Pairs trading based on statistical variability of the spread process

Article *in* Quantitative Finance · September 2013

DOI: 10.1080/14697688.2012.748934

Tim Bogomolov
University of South Australia

Quantitative Finance, Vol. 00, No. 00, July 2011, 1–36

Pairs trading based on statistical variability of the spread process

TIMOFEI BOGOMOLOV*†

†University of South Australia, Adelaide, 5000, SA

(August 23, 2011)

This research proposes a new nonparametric approach to pairs trading based on renko and kagi constructions which originated from Japanese charting indicators and were introduced to academic studies by Pastukhov (2005). The method exploits statistical information about the variability of the tradable process. The approach does not find a long-run mean of the process and trade towards it like other methods of pairs trading. The only assumption we need is that the statistical properties of the spread process volatility remain reasonably constant. The theoretical profitability of the method has been demonstrated for the Ornstein-Uhlenbeck process. Tests on the daily market data of American and Australian stock exchanges show statistically significant average excess returns ranging from 1.4% to 3.6% per month.

Keywords: Pairs Trading; Spread Process; Mean-reverting; Renko; Kagi; *h*-volatility

*Email: tim@bogomolov.net

Quantitative Finance
ISSN 1469-7688 print/ISSN 1469-7696 online © 2011 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/1469768YYxxxxxxxx

1. Introduction

Pairs trading is a form of a technical analysis strategy known since the 1990s and popular among institutional and individual investors. The strategy is believed to be market neutral and provide small but constant returns with low standard deviations. A description of pairs trading and different approaches to it can be found in many articles and books (Gatev *et al.* 2006, Vidyamurthy 2004, Elliott *et al.* 2005, Do *et al.* 2006, Herlemont 2004).

The general idea of pairs trading is simple: (1) find two assets that historically have moved together; (2) when they move apart, take a short (sell) position on the higher priced asset ('winner') and long position (buy) on the lower priced asset ('loser'); (3) unwind the positions when the assets converge together.

The pairs trading strategy might be viewed as a trading of the synthetic asset (spread process or long-short portfolio) formed by a short position on one stock and a long position on another. It is also possible to create a spread process by using more than two assets.

To implement a pairs trading strategy one should answer the following three questions:

(i) What stocks should be combined in pairs? — The stage of pairs formation.
(ii) How far should stocks deviate from each other to initiate a trade? In other words, how far should the spread process move away from its mean before we open positions? — Rules to open positions.
(iii) To what degree should stocks converge to unwind positions and what is the strategy if convergence does not occur? — Rules to close positions.

Each method of pairs trading in the literature provides its own set of rules for pairs formation and trading. Despite some differences, all methods are based on the same idea of price (or returns) equilibrium: two similar assets must provide similar returns. So, any deviations from the equilibrium are the result of market over- or under-reaction on some news and/or market mispricing of the one or both stocks in a pair. The general assumption of pairs trading strategies is that these deviations are temporary and will be corrected over time.

A similar idea of prices or returns equilibrium comes from cointegration theory (Engle and Granger 1992): a spread process between returns of the two cointegrated stocks should be a stationary process. If the spread process deviates from its long-run mean, it should return back to the mean.

However, this reversion does not always happen in real life. In practice, we observe that the parameters of a spread process may change dramatically and shift the process up or down from its previous long-term mean. This could happen as a result of some news or events related to only one of the stocks from a pair. Recovery from that shock may take longer that the investment horizon or may never happen at all.

As a result, a pairs trading strategy based only on the idea of the return to equilibrium may be unprofitable. Testing existing pairs trading strategies on the market data (Do and Faff 2011, Bogomolov 2010) confirms this observation — returns after accounting of transaction costs are minimal and not consistent over time.

Two approaches exist to target the problem of non-constant mean. One of them uses moving averages (MA) with some fixed period instead of the long-term mean. This idea is used by many practitioners, but has a limitation which is common for all models based on MAs — too large a lag between an event and reaction of the model to that event.

Another approach is a regime switching model (Bock and Mestel 2009, Wu and Elliott 2005) which allows the mean to jump between different levels. Not much research has been done in this area, and it is not clear if it is possible to recognize switches and new parameters of the spread process model quickly enough to adapt the trading strategy.

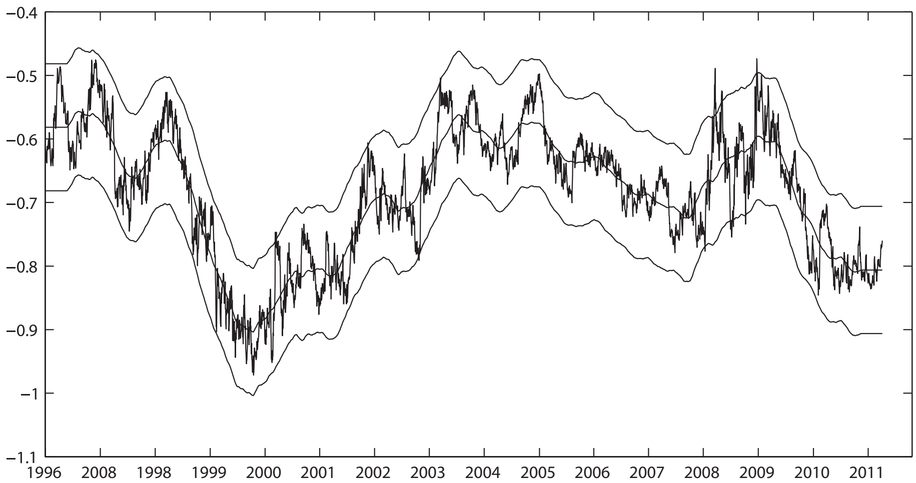


Figure 1. Log-prices spread process between two major Australian banks—Commonwealth (CBA) and Westpac (WBC). The extra lines are 200 days moving average and the same moving average shifted 10% up and down from the true location. This example illustrates an idea of relatively constant variability of the spread process.

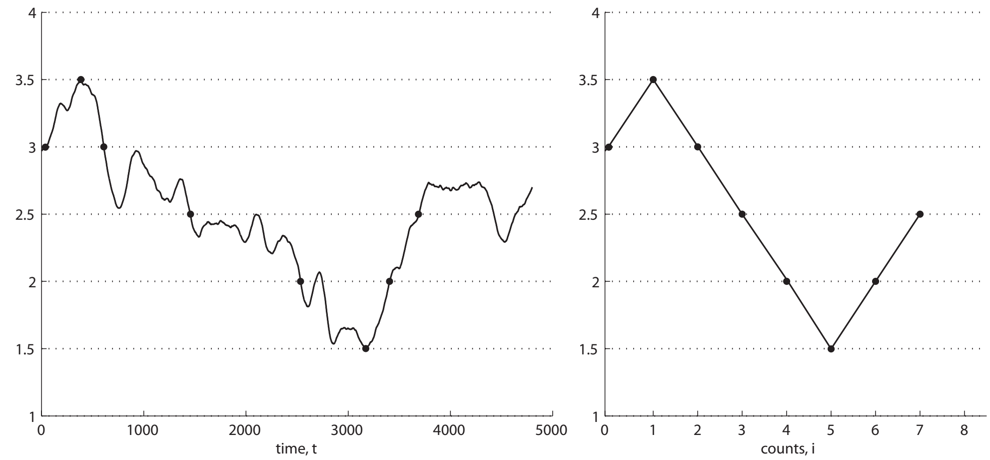
We propose a new nonparametric method of pairs trading based on some statistical properties of the spread process. It does not try to find and follow a mean of the process. Instead it utilises information about variability of the spread process, and the only assumption made is that the level of variability remains reasonably constant.

The general idea is simple. Suppose we trade an asset which we suspect has some mean-reverting property. Then, the further the asset price moves in one direction, the higher is the probability it reverses. The question is to define *how far* the price should move in one direction before trading in the opposite direction becomes potentially profitable. Obviously, this depends on a number of parameters, but the most important one is a variability of the asset price or, in the case of pairs trading, the spread process.

The tool we use to measure the variability of the spread process is based on renko and kagi constructions proposed by Pastukhov (2005). Renko and kagi are types of charts originating from 19th century Japan and are well-known to all adepts of technical analysis on the financial markets. Both charts are concerned with price movements greater than some given threshold and do not include information about time or trading volumes. This approach is believed to filter out the trading noise — small changes in assets prices — and focus only on significant price movements. Pastukhov (2005) introduced renko and kagi to the world of academic research and provided a mathematical basis for these methods of technical analysis. His research described two possible trading strategies based on the statistical properties of the renko and kagi constructions built on a real asset.

We consider the use of the renko and kagi constructions for different types of processes and extend their use from real assets to the pairs trading spread processes. We provide theoretical proofs of the profitability of the proposed method for the case of the Ornstein-Uhlenbeck process and test it on real market data from the American and Australian stock exchanges.

