Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: http://www.tandfonline.com/loi/rquf20

# Intraday pairs trading strategies on high frequency data: the case of oil companies

Bo Liu, Lo-Bin Chang & Hélyette Geman

**To cite this article:** Bo Liu, Lo-Bin Chang & Hélyette Geman (2016): Intraday pairs trading strategies on high frequency data: the case of oil companies, Quantitative Finance, DOI: 10.1080/14697688.2016.1184304

**To link to this article:** http://dx.doi.org/10.1080/14697688.2016.1184304

Published online: 07 Jun 2016.

*Quantitative Finance*, 2016
http://dx.doi.org/10.1080/14697688.2016.1184304



# Intraday pairs trading strategies on high frequency data: the case of oil companies

BO LIU†, LO-BIN CHANG*‡ and HÉLYETTE GEMAN†§

†Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 North Charles St, Baltimore, MD 21218, USA
‡Department of Statistics, Ohio State University, 1958 Neil Ave, Columbus, OH 43210, USA
§Department of Economics, Mathematics and Statistics, Birkbeck College, University of London, Malet St, London WC1E 7HX, UK

(*Received 3 November 2015; accepted 21 April 2016; published online 7 June 2016*)

This paper introduces novel 'doubly mean-reverting' processes based on conditional modelling of model spreads between pairs of stocks. Intraday trading strategies using high frequency data are proposed based on the model. This model framework and the strategies are designed to capture 'local' market inefficiencies that are elusive for traditional pairs trading strategies with daily data. Results from real data back-testing for two periods show remarkable returns, even accounting for transaction costs, with annualized Sharpe ratios of 3.9 and 7.2 over the periods June 2013–April 2015 and 2008, respectively. By choosing the particular sector of oil companies, we also confirm the observation that the commodity price is the main driver of the share prices of commodity-producing companies at times of spikes in the related commodity market.

*Keywords*: Pairs trading; Quantitative trading strategies; Conditional modelling; Doubly mean-reverting model; High frequency data; Transaction costs

*JEL Classification*: C1, C3, C6, C61, C63



## 1. Introduction

The idea of pairs trading is quite popular across various asset classes and is based on the property that, since companies within a sector are highly correlated, some pairs of price returns exhibit strong similarity. We can model the return differences of these pairs as mean-reverting processes. If they deviate too far from the mean, we short/long the pair by simultaneously buying one and short selling another. We keep the position until it reverts back to the mean level.

Let A and B be a pair of closely related stocks, $S_A(t)$ and $S_B(t)$ their prices at time $t$. The cumulative log return difference—or spread—for this pair is

$$Y(t) = \log\left(\frac{S_A(t)}{S_A(0)}\right) - \log\left(\frac{S_B(t)}{S_B(0)}\right). \tag{1}$$

In the pairs trading literature, the spread $Y(t)$ or its variation has been modelled as a mean-reverting process, oscillating either around zero or around a linear function of time. For example, Avellaneda and Lee (2010) define the spread as the difference between $\log\left(\frac{S_A(t)}{S_A(0)}\right)$ and $\beta \cdot \log\left(\frac{S_B(t)}{S_B(0)}\right)$, where $\beta$ is calculated by regressing the cumulative log return of one stock on another in a certain period. The best pairs are commonly identified as either the ones with smallest distance measures defined as the sum of squared deviations (Gatev *et al*. 2006, Bowen *et al*. 2010) or using cointegration relationships (Vidyamurthy 2004, Lin *et al*. 2006).

Both the distance method and the cointegration method have their limitations. Consider two stocks A and B with cumulative log returns both being 0 at the beginning. Suppose $\log\left(\frac{S_A(t)}{S_A(0)}\right)$ goes to a large positive value $\alpha$ in a short amount of time then stays around that level, while $\log\left(\frac{S_B(t)}{S_B(0)}\right)$ remains around 0. Then no matter how synchronized their moves afterwards, this pair would not likely be identified by the simple distance measure as they have a large average distance of $\alpha$. Now, further assume A and B co-move only in a subperiod during the whole analysis period, then the pair is not likely to be selected according to cointegration method. Yet there are clearly profits to be made in this scenario. The problem is that both those methods imply a *static* relationship between two stocks during the training period, whereas the relationship may very well be changing from one day to the next.

The strategy we propose in this paper is designed exactly to capture this kind of 'local' statistical arbitrage

*Corresponding author. Email: lobinchang@stat.osu.edu

opportunities, by searching for temporary market mispricing inefficiencies. The idea is to seek the pairs of which $Y(t)$ can be characterized by the following modelling procedure: model the long-term trend of $Y(t)$, denoted as $L(t)$, as a stochastic process, and then model $Y(t)$ via a mean-reverting process around this long-term stochastic trend $L(t)$ using the conditional modelling technique. If the mean reversion speed of $Y(t)$ is fast enough, we can make profit by making intraday pairs trades.

The utilization of two mean-reverting stochastic processes on the same series is partly inspired by Fourier series expansion. $L(t)$ can be regarded as the first term of a Fourier series with the largest period and lowest frequency; imagine that $Y(t)$ is approximated by the Fourier series, with higher frequency local oscillation being added on top of the lower frequency waves. To visualize this, imagine a long rope lying on the ground straightened out. If we hold onto one end of the rope and shake it horizontally, then it will display a wavy pattern. Now, pick two points on the rope that are close to each other, pin their locations, then shake the segment between them. The shaken part will likely become a more pronounced local wave. Repeat this for the whole rope segment-by-segment. The resulting rope would look like the $Y(t)$ process while all the pinned positions make up $L(t)$. The rational for this novel doubly mean-reverting model is that, if we can identify pairs with relative stable $L(t)$ and volatile $Y(t)$, then intraday pairs trading should perform well a priori. $Y(t)$ in these cases would return at the end of a trading day—hopefully after wild swings—to more or less the same level as daily open.

The key to *add up* local oscillation is through a framework of 'conditional modelling and conditional inference' (see the overview in Chang, 2010). This technique has been applied to analyze the dynamics of financial time series (e.g. the waiting time invariance of return sequences in Chang *et al*. (2013), aggregation theorem in Chang and Geman (2013)) as well as to other research fields (e.g. Amarasingham *et al*. 2012, Chang *et al*. 2015). Nevertheless, the technique has not been utilized for designing trading strategies in the literature.

In this study, we focus on the oil sector and look at times when the underlying commodity price is experiencing sharp moves, making it the major factor driving the share prices. Geman and Eleuterio (2013) showed, in the case of the fertilizer commodity, that shares of fertilizer-mining companies are very sensitive to the commodity price at times of high moves of this price. Geman (2015) extended this property to other types of agriculture-related companies. The literature on the subject counts as a founding paper Tufano (1998), which analyzes the price sensitivities of gold mining companies' shares. Our results confirmed the observation that the crude oil price is the main driver of oil company stock prices during market turbulence.

In the literature, most pairs trading studies use daily data and a daily trading frequency (Gatev *et al*. 2006, Lin *et al.* 2006, Avellaneda and Lee 2010, Cummins and Bucca 2012, Bogomolov 2013, Zeng and Lee 2014). Bowen *et al*. (2010), which use 60 min return series, is one of the very few that uses data with frequency higher than daily. With the availability of tick data, we are able to use five-min series for $Y(t)$ in the case of some highly liquid oil company stocks. To our best knowledge, this is the first academic study on pairs trading to use such high frequency data, and the first one on intraday pairs trading strategies.

The rest of the paper is organized as follows. Section 2 details the model and its calibration. Trading rules are discussed in section 3. We perform simulations in section 4 in order to validate the model and the strategies. Results from real data back-testing over two periods are presented and analyzed in section 5. Section 6 concludes the paper and provides directions for future research.

## 2. The model

### *2.1. Model specification*

In order to exploit intraday pairs trading profits, we use high frequency data with interval length of five min to model $Y(t)$ defined in equation (1). There are 78 five-min intervals every day during trading hours from 9:30 am to 4:00 pm, hence 79 $Y(t)$'s. We denote the 79 observed values of $Y(t)$ in day $i$ as

$$Y_{79(i-1)+1}, Y_{79(i-1)+2}, \ldots, Y_{79i} \quad i = 1, 2, \ldots, N$$

where $N$ is the number of days. The subscript in this paper refers to discretized observations of stochastic processes. Moreover, we assume that the long-term trend $L(t)$ is identified by the daily opening and closing values of the process $Y(t)$, namely for day $i$,

$$L_{2i-1} = Y_{79(i-1)+1} \quad \text{and} \quad L_{2i} = Y_{79i}. \tag{2}$$

The stochastic process $L(t)$, with two observed data points per day, is preferred to have a small variance. In this study, we model $L(t)$ as an Ornstein–Uhlenbeck (OU) process, with mean 0

$$\mathrm{d}L(t) = -\theta_L L(t)\mathrm{d}t + \sigma_L \mathrm{d}W_t^L. \tag{3}$$

Next by the definition of conditional distribution, the joint distribution of $Y_1, Y_2, \ldots, Y_{79N}$ can be written as the product of the distribution of $Y_{79(i-1)+1}$'s and $Y_{79i}$'s and the conditional distribution of $Y_i$'s given $Y_{79(i-1)+1}$'s and $Y_{79i}$'s

$$\begin{aligned} & f(Y_1, Y_2, \ldots, Y_{79N}) \\ =& f(Y_{79(i-1)+1}, Y_{79i}, i = 1, \cdots, N) \\ & \times f(Y_1, Y_2, \ldots, Y_{79N} | Y_{79(i-1)+1}, Y_{79i}, i = 1, \cdots, N) \\ =& f(L_1, L_2, \ldots, L_{2N}) f(Y_1, Y_2, \ldots, Y_{79N} | L_1, L_2, \ldots, L_{2N}) \end{aligned} \tag{4}$$

Note that due to equation (2), the last equality is valid, and the joint distribution of $L_i$'s can be obtained by discretizing the process (3).

Now to model the conditional distribution of $Y_i$'s given $L_i$'s, we use the conditional modelling technique by introducing an auxiliary process $\tilde{Y}(t)$ that follows:

$$\mathrm{d}\tilde{Y}(t) = \theta\left(\tilde{L}(t) - \tilde{Y}(t)\right)\mathrm{d}t + \sigma \mathrm{d}W_t^{\tilde{Y}} \tag{5}$$

where the mean process is $\tilde{L}(t) = \frac{L_{2i-2}+L_{2i-1}}{2}$ and $i = i(t)$ refers to the day of time $t$ ($2i - 2$ refers to the closing of day $i - 1$ and $2i - 1$ refers to the opening of day $i$). We assume the conditional distribution

$$f(Y_1, Y_2, \ldots, Y_{79N} | L_1, L_2, \ldots, L_{2N})$$

is the same as the conditional distribution of $\tilde{Y}_i$'s given $\tilde{Y}_{79(i-1)+1} = L_{2i-1}$ and $\tilde{Y}_{79i} = L_{2i}$, $i = 1, \ldots, N$, where $\tilde{Y}_i$'s are the corresponding discretization of process (5). In other words, conditional on given daily opening and closing values of the process, $Y(t)$ is the same as $\tilde{Y}(t)$ in distribution. Hence, to simplify the notation, we use $Y(t)$ and $Y_i$ in place of $\tilde{Y}(t)$ and $\tilde{Y}_i$ in the rest of the paper.

