# Portfolio Optimization in the Context of Cointelated Pairs:
## Stochastic Differential Equation vs. Machine Learning Approach
(working paper)

Babak Mahdavi-Damghani[1], Konul Mustafayeva[2], Stephen Roberts[3], Cristin Buescu[5]

*Abstract*— **The cointelation model recently introduced [11], [14] is studied in the context of portfolio optimization. The proposed models are twofold. The first model studied takes the Stochastic Differential Equation approach in which the methodology is split in a dual switching exercise using some of the classic results of Markowitz Modern portfolio theory [15] with an Ornstein-Uhlenbeck [20] optimized overlay. The methodology is then compared to what is presented as the Machine Learning mirror methodology in the form of the new Band-wise Gaussian Mixture [10] model which we expose as giving similar results while keeping the methodology simpler and more adaptable to regime change.**

***Keywords:*** Cointelation, Ornstein-Uhlenbeck, Markowitz, Modern portfolio theory, Stochastic Portfolio Theory, Band-wise Gaussian Mixture

## I. Introduction

### A. Context

Mahdavi-Damghani introduced the Cointelation model as well as the Inferred Correlation [11] measure recently. Few applications were presented but none addressed the one associated to portfolio optimization in the context of trading. At the same time the rise of Machine Learning (ML) in Finance has recently resulted in methodologies promoting disassociation to heavily parameterized mathematical models [10]. An example of these types of methodologies is the recently introduced Band-wise Gaussian Mixture [10] which seats in the family of Bayesian non parametric methodologies and proposes an interesting data focus bridge with the world of Stochastic Differential Equations (SDE) as used to quantitative finance.

### B. Problem formulation

How can we optimize a Cointelated[a] pair using classic SDEs? Is there a simpler ML methodology that would allow us to achieve the same objective? Can we recycle and adjust methodologies already introduced?

[1]babak.mahdavidamghani@oxford-man.ox.ac.uk, Oxford-Man Institute of Quantitative Finance

[2]konul.mustafayeva@kcl.ac.uk, King's College London

[3]steve.roberts@oxford-man.ox.ac.uk, Oxford-Man Institute of Quantitative Finance

[5]cristin.buescu@kcl.ac.uk, King's College London

[a]We will recall all the relevant models in the literature review including the the Cointelation model, which specific details have been summarized in section V.

### C. Structure of the paper

Bearing in mind the context of pairs trading, we will first go through a concise relevant literature review for the:

- Markowitz Modern Portfolio Theory in section II,
- Ornstein-Uhlenbeck model in section III,
- Stochastic Portfolio Theory in section IV,
- Cointelation model in section V,
- Band-wise Gaussian Mixture in section VI.

Following the literature review, we provide:

- an intuitive understanding as well as some peripheral topics around the problems we are trying to solve in section VII,
- the signal definition for our trading strategy using the SDE approach in section VIII and the ML approach in section IX,
- results in section X.

We finally conclude in section XI by providing a summary as well as thoughts regarding future research.

## II. Markowitz Portfolio Theory Review

### A. Foundations

The foundation of modern protfolio theory (MPT) was established by Harry Markowitz in 1952 with his seminal paper [15] in which he proposed expected return and variance to be criteria for portfolio selection. More specifically, the problem of an agent who wishes to build a portfolio with the maximum possible level of expected return, given a limit of variance is, considered with a focal point being the portfolio efficientness. Concepts like the ones of "efficient frontier", or "set of efficient mean-variance combinations", were introduced subsequently. Besides introducing the concept, the paper also describes the methodology in detail. Markowitz paved the way for studying theoretically the optimal portfolio choice of risk-averse agents. Based on the ideas developed by Markowitz, Tobin published his famous research work on agents' liquidity preferences and the separation theorem [18]. Later, the Capital Asset Pricing Model (CAPM) was introduced independently by Sharpe [17] and Linter [9].

### B. Optimization Methodology for a pair of assets

Let $(\Omega, (f)_{(t\geq 0)}, \mathbb{P})$, be our probability space with $(f)_{(t\geq 0)}$ generated by $(T+1)$-dimensional Brownian motion and $\mathbb{P}$, our probability measure, and $r_aS_a$, $r_bS_b$ our discounted prices. Given a set of parameters $(r_a, r_b, \sigma_a, \sigma_b, \rho)$, we can define the set of Stochastic Differential Equations' (SDE) (1).

$$\frac{dS_t^a}{S_t^a} = r_a dt + \sigma_a dW_t^a \tag{1a}$$

$$\frac{dS_t^b}{S_t^b} = r_b dt + \sigma_b dW_t^b \tag{1b}$$

$$d\langle W_t^a, W_t^b \rangle = \rho dt. \tag{1c}$$

We can then define an optimal strategy using Markowitz methodology. More specifically we consider the classic mean-variance optimization problem where we initially define the two important measures for our portfolio return and variance given respectively by equations 2a and 2b.

$$r_p = w r_a + (1-w) r_b, \tag{2a}$$

$$\begin{aligned} \sigma_p^2 &= [w\sigma_a + (1-w)\sigma_b]^2 \\ &= w^2\sigma_a^2 + 2w(1-w)\sigma_a\sigma_b\rho + (1-w)^2\sigma_b^2. \end{aligned} \tag{2b}$$

**Remark:** In this paper measured correlation will refer to Pearson's correlation coefficient, represented by equation (3),

$$\rho_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} \tag{3}$$

To clarify we define the minimization problem in which we maximize the Sharpe Ratio (SR).

**Definition:** The *Sharpe Ratio* is given by $\frac{\mathbb{E}[r_p] - \mathbb{E}[r_f]}{\sigma_p}$ and can be divided in two components:

- expected returns $\mathbb{E}[r_p] = \sum_i w_i \mathbb{E}(R_i)$ where $r_p$ is the return on the portfolio, $r_i$ is the return on asset $i$ and $w_i$ is the weighting the proportion of $i$-asset in the portfolio.
- $\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}$, is our portfolio return variance

By finding the optimal weight $w^{**}$, more formally given by equation (4b) using the methodology of equation (4a).

$$w^{**} = \arg\max_w \frac{\mathbb{E}[r_p - r_f]}{\sigma_p}, \; w \in (0,1) \tag{4a}$$

$$w^{**} = \frac{\sigma_b^2 - \sigma_a\sigma_b\rho}{\sigma_a^2 - 2\rho\sigma_a\sigma_b + \sigma_b^2} \tag{4b}$$

## III. Ornstein-Uhlenbeck Theory Review

### A. Foundations

A stochastic control approach to the problem of pairs trading was proposed by Mudchanatongsuk, Primbs and Wong [16]. More specifically, by modelling the spread of a stock prices as an Ornstein-Uhlenbeck (OU) process [20] a portfolio optimization based stochastic control problem is formulated. The optimal position to this control problem in closed form is computed by solving the corresponding Hamilton-Jacobi-Bellman equation [1]. Assuming the dynamic of risk-free asset $M(t)$ with continuously compounded risk free rate $r$ satisfies equation (5) below.

$$dM(t) = rM(t)dt. \tag{5}$$

Denote by $A(t)$ and $B(t)$ the prices of two assets at time $t$ with $B(t)$ following geometric Brownian motion,

$$dB(t) = \mu B(t)dt + \sigma B(t) dZ(t) \tag{6}$$

with drift $\mu$, diffusion $\sigma$. Here $Z(t)$ is a standard Brownian motion. The spread between the two relevant assets at time $t$ is denoted by $X(t)$ in equation (7),

$$X(t) = \log[A(t)] - \log[B(t)], \tag{7}$$

and assumed to follow the mean-reverting process of equation (8).

$$dX(t) = \kappa[\theta - X(t)]dt + \eta dW(t), \tag{8}$$

where $\theta$ is the long-term equilibrium to which the spread reverts, $\kappa$ the rate of mean reversion[b] and $\eta$ is the volatility of the spread. Let $\rho$ denotes the instantaneous correlation coefficient between $Z$ and $W$, therefore $\langle dW(t), dZ(t)\rangle = \rho dt$.

### B. Optimization Methodology for a pair of assets

Under the above assumptions and by means of Ito's lemma, the dynamics of $A(t)$ is,

$$\begin{aligned} \frac{dA(t)}{A(t)} &= \left(\mu + \kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta\right) dt \\ &\quad + \sigma dZ(t) + \eta dW(t). \end{aligned} \tag{9}$$

The wealth dynamic of the self-financing portfolio $V(t)$ is then described by equation (10).

$$dV(t) = V(t)\left[h(t)\frac{dA(t)}{A(t)} + \tilde{h}(t)\frac{dB(t)}{B(t)} + \frac{dM(t)}{M(t)}\right], \tag{10}$$

which can be rewritten as

$$\begin{aligned} \frac{dV(t)}{V(t)} &= \left[h(t)(\kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta) + r\right] dt \\ &\quad + \frac{1}{V(t)}\eta dW(t), \end{aligned}$$

where $h(t)$ is the portfolio weight of stock A and $\tilde{h}(t) = -h(t)$ is the portfolio weight of asset B at time $t$. Assuming that an investors' preference can be represented by the utility function $U(x) = \frac{1}{\gamma}x^\gamma$, with $x \geq 0$ and $\gamma < 1$, the stochastic control problem is of the form of equation (11).

$$h^*(t,x) = \sup_{h(t)} \; \mathbb{E}\left[\frac{1}{\gamma}(V(T))^\gamma\right], \tag{11}$$

subject to

- $V(0) = v_0$,
- $X(0) = x_0$,
- $dX(t) = \kappa[\theta - X(t)]dt + \eta dW(t)$ and
- $dV(t) = V(t)\left[h(t)(\kappa[\theta - X(t)] + \frac{1}{2}\eta^2 + \rho\sigma\eta) + r\right] dt + \eta dW(t)$.