We proceed as follows. Section 2 provides a brief review of the method proposed by Pastukhov (2005) — the renko and kagi constructions, their properties, constructions on the Wiener process and two possible trading strategies. Section 3 considers the renko and kagi constructions on the Ornstein-Uhlenbeck process and on discrete time processes. Section 4 provides details of the



(a) Asset price process $P(t)$ with points at τ_i (b) Classical renko process $X(i)$

Figure 2. Renko chart

practical implementation and real-data tests of the proposed pairs trading strategy. Section 5 reports the results of the testing and section 6 presents the conclusion.

2. Method of renko and kagi constructions

2.1. Renko construction

Let $P(t)$ be a time series of the actual asset prices or asset cumulative returns on the time interval $[0, T]$. At this stage we assume that $P(t)$ is continuous. Let $\tau_i, i = 0, 1, ..., N$ be an increasing sequence of time moments such that for some arbitrary $H > 0$

$$H \leq \max_{t\in[0,T]} P(t) - \min_{t\in[0,T]} P(t) \tag{1}$$

and for $\tau_0 = 0, P(\tau_0) = P(0)$

$$\tau_i = \inf\{u \in [\tau_{i-1}, T] : |P(u) - P(\tau_{i-1})| = H\}. \tag{2}$$

The process $X(i) : X(i) = P(\tau_i), i = 0, 1, ..., N$ is a 'classical' renko chart (Figure 2.1) or renko process.

Now we create another sequence of time moments $\{(\tau_n^a, \tau_n^b), n = 0, 1, ..., M\}$ based on the sequence $\{\tau_i\}$. The sequence $\{\tau_n^a\}$ defines time moments when the renko process $X(i)$ has a local maximum or minimum, that is the process $X(i) = P(\tau_i)$ changes its direction, and the sequence $\{\tau_n^b\}$ defines the time moments when the local maximum or minimum is detected.

More precisely, when take $\tau_0^a = \tau_0$ and $\tau_0^b = \tau_1$ then

$$\tau_n^b = \min\{\tau_i > \tau_{n-1}^b : (P(\tau_i) - P(\tau_{i-1}))(P(\tau_{i-1}) - P(\tau_{i-2})) < 0\}, \tag{3}$$

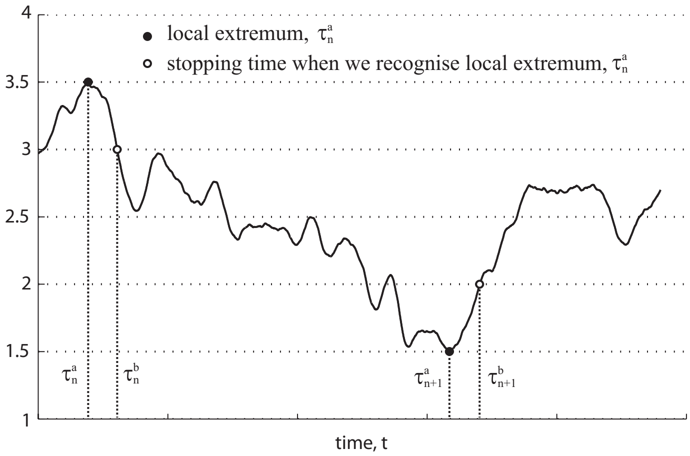


Figure 3. Renko construction

$$\tau_n^a = \{\tau_{i-1} : \tau_n^b = \tau_i\}. \tag{4}$$

If $\tau_n^b = \tau_i$ then $\tau_n^a = \tau_{i-1}$. In some cases τ_{n-1}^b and τ_n^a may be equal to each other, as they are derived from discrete process $X(i)$ and the point when we detect a local maximum may happen to be a next local minimum (Figure 3).

2.2. Kagi construction

The kagi construction is similar to the renko construction with the only difference being that to create the sequence of time moments $\{(\tau_n^a, \tau_n^b), n = 0, 1, ..., M\}$ for the kagi construction we use local maximums and minimums of the actual asset price process $P(t)$ rather than the process $X(i)$ derived from it.

The sequence $\{\tau_n^a\}$ then defines the time moments when the price process $P(t)$ has a local maximum or minimum and the sequence $\{\tau_n^b\}$ defines the time moments when that local maximum or minimum is recognized, that is, the time when the process $P(t)$ moves away from its last local maximum or minimum by a distance equal to H.

More precisely, for some arbitrary $H > 0$ satisfying (1) we define

$$\tau_0^b = \inf\{u \in [0, T] : \max_{t\in[0,u]} P(t) - \min_{t\in[0,u]} P(t) = H\} \tag{5}$$

and

$$\tau_0^a = \inf\{u < \tau_0^b : |P(u) - P(\tau_0^b)| = H\}. \tag{6}$$

It is important to know whether τ_0^a defines a local maximum or a minimum. The variable

$$S_0 = \text{sign}(P(\tau_0^a) - P(\tau_0^b)) \tag{7}$$

can take two values: 1 for a local maximum and -1 for a local minimum.

Then we define $(\tau_n^a, \tau_n^b), n > 0$ recursively. If at time τ_0^a we have a local maximum $(S_0 = 1)$

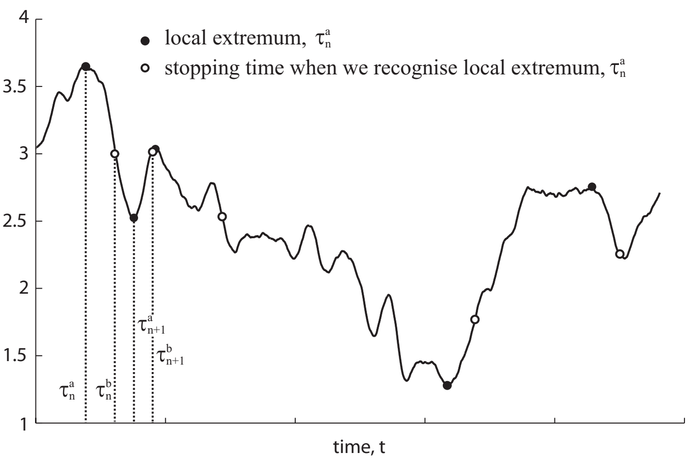


Figure 4. Kagi-construction

then all odd numbered time moments $(\tau_n^a, \tau_n^b), n = 1, 3, 5, 7, ...$ relate to local minimums where $S_n = -1, n = 1, 3, 5, 7, ...$ and should be defined by

$$\tau_n^b = \inf\{u \in [\tau_{n-1}^a, T] : P(u) - \min_{t\in[\tau_{n-1}^a, u]} P(t) = H\}$$

$$\tau_n^a = \inf\{u < \tau_n^b : P(u) = \min_{t\in[\tau_{n-1}^a, \tau_n^b]} P(t)\} \tag{8}$$

and all even numbered time moments $(\tau_n^a, \tau_n^b), n = 2, 4, 6, ...$ relate to local maximums where $S_n = 1, n = 2, 4, 6, ...$ and should be defined by

$$\tau_n^b = \inf\{u \in [\tau_{n-1}^a, T] : \max_{t\in[\tau_{n-1}^a, u]} P(t) - P(u) = H\}$$

$$\tau_n^a = \inf\{u < \tau_n^b : P(u) = \max_{t\in[\tau_{n-1}^a, \tau_n^b]} P(t)\}. \tag{9}$$

The construction of the full sequence $\{(\tau_n^a, \tau_n^b), n = 1, 2, 3, ..., N\}$ is done in the inductive manner alternating steps (8) and (9).

As the sequence $\{(\tau_n^a, \tau_n^b)\}$ is derived from the continuous process $\{P(t)\}$ the probability of $\tau_{n-1}^b = \tau_n^a$ is zero, even though they can be close to each other (Figure 4).

2.3. Some properties of renko and kagi constructions

For the following discussion we use the term *H-construction* when refering to either renko or kagi constructions as their properties are similar.

The process $P(t), t \in [0, T]$ will be defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$, taking values in $\mathbb{R}$. For some arbitrary H, we have increasing time sequence $\{(\tau_n^a, \tau_n^b), n = 0, 1, ..., N\}$ defined as above.

Obviously, τ_n^a are not stopping times as the local maximum or minimum can be defined only post-factum at times τ_n^b which are stopping times. To simplify our calculation we assume that

$T = \tau_N^b$ for some arbitrary N, which means that any stopping time τ_n^b might be considered as the end of the trading period, that is we can stop trading at this time.