By defining the mean process $\tilde{L}(t)$ as the average of $L_{2i-2}$ and $L_{2i-1}$, we assume in any trading day, the mean level that the spread process reverts to is the average of the opening value of the current day and the closing value of the previous trading day.

To recap, the distribution of the $Y(t)$ process is defined by first specifying the dynamics of $L(t)$ from equation (3), then the distributions of the in-between points are given indirectly by equation (5), via the conditional relationship (4), hence the name conditional modelling.

For $Y(t)$, there are 79 observations per day thus 79 time intervals, the lengths of which are not equal: 78 short five-min periods and a long overnight period. The similar problem exists for $L(t)$: the time span between a day's open and close is different from between the day's close and next day's open.

The lengths in real time of the trading hours per day are 6.5 h while the overnight periods are at least 17.5 h (from 4 pm market close to next day's market open 9:30 am, or longer in the case of weekends and holidays). However, the amount of information and market movements during daytime trading hours are much richer than that during overnight periods. Therefore, we estimate the lengths of both periods in *effective* time instead of real time in the following way: since the relative (effective) lengths of the trading day and the overnight period are unknown and unequal, we need two time steps. $\delta_1$ is the length of trading hours 9:30 am and 4:00 pm; $\delta_2$ is the length between market close and next day's open. Hence,

$$\delta_1 + \delta_2 = 1 \text{ day} = \frac{1}{250} \tag{6}$$

and the algorithm to estimate $\delta_1$ and $\delta_2$ is based on the ratio of variances in intraday and overnight changes as detailed in section 2.2.1.

All the model parameters $\theta_L, \sigma_L, \theta, \sigma, \delta_1$ and $\delta_2$ from equation (3)–(6) are calibrated daily using maximum likelihood estimation (MLE).

### 2.2. *Model calibration*

For better parameter estimation, the calibration is updated everyday using a moving window, also called the pairs formation period (see Gatev *et al.* 2006), the length of which is chosen properly. If the duration is too short, the calibration is unreliable due to the lack of training data; if the duration is too long, estimated parameters do not accurately reflect the present situation because of non-stationarity of market dynamics. In our model, the frequencies of $L(t)$ and $Y(t)$ are different, thus requiring training periods of different lengths. We use the past 100 days' daily open and close prices to calibrate the process $L(t)$ and use the past 30 days' five-min prices to calibrate the process $Y(t)$. Both lengths were decided at the beginning of the study and has not been tuned based on data, to avoid data-snooping biases (Lo and MacKinlay, 1990).

#### 2.2.1. Calibration for $L(t)$.

For a general OU process with constant mean

$$\mathrm{d}S(t) = \theta\,(\mu - S(t))\,\mathrm{d}t + \sigma \mathrm{d}W_t,$$

the discretization $\{S_i\}$ satisfies

$$S_{i+1} = S_i e^{-\theta\delta} + \mu(1 - e^{-\theta\delta}) + \sigma\sqrt{\frac{1 - e^{-2\theta\delta}}{2\theta}} Z_i, \tag{7}$$

for all $i$, where $\delta$ is the time step in discretization and $Z_i$'s are i.i.d. $N(0, 1)$.

For our $L(t)$, the discretized series is the combined daily opening and closing cumulative return differences. $\mu$ is assumed to be 0. Equation (7) leads to two equations.

The intraday changes

$$L_{2i} = L_{2i-1}e^{-\theta_L\delta_1} + \sigma_L\sqrt{\frac{1 - e^{-2\theta_L\delta_1}}{2\theta_L}} Z_{2i} \qquad i = 1, \ldots, N; \tag{8}$$

and the overnight changes

$$L_{2i+1} = L_{2i}e^{-\theta_L\delta_2} + \sigma_L\sqrt{\frac{1 - e^{-2\theta_L\delta_2}}{2\theta_L}} Z_{2i+1},\ i = 1, \ldots, N-1 \tag{9}$$

where $Z_i$'s are i.i.d. $N(0, 1)$.

To estimate $\delta_1$ and $\delta_2$, We use the following equation obtained from equations (8) and (9) with the variances approximated by empirical variances :

$$\frac{\mathrm{Var}(L_{2i} - L_{2i-1}e^{-\theta_L\delta_1})}{\mathrm{Var}(L_{2i+1} - L_{2i}e^{-\theta_L\delta_2})} = \frac{1 - e^{-2\theta_L\delta_1}}{1 - e^{-2\theta_L\delta_2}} \tag{10}$$

Notice that solving this equation for $\delta_1$ and $\delta_2$ requires $\theta_L$. Therefore, we develop the following algorithm to iteratively calibrate $\delta_1, \delta_2, \theta_L$ and $\sigma_L$ together:

(1) Initialize $\delta_1$ and $\delta_2$.
(2) Using $\delta_1$ and $\delta_2$ values, calibrate $\theta_L$ and $\sigma_L$ using MLE.
(3) Plug $\theta_L$ into equation (10). Then $\delta_1$ and $\delta_2$ can be solved together with equation (6).
(4) Repeat steps 2 and 3 until $\delta_1, \delta_2, \theta_L$ and $\sigma_L$ all converge.

*Step 1* Since both $\delta_1$ and $\delta_2$ are small, from equation (10),

$$\frac{\mathrm{Var}(L_{2i} - L_{2i-1})}{\mathrm{Var}(L_{2i+1} - L_{2i})} \approx \frac{1 - e^{-2\theta_L\delta_1}}{1 - e^{-2\theta_L\delta_2}} \approx \frac{\delta_1}{\delta_2} \tag{11}$$

where the variance of intraday return $\mathrm{Var}(L_{2i} - L_{2i-1})$ and the variance of overnight return $\mathrm{Var}(L_{2i+1} - L_{2i})$ are estimated empirically. Then initial $\delta_1$ and $\delta_2$† can be obtained by solving equation (6) and (11).

*Step 2* The conditional densities of $L(t)$ are

$$f(L_{2i}|L_{2i-1}; \theta_L, \hat{\sigma}_1) = \frac{1}{\hat{\sigma}_1\sqrt{2\pi}} \exp\left(-\frac{(L_{2i} - L_{2i-1}e^{-\theta_L\delta_1})^2}{2\hat{\sigma}_1^2}\right)$$

$$f(L_{2i+1}|L_{2i}; \theta_L, \hat{\sigma}_2) = \frac{1}{\hat{\sigma}_2\sqrt{2\pi}} \exp\left(-\frac{(L_{2i+1} - L_{2i}e^{-\theta_L\delta_2})^2}{2\hat{\sigma}_2^2}\right)$$

where

$$\hat{\sigma}_1 = \sigma_L\sqrt{\frac{1 - e^{-2\theta_L\delta_1}}{2\theta_L}}, \qquad \hat{\sigma}_2 = \sigma_L\sqrt{\frac{1 - e^{-2\theta_L\delta_2}}{2\theta_L}}$$

†This initial approximation turns out to be pretty close. On average, the approximates are only off by 0.3% from the final converged values.

The log-likelihood function of $(L_1, \ldots, L_{2N})$ is then

$$
\begin{aligned}
\mathcal{L}(\theta_L, \sigma_L) &= \mathcal{L}(\theta_L, \hat{\sigma}_1, \hat{\sigma}_2) \\
&= \sum_{i=1}^{N} \ln f(L_{2i}|L_{2i-1}; \theta_L, \hat{\sigma}_1) \\
&\quad + \sum_{i=1}^{N-1} \ln f(L_{2i+1}|L_{2i}; \theta_L, \hat{\sigma}_2) \\
&= -\frac{N}{2}\ln(2\pi) - N\ln(\hat{\sigma}_1) \\
&\quad - \frac{1}{2\hat{\sigma}_1^2}\sum_{i=1}^{N}(L_{2i} - L_{2i-1}e^{-\theta_L\delta_1})^2 \\
&\quad - \frac{N-1}{2}\ln(2\pi) - (N-1)\ln(\hat{\sigma}_2) \\
&\quad - \frac{1}{2\hat{\sigma}_2^2}\sum_{i=1}^{N-1}(L_{2i+1} - L_{2i}e^{-\theta_L\delta_2})^2
\end{aligned}
$$

The MLE for $\theta_L$ and $\sigma_L$ are solved numerically from this equation using a quasi-Newton optimization algorithm called limited memory BFGS (Byrd *et al*. 1995).

*Steps 3 and 4* are straightforward, and from our experiments, the algorithm converges fast (generally only 2 to 4 iterations are needed to reach a tolerance of $10^{-6}$).

**2.2.2. Calibration for $Y(t)$.** For $Y(t)$, the constant mean $\mu$ in the OU process is replaced by $\tilde{L}_t = \frac{L_{2i-2}+L_{2i-1}}{2}$. Equation (7) becomes

$$
\begin{aligned}
Y_{79(i-1)+j} &= Y_{79(i-1)+j-1}e^{-\theta\delta} + \frac{L_{2i-2}+L_{2i-1}}{2}(1-e^{-\theta\delta}) \\
&\quad + \sigma\sqrt{\frac{1-e^{-2\theta\delta}}{2\theta}}Z_{79(i-1)+j}, \quad \forall i, j
\end{aligned}
$$

where $i = 1, 2, \ldots, 30$ denotes days; $j = 2, \ldots, 79$ denotes 5-min periods; $\delta = \frac{\delta_1}{78}$ is the effective length of a five-min interval; $Z_{79(i-1)+j}$'s are i.i.d. $N(0, 1)$. Let

$$
\begin{aligned}
a &= e^{-\theta\delta} \\
b_i &= \frac{L_{2i-2}+L_{2i-1}}{2}(1-e^{-\theta\delta}) \\
&= \frac{L_{2i-2}+L_{2i-1}}{2}(1-a), \qquad i = 1, \ldots, 30 \\
\hat{\sigma} &= \sigma\sqrt{\frac{1-e^{-2\theta\delta}}{2\theta}} = \sigma\sqrt{\frac{1-a^2}{2\theta}}
\end{aligned}
$$

where $L_0$ is defined to be 0.