The optimal weight $h^*(t,x)$ is given by:

$$h^*(t,x) = \frac{1}{1-\gamma}\left[\beta(t) + 2x\alpha(t) - \frac{\kappa(x-\theta)}{\eta^2} + \frac{\rho\sigma}{\eta} + \frac{1}{2}\right]$$

[b] with $\kappa > 0$.

with

$$\alpha(t) = \frac{\kappa(1-\sqrt{1-\gamma})}{2\eta^2} \times \left(1 + \frac{2\sqrt{1-\gamma}}{1-\sqrt{1-\gamma}-(1+\sqrt{1-\gamma})\exp\left(\frac{2\kappa(T-t)}{\sqrt{1-\gamma}}\right)}\right)$$

and

$$\beta(t) = \frac{1}{2\eta^2[(1-\sqrt{1-\gamma})-(1+\sqrt{1-\gamma})\exp\left(\frac{2\kappa(T-t)}{\sqrt{1-\gamma}}\right)]} (\gamma\sqrt{1-\gamma}(\eta^2+2\rho\sigma\eta)\left[1-\exp\left(\frac{2\kappa(T-t)}{\sqrt{1-\gamma}}\right)\right]^2 - \gamma(\eta^2+2\rho\sigma\eta+2\kappa\theta)\left[1-\exp\left(\frac{2\kappa(T-t)}{\sqrt{1-\gamma}}\right)\right])$$

For its derivation, we will refer to Mudchanatongsuk, Primbs & Wong [16].

## IV. Stochastic Portfolio Theory Review

### A. Introduction

Kom Samo and Vervuurt [8] consider the inverse problem of Stochastic Portfolio Theory (SPT): learning from data an optimal investment strategy, based on any given set of of trading characteristics. Initially introduced and developed by Robert Fernholz [6], SPT is a mathematical theory for analyzing the stock market structure and portfolio behavior. Kom Samo and Vervuurt [8] labeled the methodology as "descriptive" as opposed to "normative", as well as consistent with the observed behavior of actual markets. Normative methodologies, the basis for earlier theories like modern MPT [15] and the CAPM [4], are absent from SPT. For instance SPT provides the rule-based mathematical weaponry to explain under what conditions it becomes possible to outperform a cap-weighted benchmark index.

### B. Problems

SPT has, however, several problems and limitations:

- finding relative arbitrages since they are inverse problems;
- the exclusion of possibilities of bankruptcy;
- SPT set-up being developed only for investment strategies that are driven by market capitalizations.

These limitation were more or less addressed [8] by adopting a Bayesian non-parametric approach. In this latter paper the broad range of investment strategies is considered driven by a function defined on an arbitrary space of trading characteristics, on which a Gaussian process prior is placed. One of the aims of Stochastic Portfolio Theory is to construct portfolios which outperform an index, or benchmark portfolio, over a given time-horizon with probability one, whenever this might be possible. One such investment strategy is the so-called diversity-weighted portfolio. This re-calibrates the weights of the market portfolio, by raising them all to some given power $p \in (0,1)$ and then re-normalising[c].

[c] More information on the methodology can be found on the relevant paper [5].

*1) Framework:* We will recall in this subsection some of the mathematical formalization for SPT. The latter [6] assumes the dynamics of $n$ stock capitalization processes $X_i(\cdot)$ given by (12).

$$dX_i(t) = X_i(t)\left(b_i(t)dt + \sum_{\nu=1}^{d} \sigma_{i\nu}(t)dW_{\nu(t)}\right) \tag{12}$$

where $i \in [1,n]$, $t \geq 0$, $W_1(\cdot),\ldots,W_d(\cdot)$ are independent standard Brownian motions with $d \geq n$, and $X_i(0) > 0$ are the initial capitalizations. All processes are assumed to be defined on a probability space $(\Omega.\mathcal{F},P)$, and adapted to a filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{(0\leq t<\infty)}$ that satisfies the usual conditions and contains the filtration generated by the "driving" Brownian motions. The processes of rates of return $b_i(\cdot)$ of volatilities $\sigma(\cdot) = \sigma_{i\nu}(\cdot)_{(1\leq i\leq n, 1\leq \nu\leq d)}$, are $\mathbb{F}$-progressively measurable and assumed to satisfy the integrability condition:

$$\sum_{i=1}^{n}\int_{o}^{T}\left(|b_i(t)| + \sum_{\nu=1}^{d}(\sigma_{i\nu}(t))^2\right)dt < \infty, \tag{13}$$

$\mathbb{P}$-a.s., $\forall T \in (0,\infty)$, as well as the non-degeneracy (ND) condition:

$$\exists\epsilon > 0 \;\text{ such that: }\; \zeta'\sigma(t)\zeta \geq \epsilon\|\zeta\|^2, \tag{14a}$$

$$\forall\zeta \in \mathbb{R}^n, t \geq 0; \quad \mathbb{P}\text{-a.s.} \tag{14b}$$

For the diversity-weighted portfolio with negative parameter $p$ the following no-failure (NF) condition is imposed:

$$\exists\ \phi \in (0,1/n) \text{ such that:}$$
$$P(\mu_{(n)}(t) > \phi\ \forall\ t \in [0,T]) = 1. \tag{15}$$

**Remark:** It is interesting to note that, because of the non realistic NF conditions enhancements are constantly developed, notably a negative-parameter variant ($p < 0$) [21] of the diversity-weighted portfolio [5]. The strategy is claimed to outperform the market, with probability one, over sufficiently long time-horizons, under ND assumption on the volatility structure and under the assumption that the market weights admit a positive lower bound. Several modifications of this portfolio are put forward, which outperform the market under milder versions of the latter NF condition.

A popular implemenation is the diversity-weighted portfolio with parameter $p \in \mathbb{R}^*$, defined as in [5]:

$$\pi_i^{(p)}(t) := \frac{[\mu_i(t)]^p}{\sum_i^n[\mu_j(t)]^p}, \quad i = 1,\cdots,n, \tag{16}$$

here

$$\mu_i(t) := \frac{X_i(t)}{\sum_i^n X_i(t)}, \quad i = 1,\cdots,n. \tag{17}$$

with however no sign on how to optimize $p$[d]. Our current observation for the SPT is that it is a popular theory with nevertheless absurd assumptions on the market which constraint relaxations are only partially addressed.

[d] though tables have been provided in an empirical study [5].

## V. Cointelation Model Review

### A. Definition & review

Cointelation [11] is a portmanteau neologism in finance, designed to signify a hybrid method between between the **cointe**gration and the corre**lation** models.

**Remark:** A caveat needs to be noted here: the term cointegration is differently formulated than the usual cointegration models presented in the econometrics literature. The term cointegration here represent a technical jargon introduced in the "misleading value of measured correlation" [14], [11] which aim is to specify the concept of mean reversion in the sense of the Ornstein-Uhlenbeck model [20].

**Definition (Cointelation Model):** Let $(\Omega, (f)_{(t\geq 0)}, \mathbb{P})$, be our probability space with $(f)_{(t\geq 0)}$ generated by the $T+1$ dimensional Brownian motion and $\mathbb{P}$ the historical probability measure under which the discounted price of the underlier, $rS$, is not necessarily a martingale. The Cointelation model is defined by the set of 2 SDE's given by:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t, \tag{18a}$$

$$dS_{l,t} = \theta(S_t - S_{l,t})dt + \sigma_l S_{l,t} dW_t^l, \tag{18b}$$

$$d\langle W_t, W_t^l \rangle = \rho dt, \tag{18c}$$

where $\rho$ is alternatively called the Correlation of the Cointelation, instantaneous Correlation or the infinitesimal Correlation. The process $(S)_{t\geq 0}$ is called the leading process, $(S)_{l,t\geq 0}$ the lagging process, $r$ the drift, the $\sigma$ the volatility and $\theta$ the speed of mean reversion.

We note that setting $\theta = 0$ in sub-equation (18b) yields the classic correlation model. Conversely, setting $\rho = 0$ in sub-equation (18c) yields our version of the cointegration model (essentially an Ornstein-Uhlenbeck process [20] with an $S_{l,t}$ in front of the stochastic part of the SDE to enforce positivety).

**Proposition:** Using Pearson's correlation measure of equation (3), therefore independent from the classic ways of tempering in with measuring correlation, one notable interesting property of the Cointelation model is that its inferred correlation mirror may hit the whole measured correlation spectrum $[-1, 1]$ depending on the choice of $\rho = -1$ and $\theta$. Figure 1 illustrates this claim.

### B. Parameter estimation

*1) Estimating $\theta$ by rearranging the SDE:* Using the methodology of the original paper [11], we can use sequential estimation in order to estimate the key parameters.

$$\hat{\rho}_1 = \mathbb{E}\left[\rho_1\right]$$

$$\hat{\theta} = \mathbb{E}\left[\frac{dS_{l,t}}{(S_t - S_{l,t})dt + \sigma S_{l,t}\left(\hat{\rho}_1 dW_t + \sqrt{1-\hat{\rho}_1^2}dW_t^{\perp}\right)}\right]$$

$$\hat{\sigma} = \mathbb{E}\left[\frac{dS_{l,t} - \hat{\theta}(S_t - S_{l,t})dt}{\hat{\rho}_1 dW_t + \sqrt{1-\hat{\rho}_1^2}dW_t^{\perp}}\right]$$

*2) Variance reduction technique:* Similarly to the variance reduction methodology described by [14], [11], we can define $B_+ = |\frac{\max(S_t - S_{l,t}, t\in[0,T])}{2}|$, and $B_- = |\frac{\inf(S_t - S_{l,t}, t\in[0,T])}{2}|$). We note that the estimation of $\theta$ has a higher variance when $Z_\sigma = B_+ > |S_t - S_{l,t}| > B_-$ where $\sigma$, on the other hand has quality samples. The reverse is true when $Z_\theta = |S_t - S_{l,t}| > B_+ \bigcup |S_t - S_{l,t}| < B_-$. We can therefore sample $\theta$ in $Z_\theta$ and $\sigma$ in $Z_\sigma$. Figure 2 gives a representation of these sampling zones.

### C. Inferred Correlation

*1) Concept & definition:* The proposed Cointelation test [11] is subdivided in 4 steps including the inferred correlation conjecture. The idea being that if one takes the discrete version of equation (18a), the correlation that you measure as a function of the timescale would increase faster as $\theta$ increases.