We now list some useful variables introduced by Pastukhov (2005):

H-inversion counts the number of times the process $P(t)$ changes its direction for selected H, T and $P(t)$ and is given by

$$N_T(H, P) = \max\{n : \tau_n^b = T\} = N. \tag{10}$$

H-volatility of order p is a measure of the variability of the process $P(t)$ for selected H and T and is given by

$$\xi_T^p(H, P) = \frac{V_T^p(H, P)}{N_T(H, P)} \tag{11}$$

where $V_T^p(H, P)$ is a sum of vertical distances between local maximums and minimums to the power p

$$V_T^p(H, P) = \sum_{n=1}^{N} |P(\tau_n^a) - P(\tau_{n-1}^a)|^p \tag{12}$$

H-volatility of order 2 is similar to variance and can be used to describe the process $P(t)$. However, for our purpose it is sufficient to have H-volatility of order 1 only:

$$\xi_T(H, P) = \frac{V_T(H, P)}{N_T(H, P)} \tag{13}$$

Pastukhov (2005) shows that for a Wiener process $\{W(t)\}$ the condition $\xi_T(H, W) = 2H$ holds for any value of H, subject to (1). More specifically

$$\lim_{T\to\infty} \xi_T^{(p)}(H, \sigma W) = R_W(p)H^p \tag{14}$$

where

$$R_W(p) = \begin{cases} \sum_{n=1}^{\infty} \frac{n^p}{2^n}, & \text{for renko construction;} \\ \int_0^{\infty} (1 + x)^p e^{-x}\, dx, & \text{for kagi construction.} \end{cases}$$

For both constructions $R_W(1) = 2$. So, $\xi_T(H, W) = 2H$.

2.4. Trading strategies

Corresponding to the definition of H-construction, we define the term *H-strategy* without specifying renko or kagi H-construction, as the differences between renko and kagi strategies are minor and not important for our research.

There are two possible H-strategies — momentum and contrarian.

1. The trend following or momentum strategy. Here, the investor buys (sells) an asset at a stopping time τ_n^b when he recognizes that the process passed its previous local minimum (maximum) and the investor expects a continuation of the movement. There are two types of trading signals which are equivalent:

$$P(\tau_n^b) - P(\tau_n^a) > 0 \text{ or } P(\tau_{n-1}^a) - P(\tau_n^a) > 0 \quad \text{buy signal}$$
$$P(\tau_n^b) - P(\tau_n^a) < 0 \text{ or } P(\tau_{n-1}^a) - P(\tau_n^a) < 0 \quad \text{sell signal.}$$

The profit from one trade according to the trend following H-strategy over time from τ_{n-1}^b to τ_n^b is

$$Y_{\tau_n^b} = (P(\tau_n^b) - P(\tau_{n-1}^b)) \cdot \operatorname{sign}(P(\tau_n^a) - P(\tau_{n-1}^a)) \tag{15}$$

and the total profit from time 0 till time T is

$$Y_T(H,P) = (\xi_T(H,P) - 2H) \cdot N_T(H,P). \tag{16}$$

2. The contrarian strategy. Here the investor sells (buys) an asset at a stopping time τ_n^b when he decides that the process has passed *far enough* from its previous local minimum (maximum), and the investor expects a movement reversion. The trading signals are

$$P(\tau_n^b) - P(\tau_n^a) > 0 \text{ or } P(\tau_{n-1}^a) - P(\tau_n^a) > 0 \quad \text{sell signal}$$
$$P(\tau_n^b) - P(\tau_n^a) < 0 \text{ or } P(\tau_{n-1}^a) - P(\tau_n^a) < 0 \quad \text{buy signal.}$$

The profit from the one trade according to the contrarian H-strategy over time from τ_{n-1}^b to τ_n^b is the same as (16) but with negative sign.

$$Y_{\tau_n^b} = (P(\tau_n^b) - P(\tau_{n-1}^b)) \cdot \operatorname{sign}(P(\tau_{n-1}^a) - P(\tau_n^a)) \tag{17}$$

and the total profit till time T is

$$Y_T(H,P) = (2H - \xi_T(H,P)) \cdot N_T(H,P) \tag{18}$$

As we can see, trading signals for both strategies are the same but point in different directions. The investor constantly stays in the market for any strategy, just changing the direction of the trade. On this stage we assume that the investor can trade long and short with no restrictions, and that the transaction costs are zero.

It clearly follows from (16) and (18) that the choice of H-strategy depends on the value of H-volatility, $\xi_T(H,P)$. If $\xi_T(H,P) > 2H$, then to achieve a positive profit the investor should employ a trend following H-strategy; if $\xi_T(H,P) < 2H$ then the investor should use a contrarian H-strategy.

From Pastukhov (2005) we know that for the Wiener process the H-volatility $\xi_T(H,W) = 2H$ and, as a result, it is impossible to profit from the trading on the Wiener process. It seems that the same result is true for any Lévy process with independent increments symmetrically distributed around zero regardless of the shape of their distribution.

We can see that H-volatility $\xi_T(H,P) = 2H$ is a property of a martingale. Likewise $\xi_T(H,P) > 2H$ could be a property of a sub-martingale or a super-martingale or a process regularly switching over time from a sub-martingale to a super-martingale and back. It is unlikely that these sort of processes exist on the financial markets. Pastukhov (2005) does not provide examples of processes for which $\xi_T(H,P) > 2H$.

From the practical point of view, a more interesting situation occurs if $\xi_T(H,P) < 2H$. The obvious example of such process could be an Ornstein-Uhlenbeck process (Finch 2004) and by extension any mean-reverting process regardless of the distribution of its innovations.

The condition H-volatility less than $2H$ is a statistical property of a process $P(t)$ and can be considered a very mild restriction. It does not require the process $P(t)$ to be mean-reverting in the formal definition and have a constant mean and variance. To create a profitable trading strategy we simply need the process $P(t)$ to have a mean-reverting property over certain time intervals. A perfect candidate for such process $P(t)$ is a pairs trading stochastic spread process (Elliott *et al.* 2005).

3. More on the renko and kagi constructions

3.1. *Properties of* H*-constructions on the Ornstein-Uhlenbeck process*

We now consider H-constructions made over the Ornstein-Uhlenbeck (OU) process (Finch 2004):

$$dX_t = -\rho(X_t - \mu)\,dt + \sigma\,dB_t \tag{19}$$

where $\{B_t : t \geq 0\}$ is a standard Brownian motion and $\rho > 0, \sigma > 0, \mu$ are constants.

In most situations without loss of generality we can assume $\mu = 0$ and $\sigma = 1$ (by using $X_t - \mu$ rather then X_t and by time scaling). Then (19) takes the form

$$dX_t = -\rho X_t\,dt + dB_t \tag{20}$$

The Ornstein-Uhlenbeck process has finite variance $\mathbf{Var}(P_t) = \sigma^2/2\rho$. Then, if we take $H = \max P(t) - \min P(t)$, we get not more than one swing between maximum and minimum and it equals H. So the H-volatility $\xi_T(H,P) = H$. If we increase the value of H, then (1) does not hold and we can not built H-construction. Furthermore, if we take $H \to 0$ then it is equivalent to $\rho \to 0$ for fixed H. As a result, the Ornstein-Uhlenbeck process converges to the Wiener process and the H-volatility $\xi_T(H,P) \to 2H$. Hence, for the Ornstein-Uhlenbeck process

$$\frac{\xi_T(H,P)}{H} \in [1,2). \tag{21}$$

Theorem 1. (Proof is provided in the Appendix) Let P be an Ornstein-Uhlenbeck process. Then for any positive H satisfying (1), the H-volatility is less than $2H$

$$\lim_{T\to\infty} \xi_T(H,P) < 2H \tag{22}$$

Hence, trading the Ornstein-Uhlenbeck process by the contrarian H-strategy is profitable for any choice of H (18). The same is true for any mean-reverting process regardless the distribution of its innovations.

3.2. H*-construction on the discrete process*

Most financial data are discrete. So, it is important to consider the properties of H-construction on the discrete process.

3.2.1. Random Walk

Let $\{X(t)\}, t = 0, 1, 2, ...$ be an independent and identically distributed increments process, $X(t) \sim N(0, \sigma^2)$. Then the sum of increments is a random walk

$$Y(t) = \sum_{i=0}^{t} X(i).$$

If we build the H-construction on the discrete process $Y(t)$ then conditions (2), (8), (9) do not hold, as

$$\mathrm{P}(|Y(t) - Y(t+n)| = H) = 0 \quad \forall\, n, t \geq 0.$$

For the discrete process $Y(t)$ we have the following condition at the stopping time τ_n^b

$$|Y(\tau_n^a) - Y(\tau_n^b)| = \tilde{H}_n \geq H \tag{23}$$

where $\tilde{H}_n$ is an independent random variable.

This means that we get an overshot and as a result the ratio of H-volatility to the parameter H gets inflated.

$$\xi_T(H, Y) = 2\mathbf{E}[\tilde{H}_n] \geq 2H \tag{24}$$

Hence, when applied to real life it is quite possible to observe H-volatility greater than $2H$. However, this does not imply that the underlying process is not a martingale and that it would be possible to trade it with a trend-following H-strategy.

The size of overshot depends on the value of H and on the standard deviation of the increments process $X(t)$. If $H/\sigma \to \infty$ then $\mathbf{E}[\tilde{H}_n] \to H$.

3.2.2. AR(1) process

Now we consider AR(1) process which is a discrete representation of the Ornstein-Uhlenbeck process.

$$Y(t) = \alpha Y(t-1) + X(t)$$

where $\alpha \in [0, 1)$ and $X(t) \sim N(0, \sigma^2)$.