Then

$$
\begin{aligned}
&Y_{79(i-1)+j} - aY_{79(i-1)+j-1} = b_i + \hat{\sigma}Z_{79(i-1)+j}, \\
&\forall i = 1, \ldots, 30, \forall j = 2, \ldots, 79
\end{aligned}
$$

Since $\{L_i | i = 1, \ldots, 2N\}$ is a subsequence of $\{Y_i | i = 1, \ldots, 79N\}$, we have

$$
f(\vec{Y}; \theta, \sigma) = f(\vec{L}, \vec{Y}) = f(\vec{L})f(\vec{Y}|\vec{L})
$$

The log-likelihood for $Y(t)$ is $\ln f(\vec{L}) + \ln f(\vec{Y}|\vec{L})$. The first term $\ln f(\vec{L})$ does not depend on $\theta$ and $\sigma$. We only focus on the second term

$$
\begin{aligned}
&\ln f(\vec{Y}|\vec{L}) \\
&= \ln f(\vec{Y}|Y_{79(i-1)+1} = L_{2i-1}, Y_{79i} = L_{2i}, \forall i = 1, \ldots, 30) \qquad (12) \\
&= \ln f(Y_1, Y_2, \ldots, Y_{79}|Y_1 = L_1, Y_{79} = L_2) \\
&\quad + \ln f(Y_{80}, Y_{81}, \ldots, Y_{79\times 2}|Y_{79} = L_2, \\
&\quad Y_{80} = L_3, Y_{79\times 2} = L_4) \\
&\quad + \cdots\cdots \\
&\quad + \ln f(Y_{79\times 29+1}, \ldots, Y_{79\times 30}|Y_{79\times 29} = L_{58}, \\
&\quad Y_{79\times 29+1} = L_{59}, Y_{79\times 30} = L_{60})
\end{aligned}
$$

The last equality is due to the definition of $Y_i's$. The remaining derivation of MLE formula is rather cumbersome, thus given in appendix 1.

## 3. Trading rules

As mentioned before, training periods of 100 days and 30 days are fixed for the calibration of $L(t)$ and $Y(t)$, respectively. For an intraday trading strategy, the trading period is one day. The three periods are illustrated in figure 1.

After getting the parameters and consequently the variance estimations of both $L(t)$ and $Y(t)$, we select a set of 'best' pairs to be trading candidates for that day. Our ideal trading candidate pair will have a large $Y(t)$ variance and a small $L(t)$ variance. A large $Y(t)$ variance is preferred because more volatile intraday movements lead to more trading opportunities. The preference of small $L(t)$ variances is to ensure that the long-term value of the spread is not volatile. The most desirable situation would be $L(t)$ remaining constant over time while $Y(t)$ fluctuating a lot during the day but always coming back to the constant level.

The procedure to select the 'best' pairs is: first remove all the pairs with negative $\theta_L$, then rank all remaining pairs by $L(t)$'s short-term variance

$$
\frac{\sigma_L^2}{2\theta_L}\left(1 - e^{-2\theta_L(\delta_1+\delta_2)}\right) \qquad (13)
$$

in *ascending* order, record the ranking $r_L$, then rank them again by $Y(t)$'s short-term variance

$$
\frac{\sigma^2}{2\theta}\left(1 - e^{-\frac{2\theta\delta_1}{78}}\right) \qquad (14)
$$

in *descending* order, record the ranking $r_Y$ and finally select the top 25 or 50 or 100 pairs with smallest $r_L + r_Y$. In section 5, we test different pair selection criteria on real data by varying the number of pairs selected.

During day $i$, for each candidate pair stock A and stock B, we make a trade immediately when $Y(t)$, the cumulative return difference between stock A and B, goes out of a 'confidence band'. More specifically, we

(i) short the pair (simultaneously short A and long B) if $Y(t)$ exceeds $\frac{L_{2i-2}+L_{2i-1}}{2} + \epsilon$;
(ii) long the pair (simultaneously long A and short B) if $Y(t)$ drops below $\frac{L_{2i-2}+L_{2i-1}}{2} - \epsilon$.

The value $\epsilon$ is the 98% percentile of the absolute daily change in $L(t)$ values in the past 100 days. If $\epsilon$ is too large, we miss out trading opportunities by executing only few trades; if $\epsilon$

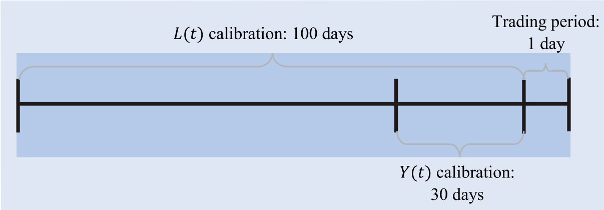



Figure 1. Illustration of formation period and trading period.

is too small, excessive trading leads to the profits of many trades being dwarfed by transaction costs. In simulations, we also used two other $\epsilon$ levels (95 and 90% percentiles) for comparison. In real data back-testing, we stick with the 98% percentile for better performances.†

For each pair-trade, we buy \$1 worth of one stock and short \$1 worth of the other stock. For example, when we long the pair, we buy $\frac{\$1}{P_A}$ shares of stock A and simultaneously short $\frac{\$1}{P_B}$ shares of stock B, where $P_A$ and $P_B$ are their respective prices. Our net position at the outset of each pair-trade is zero.

The open position is closed by making the opposite trades (selling the stock bought, buying back and returning the stock shorted) when either (a) $Y(t)$ reverts back to $\frac{L_{2i-2}+L_{2i-1}}{2}$, or (b) the market closes for the day at 4pm, whichever happens first.

## 4. Simulation

In this section, we demonstrate the validity of our doubly mean-reverting model and the proposed trading strategy using simulation. The goal is to simulate $L_i$'s and $Y_i$'s for a whole year using a set of parameters $\theta_L, \sigma_L, \delta_1, \theta, \sigma$, then apply the strategy on the simulated data. For simplicity, we assume the parameters remain constant in the simulation, although we update them daily for trading on real data. If the model and strategy are well-designed, the profit should be robust for a 'good' set of parameters but not for a 'bad' one.

### *4.1. Simulating $L_i$'s and $Y_i$'s*

First we simulate $L_k$, $\forall k = 1, \ldots, 2N$ by equation (8) and (9) in section 2.2.1.

For each day $i$, define $X_{78(i-1)+j} = Y_{79(i-1)+j+1} - a \times Y_{79(i-1)+j}$ for all $j = 1, \cdots, 78$, where $a = e^{-\theta\delta}$. Now given $L_k$'s, we first generate $X_{78(i-1)+j}$, $j = 1, \cdots 78$ using the conditional distribution [derived in appendix equation (A1)–(A3)]:

$$f(X_{78(i-1)+1}, \ldots, X_{78i} | Y_{79(i-1)} = L_{2(i-1)}, Y_{79(i-1)+1} = L_{2i-1}, Y_{79i} = L_{2i})$$
$$= \frac{\left(\frac{1}{\hat\sigma\sqrt{2\pi}}\right)^{77} \exp\left[-\frac{1}{2\hat\sigma^2}\left(\sum_{j=1}^{78}(X_{79(i-1)+j} - b_i)^2 + (L_{2i} - a^{78}L_{2i-1} - \sum_{j=1}^{78} a^{78-j}X_{78(i-1)+j} - b_i)^2\right)\right]}{\sqrt{\frac{1-a^2}{1-a^{156}}}\exp\left[-\frac{1}{2\hat\sigma^2}\frac{1-a^2}{1-a^{156}}\left(L_{2i} - \frac{1+a^{78}}{2}L_{2i-1} - \frac{1-a^{78}}{2}L_{2i-2}\right)^2\right]}$$

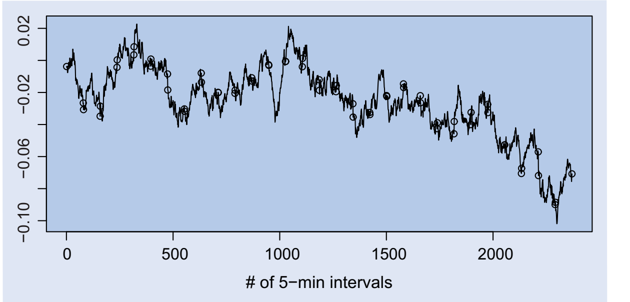



Figure 2. Simulated 30 days' $Y_i$'s and $L_i$'s (circles) for the good pair.

where $b_i$ and $\hat\sigma$ are given in section 2.2.2. In fact, the above conditional density is a multivariate normal density with mean

$$\boldsymbol{\mu} = \begin{bmatrix} 1+a^{154} & a^{153} & \ldots & a^{78} \\ a^{153} & 1+a^{152} & \ldots & a^{77} \\ \vdots & \vdots & \ddots & \vdots \\ a^{78} & a^{77} & \ldots & 1+a^2 \end{bmatrix}^{-1} \times \begin{pmatrix} b_i + a^{77}(L_{2i} - a^{78}L_{2i-1} - b_i) \\ b_i + a^{76}(L_{2i} - a^{78}L_{2i-1} - b_i) \\ \vdots \\ b_i + a(L_{2i} - a^{78}L_{2i-1} - b_i) \end{pmatrix}$$

and variance

$$\boldsymbol{\Sigma} = \hat\sigma^2 \begin{bmatrix} 1+a^{154} & a^{153} & \ldots & a^{78} \\ a^{153} & 1+a^{152} & \ldots & a^{77} \\ \vdots & \vdots & \ddots & \vdots \\ a^{78} & a^{77} & \ldots & 1+a^2 \end{bmatrix}^{-1}$$

Hence, we can generate multivariate normal random variables $X_i$'s, from which $Y_i$'s can be computed straightforwardly.

The simulated $L_i$'s and $Y_i$'s of one sample of 30 days are shown in figure 2 with $L_i$'s indicated by circles.

### *4.2. Choosing parameters*

Choosing particular parameters to make the variances in $L(t)$ and $Y(t)$ small and large, respectively, can easily result in astronomically high profits. However, the simulated spread process trajectory may simply not be achievable by real stock pairs. Therefore, to be more realistic, we use calibrated parameters $\theta_L, \sigma_L, \delta_1, \theta, \sigma$ from real data for simulation.

The data-set is described in detail in section 5. We used the ranking method described in the trading rules to select a 'good pair' and a 'bad pair' as the ones with best and worst parameters respectively in the first trading day.

†It is possible that the optimal threshold is not 98%. We did not experiment on real data to find out the exact optimal value. Zeng and Lee (2014) derived optimal thresholds for maximum profitability per unit of time in a single OU process model framework.