**Conjecture (Inferred Correlation):** We consider a stock price process $(S_t)_{t\geq 0}$ with natural filtration $(\mathcal{F}_t)_{t\geq 0}$, and we define the forward price process $(F_t)_{t\geq 0}$ by $F_t := \mathbb{E}(S_t|\mathcal{F}_0)$ then considering, the dynamics of equation (18), we have:

$$\rho_\tau^* \approx \rho + (1-\rho)\left[1 - \exp\left(-\theta(\tau - 1)\right)\right] \tag{20}$$

where $\rho_\tau^* = \mathbb{E}[\sup_{0<t\leq\tau} \rho_t]$, $\tau \in \mathbb{Z}^*, \theta \in [0, 1]$

**Remark:** The author [11] has set $\lambda \approx 1.75$ *for "regular financial data"*. The author explains that $\lambda$ is actually itself a function of the other parameters. Though this approximation provides interesting empirical results, this latter point is an open problem with the Cointelation test.

### D. Number of Crosses

*1) Concept:* Like for the case of the inferred correlation estimator, the second steps of the Cointelation test introduces the concept of *Number of Crosses* [11] which precise formulation is an open problem but which empirical formula with approximate constants yields good results. Its intuitive rational is that, compared to the number of time purely correlated SDEs (eg: without the mean reversion component[e].) then the number of times the descrete version of the cointelated SDEs cross is more than if they were random and the bigger the $\theta$ the more often the discretized SDEs *cross path* per unit of time[f].

**Conjecture (Number of Crosses):** If we discretize equation (18), then we can approximate the number of time the 2 stochastic process, $x = S_{i\in[1,2,...L]}$ and $y = S_{l,i\in[1,2,...L]}$, cross paths, by equation (21)

$$\mathbb{E}[\Gamma_{x,y}(\theta, L)] \approx L\left[\gamma(1-\theta) + \frac{1}{2}\sqrt{\theta}\right] \tag{21}$$

with $L$, the length of the data.

**Remark:** This Conjecture (Number of Crosses), though presents good empirical results, its proof however is an open

[e] $\theta = 0$

[f] the mirror concept in continuous time can be thought of a version of local time

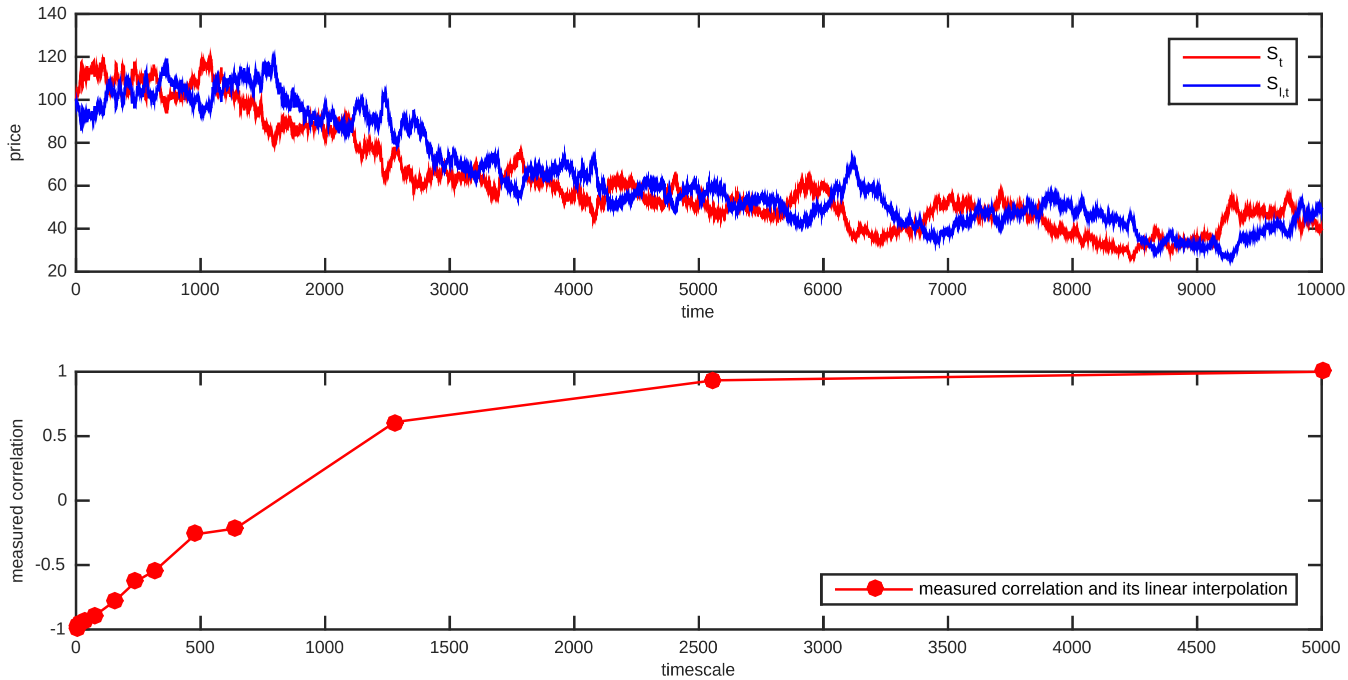


Fig. 1. Example of Cointelation model with a $\rho = -1$, $\theta = 0.1$, $\sigma = 0.01$ and its resulting mirror measured correlation at random increasing timescale.

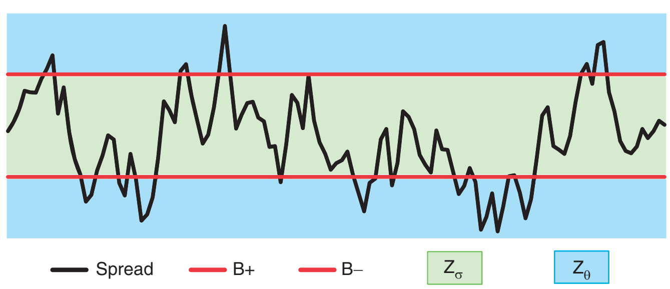


Fig. 2. Visual representation for the sampling zones for $\theta$ [14].

problem. The author mentions [11] that with *"reasonable financial data"* [11] $\gamma \approx 0.01$. Like we have seen when we discussed the inferred correlation concept with $\lambda$, $\gamma$ is itself function of the other parameters but its precise formulation, an open problem currently.

*2) Estimating $\theta$ via the number of crosses formula:* Another estimation of $\theta$ can be give by equation 22

$$\theta \approx \lambda \left( \frac{\hat{\Gamma}_{x,y}(\theta, l)}{l} - \gamma \right)^2 \tag{22}$$

### *E. Cointelation Test*

We briefly define the Cointelation test.

**Definition (Cointelation Test):** Two stochastic processes representing financial data, will be Cointelated if:

- Conjecture (Inferred Correlation) is verified
- Conjecture (Number of Crosses) is verified
- The underlying assets have reasonable physical connection that would suggest that their spread should mean revert
- In the instances where the first two bullet points are not verified exactly, the correlation model cannot possibly be a substitute as correlation is a special case of cointelation (where $\theta = 0$)

**Remark:** Note that the third point, above is a vague statement, not a mathematical definition.

## VI. Band-wise Gaussian Mixture Review

The cointelation model is a special case of the generalized bumping equation recently introduced in [10], and can therefore use the corresponding Band-wise Gaussian Mixture model. The generalized bumping methodology [10] introduced refered to secondary parameters[g] which objective is empirical manual fitting which are relevant to this paper. Disregarding those, we have equation (23).

$$dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^\alpha (1 - X_t^2)^\beta dW_t \tag{23}$$

Where $\theta_t$ is the speed of mean reversion, $\mu_t$, the long term mean, $\alpha$ the positivity flag enforcer, $\beta$, the $[-1, +1]$ boundary flag enforcer and $\{\bigcup dW_i\}_{i=t-\tau}^{t}$, the set of historical deviations of the assumed model's distribution (e.g.: all the historical absolute returns in the context of a normal diffusion). This stochastic process can model:

- Proportional returns (log-normal diffusion). Simply enforce $\theta = 0$, $\alpha = 1$, $\beta = 0$.

[g] Please see the original paper for more information [10]

- Absolute returns (normal diffusion). Simply enforce $\theta = 0,\ \alpha = 0,\ \beta = 0$,
- Mean reverting returns where we enforce positivity (such as in the case of the CIR [2] diffusion),
- Mean reverting returns where we do not enforce positivity (such as in the case of the OU [20] diffusion),
- Mean reverting returns bounded in $[-1, 1]$. For example the dynamics of the $\rho$ parameter in the SVI/gSVI/IVP [7], [13], [12] implied volatility parametrization.

**Lemma 1a:** Let $R = \{x_1, \dots, x_n\}$ be a set of empirical random variables taken from equation (23) with cumulative distribution function $F(x)$ and density $f(x)$. Denote $O = \{x_{(1)}, \dots, x_{(n)}\}$ the ordered set of $R$ such that $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ and $O_p^i = \{x_{(\lceil n((i-1)+1)/p \rceil)}, \dots, x_{(\lfloor n(i)/p \rfloor)}\}$. Then the empirical distribution function for an SDE of the form of equation (23) can be approximated by a union of band-wise Bernoulli process given by:

$$\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n} \sum_{j=1}^{p} \sum_{i=\eta}^{\zeta} \mathbf{1}_{x_i \in O_p^j} \tag{24}$$

with $\eta = \lceil n((i-1)+1)/p \rceil$ and $\zeta = \lfloor n(i)/p \rfloor$.

**Remark:** In the case $p = 3$, using a Gaussian Mixture such that $\hat{F}_n(x_i|\mathcal{F}_t) = \mathcal{N}(-3, 1)\mathbf{1}_{x_t \in O_3^1} + \mathcal{N}(0, 1)\mathbf{1}_{x_t \in O_3^2} + \mathcal{N}(3, 1)\mathbf{1}_{x_t \in O_3^3}$, we obtain the approximate stratification of figure 2. The stratification in our case being made so that the cardinality in each $O_p^j$ region remains approximately the same, as opposed to being the result of a geometrical separation function of $x_{(1)}$ and $x_{(n)}$. Figure 3 illustrates this remark.