The AR(1) process has the same problem with overshot as a random walk.

$$|Y(\tau_n^a) - Y(\tau_n^b)| = \tilde{H}_n \geq H \tag{25}$$

The difference is: the value of H for AR(1) can not go to infinity as the AR(1) process is bounded (Novikov and Kordzakhia 2008). Hence, similar to a random walk case, the value of $\mathbf{E}[\tilde{H}_n]$ gets closer to H as H increases but $\mathbf{E}[\tilde{H}_n]$ never converges to H. It follows that it is possible to observe a discrete mean-reverting process with $\xi_T(H, Y) \geq 2H$ and it does not contradict Theorem 1.

The true value of the ratio of the H-volatility to the parameter H is

$$R(H, P) = \frac{\xi(H, P)}{\mathbf{E}[\tilde{H}_n(H, P)]} \in [1, 2) \tag{26}$$

The renko H-strategy is the mostly effected by the overshot as it has to generate a renko-process—a sequence of stopping times with the fixed price step H. Obviously, it is quite problematic to do it with the discrete time series, especially if the standard deviation of the increments is comparable to the value of H. That is why we run the test for kagi H-strategy only which is more appropriate for the daily data we have.

However, overshot is not always a bad thing. It might improve the profitability of the contrarian trading strategy. The overshot always happens in the direction of the price movement and in contrarian strategy we always trade in the direction opposite to the last movement of the spread process. It means that in the case of the overshot we sell at the higher price and we buy at lower price than we would if we trade as per the continuous process.

3.2.3. Choice of parameters **H** *and* **T** *for AR(1) process*

It is clear that for any given process $P(t)$ the investor can control only two parameters: H and T. There are several important considerations related to their choice.

1. Value of H should be reasonably large to minimize the overshot problem and improve ration $\xi_T(H,P)/H$, which we would like to be close to one.

2. At the same time H can not be too large. We have demonstrated the profit from the trading of the Ornstein-Uhlenbeck process $\{P(t)\}$ by the contrarian H-strategy (18). If we consider the discrete process and some transaction cost λ then the profit is

$$
\begin{aligned}
Y_T(H,P) &= (2\mathbf{E}[\tilde{H}_n(H,P)] - \xi_T(H,P) - \lambda)\cdot N_T(H,P) \\
&= \left(2 - R(H,P) - \frac{\lambda}{\mathbf{E}[\tilde{H}_n(H,P)]}\right)\cdot N_T(H,P)\cdot \mathbf{E}[\tilde{H}_n(H,P)] \qquad (27)
\end{aligned}
$$

On the one hand, from the equation (27) it follows that if the value of H increases then the expectation $\mathbf{E}[\tilde{H}_n(H,P)]$ increases, the ratio $R(H,P)$ decreases (26) and the ratio of transaction costs to the expectation of H decreases. As a result the profit increases.

On the other hand, as H increases the H-inversion $N_T(H,P)$ — the number of trades — decreases and dramatically reduces profit. So, there exists some optimal H which maximizes the profit for any given $P(t),T$ and transaction costs λ.

3. The H-strategy exploits a statistical property of the process, so we need a sample size, that is H-volatility $N_T(P,H)$, to be large enough to give us some confidence that observed process behavior is not a result of random fluctuation. Typically, all statistics textbooks recommend a sample size greater than 30. So, the length of the history period T used to calibrate the trading strategy needs to be quite large and/or the value of H small enough to guarantee $N_T(P,H) > 30$.

For testing, we arbitrarily choose H equal to one standard deviation of the spread process and the length of the calibration period of one year. This provides just around 30 trades over the history period for the best pairs.

A more rigorous approach to the selection of H is possible. The first idea is to use not the constant but adaptive H. As the H-volatility $\xi_T(P,H)$ is a measure of variability of the process $P(t)$, it looks promising to use a GARCH model to predict the size of the swing between the next local maximum and minimum and adjust H accordingly.

Another possibility is to analyse the distribution of $\{X_n\}$

$$X_n = |P(\tau_n^a) - P(\tau_{n-1}^a)| \qquad (28)$$

It is clear that $\{X_n\} \geq H$ and the distribution is skewed to the right. It follows from (18) that $X_n > 2H$ means negative profit in the nth trade. So, during pairs formation we can select pairs with 'lighter' right tale and minimal number of observations with $X_n > 2H$.

4. Pairs trading by the contrarian *H*-strategy

Each pairs trading strategy defines three steps: pairs formation, rules to open position on the spread and rules to close position. For the H-strategy the last two points are combined into one. The signal to close position acts as the signal to open a new one in the opposite direction.

This section provides a description of the trading strategy based on the kagi constructions as well as details of the data set and testing method.

4.1. Data sets

We use the Australian Stock Exchange (ASX) and American Stock Exchanges (NYSE, NASDAQ) daily closing prices to test H-strategies of pairs trading. The data are obtained from the Securities Industry Research Centre of Asia-Pacific (SIRCA). There are four data sets:

1. ASX data set covers 3863 trading days starting from January 2, 1996 and finishing on April 6, 2011 and includes more than 3,000 shares traded at different times. We use an index S&P/ASX 200 as a benchmark for this data set.

2. The top 500 American companies by market capitalisation included in the S&P 500 index, which is also used as a benchmark. The data set covers 3853 trading days from January 2, 1996 to April 29, 2011.

3. The medium capitalisation American companies included in the S&P 400 Mid Cap index, which is also used as a benchmark. The data set covers 2592 trading days from January 2, 2001 to April 29, 2011.

4. The small capitalisation American companies included in the S&P 600 Small Cap index, which is also used as a benchmark. The data set covers 2842 trading days from January 3, 2000 to April 29, 2011.

Following the testing methodology in Gatev *et al.* (2006), we use 12 months history to calibrate the system and construct pairs and the next 6 months to trade selected pairs. We start trading at the first working day of each month and trade until the last working day of the trading period. So each month, except for the first and the last five months, we get six different estimations of the monthly returns which are averaged to get the final estimation. Hence the actual time interval used for testing is shorter than the length of the data sets due to the 12-month historical data used for strategy calibration and the first and last five months of the actual testing period disregarded due to averaging.

The testing interval includes the period of the Global Financial Crisis (GFC). Short selling was banned at some period of time during GFC. However we run the test of pairs trading over that period as usual. Institutional investors, who hold large diversified portfolios, could still use pairs trading as a part of the tactical asset allocation strategy. They do not need short selling to fulfill the rules of pairs trading; they can sell some shares from the existing portfolio and buy them back when the strategy signals to close position on the pair.

4.2. Stocks pre-selection

From the ASX data set we pre-select only the top 30% companies by their dollar valued trading volume during the 12 months history period used for system calibration. The S&P 500 data set also includes only large cap stocks. That ensures that we run the strategy test over the most liquid companies at the day of the start of trading period, which, most probably, could be traded at the prices we used for testing. Thus we expect to avoid a potential liquidity problem.

On the downside, this approach makes our testing biased towards large cap companies. We could expect medium and small cap companies to have a higher probability of being mispriced than the large cap companies which attract greater attention of institutional and individual traders. As a result, potential profit could be higher. Testing of the same pairs trading strategy on the ASX data with inclusion of a broader range of stocks shows much higher returns. However, the results can be unreliable due to liquidity problems thus they are not reported here.

The last two data sets — companies from S&P 400 Mid Cap and S&P 600 Small Cap — are included to test if medium and small capitalisation companies provide more opportunities for pairs trading. It looks to be a reasonable compromise between the opportunity to test the strategy on stocks with smaller capitalisation and possible liquidity problem. As the American market is the most liquid stock market in the world, we expect that the effect of the liquidity problem is minor if any. However, one should be aware of the possible limitation of the presented results.

We put a restriction on not more than 10 non-trading days during the calibration period. It is a milder condition than 'zero non-trading days' in Gatev *et al.* (2006). We believe it is reasonable as even the strongest and most liquid stocks can halt trading once or twice a year. We aim to employ a pure quantitative approach to the testing with a minimal number of constrains, so we do not put any extra restrictions on the pairs selection, for example, sectors.

The opening and closing prices are the results of auctions, which usually attract a large trading volume. In many instances, the volume of the opening and closing auctions exceeds 50% of the total daily trading volume. By using the closing and/or opening prices we can be sure that we could make a trade at a given price and at the same time avoid bid-ask bounce bias.

If a stock has a non-trading day during the trading period (price and/or volume equals to zero), we use the closing price of the previous day to create the spread. However, the trading positions on the pairs having that stock cannot be opened or closed on that day, even if the spread process signals to do so.

4.3. Pairs formation

We take log-prices of all stocks pre-selected for pairs trading based on the 12-month history, combine them in all possible pairs and build spread process for each pair.

$$y_{i,j}(t) = \log P_i(t) - \log P_j(t)$$

where $P_i(t), P_j(t)$ are prices of stocks i and j on day t.