The good pair's parameters are

$$\theta_L = 1.626237, \sigma_L = 0.229202, \delta_1 = 0.0032902,$$

$$\theta = 123.22211, \sigma = 0.341101;$$

and the bad pair's parameters are

$$\theta_L = 9.328820, \sigma_L = 0.198623, \delta_1 = 0.0014689,$$

$$\theta = 482.18402, \sigma = 0.205307.$$

We simulate separately using both sets of parameters and compare results.

### 4.3. *Simulation trading results*

Following the procedure described in section 4.1, we simulate the spread series of one year (250 days) for a large number of simulations. Each time, we apply the same strategy and calculate the profit and loss (PNL). In table 1, the trading results are shown for both the good pair (top half) and the bad pair (bottom half), each pair for three different threshold $\epsilon$ levels, each level for 400 simulations.

For the good pair, in every one of the 1200 simulations, the whole year's profit is positive. On average, there are 29/55/92 trades in the whole 250-day period, for the three $\epsilon$ levels respectively; 23/41/68 of them being profitable. The annual Sharpe ratios and annual returns are 3.325/4.447/5.668 and 80.2%/180.9%/409.5%.

Meanwhile for the bad pair, although the number of trades triggered is comparable to the good pair, the profitability is much worse. Slightly less than half the trades are winning ones compared with well over 70% for the good pair. As a result, the Sharpe ratios and annual returns are close to zero.

These simulation results validated our model by showing that a good pair identified by the model can indeed provide stable profits while a bad pair cannot.

For the good pair, as the threshold $\epsilon$ is lowered from 98 to 90% percentile, the average number of trades per simulation is more than tripled from 29 to 92. Annual Sharpe ratio and annual return increased significantly due to the larger number of trades triggered. However, both the winning percentage and the profit per trade dropped slightly, from 77.5% and 78 bps to 74.1% and 70 bps, respectively. This is expected since lowering the threshold means lowering the 'standard' in identifying trading opportunities.

The fact that the good pair provides highly profitable results regardless of threshold level and the bad pair does not—also regardless of $\epsilon$ level, proves that (1) identifying correct pairs are much more important than choosing threshold level and (2) the trading results (especially in terms of profit per trade) are not sensitive to the selection of $\epsilon$. The second point is also verified on real data.

The results in this section are presented without transaction costs. But as will be discussed and analyzed in detail in section 5, the profits for the good pair are large enough to cover any reasonable transaction costs estimation. Note that there is only one pair in the simulation trading for simplicity, but in real data trading strategy we have at least 25 potential pairs every day.

## 5. Back-testing on real data

### 5.1. *The data*

The data for this study are from the NYSE Trade and Quote database on Wharton Research Data Services (WRDS) platform. Tick data for 26 oil company stocks during trading hours 9:30 AM to 4:00 PM are downloaded, then processed to five-min time series by extracting the first tick price right after each five-min mark (i.e. 9:30:00, 9:35:00 etc.).

Initially, 31 stocks with the largest market capital in the Oil Refining & Marketing industry group traded on NYSE and/or NASDAQ were downloaded. After processing tick prices, we removed five stocks with too many missing data,† all of them non-US companies. The information of the 26 stocks is shown in table 2. Notice that our trading universe comprises stocks with extremely large market caps and liquidity, compared with most studies in the pairs trading literature. All the companies have market caps over \$12 billion. On average, there are 68.0 tick prices per minute for each stock. The availability of such a high frequency database is critical for this study.

We first did the back-testing for the period 2 January 2013–29 April 2015 (579 business days). Then, in order to examine the performance of the strategy during market turmoil, we back-tested on an earlier period from 2 July 2007 to 31 December 2008 (374 business days). It is worth pointing out that although the data periods in this study may seem short compared with the prior literature, the high-frequency nature of the data-set makes it actually larger, in terms of numbers of data points per stock. Five hundred and seventy-nine days with 79 data points per day are equivalent to $579 \times 79/252 = 181$ years of daily data.

For the more recent 2013–2015 period, we used the 26 stocks described before. For the earlier period however, five stocks (CLR, CVE, MPC, PAA and PSX) either had not started trading or did not have enough liquidity. We used the other 21 stocks for the 2007–2008 period.

### 5.2. *Transaction costs and return calculation*

Pairs trading strategies aim to capture stable and modest profits from market mispricing. As a result, transaction costs can have a major impact on the profitability. Bowen *et al.* (2010) found that a moderate level of 15 basis points (bps) transaction costs‡ would reduce the excess returns by more than 50% on one year's data of 100 UK stocks. Testing on the US equity market in the period 1963–2009, Do and Faff (2012) found that profitability of the simple algorithm from the original Gatev *et al.* (2006) paper was largely diminished after various transaction costs.

The magnitude of transaction costs depends on many factors such as the type and size of the investor (institutional vs. retail), liquidity of the particular security and the size and timing of the order. Gatev *et al.* (2006) estimated a large transaction

†A missing data refers to the situation where there is no trade in a five-min interval. There are only 11 total missing data points among the remaining 26 stocks in the whole analysis period. We interpolate these 11 missing points.

‡It was unclear to us whether these transaction costs were per round trip or per trade.

Table 1. Simulation Trading Results (Top half: the good pair; bottom half: the bad pair).

| $\epsilon$ level | Avg num of trades in 250 days | Avg num/ percentage of profitable trades | | Avg num/ percentage of trades that reverts back to mean | | Avg pnl | Max pnl in 400 simulations | Min pnl in 400 simulations | Annual Sharpe ratio | Annual return | Profit per trade in bp (breakeven transac. cost) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 98% | 29.4 | 22.8 | 77.5% | 1.4 | 4.8% | 0.230 | 0.610 | 0.028 | 3.325 | 80.2% | **78** |
| 95% | 54.5 | 41.4 | 75.9% | 4.6 | 8.4% | 0.404 | 0.800 | 0.052 | 4.447 | 180.9% | 74 |
| 90% | 91.9 | 68.1 | 74.1% | 13.1 | 14.3% | 0.644 | 1.111 | 0.320 | **5.668** | **409.5%** | 70 |
| 98% | 27.5 | 13.4 | 48.8% | 0.1 | 0.5% | -0.003 | 0.095 | −0.063 | −0.165 | −0.7% | −1.0 |
| 95% | 48.2 | 23.7 | 49.3% | 0.6 | 1.3% | 0.000 | 0.107 | −0.112 | −0.039 | 0.0% | 0.0 |
| 90% | 77.7 | 38.6 | 49.7% | 2.6 | 3.4% | 0.005 | 0.134 | −0.136 | 0.074 | 1.2% | 0.6 |

Table 2. Oil Company Stocks Descriptions and Statistics (as of December 2014).

| | NYSE ticker | mkt cap in billion | Annual revenue | Avg daily volume from Google finance (in M) | Avg daily # of trades from tick data | Min daily # of trades from tick data |
|---|---|---|---|---|---|---|
| Exxon Mobil Corporation | XOM | 393.87 | 420,836 | 12.71 | 62977 | 30546 |
| Royal Dutch Shell plc (ADR) | RDSA | 215.73 | 451,235 | 2.5 | 10478 | 4281 |
| Chevron Corporation | CVX | 207.98 | 220,264 | 7.29 | 42279 | 20506 |
| Total SA (ADR) | TOT | 132.18 | 227,969 | 1.4 | 5773 | 2749 |
| BP plc (ADR) | BP | 119.95 | 379,136 | 6.78 | 22895 | 8966 |
| ConocoPhillips | COP | 82.85 | 56,185 | 8.07 | 34616 | 15317 |
| Occidental Petroleum Corporation | OXY | 61.82 | 24,561 | 6.6 | 28982 | 14690 |
| Statoil ASA(ADR) | STO | 57.53 | 87,781 | 2.85 | 7061 | 3272 |
| Petroleo Brasileiro Petrobras SA (ADR) | PBR | 57.49 | 141,462 | 48.82 | 65939 | 27431 |
| EOG Resources Inc | EOG | 49.35 | 14,290 | 6.21 | 24409 | 11296 |
| Suncor Energy Inc. (USA) | SU | 44.34 | 35,398 | 4.56 | 19876 | 10121 |
| Anadarko Petroleum Corporation | APC | 39.74 | 14,581 | 6.05 | 30750 | 10493 |
| Phillips 66 | PSX | 38.83 | 171,596 | 4.71 | 23828 | 10169 |
| Canadian Natural Resource Ltd (USA) | CNQ | 35.85 | 14,182 | 4.73 | 16446 | 5345 |
| Valero Energy Corporation | VLO | 25.47 | 138,074 | 7.05 | 46740 | 20537 |
| Marathon Petroleum Corp | MPC | 25.19 | 100,248 | 3.62 | 26820 | 13113 |
| Devon Energy Corp | DVN | 24.16 | 10,397 | 4.34 | 22804 | 8650 |
| Apache Corporation | APA | 22.31 | 16,054 | 4.53 | 22623 | 7876 |
| Hess Corp. | HES | 21.87 | 22,247 | 3.75 | 18808 | 7921 |
| Pioneer Natural Resources | PXD | 20.69 | 3,506 | 2.63 | 14786 | 7100 |
| Marathon Oil Corporation | MRO | 18.88 | 14,959 | 7.75 | 32701 | 16308 |
| Plains All American Pipeline, L.P. | PAA | 18.27 | 42,249 | 1.66 | 6058 | 3627 |
| Cenovus Energy Inc (USA) | CVE | 15.60 | 16,389 | 1.85 | 7051 | 3114 |
| Continental Resources, Inc. | CLR | 13.18 | 3,455 | 3.72 | 10450 | 4899 |
| EQT Corporation | EQT | 12.91 | 1,862 | 1.82 | 12835 | 5003 |
| Cabot Oil & Gas Corporation | COG | 12.83 | 1,746 | 6.11 | 35182 | 16146 |

cost of 162 bps per pair per round-trip for the period 1962–2002. However, the figure has been vastly reduced in recent years due to technology advances. Avellaneda and Lee (2010) chose a transaction cost of 10 bps per round-trip pair-trade. Bogomolov (2013) used a more conservative estimation of 40 bps per round-trip pair-trade.

Transaction costs mainly consist of commissions, bid-ask spreads and short selling costs (Do and Faff, 2012). In this paper, where we consider trading a pool of highly liquid large cap US stocks from the perspective of hedge funds, commissions and short selling costs are negligible. The bid-ask spread—also known as bid-ask bounce, slippage, or market impact—can be estimated both directly and indirectly. Since all stocks in our investment universe are highly liquid, we used one of the lower estimates in literature as our baseline number, 10 bps per round-trip per pair-trade as in Avellaneda and Lee (2010). An alternative way to estimate bid-ask spreads is to use delayed trading as a proxy. As argued by Gatev *et al.* (2006), if a trade is made one period (one day in their case) after the divergence signal is identified, instead of immediately, the drop in return would be a rough estimate of half the round-trip transaction cost. As will be seen in detail in the next subsection, the proxy result is consistent with our selection of 10 bps.