**Lemma 1b:** The distribution given by equation (24) converges towards a p-Gaussian Mixture.

*Proof:* $\mathbf{1}_{x_i \in O_p^j}$ is a Bernoulli random variable with parameter $p$, and since the sum of Bernoulli random variable is also Bernoulli, $\hat{F}_n(x_i|\mathcal{F}_t) = \frac{1}{n} \sum_{j=1}^{p} \sum_{i=\eta}^{\zeta} \mathbf{1}_{x_i \in O_p^j}$ is Bernoulli distributed. We can also see that in equation (18) $\lim_{n \to \infty, p \to \infty} (\mu_{t,\tau} - X_t) = \lambda_{t,\tau}$ and therefore $dX_t - \lambda_{t,\tau} = \sigma X_t^{\alpha}(1 - X_t^2)^{\beta} dW_t$ becomes a locale martingale. Using the Glivenko-Cantelli theorem [19], [3], $\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0$. The distribution generated by equation (18) can be mimicked by equation (24) therefore approximates $\cup_{i=1}^{p} \mathcal{N}(\lambda_i, \sigma_i)$. ∎

**Remark:** Empirically, we can see from figure 3 and figure 4 that increasing $p$ can lead to transition probabilities that are more smooth (though less data will be available per band and therefore the significance of the empirical probability driving that particular band will be less significant).

**Theorem (SDE to Band-wise Gaussian Mixture):** Let $(\Omega, (\mathcal{F})_{(t \geq 0)}, \mathbb{Q})$ be our probability space, with $(\mathcal{F})_{(t \geq 0)}$ generated by the $T + 1$ dimensional Brownian motion and $\mathbb{Q}$ is the probability measure. The probability distribution $f(x|\mathcal{F}_t)$ induced by the Stochastic Differential Equation $dX_t = \theta_{t,\tau}(\mu_{t,\tau} - X_t)dt + \sigma X_t^{\alpha}(1 - X_t^2)^{\beta} dW_t$ converges almost surely towards a p-Gaussian mixture as $n$ and $p$ converge towards $\infty$.

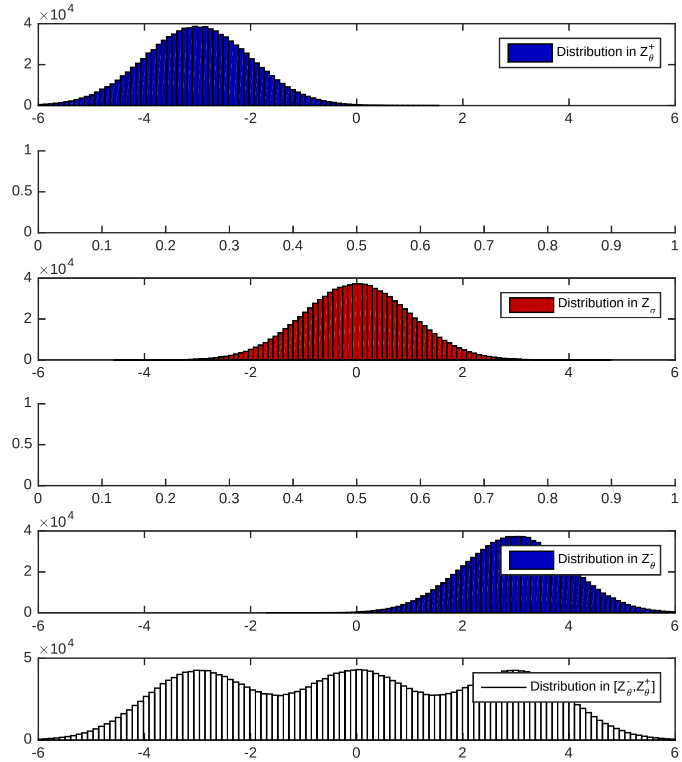

Fig. 3. Gaussian distribution confined to 3 different bands mimicking the ones from figure 2.

*Proof:* The proof can be split in 2 steps using Lemma 1a and Lemma 1b. ∎

The calibration for the Band-Wise Gaussian Mixture can be found in algorithm 1.

## VII. Trading Big Picture

### A. Problem formulation

*1) Portfolio Definition:* We propose an optimization of our portfolio of cointelated pair described in equations (18a) and (18b) with an analog[h] switching signal between a modified Markowitz methodology of section II-B and the OU optimization process methodology described in section III-B.

**Definition (Trading Signal):** Let $s_t^f \in \{1, 0, -1\}$ and $f \in \{\theta, \rho\}$ be the *Trading Signal* at time $t$ of the $\theta$ and the $\rho$ centered sub-strategies.

**Definition ($\rho$-centered sub-strategy):** We will informally call the $\rho$-centered sub-strategies, the strategy described by section II-B and with P&L $V^\rho$ given by equation (25).

$$V_t^\rho = \sum_{0}^{T} s_{t-1}^\rho [w \Delta S_t - w_l \Delta S_{l,t}] \tag{25}$$

[h] To understand to mean in opposition to a discrete (non continuous) signal.

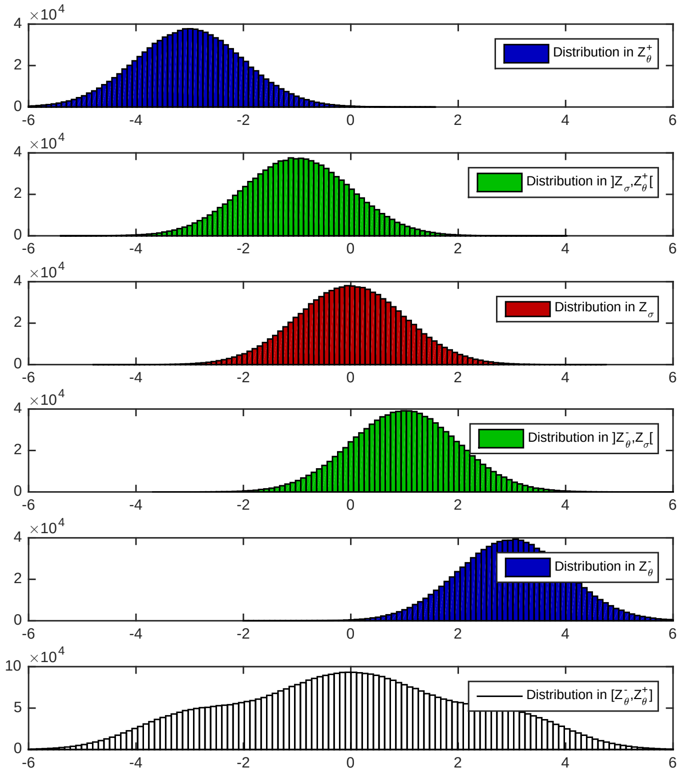


Fig. 4. Gaussian distribution in 5 different zones

**Algorithm 1** BAND-WISE GAUSSIAN MIXTURE$(X, p)$

**Require:** array $X_{1:n}$ and number of bands $p$
**Ensure:** $\Omega^{(1:p)}$, $[B^+_{(1:p)}, B^-_{(1:p)}]$ are returned.

**Sorting state**:
$X_{(1:n)} \leftarrow$ QuickSort$(X_{1:n})$
$[B^+_{(1:p)}, B^-_{(1:p)}] \leftarrow$ FindPercentileBands$(X_{(1:n)}, p)$
$\Omega^{(1:\lceil n/p \rceil)} \leftarrow []$

**Allocation state**:
**for** $j = 1$ to $p$ **do**
  **for** $i = 1$ to $n$ **do**
    **if** $B^-_{(1:p)} \leq X_{(i)} < B^+_{(1:p)}$ **then**
      Amend$(\Omega^{(j)}, X_{(i)})$
    **end if**
  **end for**
**end for**

**Checking Approximation state**:
$\hat{\mu}_{1:p} \leftarrow$ mean$(\Omega^{(1:p)})$
$\hat{\sigma}_{1:p} \leftarrow$ stdev$(\Omega^{(1:p)})$
Print$(\cup_{i=1}^{p} \mathcal{N}(\hat{\mu}_i, \hat{\sigma}_i))$

**Return state**:
$\Omega^{(1:p)}$, $[B^+_{(1:p)}, B^-_{(1:p)}]$

**Definition ($\theta$-centered sub-strategy):** We will informally call the $\theta$-centered sub-strategies, the strategy described by section III and with P&L $V^\theta$ given by equation (26).

$$V_t^\theta = \sum_0^T s_{t-1}^\theta (\Delta S_t - \Delta S_{l,t}) \tag{26}$$

**Definition (Multi-strategy portfolio):** Let $V_t$ be the P&L of the combined $\theta$ and $\rho$ *multi-strategy* at time $t$ given by equation (27).

$$\begin{aligned} V_T = \quad & + \textstyle\sum_0^T (s_{t-1}^\theta + w s_{t-1}^\rho) \Delta S_t \\ & - \textstyle\sum_0^T [s_{t-1}^\rho + (1-w) s_{t-1}^\rho] \Delta S_{l,t} \end{aligned} \tag{27}$$

*Proof:* Equation (27), can be split in half, we have $V_t^\theta = \sum_0^T s_{t-1}^\theta (\Delta S_t - \Delta S_{l,t})$, $V_t^\rho = \sum_0^T s_{t-1}^\rho [w \Delta S_t - w_l \Delta S_{l,t}]$ and we have $V_t = V_t^\rho + V_t^\theta$, by simply rearranging and noticing that $w_l = (1-w)$ we obtain equation (27). ■