For each spread process we calculate its standard deviation $C_{i,j}$. We arbitrarily set parameter $H_{i,j}$ for our H-strategy equal parameter $C_{i,j}$.

$$H_{i,j} = C_{i,j}$$

We make H-construction for each spread process and calculated H-volatility $\xi_{i,j}(H_{i,j})$ and H-inversion $N_{i,j}(H_{i,j})$. Then all pairs are sorted in descending order by the size of the H-inversion $N_{i,j}(H_{i,j})$. We hope that pairs with the highest H-inversion over the period of history used for calibration will tend to have statistically similar behavior in the future and provide higher profit.

H-inversion acts as a good proxy for a number of parameters. The spread process with smaller standard deviation (which is equivalent to a smaller squared distance in (Gatev *et al.* 2006)) has smaller H and as a result tends to have higher H-inversion. For two spread processes with the same H and $\xi(H)$ the higher value of H-inversion means higher profit by (27). The same time, higher H-inversion means a larger sample size. That provides us more confidence in the statistical power of the calibration results.

The top 5 and 20 pairs with the highest H-inversion $N_{i,j}(H_{i,j})$ are used for pairs trading. In contrast to previous researches on pairs trading we remove the company selected for the pair from the pool of the pre-selected stocks. So, each company could be selected just once and be a part of the one pair only. We expect this approach improves the diversification of the portfolio.

4.4. Trading rules

We start trading all pairs selected during the stage of pairs formation from the first day of trading period and constantly stay in the market until the very last day of the trading period when we close all positions.

To define the direction of trades for each pair on the first day we make the H-construction over the history using the parameter $H_{i,j}$ defined during calibration and take the direction of the trade at the end of the calibration period and the value of the last local extremum. Hence, the virtual trading on the history extends beyond the end of calibration period and becomes real trading on the 6-month trading period.

On the first trading day we know from the history if we have a local maximum or minimum just before the start of real trading. Let's say that it is a maximum. It means the following:

(i) on the first day of the trading period we should have a long position on the spread process $y_{i,j}(t)$, so we buy long stock i and sell short stock j regardless of the current price levels or the spread process current value;

(ii) then we follow the spread process waiting for the sell signal – first moment t after time τ_0^b (the last stopping time of the calibration period) such that

$$y_{i,j}(t) - \min_{\tau_0^b \le n \le t} y_{i,j}(n) \ge H_{i,j}$$

that is the time when we recognize the next local minimum — the spread process moves away from the previous local minimum on the distance greater than $H_{i,j}$.

When it happens, we set $\tau_1^b = t$ and reverse position on the spread from 'long on the spread' to 'short on the spread'. To do so, we close existing positions and sell short stock i and buy stock j. Then we keep following the spread process $y_{i,j}(t)$ waiting for the signal to buy the spread again — the first moment t after time τ_1^b such that

$$\max_{\tau_1^b \le n \le t} y_{i,j}(n) - y_{i,j}(t) \ge H_{i,j}.$$

We repeat this procedure again and again constantly staying in the market and alternating 'buying the spread' and 'selling the spread' until the last day of the predefined trading period when we just close all open positions.

If on the first trading day the spread process has a minimum as its last historical local extremum, then we do the opposite for the first trade — sell the spread, that is sell short the stock i and buy the stock j. After that we follow the same way as above changing 'selling the spread' and 'buying the spread'.

4.5. Excess return calculation and transaction costs

To calculate strategy excess return we follow the procedure common for pairs trading literature (Gatev *et al.* 2006, Do and Faff 2010, 2011, Bogomolov 2010). The proposed strategy is dollar neutral, so we trade \$1 in each leg of the pair. We calculate value-weighted daily market-to-market cash flows from each pair which are considered as excess return:

$$r_{P,t} = \frac{\sum_{i\in P} w_{i,t}\, c_{i,t}}{\sum_{i\in P} w_{i,t}} \tag{29}$$

where:

$c_{i,t}$ is the daily cash flow from the two positions from the pair i;
$w_{i,t}$ is the weight of each pair. For each newly opened position on the pair initial weight equals to 1 and then evolves by the formula $w_{i,t} = w_{i,t-1}(1 + c_{i,t-1}) = (1 + c_{i,1}) \cdots (1 + c_{i,t-1})$

The daily cash flow from the pair or a daily return of the pair is

$$c_i(t) = \sum_{j=1}^{2} I_j(t)\, v_j(t)\, r_j(t) \tag{30}$$

where:

$I_j(t)$ is a dummy variable which is equal to 1 if a long position on stock j is open and -1 if a short position on stock j is open;
$r_j(t)$ is a daily return on stock j;
$v_j(t)$ is a weight of stock j and is used to calculate daily cash flows $v_j(t) = v_j(t-1)\cdot(1 + r_j(t-1)) = (1 + r_j(1)) \cdots (1 + r_j(t-1))$

Then the strategies' daily returns are compounded to obtain monthly returns.

It is reported in the academic literature that transaction costs make a serious impact on the profitability of the pairs trading strategies. Bowen *et al.* (2010) record more than a 50% reduction in the excess returns of the high frequency pairs trading strategy after applying 15 basis point transaction fee. Do and Faff (2011) replicated the research by Gatev *et al.* (2006) and demonstrated that the strategy became unprofitable after detailed accounting for all transaction costs.

For our tests we choose transaction costs equal to 0.10% (10 basis points) per transaction. It is the average brokerage fee on April 2011 on the Australian market for retail investors (Interactive Brokers 0.08%, CommSec 0.12%, E-Trade 0.11%, Macquare Edge 0.10%). Commission on the trading American stocks is calculated based on the number of stocks traded rather than dollar volume and starts from US$0.005 per share. In most cases it is cheaper than 0.10% per trade as most companies in S&P 500 and even in S&P 600 Small Cap are priced higher than 5$ per share. The selected level of transaction costs of 0.10% per trade means about 0.20% per round trip per stock and about 0.40% per round trip for the pair or the spread as a synthetic asset.

To account for transaction costs we employ the following rules. At stopping time τ_n^b, when we change direction of the trade on the spread process, we reduce the cash flow from the current day by the weighted size of transaction costs

$$c_i(\tau_n^b) = \sum_{j=1}^{2} \left(I_{j,t} v_{j,t} r_{j,t} - \lambda\, v_{j,t}(1 + r_{j,t}) \right), \quad \tau_n^b = t;$$

and reduce the next day's cash flow from the new position on the spread by the doubling the size of the transaction costs (as we trade $1 short and $1 long — total trading volume $2)

$$c_i(\tau_n^b + 1) = c_i(\tau_n^b + 1) - 2\lambda$$

where:

$c_{i,t}$ is a cash flow or excess return on the pair i calculated as (30);
λ $= 0.001$ is a brokerage fee;
$I_{j,t}$ is a dummy variable which is equal to 1 if a long position on stock j is open and -1 if a short position on stock j is open;
$r_{j,\,t}$ is a daily return on the stock j;
$v_{j,\,t}$ $= v_{j,\,t-1}(1 - r_{j,\,t-1})$ weight of the stock j.

Our estimation of transaction costs is quite conservative. Nowadays the price of trading for the most institutional investors has reduced dramatically. We present the strategy performance before transaction costs as well. One can make his own estimation of the impact of transaction cost by multiplying the chosen level of transaction cost per trade per stock by four and by the average number of trades per month as reported in the results section.

It is important to remember that pairs trading is a naturally leveraged product and the above method of excess return computation uses 2:1 leverage. One should be very careful when comparing the results of the pairs trading strategy with possible transaction costs or performance of the non-leveraged strategies, for example, the naïve *buy-and-hold* strategy. For example, the brokerage fee applies to the full traded volume, which is \$2 for pairs trading, while reported excess returns are based on a \$1 dollar investment.

5. Results

The results of the testing presented in Table 1 and on Figures 5–12 show monthly excess returns for the strategy and its historical performance together with proper benchmark indexes. Monthly excess returns are statistically significant at 99% significance level for all scenarios except one — S&P 400 MidCaps top 20 pairs after transaction costs. Strategy excess returns before transaction costs vary from 1.42% to 3.65% per month at standard deviations from 2.0% to 5.67%. For the same period of time the benchmark performances are: ASX 200 — 0.31% per month at 3.95% standard deviations; S&P 500 — 0.11% per month at 4.92%; S&P 400 MidCaps — 0.44% per month at 5.49%; S&P 600 SmallCaps — 0.34% per month at 5.97%. The proposed strategy outperforms the market indexes for each data set for the top 5 and top 20 pairs portfolios.

Increased numbers of pairs traded simultaneously from 5 to 20 pairs reduces the profit but the same time reduces the standard deviations of returns. That is exactly what we expect from increased diversification of the portfolio. As a result, the Modigliani Risk-Adjusted Performance and Sharpe ratio for the portfolio of the top 20 pairs are higher than for the top 5 pairs for all data sets.

Also as we expected, testing on the capitalisation stocks demonstrates higher returns than the large cap stocks. One possible explanation for this is the different levels of market efficiency. The stocks from S&P 500 are the most liquid and most efficient stock market in the world attracting great attention of domestic and international institutional and individual investors. There are very low chances of mispricing any companies under normal market conditions. It can be clearly seen on the Figures 7 and 8 that the strategy is highly profitable from 2000 to 2003 (dot-com crash, September 11) and from mid 2008 to 2010 (Global Financial Crisis) — the returns are from 3% to 6% per month. But the strategy barely covers transaction costs between those two periods. So, the strategy generates more profit in times of uncertainty and high volatility, which is not a surprise for a contrarian strategy.