Another tricky issue in comparing pairs trading studies is the return calculation, which warrants two considerations. The first is the leverage ratio. Pair trading, as a market neutral strategy, has a zero net initial investment (long \$1 and short \$1 for example) in theory. But it is not zero in practice. In order to short stocks, we need to put margin deposits in the brokerage account. The number of dollars of market exposure allowed for every dollar in the margin account is called the margin leverage ratio. We compute our returns as profit or loss divided by the margin, as does the literature.

Avellaneda and Lee (2010) used a 4:1 leverage, which means \$2 long and \$2 short is permitted for every dollar deposited. Gatev *et al.* (2006) defined excess return as the profit/loss for each \$1 long–\$1 short pair trade. This implied a 2:1 leverage. Large institutional investors can generally get large leverages. In this paper, we assumed a 5:1 leverage. Note that although return numbers largely depend on the leverage selection, Sharpe ratios do not, hence are more suited to be compared across studies. Annualized Sharpe ratio is calculated as

$$\text{Annualized Sharpe ratio} = \frac{\text{Annualized return}}{\text{Annualized volatility}} = \frac{E(\text{daily return}) \times 252}{sd(\text{daily return}) \times \sqrt{252}}$$

The second consideration in return calculation is return on *committed* capital vs. return on *actual employed* capital (see Gatev *et al.* 2006). The former uses capital related to all the selected pairs for a day; the latter only uses capital related to those pairs that are traded in the day. Gatev *et al.* argue that '... to the extent that hedge funds are flexible in their sources and uses of funds, computing excess return relative to the actual capital employed may give a more realistic measure of the trading profits.' When discussing results in the next subsection, we refer to the return on actual employed capital (but we displayed both types of returns and corresponding Sharpe ratios in tables 3–6).

To illustrate all the above points, consider an example where we selected 20 pairs each trading day based on calibration results. The margin deposit required for each \$1 long–\$1 short trade is

$$\frac{\text{Gross market exposure}}{\text{leverage}} = \frac{\$2}{5} = \$0.4$$

Assuming a 10-bp transaction cost per round trip pair trade, then

(a) daily return on ***committed*** capital
$$= \frac{\text{daily net PNL}}{\text{margin position}} = \frac{(\text{daily PNL}) - \$0.001 \times (\#\text{ of trades})}{\$0.4 \times (\#\text{ of } \boldsymbol{pairs})}$$

(b) daily return on ***actual employed*** capital
$$= \frac{\text{daily net PNL}}{\text{margin position}} = \frac{(\text{daily PNL}) - \$0.001 \times (\#\text{ of trades})}{\$0.4 \times (\#\text{ of } \boldsymbol{trades})}$$

Consider the following three scenarios:

(i) If in one day the daily profit is \$0.3 with 2 trades, return (a) is $\frac{0.3-0.002}{0.4\times 20} = 3.725\%$; return (b) is $\frac{0.3-0.002}{0.4\times 2} = 37.25\%$

(ii) If the daily loss is $-\$1.2$ with 5 trades, return (a) is $\frac{-1.2-0.005}{0.4\times 20} = -15.06\%$; return (b) is $\frac{-1.2-.005}{0.4\times 5} = -60.25\%$

(iii) If no trade in one day, the daily return is 0 for both returns (a) and (b).

### 5.3. *Empirical results*

**5.3.1. June 2013–April 2015.** As described in section 3, for each day, we use the previous 100 days' data to select best pairs (i.e. the formation period is 100 days, trading period is one day). Therefore the first five months (January to May 2013) in the data is left out for calibration and the trading starts from June 2013. We report the trading results for different pair selection criteria and transaction cost levels in table 3. As in the simulation, among three threshold $\epsilon$ levels (98, 95 and 90% percentiles), the 98% threshold yields the highest profit per trade. Thus, the results reported in this section are based on threshold $\epsilon = 98\%$.

The top part of table 3 shows the results for four different pair selection criteria. As discussed, for each trading day we skipped the pairs with negative estimated $\theta_L$ and then selected top 25/50/75/100 pairs as trading candidates according to their rankings.

As expected, the average number of trades per day depends on the number of pairs selected: the more pairs we select, the higher number of trades per day. The profit per trade ranges from 14 to 20 bps after deducting the 10 bps transaction costs. We selected the 'top 50' as our baseline strategy since it has the best overall performance metrics.

The middle part of table 3 shows the impact of transaction costs for the baseline 'top 50' selection method. Without transaction costs, the profit per trade is 30 bps. In other words, the break-even transaction cost is 30 bps on this data-set period. Even if we relax the estimate to a more conservative 20 bps per trade, we still have a 10 bps per trade profit and a 2.338 annualized Sharpe ratio, compared with the Sharpe ratio of 1.51 from 2003 to 2007 by Avellaneda and Lee (2010).

Finally, we rerun the baseline strategy but with the wait-one-period constraint mentioned before. Gatev *et al.* (2006) argued that when a spread is identified, it is more likely that the winner stock price is an ask price and the loser stock price is a bid price. After waiting a period, five min in our strategies, the prices are presumably equally likely to be bid or ask prices. Therefore, the drop in PNL after waiting for five min as opposed to making the trade at the moment of divergence signal, would be a proxy of half the bid ask bounce—the other half happening at convergence, in the same vein. Of course, part of this drop could also be attributed to the natural mean-reversion in prices. Comparing the second and the bottom lines in table 3, the drop in profit per trade is 6.5 bps. If the drop was exclusively due to the bid-ask bounce, the proxy would be $6.5 \times 2 = 13$ bps,†, which is consistent with our direct estimation of 10 bps.

To see the stability of the strategies over time, we plotted the PNL and returns over the trading period of almost two years, and reported the results by quarter. As seen from figure 3, both the PNL and returns increase quite stably over the period.

In table 4, quarterly results are reported for the baseline 'top 50' strategy with 10 bps transaction costs. Out of the seven

†Out of the 3.97 average trades per day, only 0.21 or about 5% trades converged, i.e. reverting back to $\frac{L_{2i-2}+L_{2i-1}}{2}$ level before day's close. Most other spreads were on their way toward $\frac{L_{2i-2}+L_{2i-1}}{2}$ when market closed. As stated in the trading rules, we close all pairs at market's close. Hence, the waiting-one-period proxy applies to most pairs for only the opening half of the trade.

Table 3. June 2013–April 2015 trading results for different pair selections and transaction costs.

| Pair selection criteria | Transaction cost | # of total trades | Avg # of trades per day | % of winning trades | Total PNL($) after TC | Ann. Sharpe based on employed capital | Ann. return on employed capital | Ann. Sharpe based on committed capital | Ann. return on commited capital | PNL per trade in bp (breakeven TC) |
|---|---|---|---|---|---|---|---|---|---|---|
| **top 25** | 10 bp | 1050 | 2.20 | 53.9% | 1.946 | 3.353 | 148.1% | 1.256 | 10.4% | 17.6 |
| **top 50** | 10 bp | 1897 | 3.97 | 54.2% | 3.710 | 3.885 | 187.8% | 1.738 | 10.1% | 19.6 |
| **top 75** | 10 bp | 2694 | 5.64 | 53.2% | 4.371 | 4.138 | 192.7% | 1.694 | 7.9% | 16.2 |
| **top 100** | 10 bp | 3507 | 7.34 | 52.9% | 4.879 | 4.078 | 190.8% | 1.592 | 6.6% | 13.9 |
| top 50 | **0 bp** | 1897 | 3.97 | 59.7% | 5.607 | 5.386 | 346.8% | 2.609 | 15.7% | 29.6 |
| top 50 | **10 bp** | 1897 | 3.97 | 54.2% | 3.710 | 3.885 | 187.8% | 1.738 | 10.1% | 19.6 |
| top 50 | **20 bp** | 1897 | 3.97 | 49.0% | 1.813 | 2.338 | 85.2% | 0.850 | 4.7% | 9.6 |
| top 50 wait one period | 10 bp | 1897 | 3.97 | 53.1% | 2.481 | 3.129 | 125.3% | 1.244 | 6.6% | 13.1 |

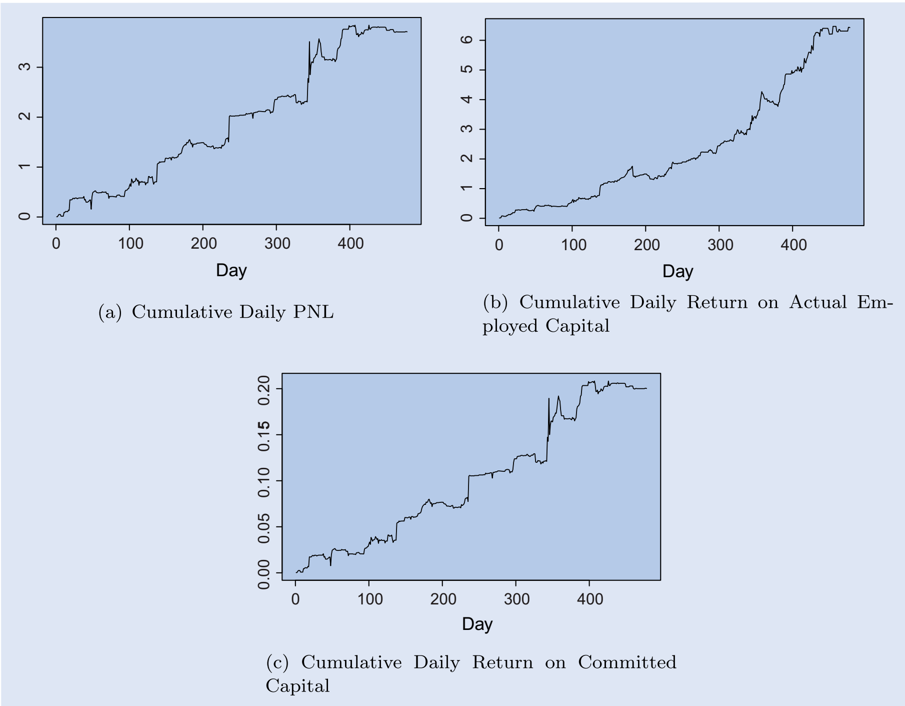



Figure 3. Back-testing performance for period June 2013–April 2015 (top 50 pairs, threshold = 98%, transaction costs = 10 bps).



quarters in the period, six are winning quarters and one breaks even.