**Remark:** Note that $\forall t \in \{1, T\}$ $s_t^\theta$ is not necessarily equal to $1 - s_t^\rho$.

### B. Performance Measures

SR (from equation (4a)) may perhaps be an acceptable performance measure for most financial strategies but its appropriateness for the spread model of section III does not work as expected. Indeed, the absence of SR in the latter methodology in the authors original paper [16] is due to its contrarian performance when used with a mean reverting strategy. To clarify, when spread is really big the foretasted infinitesimal return over risk is also very big but the likelihood of such event is also rare so the many opportunities with positive returns is neglected for no good reason. We therefore, instead, use the utility function of equation (11) for the spread dynamic given by $\theta$-centered sub-strategy and the SR performance measure for the $\rho$-centered sub-strategy. We write our P&L function by equation (28c), where sub-equation (28a) represents the P&L driven by the $\theta$-centered sub-strategy, (28b) represents the P&L driven by the $\rho$-centered sub-strategy and equation (28c) represent the P&L of the multi-strategy in which the switching is done based on a simple threshold function for the spread.

$$V^\theta = w^\theta(t,x)(\Delta S_t - \Delta S_{l,t}) \tag{28a}$$
$$V^\rho = w^\rho(t,x)\Delta S_t + [1 - w^\rho(t,x)]\Delta S_{l,t} \tag{28b}$$
$$V = \lambda(t,x)V^\theta + [1 - \lambda(t,x)]V^\rho \tag{28c}$$
$$\lambda(t,x) = 1_{x>\mu} \tag{28d}$$

**Remark:** Please note that the re-balancing of the portfolio is done through a simple on and off switch given by equation (28d)

### C. Finding the right phase

One other important aspect to note is that when dealing with cointelated pairs, after estimating $\theta$, $\rho$, $\sigma$, an understanding of phase is still required. The latter point is discussed in this subsection. We have seen in equation (21) that the expectation of the number of crosses $\mathbb{E}[\Gamma_{x,y}]$ can be approximated by $L\big[\gamma(1-\theta) + \frac{1}{2}\sqrt{\theta}\big]$.

**Definition (Beginning of Time):** We call the *Beginning of Time* $\tau_0$, the inception time of the discrete version of two

Cointelated pairs in which, $\Delta S_t = S_t - S_{t-1}$ with arbitrary, $t > \tau_0$, $\Delta x_{\tau_0} := 0$, $\tau_0 := 0$.

**Definition (Equilibrium & Disequilibrium):** A Cointelated pair $(S_t, S_{l,t})_{t>0}$, is said to be in *Disequilibrium* at time $\tau$ if

$$|S_\tau - S_{l,\tau}| > 0$$

Likewise, the same pair is said to be in *Equilibrium* if

$$S_\tau - S_{l,\tau} = 0$$

In the case in which our cointelated pair's $\tau_0$ is unknown or two cointelated pairs, $(S_t, S_{l,t})_{t>0}$ started in *disequilibrium* (eg: $|S_{\tau_0} - S_{l,\tau_0}| > 0$), then it is primordial to start the strategy in a case where they are in *equilibrium*.

**Definition (Cointelated Pairs' Phase Corrector):** The *Phase Corrector* of two cointelated pairs, $(x, y)$ where $x = (S_t)_{t>0}$ and $y = (S_{l,t})_{t>0}$ is a constant $c \in \mathbb{R}$ which maximize historical instance in which the cointelated pair is in equilibrium and is given by:

$$c = \arg\max_c \Gamma_{x+c,y}(\theta, L) \tag{29}$$

where $\Gamma_{x+c,y}(\theta, L)$ is the number of cross function of equation (21) phase corrected. We may need to use the Phase Corrector methodology described above to address the reservations described in this section.

## VIII. Signal Definition: the Stochastic Differential Equation Approach

### A. The $\theta$-centered sub-strategy

Applying the methodology from section III to the naming of the parameters in section I, we have:

$$w^\theta(t,x) = \frac{1}{1-\gamma}\left[\beta(t) + 2x\alpha(t) - \frac{\theta x}{\sigma_l^2} + \frac{\rho\sigma}{\sigma_l} + \frac{1}{2}\right]$$

with $x = \log(S_t) - \log(S_{l,t})$,

$$\alpha(t) = \frac{\theta(1-\sqrt{1-\gamma})}{2\sigma_l^2} \times \left(1 + \frac{2\sqrt{1-\gamma}}{1-\sqrt{1-\gamma} - (1+\sqrt{1-\gamma})\exp\left(\frac{2\theta(T-t)}{\sqrt{1-\gamma}}\right)}\right)$$

and

$$\beta(t) = \frac{1}{2\sigma_l^2[(1-\sqrt{1-\gamma}) - (1+\sqrt{1-\gamma})\exp\left(\frac{2\theta(T-t)}{\sqrt{1-\gamma}}\right)]}$$
$$(\gamma\sqrt{1-\gamma}(\sigma_l^2 + 2\rho\sigma\eta)\left[1 - \exp\left(\frac{2\theta(T-t)}{\sqrt{1-\gamma}}\right)\right]^2$$
$$-\gamma(\eta^2 + 2\rho\sigma\sigma_l + 2\theta)\left[1 - \exp\left(\frac{2\theta(T-t)}{\sqrt{1-\gamma}}\right)\right])$$

### B. The $\rho$-centered sub-strategy

Applying the methodology from section II-B to the naming of the parameters in section I, we have:

$$w^\rho(t,x) = \frac{\sigma_l^2 - \sigma\sigma_l\rho}{\sigma^2 - 2\rho\sigma\sigma_l + \sigma_l^2} \tag{30}$$

## IX. Signal Definition: the Machine Learning Approach

### A. The Bayesian set-up

We set, from equation (18) $\Delta_t = S_t - S_{l,t}$, assume for now that $r > 0$ and $B_t = \{B^+_{n,t}, B^+_{n-1,t}, \ldots, B^+_{1,t}, B^1_{1,t}, \ldots, B^-_{n-1,t}, B^-_{n,t}\}$, such that $B^+_{n,t} > B^+_{n-1,t} > \ldots > B^+_{1,t} > 0 > B^-_{1,t} > \ldots > B^-_{n-1,t} > B^-_{n,t}$. We have seen that depending on the spread, the resulting approximated distribution of the samples differ [10]. The calibration algorithm will then consist of creating as many zones as possible whilst converging to the results from Theorem (SDE to Band-wise Gaussian Mixture) of page 6 introduced originally in [10]. We first define the four strategies as well as their cumulative P&Ls.

- Strategy $S^{++}$ in which we are long both $S$ and $S_l$ at time $t$ in between bands $[a,b]$ and with P&L $V^{++}_{[a,b],t}$.
- Strategy $S^{+-}$ in which we are long $S$ and short $S_l$ at time $t$ in between bands $[a,b]$ and with P&L $V^{+-}_{[a,b],t}$.
- Strategy $S^{-+}$ in which we are short $S$ and long $S_l$ at time $t$ in between bands $[a,b]$ and with P&L $V^{-+}_{[a,b],t}$.
- Strategy $S^{--}$ in which we are short both $S$ and $S_l$ at time $t$ in between bands $[a,b]$ and with P&L $V^{--}_{[a,b],t}$.

$$V^{++}_{[a,b],T} = \sum_{t=0}^{T}[w^{++}_{[a,b],t}\Delta S_t + (1 - w^{++}_{[a,b],t})\Delta S_{l,t}]1_{a<\Delta_t\le b}$$
$$V^{+-}_{[a,b],T} = \sum_{t=0}^{T}[w^{+-}_{[a,b],t}\Delta S_t - (1 - w^{+-}_{[a,b],t})\Delta S_{l,t}]1_{a<\Delta_t\le b}$$
$$V^{-+}_{[a,b],T} = \sum_{t=0}^{T}[-w^{-+}_{[a,b],t}\Delta S_t + (1 - w^{-+}_{[a,b],t})\Delta S_{l,t}]1_{a<\Delta_t\le b}$$
$$V^{--}_{[a,b],T} = \sum_{t=0}^{T}[-w^{--}_{[a,b],t}\Delta S_t - (1 - w^{--}_{[a,b],t})\Delta S_{l,t}]1_{a<\Delta_t\le b}$$

We define the maximum P&L achieved by each of these strategies by $V^{\mp\mp,*}_{[a,b],T}$, as given by equation (32) and define $S^{**}_{[a,b],T}$ of P&L $V^{**}_{[a,b],T}$ (equation (33)), the optimal strategy using Gaussian Learning in band $[a,b]$.

$$V^{\mp\mp,*}_{[a,b],T} = \underset{w^{\mp\mp}_{[a,b],t\in[0,T]}}{\arg\max}\ V^{\mp}_{[a,b],T},\ w^{\mp\mp}_{[a,b],t} \in [0,1] \tag{32}$$

$$V^{**}_{[a,b],T} = \max(V^{++,*}_{[a,b],T}, V^{+-,*}_{[a,b],T}, V^{-+,*}_{[a,b],T}, V^{--,*}_{[a,b],T}) \tag{33}$$

Table I is an intuitive representation of what the SDE and the Bayesian set-up should capture. We further provide algorithm 2 as the pseudo-code for the calibration process.

**Remark:** Note that in both algorithm 1 and 2, we have used a QuickSort which can be substituted by other sorting algorithm. We invite the motivated readers to investigate on their own this idiosyncratic issue. Also note that this algorithm has neither been optimized nor checked for data quality (eg: the combination of $n$ and $p$ should be such that each band has enough data (eg: minimum 30) for the statistical estimators to be significant. Also note the use of self explanatory functions such as:

| $\Delta_t \geq B^+_{n,t}$ | $\dots$ | $B^+_{2,t} \geq \Delta_t > B^+_{1,t}$ | $B^+_{1,t} \geq \Delta_t > B^-_{1,t}$ | $B^-_{1,t} \geq \Delta_t > B^-_{2,t}$ | $\dots$ | $B^-_{n,t} \geq \Delta_t$ |
|---|---|---|---|---|---|---|
| $\mathbb{E}[w^\theta(t,x)] \gg 0$ | $\dots$ | $\mathbb{E}[w^\theta(t,x)] > 0$ | $\mathbb{E}[w^\theta(t,x)] = 0$ | $\mathbb{E}[w^\theta(t,x)] < 0$ | $\dots$ | $\mathbb{E}[w^\theta(t,x)] \ll 0$ |
| $\mathbb{E}[w^\rho(t,x)] = r$ | $\dots$ | $\mathbb{E}[w^\rho(t,x)] = r$ | $\mathbb{E}[w^\rho(t,x)] = r$ | $\mathbb{E}[w^\rho(t,x)] = r$ | $\dots$ | $\mathbb{E}[w^\rho(t,x)] = r$ |
| $1 > \mathbb{E}[\lambda(t,x)] \gg .5$ | $\dots$ | $1 > \mathbb{E}[\lambda(t,x)] > .5$ | $\mathbb{E}[\lambda(t,x)] = 0$ | $\frac{1}{2} > \mathbb{E}[\lambda(t,x)] > 0$ | $\dots$ | $\frac{1}{2} \gg \mathbb{E}[\lambda(t,x)] > 0$ |
| $V^{**}_{B_i,T} = V^{+-,*}_{B_i,T}$ | $\dots$ | $V^{**}_{B_i,T} = V^{+-,*}_{B_i,T}$ | $V^{**}_{B_i,T} = V^{++,*}_{B_i,T}$ | $V^{**}_{B_i,T} = V^{-+,*}_{B_i,T}$ | $\dots$ | $V^{**}_{B_i,T} = V^{-+,*}_{B_i,T}$ |

TABLE I

INTUITIVE REPRESENTATION OF WHAT THE SDE AND BAYESIAN SET-UP SHOULD CAPTURE.