The small capitalisation companies attract less attention and less research. It is more difficult to estimate their future risks and returns. So, they have higher chances of being mispriced. That provides more opportunities for pairs trading. Strategy testing on the S&P 600 SmallCap data set shows the highest and the most stable returns. Even the GFC does not change anything in the pattern of returns compared to the 'normal' periods.

The strategy performance on all data sets has very low correlation with chosen benchmarks and market betas close to zero. The plot of historical performance for each data set is very close to a straight line regardless the market conditions for the same period of time. The only exception is an increase in the slope of that line during the periods of higher financial uncertainty. All of this allows us to say that the proposed strategy is truly market neutral.

Transaction cost is a big issue for all methods of pairs trading. The strategy makes about 2.5 trades on the spread per month. It means that the average transaction cost per month at the chosen level of 10 basis points per trade per stock is about 1%

$$0.1\% \cdot 2.5 \text{ (trades) } \cdot 2 \text{ (stocks in pair) } \cdot 2 \text{ (in and out) } \approx 1\%$$

As a result, the strategy loses from 40% to 50% of its monthly excess returns. However, even after accounting for the transaction costs the strategy demonstrates quite an impressive performance of about 1% per month.

6. Conclusions

This research proposes a new method of pairs trading based on the volatility of the spread process. The novelty of this approach is in its flexibility and the less restrictive nature of the method, as compared to other methods in the academic literature. We do not expect the spread process to be mean-reverting in its formal definition with constant mean, variance and coefficient of mean-reversion. To generate the profit from pairs trading it is sufficient if the process has some mean-reverting property statistically more often than it has not. As it is a non-parametric method, it is free from possible problems with model misspecification.

Testing on real market data shows statistically significant profits from using the same strategy across different markets. Differences in levels of returns can be explained in the framework of the efficient-market hypothesis.

The H-constructions are very effective way to measure the variability of the process and can be used successfully as a basis of the pairs trading strategy. However, it is not the only method to tackle variability and other approaches can be developed.

References

Billingsley, P., *Probability and Measure*, 3 1995, Wiley-Interscience.

Bock, M. and Mestel, R., A Regime-Switching Relative Value Arbitrage Rule. In *Operations Research Proceedings 2008*, edited by B. Fleischmann, K.H. Borgwardt, R. Klein and A. Tuma, pp. 9 – 14, 2009, Springer Berlin Heidelberg.

Bogomolov, T., Pairs Trading in the Land Down Under. *SSRN eLibrary*, 2010.

Bowen, D., Hutchinson, M.C. and O'Sullivan, N., High Frequency Equity Pairs Trading: Transaction Costs, Speed of Execution and Patterns in Returns. *Journal of Trading, Forthcoming*, 2010.

Bradley, R.C., Basic properties of strong mixing conditions. A survey and some open questions.. *Probability Surveys*, 2005, **2**, 107 – 144.

Do, B. and Faff, R., Does simple pairs trading still work?. *Financial Analysts Journal*, 2010, p. 83.

Do, B. and Faff, R., Are Pairs Trading Profits Robust to Trading Costs?. , 2011.

Do, B., Faff, R. and Hamza, K., A New Approach to Modeling and Estimation for Pairs Trading. *Monash University, Working Paper*, 2006.

Elliott, R.J., Hoek, J.V.D. and Malcolm, W.P., Pairs trading.. *Quantitative Finance*, 2005, **5**, 271 – 276.

Engle, R.F. and Granger, C.W.J., *Long-Run Economic Relationships Readings in Cointegration*, Advanced Texts in Econometrics 1992, Oxford University Press.

Finch, S., Ornstein-Uhlenbeck Process. 2004.

Gatev, E., Goetzmann, W. and Rouwenhorst, K., Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Rev. Financ. Stud.*, 2006, **19**, 797 – 827.

Graversen, S.E. and Peskir, G., Maximal Inequalities for the Ornstein-Uhlenbeck Process. *Proceedings of the American Mathematical Society*, 2000, **128**, 3035 – 3041.

Herlemont, D., Pairs Trading, Convergence Trading, Cointegration. 2004.

Hoek, J.V.D. and Elliott, R.J., *Binomial Models in Finance* 2006, Springer.

Kolmogorov, A.N. and Rozanov, Y.A., On Strong Mixing Conditions for Stationary Gaussian Processes. *Theory of Probability and its Applications*, 1960, **5**, 204 – 208.

Novikov, A. and Kordzakhia, N., Martingales and first passage times of AR(1) sequences.. *Stochastics: An International Journal of Probability & Stochastic Processes*, 2008, **80**, 197 – 210.

Pastukhov, S.V., On Some Probabilistic-Statistical Methods in Technical Analysis.. *Theory of Probability & Its Applications*, 2005, **49**, 245 – 260.

Rosenblatt, M., A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Science*, 1956, **42**, 43 – 47.

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

Wu, P. and Elliott, R.J., Parameter estimation for a regime-switching mean-reverting model with jumps.. *International Journal of Theoretical & Applied Finance*, 2005, **8**, 791 – 806.

Market	ASX		S&P 500		S&P 400 MidCap		S&P 600 SmallCap	
Number of pairs traded	5 pairs	20 pairs	5 pairs	20 pairs	5 pairs	20 pairs	5 pairs	20 pairs
Distribution of monthly excess returns before transaction costs								
Mean	0.0269	0.0185	0.0228	0.0190	0.0239	0.0142	0.0365	0.0238
Standard error	0.0027	0.0016	0.0032	0.0023	0.0057	0.0032	0.0047	0.0024
t-Statistics	10.0348	11.6954	7.0183	8.3539	4.2161	4.3956	7.8292	9.8084
P-value	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
Median	0.0264	0.0184	0.0160	0.0132	0.0134	0.0116	0.0328	0.0242
Standard deviations	0.0339	0.0200	0.0410	0.0288	0.0567	0.0324	0.0493	0.0257
Skewness	0.9833	0.5514	1.6244	1.2752	2.6557	2.6096	1.1913	0.9283
Kurtosis	8.0619	5.8379	7.6551	6.2938	12.5225	13.0219	6.8876	5.0155
Minimum	-0.0761	-0.0377	-0.0751	-0.0565	-0.0507	-0.0599	-0.0667	-0.0259
Maximum	0.1931	0.1016	0.2124	0.1373	0.3225	0.1802	0.2686	0.1116
Average profitable month	0.0359	0.0236	0.0401	0.0278	0.0436	0.0253	0.0535	0.0309
Average losing month	-0.0196	-0.0108	-0.0142	-0.0113	-0.0200	-0.0116	-0.0201	-0.0107
Negative observations	16.3%	15.0%	31.9%	22.5%	31.0%	30.0%	23.2%	17.0%
Distribution of the monthly excess returns after transaction costs								
Mean	0.0161	0.0098	0.0124	0.0094	0.0143	0.0057	0.0253	0.0140
Standard error	0.0026	0.0015	0.0030	0.0021	0.0053	0.0030	0.0044	0.0023
t-Statistics	6.3098	6.5384	4.0824	4.4415	2.7048	1.9016	5.7690	6.0861
P-value	0.0000	0.0000	0.0001	0.0000	0.0080	0.0601	0.0000	0.0000
Median	0.0168	0.0097	0.0063	0.0037	0.0049	0.0028	0.0224	0.0135
Standard deviations	0.0323	0.0190	0.0384	0.0267	0.0528	0.0298	0.0464	0.0243
Skewness	0.8772	0.3251	1.5523	1.0882	2.5974	2.4183	1.1020	0.7981
Kurtosis	8.0292	5.5916	7.4354	6.0756	12.3564	12.4902	6.3577	4.7958
Risk measures based on the monthly excess returns before transaction costs								
Average number of trades per month per pair	2.6	2.1	2.5	2.3	2.2	2.0	2.6	2.3
Average holding time, days	8.76	10.79	9.16	9.71	10.17	11.05	8.72	9.48
Correlation with the benchmark	0.13	0.06	0.03	0.02	-0.13	-0.15	-0.12	-0.03
Sharpe ratio	0.7933	0.9246	0.5548	0.6604	0.4216	0.4396	0.7398	0.9268
Modigliani Risk-Adjusted Performance	0.0313	0.0365	0.0273	0.0325	0.0231	0.0241	0.0442	0.0553
Jensen's alpha	0.0266	0.0184	0.0227	0.0190	0.0245	0.0146	0.0368	0.0239
Market beta	0.1099	0.0289	0.0246	0.0128	-0.1351	-0.0886	-0.1030	-0.0116

Table 1. Monthly excess returns of the kagi pairs trading strategy with and without transaction costs (0.10% per one trade)