**5.3.2. The year 2008.** As a contrarian strategy, pairs trading tends to perform better during markets downturns (Do and Faff, 2010). The results during the months of recent oil market crash (July 2014–January 2015) show a promising performance: the average monthly PNL without transaction costs during this seven-month span is 43% higher than the whole period. To further test this hypothesis and verify our strategies, we repeated the analysis on the whole year 2008, during which the oil market spiked to an all-time high of $145 per barrel in July then crashed to $30 in December amid global financial crisis, as shown in the US crude oil benchmark index West Texas Intermediate (WTI) history price chart (figure 4).

The original 26 stocks we selected were not all available for the period July 2007–December 2008 (the last five months of 2007 were needed for calibration). Three (CVE, MPC and PSX) had not started trading; two (CLR and PAA) had too many missing data due to low liquidity. Therefore we used 21 stocks and $\binom{21}{2} = 210$ total pairs for this period.

The results for 2008 are presented in table 5. The average number of trades per day for the baseline 'top 50' strategy remarkably increases to 6.26, from 3.97 in the 2013–2015 period. The higher numbers of trading opportunities were driven by higher volatilities in stock prices.† The annual volatilities of all stocks in the 2013–2015 period range from 11.1 to 37.7%

†We recorded the number of times each stock is selected and traded over the whole period, to see if there are any discrepancies. Not surprisingly, they have a strong relationship with stocks volatility. The correlation between a stock's volatility and number of times it being selected and traded are 88.2% and 86.6% respectively.

Table 4. June 2013–April 2015 Trading Results by Quarter (Top 50 pairs, transaction costs = 10 bps).

| Quarter | # of total trades | % of winning trades | Total PNL ($) | Annl Sharpe based on employed capital | Annl return on employed capital | Annl Sharpe based on committed capital | Annl return on commited capital | PNL per trade in bp |
|---|---|---|---|---|---|---|---|---|
| 2013 Q3 | 209 | 53.1% | 0.070 | 1.96 | 52.4% | 0.39 | 1.3% | 3 |
| 2013 Q4 | 277 | 49.5% | 0.661 | 4.48 | 475.6% | 2.82 | 14.3% | 24 |
| 2014 Q1 | 212 | 55.2% | 0.304 | 0.96 | 27.1% | 3.06 | 6.5% | 14 |
| 2014 Q2 | 188 | 54.8% | 0.690 | 5.56 | 222.8% | 2.50 | 14.6% | 37 |
| 2014 Q3 | 230 | 53.0% | 0.174 | 4.43 | 132.8% | 1.34 | 3.5% | 8 |
| 2014 Q4 | 500 | 56.8% | 1.489 | 5.41 | 459.9% | 2.43 | 34.3% | 30 |
| 2015 Q1 | 204 | 50.0% | −0.007 | 3.91 | 172.0% | −0.06 | −0.2% | 0 |

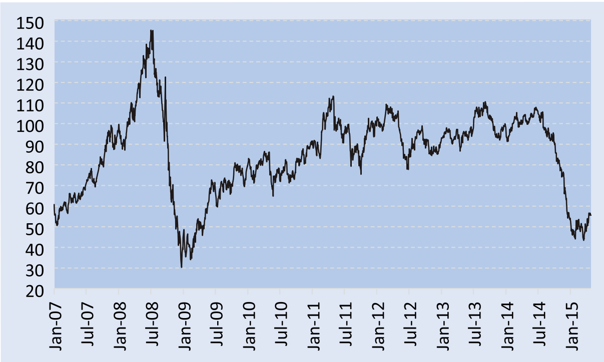



Figure 4. WTI Spot Price (dollar per barrel) 2007–2015.

with mean 23.1%. In 2008, they range from 31.2 to 63.0% with mean 50.3%. The higher volatilities in stock prices were driven by crude oil's volatile movements (see Geman 2015). WTI's volatilities were 27% in June 2013–April 2015 and 55% in 2008.

Furthermore, the strategies' returns are much larger in 2008. For the baseline 'top 50' strategy with a 10 bps transaction cost, the annualized return is 187.8% for the recent period and 1787.6% for 2008. The 7.17 Sharpe ratio of 2008 also dominates the 3.89 in 2013–2015. The breakeven transaction cost is 88 bps for 2008 compared with 30 bps for 2013–2015. More impressively, the quality performance in 2008 is consistent throughout the year as shown in the quarterly breakdown in table 6. In fact, the returns before transaction costs are positive in every *month*.

Lastly in table 5, the drop in profit per trade when we apply the wait-one-period constraint is $78 - 54 = 24$ bps. As in section 5.3.1, less than 13 bps of the drop is likely due to the bid-ask bounce, while the rest is presumably caused by the convergence of the spread, which is more prominent in the more volatile 2008. The fact that our strategy is still profitable in both periods after posing the wait-one-period constraint shows its robustness to the speed of execution (see Bowen *et al*. 2010).

Within 2008, the trading strategy performed extremely well in the second half of year (figure 5), coinciding with the nose-dive of WTI price. This is clear from table 6: the order of the strategy performance of the four quarters is the exact reverse order of WTI's quarterly performance. The monthly returns of the baseline trading strategy (not reported) and WTI index are strongly negatively correlated, with a correlation coefficient −0.78. This figure is only −0.04 in the 2013–2015 period. We identified two reasons for this significant difference. The first is again volatility: higher WTI and stock volatilities can much better translate the plummeting prices into trading profits through more trading opportunities. After all, pairs trading fundamentally relies on temporary relative mispricing of two stocks. The second reason is that the oil market crash in 2008 was more dramatic and more impactful. In 2008, the WTI index plunged almost 80% in less than six months, compared with a drop of almost 60% in seven months from mid-2014 to early 2015. The stocks in our trading universe lost 49% on average during the 2008 crash and only 29% in the recent market turmoil.

## 6. Conclusion and discussion

This paper introduces a doubly mean-reverting process to model stock price spreads. We developed intraday pairs trading strategies using high frequency data with five-min intervals on oil company stocks. Results from both simulations and real data back-testing display significant realized profits. In particular, we are able to achieve a 3.9 annualized Sharpe ratio and a 188% annualized return after transaction costs for the period June 2013–April 2015. We also tested the hypothesis that pairs trading strategies perform better in market turmoil by back-testing on 2008 data. The impressive Sharpe ratio and annualized return of 7.2 and 1788%, respectively, in that year underpin this theory as well as the fundamental relationship that oil company stocks are driven by crude oil price. We also showed that the strategy is robust to both speed of execution and reasonable transaction costs.

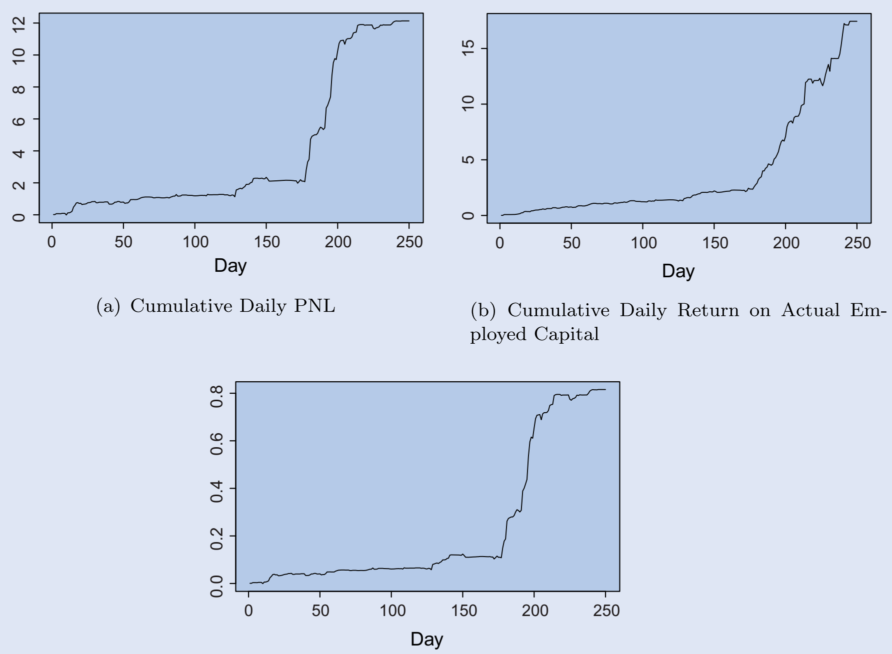
The pair selection procedure proposed in the paper is based on our model design and calibration. However, what is the distinct value added by the model and calibration? In other words, what would happen if we replace *model* variances in equation (13) and (14) with *sample* variances and perform the same strategy on pairs selected by ranking sample variances (without calibration)? The idea of this alternative algorithm is similar: to identify pairs with stable long-term trends and relatively volatile intraday movements. However, experiments with real data on this alternative algorithm show inferior performance

Table 5. 2008 Trading results for different pair selections and transaction costs.

| Pair selection criteria | Trans-action cost | # of total trades | avg # of trades per day | % of winning trades | Total PNL($) after TC | Ann. Sharpe based on employed capital | Ann. return on employed capital | Ann. Sharpe based on committed capital | Ann. return on commited capital | PNL per trade in bp (breakeven TC) |
|---|---|---|---|---|---|---|---|---|---|---|
| **top 25** | 10 bp | 794 | 3.18 | 60.2% | 6.260 | 6.103 | 1571.5% | 4.370 | 85.9% | 78.8 |
| **top 50** | 10 bp | 1566 | 6.26 | 60.2% | 12.133 | 7.169 | 1787.6% | 4.333 | 82.4% | 77.5 |
| **top 75** | 10 bp | 2279 | 9.12 | 60.0% | 17.406 | 7.679 | 1922.5% | 4.164 | 77.7% | 76.4 |
| **top 100** | 10 bp | 2980 | 11.92 | 60.1% | 23.405 | 7.311 | 1975.3% | 4.005 | 78.4% | 78.5 |
| top 50 | **0 bp** | 1566 | 6.26 | 62.7% | 13.699 | 8.223 | 2938.9% | 4.763 | 97.3% | 87.5 |
| top 50 | **10 bp** | 1566 | 6.26 | 60.2% | 12.133 | 7.169 | 1787.6% | 4.333 | 82.4% | 77.5 |
| top 50 | **20 bp** | 1566 | 6.26 | 57.0% | 10.567 | 6.084 | 1071.2% | 3.874 | 68.7% | 67.5 |
| top 50 | **40 bp** | 1566 | 6.26 | 51.6% | 7.435 | 3.834 | 349.2% | 2.867 | 44.2% | 47.5 |
| top 50 wait one period | 10 bp | 1566 | 6.26 | 58.0% | 8.415 | 6.476 | 754.6% | 4.002 | 51.9% | 53.7 |

Table 6. 2008 Trading Results by Quarter (Top 50 pairs, transaction costs = 10 bps).