**Algorithm 2** BAND-WISE ML FOR COINTELATION($X, p$)

**Require:** array $X_{1:n}$ and number of bands $p$

**Ensure:** $\Omega^{(1:p)}$, $[B^+_{(1:p)}, B^-_{(1:p)}]$ are returned.

**Sorting state**:

1: $X_{(1:n)} \leftarrow$ QuickSort($X_{1:n}$)

2: $[B^+_{(1:\frac{p}{2})}, B^-_{(1:\frac{p}{2})}] \leftarrow$ FindPercentileBands($X_{(1:n)}$, $p$)

3: $B_{(1:p)} \leftarrow [B^+_{(1:\frac{p}{2})}, B^-_{(1:\frac{p}{2})}]$

4: $\Omega^{(1:\lceil n/p \rceil)} \leftarrow []$

**Allocation state**:

5: **for** $j = 1$ to $p$ **do**

6: **for** $i = 1$ to $n$ **do**

7: **if** $X_{(i)} \in B^i$ **then**

8: Amend($\Omega^{(j)}, X_{(i)}$)

9: **end if**

10: **end for**

11: **end for**

**Optimize the 4 types of P&L for each band**:

12: **for** $i = 1$ to $p$ **do**

13: $V^{++,*}_{B_i,T} \leftarrow \arg\max_{w^{++}_{B_i,t\in[0,T]}} V^{++}_{B_i,T}$

14: $V^{+-,*}_{B_i,T} \leftarrow \arg\max_{w^{+-}_{B_i,t\in[0,T]}} V^{+-}_{B_i,T}$

15: $V^{-+,*}_{B_i,T} \leftarrow \arg\max_{w^{-+}_{B_i,t\in[0,T]}} V^{-+}_{B_i,T}$

16: $V^{--,*}_{B_i,T} \leftarrow \arg\max_{w^{--}_{B_i,t\in[0,T]}} V^{--}_{B_i,T}$

17: **end for**

**Rank and return best strategy for each band**:

18: **for** $i = 1$ to $p$ **do**

19: $V^{**}_{B_i,T} \leftarrow \max(V^{++,*}_{B_i,T}, V^{+-,*}_{B_i,T}, V^{-+,*}_{B_i,T}, V^{--,*}_{B_i,T})$

20: $S^*_T \leftarrow (S^{++,*}_{B_i,T}, S^{+-,*}_{B_i,T}, S^{-+,*}_{B_i,T}, S^{--,*}_{B_i,T})$

21: $S^{**}_{B_i,T} \leftarrow$ returnCorrespondingStrat($V^{**}_{B_i,T}, S^*_T$),

22: **end for**

**Forecasting** :

23: signal$^S$, signal$^{S_l}$ $\leftarrow$ forecast($S^{**}_{B_i,T}, S_t, S_{l,t}$)

**Return buy/sell signals**:

24: signal$^S$, signal$^S_l$

- returnCorrespondingStrat(x,y) which given the set of strategies and the P&L that maximized that strategy returns as its name indicates the corresponding strategy.
- forecast(x,y,z) which given the set of trained strategies and the current level of $S_t$ and $S_{l,t}$ returns a prediction of where the signals for the latter two should be.

Finally the use of the $\arg\max$ function in lines 13-16 can be replaced by a simple for loop but in the interest of not making the pseudocode too crowded we have kept it this way.

### *B. A model that can dynamically accommodate regime changes*

One important point to note about the additional benefits of a data-driven Learning approach over the SDE approach[i] is to take a look at the example associated to the bizarre world of interest rates. Indeed up to 2014, it was assumed that interest rates could never become negative[j].

**Remark:** We have seen [10] that a reasonable risk manager or trader can assume that the generalized SDE, in which the cointelation model is a special case, would have chosen a $\beta = 0$ and an $\alpha = 1$, the latter enforcing positivity for the simulated scenarios of our risk factor. This very reasonable assumption would have crashed the whole risk engine. The approach we advocate would have, however, been able to continue its dynamical learning scenario without any problem.

## X. RESULTS

### *A. Observation*

Following the strategy optimization methodology described in the previous sections we get the results of figures 5, 6, 7, 8, and 9 in which in can be seen that the the addition of each strategy $\theta$ and $\rho$ into a global/total portfolio enhances our performance.

### *B. Analysis*

The SDE and ML methodologies provide similar results. However, couple important points are worthwhile to note about the additional benefits of the ML approach over the SDE approach

[i]Beyond the obvious benefits associated to achieving the same results though through a simpler channel and also bypassing convoluted SDE calibration issues in the process.

[j]Why would you pay to put your money in the bank? That would essentially be the physical question one may ask oneself.

*1) Simplicity:* The first obvious benefits from the approach presented here over the SDE approach is that we were able to achieve the same results though through a simpler channel[k].

*2) flexibility:* The second advantage that the approach presented here provides flexibility to regime change. The easiest way to understand this is to take a look at the bizarre world of interest rates. Indeed up to 2014, it was assumed that interest rates could never become negative and a similar SDE approach would have enforced a CIR like model [2] and would not have therefore been able to accommodate the regime change towards negative interest rates. These types of transitions are easy to handle for the ML methodology and the adaptation is almost immediate through, for example, a simple filtering process.

## XI. Conclusion

### A. Summary

We have studied the Cointelation model in the context of a dual portfolio optimization exercise taking 2 approaches. The first one consisted of a classic optimal control problem in the form of a switching Markowitz/Ornstein-Uhlenbeck switching methodology in which one strategy would act as an overlay to the other. The second approach we took was both simpler and more adaptable thanks to a more powerful methodology in the form of a band-wise Gaussian Mixture model [10].

### B. Next step

There are few extensions or improvement that can be performed on this work.

*1) The n-Cointelated case:* We can first ask ourselves the question of the n-Cointelated case for which the trio has been specified by equation (34).

$$\begin{array}{rcl} \frac{dS_t^a}{S_t^a} & = & \sigma dW_t^a \\ dS_t^b & = & \theta(S_t^a - S_t^b)dt + \sigma S_t^a dW_t^b \\ dS_t^c & = & \theta(S_t^a - S_t^c)dt + \sigma S_t^c dW_t^c \end{array} \tag{34}$$

One natural question would first be about how to define this trio? For instance would equation (34) be more in-line with the pair from equation (18) or would equation (35) be better? Can they be turned into equivalent forms?

$$\begin{array}{rcl} \frac{dS_t^a}{S_t^a} & = & \sigma dW_t \\ dS_t^b & = & \theta(S_t^a - S_t^b)dt + \sigma S_t^a dW_t^b \\ dS_t^c & = & \theta(S_t^b - S_t^c)dt + \sigma S_t^c dW_t^c \end{array} \tag{35}$$

*2) Performance measure:* We have seen that there are few performance functions that can be used (SPT, Sharpe Ratio, Utility Function). Can we come up with a more robust all encompassing methodology?

*3) Band-wise Gaussian mixture:* We have used the Band-wise Gaussian mixture. Is this methodology optimal or can we find better? We will consider these few questions and more in a subsequent paper.

[k] We bypassed convoluted SDE calibration issues in the process.

## References


[1] R. Bellman. *Dynamic programming*. Princeton University Press, Princeton, 1957.
[2] J.E. Ingersoll Cox, J.C. and S.A. Ross. A theory of the term structure of interest rates. pages 385–407, 1985.
[3] A. W. Van der Vaart. Asymptotic statistics. page 265, 1998.
[4] Kenneth F. Fama, E. The capital asset pricing model: Theory and evidence. 18:25–46, 2004.
[5] Karatzas I. Kardaras C. Fernholz, R. Diversity and relative arbitrage in equity markets. 9:1–27, 2005.
[6] R. Fernholz. *Stochastic Portfolio Theory*. Springer New York, New York, 2002.
[7] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. page 8, 2012.
[8] Vervuurt A. Kom Samo, Y. Stochastic portfolio theory: A machine learning perspective.
[9] J. Linter.
[10] Babak Mahdavi-Damghani.
[11] Babak Mahdavi-Damghani. Introduction to the cointelation model and inferred correlation. 2013.
[12] Babak Mahdavi-Damghani. Introducing the implied volatility surface parametrisation (IVP):application to the fx market. 2015.
[13] Babak Mahdavi-Damghani and Andrew Kos. De-arbitraging with a weak smile. 2013.
[14] Babak Mahdavi-Damghani, Daniella Welch, Ciaran O'Malley, and Stephen Knights. The misleading value of measured correlation. 2012.
[15] H. Markowitz. Portfolio selection. 7:77–91, 1952.
[16] Primbs J. Wong W. Mudchanatongsuk, S. Optimal pairs trading: A stochastic control approach. In *American Control Conference*, Seattle, Washington, USA, 6 2008. IEEE.
[17] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. 19:425–442, 1964.
[18] J. Tobin. Liquidity preference as behavior towards risk. the review of economic studies. 25:65–86, 1958.
[19] Howard Tucker. A generalization of the glivenko-cantelli theorem. page 828830, 1959.
[20] G.E. Uhlenbeck and L.S. Ornstein. On the theory of brownian motion. *Phys.Rev*, 36, 1930.
[21] Karatzas I. Vervuurt, A. Diversity-weighted portfolios with negative parameter. 11:411–432, 2015.


