac{\mathbb{E}[N_t]}{t} = \frac{1}{\mathbb{E}[U_1]}.$$

To consider a suitable "mean–variance" optimization of profit is an interesting and important challenge as the fluctuation from the mean value seems to have a considerable effect in realistic situations with a finite time horizon. One such mean–variance optimization is to consider the criterion function

$$\mathrm{MV}(a, b) = L(a, b) - \alpha V(a, b),$$

where

$$L(a, b) = \lim_{t\to\infty} \frac{\mathbb{E}[(a - b - c)N_t]}{t} = \frac{(a - b - c)}{\mathbb{E}[U_1]} = \frac{2(a - b - c)}{m(E)\{s(a) - s(b)\}}$$

is the long-time limit of the mean-value of profit, which was analysed in Section 4. Note that

$$V(a,b) = \lim_{t\to\infty} \frac{\mathbb{V}[(a-b-c)N_t]}{t} = \frac{(a-b-c)^2\mathbb{V}[U_1]}{\mathbb{E}[U_1]^3} \tag{7.1}$$

is the long-time limit of the variance of profit, and $\alpha > 0$ is the risk-aversion parameter. The limiting variance (7.1) is obtained from the central limit theorem

$$\frac{\sqrt{t}\Big(\dfrac{N_t}{t} - \dfrac{1}{\mathbb{E}[U_1]}\Big)}{\sqrt{\dfrac{\mathbb{V}[U_1]}{\mathbb{E}[U_1]^3}}} \Rightarrow N(0,1)$$

for the scaled and centred renewal process (see Borovkov [2, Chapter 10]). Defining $\mathbb{V}_x[(\cdot)] = \mathbb{E}_x[(\cdot)^2] - \{\mathbb{E}_x[(\cdot)]\}^2$, the variance $\mathbb{V}[U_1]$ in (7.1), rewritten as

$$\mathbb{V}[U_1] = \mathbb{V}_a[\tau_b] + \mathbb{V}_b[\tau_a \wedge \tau_{-a}],$$

can be computed by using Kac's moment formula [5],

$$\mathbb{E}_x[(\tau_\alpha \wedge \tau_\beta)^2] = 2\int_E G_{\alpha,\beta}(x,y)\mathbb{E}_y[\tau_\alpha \wedge \tau_\beta]m(dy),$$

where we use the Green function (3.1), combined with the expected value formula (3.2). Although we have an analytic representation of $V(a,b)$ in closed form, the maximization of $\mathrm{MV}(a,b)$ does not seem to be straightforward, so this is left as a future research topic. For a different but related optimization for determining thresholds to trigger trading signals, we refer the reader to the Sharpe ratio maximization problem, discussed by Bertram [1].

## Acknowledgments

The authors are grateful to Chiaki Hara for fruitful discussions, and to the anonymous referees for helpful feedback leading to the improvements of this paper. Masaaki Fukasawa's research is supported by a Grant-in-Aid for Scientific Research (A), no. 25245046, from the Japan Society for the Promotion of Science. Jun Sekine's research is supported by a Grant-in-Aid for Scientific Research (C), no. 15K03540, from the Japan Society for the Promotion of Science.

## References

[1] W. K. Bertram, "Analytic solutions for optimal statistical arbitrage trading", *Phys. A* **389** (2010) 2234–2243; doi:10.1016/j.physa.2010.01.045.

[2] A. A. Borovkov, *Probability theory*, translation from the 5th Russian language edition, *Universitext,* (Springer-Verlag, London, 2013); doi:10.1007/978-1-4471-5201-9.

[3] J. Chambers, W. Cleveland, B. Kleiner and P. Tukey, *Graphical methods for data analysis* (Chapman and Hall/CRC, Boca Raton, FL, 2017); doi:10.1201/9781351072304.

[4] R. Durret, *Stochastic calculus: a practical introduction, Probab. Stoch. Ser.*, 1st edn (CRC Press, Boca Raton, FL, 1996); doi:10.1201/9780203738283.

[5] P. J. Fitzsimmons and J. Pitman, "Kac's moment formula and the Feynman–Kac formula for additive functionals of a Markov process", *Stochastic Process. Appl.* **79** (1999) 117–134; doi:10.1016/S0304-4149(98)00081-7.

[6] H. Föllmer and W. Schachermayer, "Asymptotic arbitrage and large deviations", *Math. Finance Econ.* **1** (2008) 213–249; doi:10.1007/s11579-008-0009-3.

[7] J. L. Forman and M. Sørensen, "The Pearson diffusions: a class of statistically tractable diffusion processes", *Scand. J. Stat.* **35** (2008) 438–465; doi:10.1111/j.1467-9469.2007.00592.x.

[8] S. Hogan, R. Jarrow, M. Teo and M. Warachka, "Testing market efficiency using statistical arbitrage with applications to momentum and value trading strategies", *J. Finance Econ.* **73** (2004) 525–565; doi:10.1016/j.jfineco.2003.10.004.

[9] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, 2nd edn (Springer, New York, 1998); doi:10.1007/978-1-4612-0949-2.

[10] S. Karlin and H. M. Taylor, *A second course in stochastic processes* (Academic Press, San Diego, CA, 1981).

[11] C. Krauss, "Statistical arbitrage pairs trading strategies: review and outlook", *J. Econ. Surv.* **31** (2017) 513–545; doi:10.1111/joes.12153.

[12] M. Pollak and D. Siegmund, "A diffusion process and its applications to detecting a change in the drift of Brownian motion", *Biometrika* **72** (1985) 267–280; doi:10.1093/biomet/72.2.267.

[13] E. Wong, "The construction of a class of stationary Markoff processes", in: *Stochastic processes in mathematical physics and engineering* (ed R. Bellman) (American Mathematical Society, Providence, RI, 1964) 264–276; doi:10.1090/psapm/016/0161375.

[14] Z. Zeng and C.-G. Lee, "Pairs trading: optimal thresholds and profitability", *Quant. Finance* **14** (2014) 1881–1893; doi:10.1080/14697688.2014.917806.