| Quarter | # of total trades | % of winning trades | Total PNL ($) | Annl Sharpe based on employed capital | Annl return on employed capital | Annl Sharpe based on committed capital | Annl return on commited capital | PNL per trade in bp | WTI quarterly return |
|---|---|---|---|---|---|---|---|---|---|
| 2008 Q1 | 302 | 62.3% | 1.021 | 8.05 | 1380.8% | 4.43 | 23.3% | 34 | 5.2% |
| 2008 Q2 | 193 | 51.8% | 0.179 | 3.17 | 100.9% | 1.56 | 3.6% | 9 | 35.6% |
| 2008 Q3 | 511 | 61.1% | 4.286 | 8.36 | 3541.4% | 5.25 | 132.3% | 84 | −29.5% |
| 2008 Q4 | 560 | 61.3% | 6.648 | 8.83 | 12314.0% | 6.23 | 276.3% | 119 | −62.2% |




(a) Cumulative Daily PNL

(b) Cumulative Daily Return on Actual Employed Capital

(c) Cumulative Daily Return on Committed Capital

Figure 5. Back-testing performance for period 2008 (top 50 pairs, threshold = 98%, transaction costs = 10 bps).

to ours, possibly due to the following beneficial features of our methodology: (1) the conditional variance, based on which we select pairs, can be very different from (unconditional) sample variance; (2) choosing pairs of the mean-reverting patterns, our algorithm eliminates pairs with negative $\theta_L$. Yet, sample variance (without calibration) does not provide information about which pairs have negative $\theta_L$, i.e. not mean-reverting. In other words, if we use only sample variance, then only the magnitude—not the direction—of the changes at each step would matter; (3) based on our model, the lengths of trading daytime ($\delta_1$) and overnight period ($\delta_2$) can be estimated (see section 2.2.1), and the $Y_t$ variance also depends on $\delta_1$. Using only the sample variance without models and calibration, this information would not be available and utilized; (4) by modelling the stock price movements as a stochastic process, we are able to simulate prices using different parameters, and design and test strategies without using real data, to avoid overfitting.

Concerning the normality assumption in the model, it is well known that stock prices as well as stock returns exhibit fat tails, especially on small scales such as the 5 min used in this paper (Cont 2001, Schoutens 2003, Leoni 2014), yet we used Brownian motion to describe the diffusions. To investigate how fat-tailed the data really are, we computed the excess kurtosis of residuals from the calibration of $L_t$ and $Y_t$. For the daily series $L_t$, the excess kurtosis of all residuals is 2.2 and 5.5, respectively, for the two periods. The five-min series $Y_t$ indeed has much fatter tails, with excess kurtosis 14.0 and 13.9 for the two periods. There are two potential issues with the existence of fat tails in the data: (1) the fat tails may affect the profit of the strategy; (2) the existence of fat tails may render the calibration unreliable.

For the first issue, intuitively the fat tails may have a small positive impact on the profitability of the strategy. Yet we have not found evidence from the $Y_t$ residuals. The pairs selected as trade candidates have an average excess kurtosis of 16.2 for 2008 and 17.0 for 2013–2015; for all the other pairs, the numbers are 12.6 and 12.9. The slightly fatter tails in the trade candidate pairs are probably due to their large $Y_t$ variances. In fact, the fat tails are not the main source of the strategy's profitability, though they may help a little. As discussed in the simulation section, the strategy works extremely well on processes that follow our model, i.e. in the absolute absence of fat tails.

As for the second issue, we tested the reliability of the calibration by simulation. We simulated sample paths of an OU process according to equation (7) using known values for parameters $\mu$, $\theta$ and $\delta$, but replaced $Z_i$ with a fat tail distributed random number with mean 0, variance 1, skewness 0, and excess kurtosis 14. Then we calibrated the parameters assuming that the $Z_i$'s are standard normal random numbers. Simulation results show virtually no bias in the calibration, at least at this moderate level of excess kurtosis. Even if there were small bias introduced, the impact on our algorithm is minimal, since what matters is the relative magnitude of parameters rather than their absolute values. If a certain parameter is biased due to fat tails, it is likely biased for all the pairs to some degree, as the excess kurtosis is observed in all pairs at similar levels.

There are several possible directions for future research. First, the frequencies of the two processes may be changed. To utilize stock price data with high liquidity, intervals smaller than five min could be used as the frequency for $Y(t)$. On the other hand, we can increase the interval length of $L(t)$, to make the holding period longer, enabling overnight positions. Second, some details in the strategy implementation may be refined to achieve higher returns, such as the optimal thresholds to enter and exit a trade, and the training windows of 100 and 30 days. This has to be done in a careful manner to avoid over-fitting and data-snooping biases. Third, other types of data available on the NYSE Trade and Quote database can be included in the model. In particular, volume data may be used to adjust for different trading intensities throughout the day. Lastly, the model can be extended (a) from pairs trading to groups trading (also known as generalized pairs trading) with the simultaneous buying and selling of more than two stocks which co-move in some pattern; (b) from stock pairs within the Oil Refining industry group to cross-industry pairs, e.g. those in the highly related Oil Services & Equipment industry group and (c) from Brownian motion to the more general Lévy process to address the fat tail issue discussed above.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References


Amarasingham, A., Harrison, M.T., Hatsopoulos, N.G. and Geman, S., Conditional modeling and the jitter method of spike resampling. *J. Neurophysiol.*, 2012, **107**, 517–531.

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

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

Bowen, D., Hutchinson, M.C. and O'Sullivan, N., High frequency equity pairs trading: Transaction costs, speed of execution and patterns in returns. *J. Trad.*, 2010, **5**, 31–38.

Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C., A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 1995, **16**, 1190–1208.

Chang, L.-B., Conditional modeling and conditional inference. PhD Thesis, Brown University, 2010.

Chang, L.-B., Borenstein, E., Zhang, W. and Geman, S., Maximum likelihood features for generative image models. *Ann. Appl. Stat.*, 2015. In revision.

Chang, L.-B. and Geman, S., Empirical scaling laws and the aggregation of non-stationary data. *Physica A*, 2013, **392**, 5046–5052.

Chang, L.-B., Geman, S., Hsieh, F. and Hwang, C.-R., Invariance in the recurrence of large returns and the validation of models of price dynamics. *Phys. Rev. E*, 2013, **88**, 022116.

Cont, R., Empirical properties of asset returns: Stylized facts and statistical issues. *Quant. Finance*, 2001, **1**, 223–236.

Cummins, M. and Bucca, A., Quantitative spread trading on crude oil and refined products markets. *Quant. Finance*, 2012, **12**, 1857–1875.

Do, B. and Faff, R., Does simple pairs trading still work? *Financ. Anal. J.*, 2010, **66**, 83–95.

Do, B. and Faff, R., Are pairs trading profits robust to trading costs? *J. Financ. Res.*, 2012, **35**, 261–287.

Gatev, E., Goetzmann, W.N. and Rouwenhorst, K.G., Pairs trading: Performance of a relative-value arbitrage rule. *Rev. Financ. Stud.*, 2006, **19**, 797–827.

Geman, H., *Agricultural Finance: From Crops to Land, Water and Infrastructure*, 2015 (Wiley: Chichester, UK).

Geman, H. and Eleuterio, Vergel, P., Investing in fertilizer-mining companies in times of food scarcity. *Resour. Policy*, 2013, **38**, 470–480.

Leoni, P., *The Greeks and Hedging Explained*, 2014 (Palgrave Macmillan: London, UK).

Lin, Y.X., McCrae, M. and Gulati, C., Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *J. Appl. Math. Decis. Sci.*, 2006, **6**, 1–14.

Lo, A.W. and MacKinlay, A.C., Data-snooping biases in tests of financial asset pricing models. *Rev. Financ. Stud.*, 1990, **3**, 431–467.

Schoutens, W., *Lévy Processes in Finance: Pricing Financial Derivatives*, 2003 (Wiley: Chichester, UK).

Tufano, P., The determinants of stock price exposure: Financial engineering and the gold mining industry. *J. Finance*, 1998, **53**, 1015–1052.

Vidyamurthy, G., *Pairs Trading: Quantitative Methods and Analysis*, 2004 (Wiley: Hoboken, NJ).

Zeng, Z. and Lee, C.G., Pairs trading: Optimal thresholds and profitability. *Quant. Finance*, 2014, **14**, 1881–1893.


## Appendix 1. Derivation of likelihood function for $Y(t)$ and the MLE

The 30 summands in equation (12) are similar. Define $Y_0 = L_0 = 0$ so that all 30 terms have the form $\ln f(Y_{79(i-1)+1}, Y_{79(i-1)+2}, \ldots, Y_{79i} | Y_{79(i-1)} = L_{2(i-1)}, Y_{79(i-1)+1} = L_{2i-1}, Y_{79i} = L_{2i})$. We work on the first term for now. Since $(Y_1, Y_2, \ldots, Y_{79})$ are not jointly normal, we perform change of variables. Let

$$\begin{aligned} X_1 &= Y_2 - aY_1 \\ X_2 &= Y_3 - aY_2 \\ &\cdots \\ X_{77} &= Y_{78} - aY_{77} \\ X_{78} &= Y_{79} - aY_{78} \end{aligned}$$

Then $X_1, \ldots, X_{78}$ are i.i.d. $N(b_1, \hat{\sigma}^2)$, where

$$\begin{aligned} a &= e^{-\theta\delta} \\ b_i &= \frac{L_{2i-2} + L_{2i-1}}{2}(1 - e^{-\theta\delta}) \\ &= \frac{L_{2i-2} + L_{2i-1}}{2}(1 - a), \qquad i = 1, \ldots, 30 \\ \hat{\sigma} &= \sigma\sqrt{\frac{1 - e^{-2\theta\delta}}{2\theta}} = \sigma\sqrt{\frac{1 - a^2}{2\theta}} \end{aligned}$$

Use these 78 equations to recursively cancel out $Y_2$ to $Y_{78}$, and express $Y_{79}$ using X's

$$\begin{aligned} L_2 = Y_{79} &= X_{78} + aX_{77} + a^2 X_{76} + \cdots + a^{77} X_1 + a^{78} Y_1 \\ &= X_{78} + aX_{77} + a^2 X_{76} + \cdots + a^{77} X_1 + a^{78} L_1 \end{aligned}$$

Then, the conditional likelihood

$$\begin{aligned} &\ln f(X_1, X_2, \ldots, X_{78} | Y_0 = L_0, Y_1 = L_1, Y_{79} = L_2) \\ &= \ln \frac{f(X_1) \cdots f(X_{77}) f(L_2 - a^{78} L_1 - a^{77} X_1 - \cdots - aX^{77})}{f_U(L_2 - a^{78} L_1)} \\ &= \ln(*) \end{aligned} \tag{A1}$$

where

$$\begin{aligned} U &= X_{78} + aX_{77} + a^2 X_{76} + \cdots + a^{77} X_1 \\ &\sim N\left(b_1\left(1 + a + \cdots + a^{77}\right), \hat{\sigma}^2\left(1 + a^2 + \cdots + a^{154}\right)\right) \\ &= N\left(b_1 \frac{1 - a^{78}}{1 - a}, \hat{\sigma}^2 \frac{1 - a^{156}}{1 - a^2}\right) \end{aligned}$$

The $f$ without subscript is the density for $N(b_1, \hat{\sigma}^2)$.