## Appendix

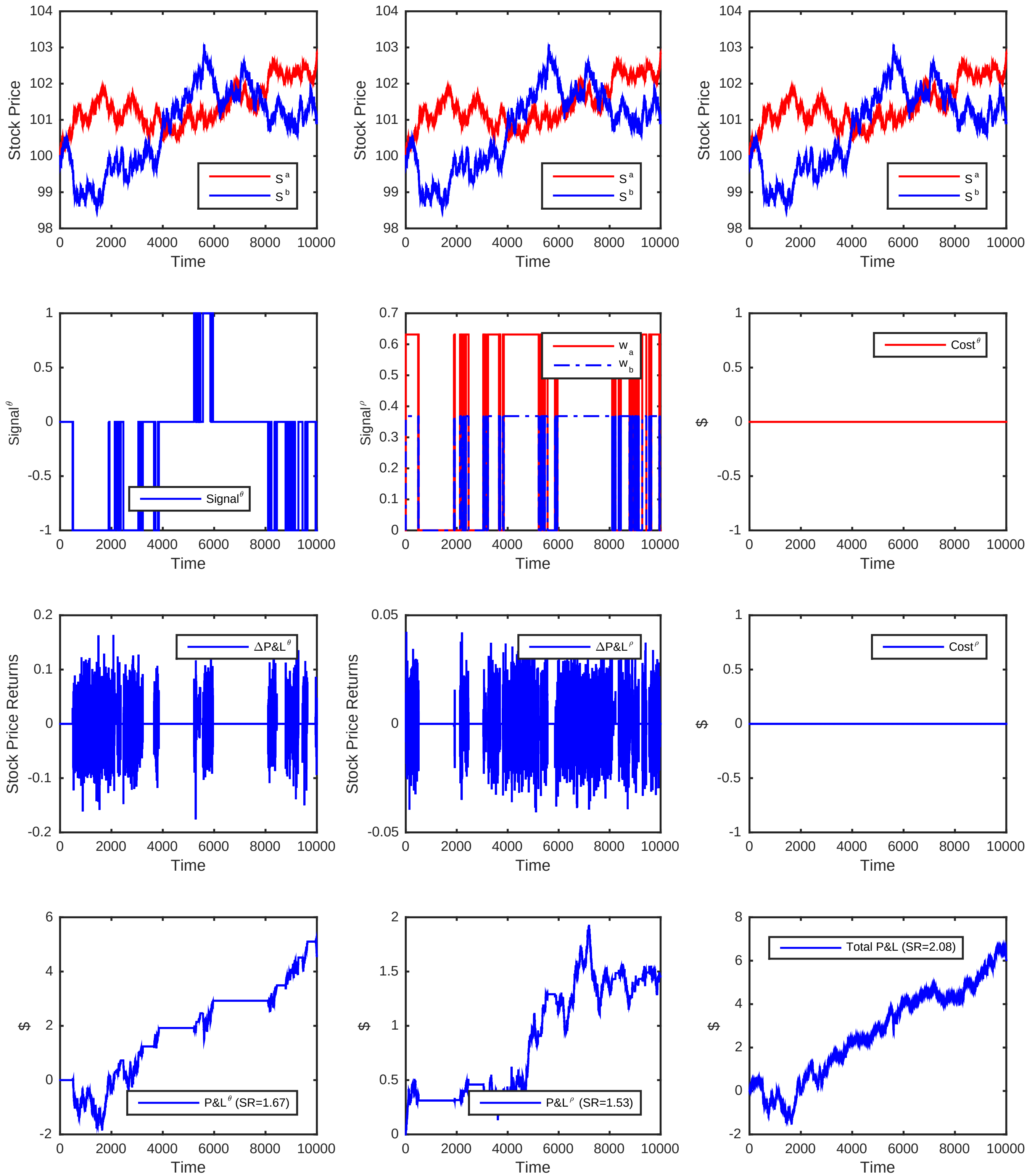


Fig. 5. Example of Pairs Trading performance charts in the context of the dual Cointelation model strategy

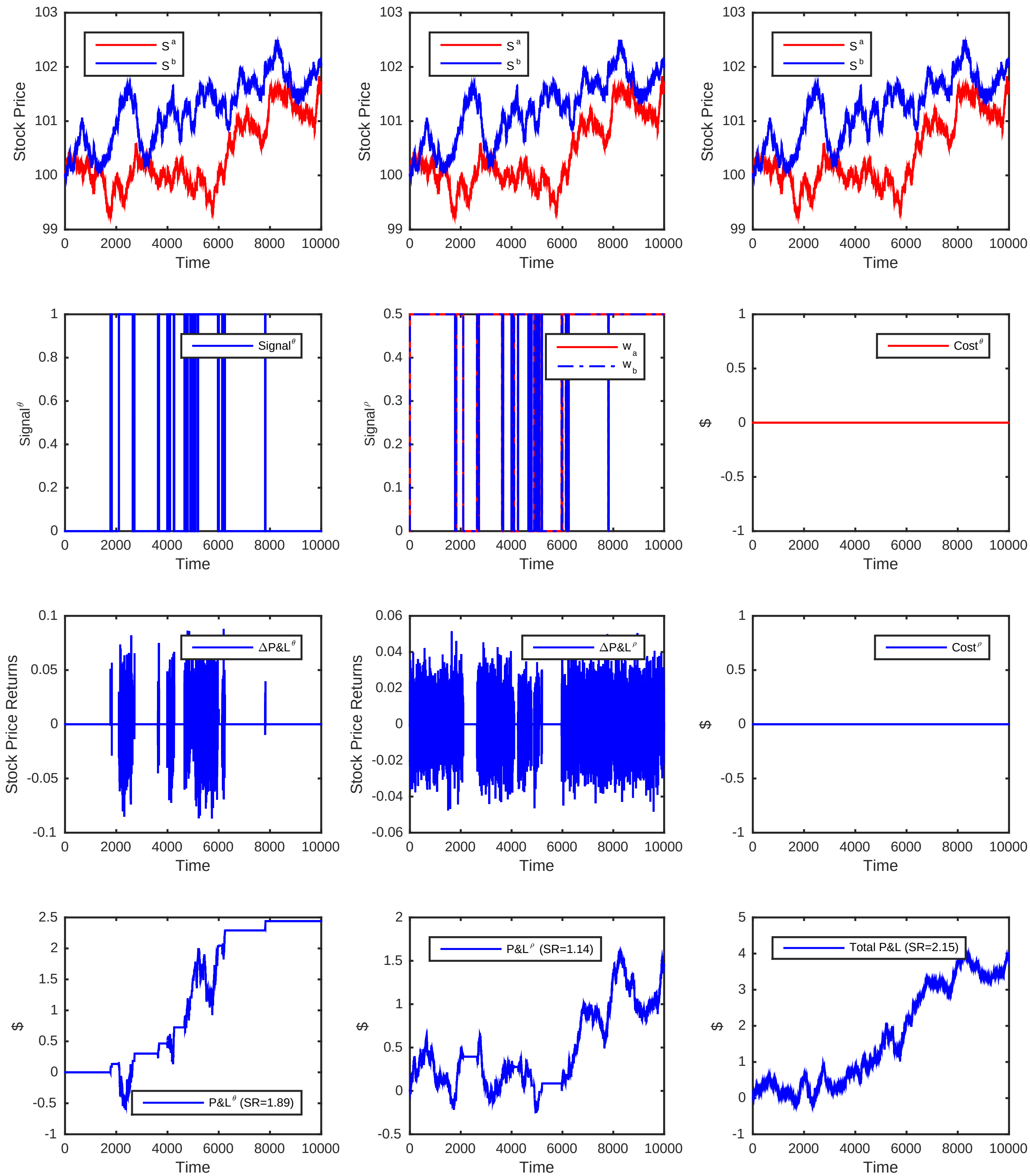


Fig. 6. Example of Pairs Trading performance charts in the context of the dual Cointelation model strategy