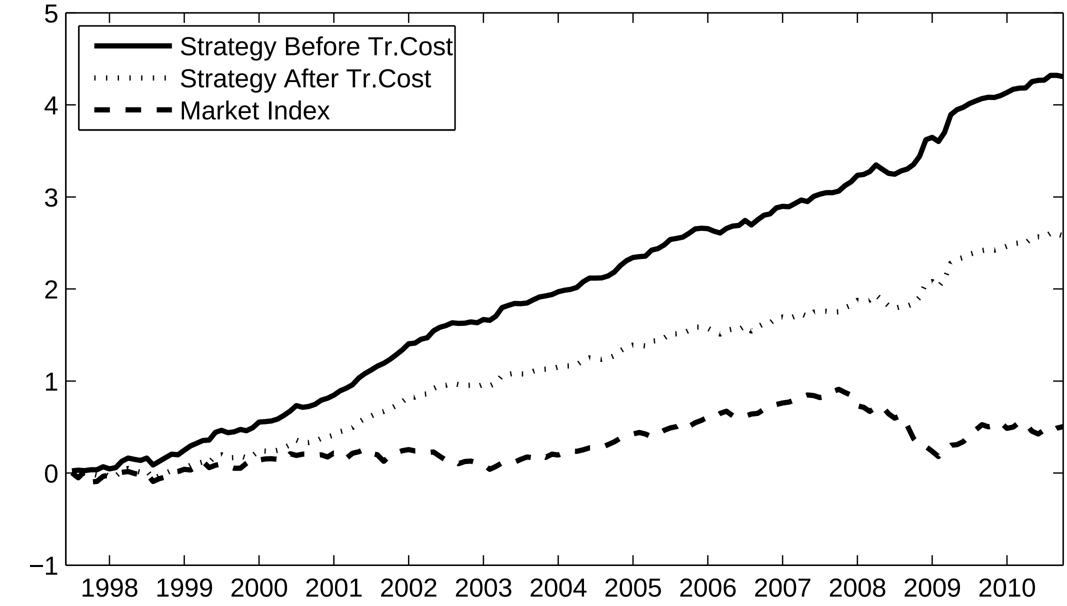


Figure 5. Strategy historical performance on the Australian market data set for top 5 pairs portfolio before and after transaction cost

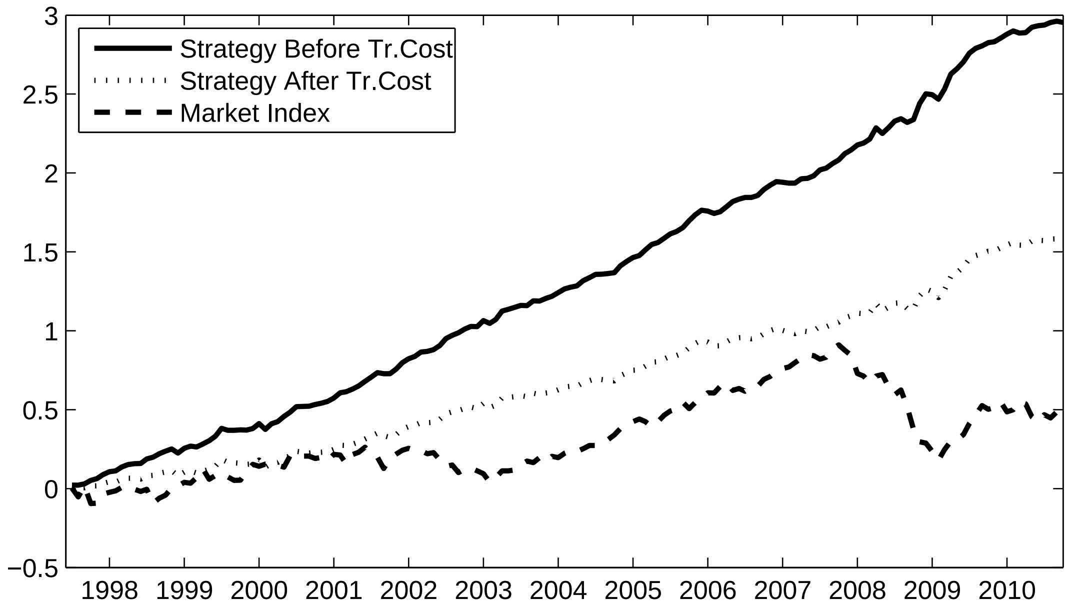


Figure 6. Strategy historical performance on the Australian market data set for top 20 pairs portfolio before and after transaction cost

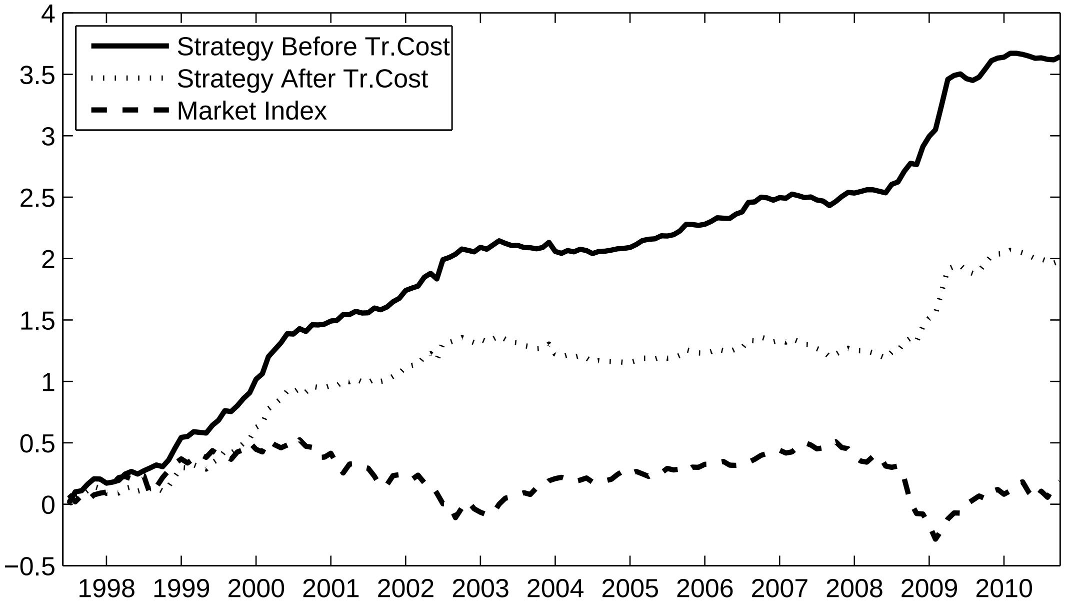


Figure 7. Strategy historical performance on the S&P 500 data set for top 5 pairs portfolio before and after transaction cost

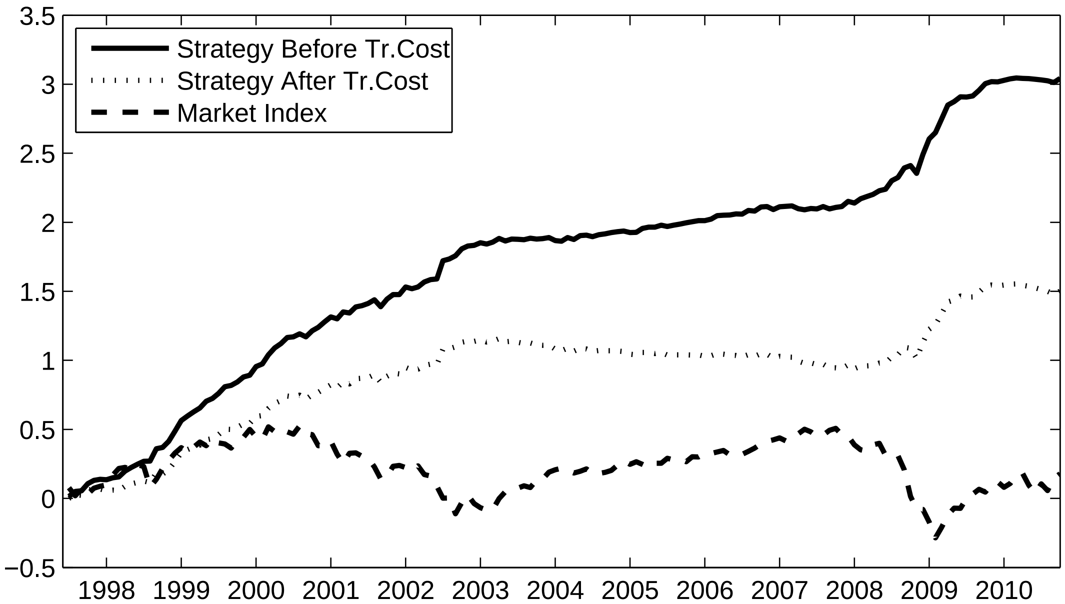


Figure 8. Strategy historical performance on the S&P 500 data set for top 20 pairs portfolio before and after transaction cost