The numerator in $(*)$ is

$$\left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\right)^{78} \exp\left[-\frac{1}{2\hat{\sigma}^2}\Big((X_1 - b_1)^2 + \cdots + (X_{77} - b_1)^2 + (L_2 - a^{78} L_1 - a^{77} X_1 - \cdots - aX^{77} - b_1)^2\Big)\right] \tag{A2}$$

The denominator in $(*)$ is

$$\begin{aligned} &\frac{1}{\hat{\sigma}\sqrt{2\pi}}\sqrt{\frac{1 - a^2}{1 - a^{156}}} \\ &\quad\times \exp\left[-\frac{1}{2\hat{\sigma}^2}\frac{1 - a^2}{1 - a^{156}}\left(L_2 - a^{78} L_1 - b_1 \frac{1 - a^{78}}{1 - a}\right)^2\right] \\ &= \frac{1}{\hat{\sigma}\sqrt{2\pi}}\sqrt{\frac{1 - a^2}{1 - a^{156}}} \\ &\quad\times \exp\left[-\frac{1}{2\hat{\sigma}^2}\frac{1 - a^2}{1 - a^{156}}\left(L_2 - \frac{1 + a^{78}}{2} L_1 - \frac{1 - a^{78}}{2} L_0\right)^2\right] \end{aligned} \tag{A3}$$

Plug equation (A2) and (A3) into (A1). Then plug (A1) and 29 other similar terms into equation (12),

$$\begin{aligned} \mathcal{L} &= \ln f(\vec{Y} | \vec{L}) = \sum_{i=1}^{30} \ln f(Y_{79(i-1)+1}, \ldots, Y_{79i} | Y_{79(i-1)} \\ &= L_{2(i-1)}, Y_{79(i-1)+1} = L_{2i-1}, Y_{79i} = L_{2i}) \\ &= \sum_{i=1}^{30} \left\{ \begin{array}{c} -77\ln(\hat{\sigma}) - \frac{77}{2}\ln(2\pi) - \frac{1}{2\hat{\sigma}^2} \\ \Big[(X_{78(i-1)+1} - b_i)^2 + \cdots \\ + (X_{78i-1} - b_i)^2 + (L_{2i} - a^{78} L_{2i-1} \\ - a^{77} X_{79(i-1)+1} - \cdots - aX_{79i-1} - b_i)^2\Big] \\ + \frac{1}{2}\ln\left(\frac{1-a^{156}}{1-a^2}\right) + \frac{1}{2\hat{\sigma}^2}\frac{1-a^2}{1-a^{156}} \\ \left(L_{2i} - \frac{1+a^{78}}{2} L_{2i-1} - \frac{1-a^{78}}{2} L_{2i-2}\right)^2 \end{array} \right\} \\ &\text{(plug in } \hat{\sigma}) \\ &= \sum_{i=1}^{30} \left\{ \begin{array}{c} -77\ln(\sigma) - \frac{77}{2}\ln\left(\frac{1-a^2}{\theta}\right) - \frac{77}{2}\ln(\pi) \\ - \frac{\theta}{\sigma^2(1-a^2)}\left[\sum_{j=1}^{78}(X_{78(i-1)+j} - b_i)^2\right] \\ + \frac{1}{2}\ln\left(\frac{1-a^{156}}{1-a^2}\right) + \frac{\theta}{\sigma^2(1-a^{156})} \\ \left(L_{2i} - \frac{1+a^{78}}{2} L_{2i-1} - \frac{1-a^{78}}{2} L_{2i-2}\right)^2 \end{array} \right\} \\ &\text{(plug in } b_{i-1} \text{ and } x) \\ &= \sum_{i=1}^{30} \left\{ \begin{array}{c} -77\ln(\sigma) - \frac{77}{2}\ln\left(\frac{1-a^2}{\theta}\right) - \frac{77}{2}\ln(\pi) \\ - \frac{\theta}{\sigma^2(1-a^2)}(**) \\ + \frac{1}{2}\ln\left(\frac{1-a^{156}}{1-a^2}\right) + \frac{\theta}{\sigma^2(1-a^{156})} \\ \left(L_{2i} - \frac{1+a^{78}}{2} L_{2i-1} - \frac{1-a^{78}}{2} L_{2i-2}\right)^2 \end{array} \right\} \end{aligned} \tag{A4}$$

where

$$\begin{aligned} (**) &= \sum_{j=1}^{78}\Bigg(Y_{79(i-1)+j+1} - aY_{79(i-1)+j} \\ &\qquad + (a-1)\frac{L_{2i-2} + L_{2i-1}}{2}\Bigg)^2 \\ &= \sum_{j=1}^{78} Y^2_{79(i-1)+j+1} + a^2 \sum_{j=1}^{78} Y^2_{79(i-1)+j} \\ &\qquad + 78(a-1)^2 \frac{(L_{2i-2} + L_{2i-1})^2}{4} \\ &\qquad + (a-1)(L_{2i-2} + L_{2i-1}) \sum_{j=1}^{78} Y_{79(i-1)+j+1} \\ &\qquad - a(a-1)(L_{2i-2} + L_{2i-1}) \sum_{j=1}^{78} Y_{79(i-1)+j} \end{aligned}$$

$$-2a\sum_{j=1}^{78} Y_{79(i-1)+j+1}Y_{79(i-1)+j} \tag{A5}$$

Plug (A5) into (A4) and expand the summations,

$$\begin{aligned}\mathcal{L}(\theta,\sigma) = & -77\times 30\ln(\sigma) - \frac{77\times 30}{2}\ln\left(\frac{1-a^2}{\theta}\right)\\ & - \frac{77\times 30}{2}\ln(\pi) + \frac{30}{2}\ln\left(\frac{1-a^{156}}{1-a^2}\right)\\ & - \frac{\theta}{\sigma^2(1-a^2)}(Aa^2+Ba+C)\\ & + \frac{\theta}{\sigma^2(1-a^{156})}(Da^{156}+Ea^{78}+F)\end{aligned} \tag{A6}$$

where

$$\begin{aligned}A = & \sum_{i=1}^{30}\left[\sum_{j=1}^{78} Y^2_{79(i-1)+j} + 78\frac{(L_{2i-2}+L_{2i-1})^2}{4}\right.\\ & \left. -(L_{2i-2}+L_{2i-1})\sum_{j=1}^{78} Y_{79(i-1)+j}\right]\\ B = & \sum_{i=1}^{30}\left[-156\frac{(L_{2i-2}+L_{2i-1})^2}{4}\right.\\ & +(L_{2i-2}+L_{2i-1})\sum_{j=1}^{78} Y_{79(i-1)+j+1}\\ & +(L_{2i-2}+L_{2i-1})\sum_{j=1}^{78} Y_{79(i-1)+j}\\ & \left. -2\sum_{j=1}^{78} Y_{79(i-1)+j+1}Y_{79(i-1)+j}\right]\\ C = & \sum_{i=1}^{30}\left[\sum_{j=1}^{78} Y^2_{79(i-1)+j+1} + 78\frac{(L_{2i-2}+L_{2i-1})^2}{4}\right.\\ & \left. -(L_{2i-2}+L_{2i-1})\sum_{j=1}^{78} Y_{79(i-1)+j+1}\right]\end{aligned}$$

$$\begin{aligned}D = & \sum_{i=1}^{30}\left[\frac{1}{4}L^2_{2i-1}+\frac{1}{4}L^2_{2i-2}-\frac{1}{2}L_{2i-1}L_{2i-2}\right]\\ E = & \sum_{i=1}^{30}\left[\frac{1}{2}L^2_{2i-1}-\frac{1}{2}L^2_{2i-2}-L_{2i}L_{2i-1}+L_{2i}L_{2i-2}\right]\\ F = & \sum_{i=1}^{30}\left[L^2_{2i}+\frac{1}{4}L^2_{2i-1}+\frac{1}{4}L^2_{2i-2}-L_{2i}L_{2i-1}\right.\\ & \left. -L_{2i}L_{2i-2}+\frac{1}{2}L_{2i-1}L_{2i-2}\right]\end{aligned}$$

In the expression of $\mathcal{L}$ in (A6), the two parameters are $\theta$ and $\sigma$. $A, B, C, D, E$ and $F$ are functions of $L_i$ and $Y_i$; $a = e^{-\theta\delta}$ contains parameter $\theta$; $\delta = \frac{\delta_1}{78}$ is the discretization step size.

Setting first-order derivatives of $\mathcal{L}$ with respect to $\sigma$ to zero

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\sigma} = & -77\times 30\frac{1}{\sigma} - \frac{2}{\sigma^3}\left[\frac{\theta}{1-a^{156}}(Da^{156}+Ea^{78}+F)\right.\\ & \left. -\frac{\theta}{1-a^2}(Aa^2+Ba+C)\right] = 0\end{aligned} \tag{A7}$$

The optimal pair $(\theta,\sigma)$ that maximizes $\mathcal{L}$ satisfies $\frac{\partial\mathcal{L}}{\partial\sigma} = 0$. From (A7), we can express $\sigma$ using $\theta$

$$\begin{aligned}&\sigma(\theta)\\ &= \sqrt{\frac{2}{77\times 30}\left[\frac{\theta}{1-a^2}(Aa^2+Ba+C) - \frac{\theta}{1-a^{156}}(Da^{156}+Ea^{78}+F)\right]}\end{aligned} \tag{A8}$$

Plug (A8) into (A6), $\mathcal{L}(\theta,\sigma)$ becomes $\mathcal{L}^*(\theta)$:

$$\begin{aligned}\mathcal{L}^*(\theta) = & -77\times 30\ln\left(\sigma(\theta)\right) - \frac{77\times 30}{2}\ln\left(\frac{1-a^2}{\theta}\right)\\ & -\frac{77\times 30}{2}\ln(\pi) + \frac{30}{2}\ln\left(\frac{1-a^{156}}{1-a^2}\right) - \frac{77\times 30}{2}\end{aligned}$$

The maximal solution for $\mathcal{L}(\theta,\sigma)$ is the maximal solution for $\mathcal{L}^*(\theta)$, which is solved numerically.