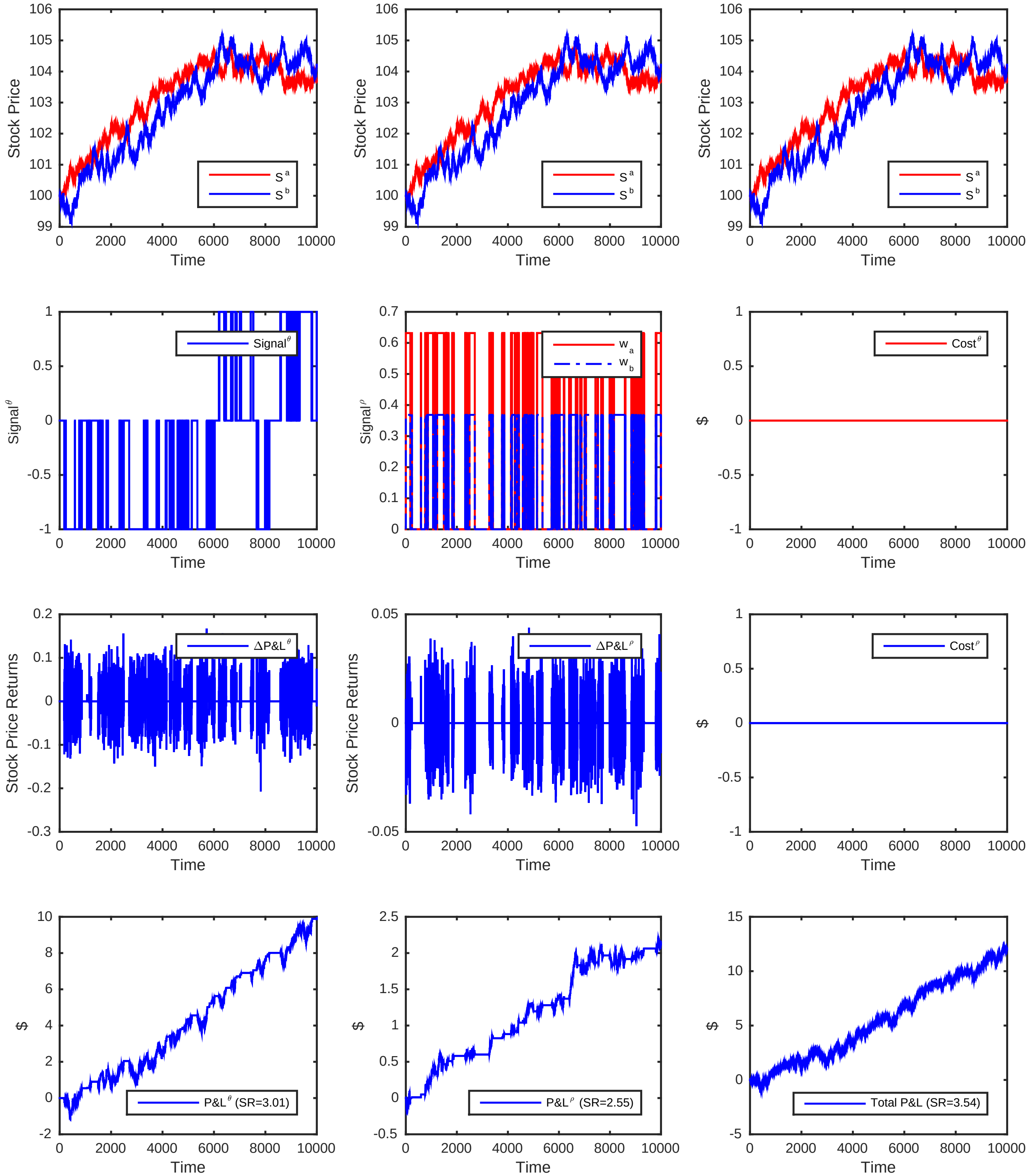


Fig. 7. Example of Pairs Trading performance charts in the context of the dual Cointelation model strategy

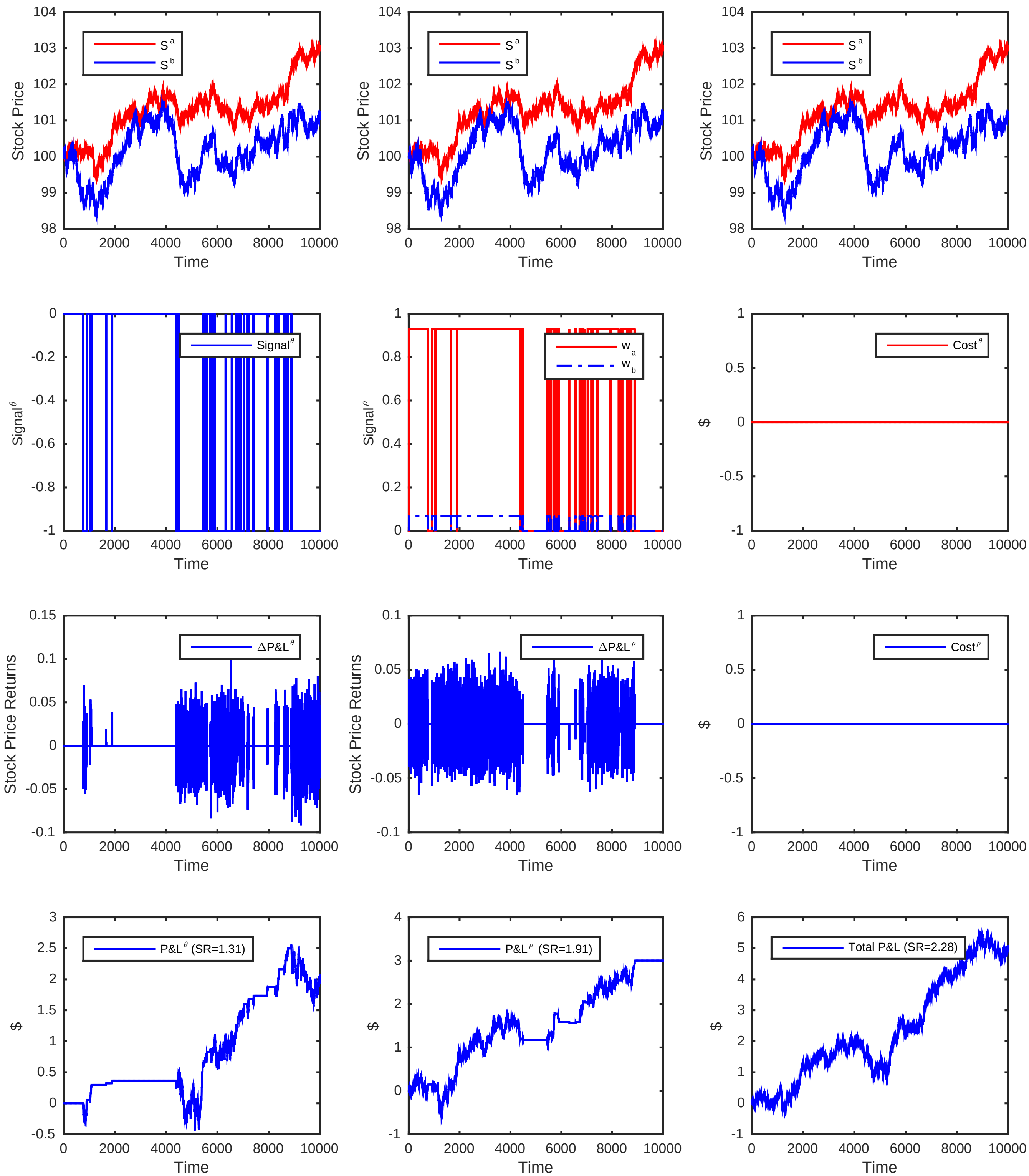


Fig. 8. Example of Pairs Trading performance charts in the context of the dual Cointelation model strategy

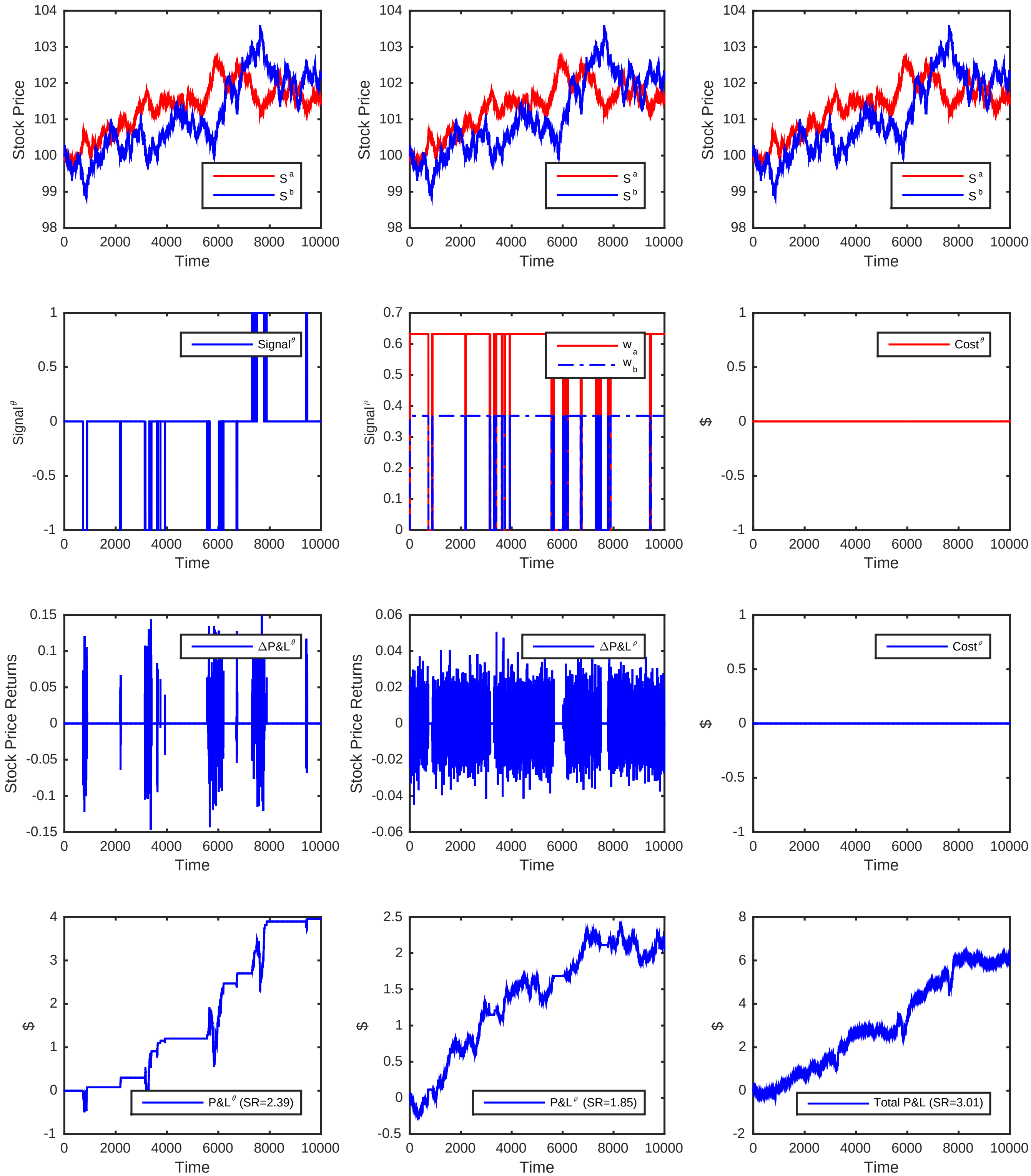


Fig. 9. Example of Pairs Trading performance charts in the context of the dual Cointelation model strategy