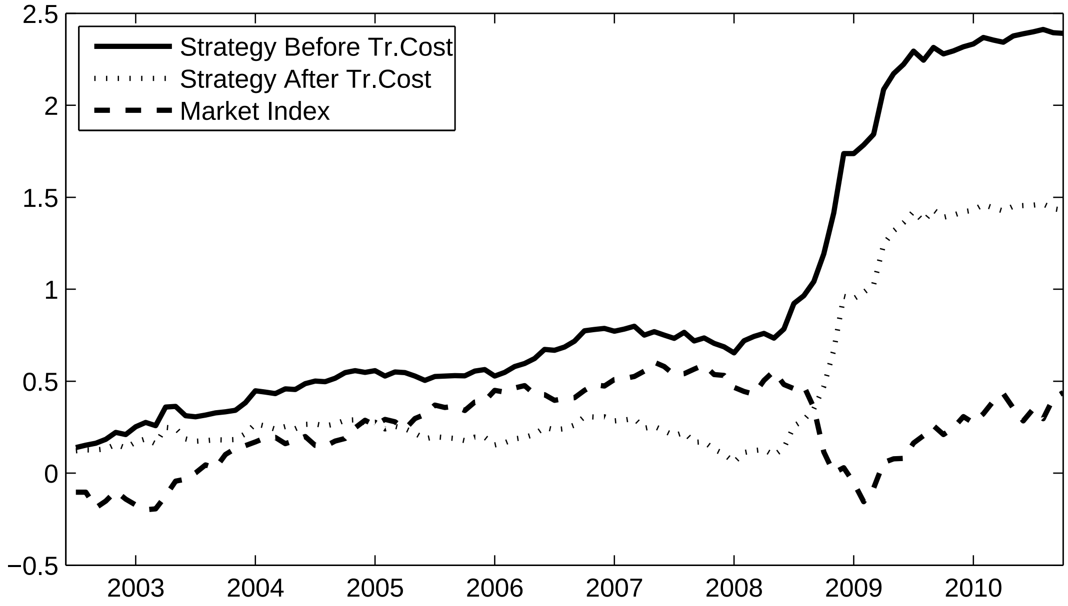


Figure 9. Strategy historical performance on the S&P 400 MidCap data set for top 5 pairs portfolio before and after transaction cost

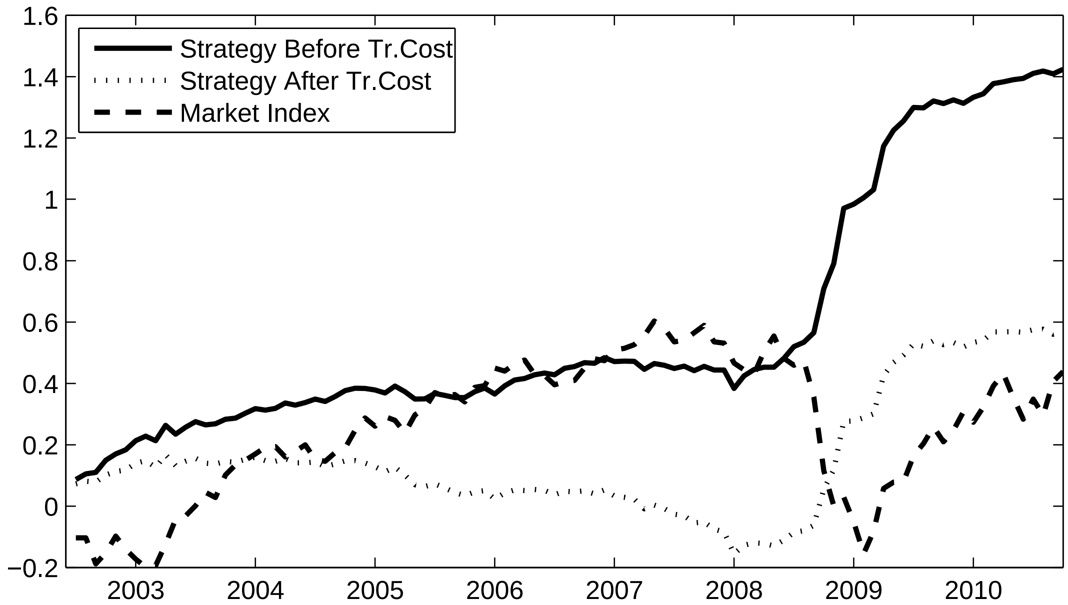


Figure 10. Strategy historical performance on the S&P 400 MidCap data set for top 20 pairs portfolio before and after transaction cost

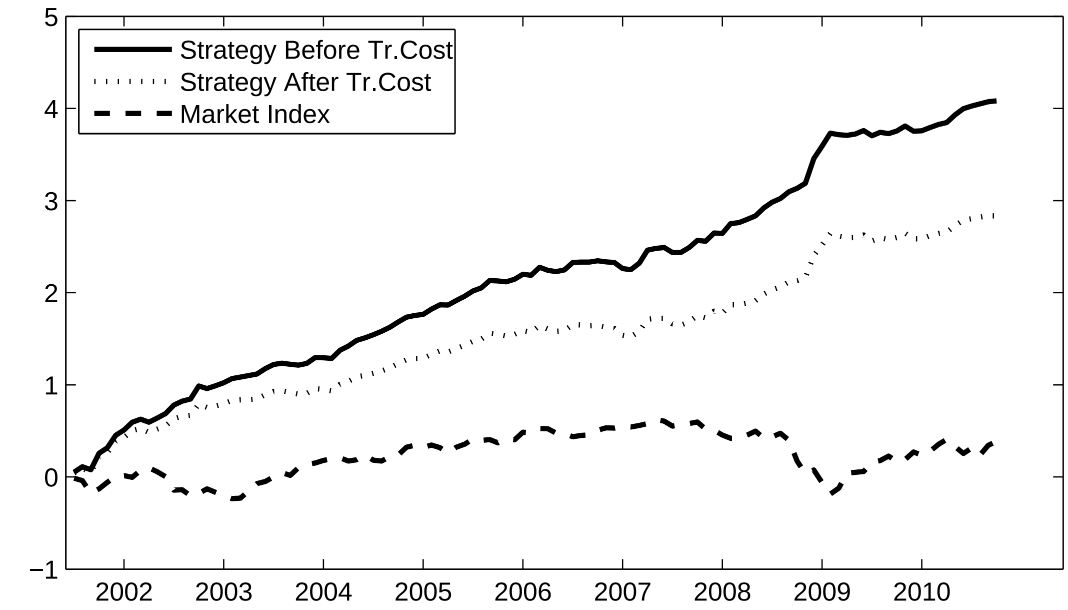


Figure 11. Strategy historical performance on the S&P 600 SmallCap data set for top 5 pairs portfolio before and after transaction cost

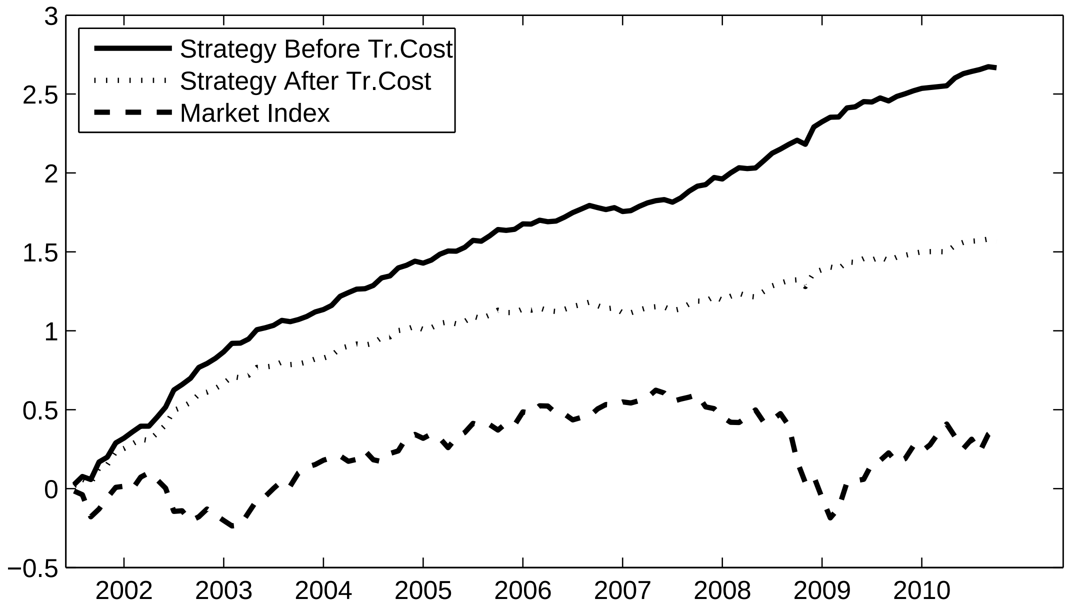


Figure 12. Strategy historical performance on the S&P 600 SmallCap data set for top 20 pairs portfolio before and after transaction cost

Appendix A: Proof of the Theorem

Before proving the theorem we need some axillary lemmas.

Lemma A.1: *Let $\{x_t\}$ be the Ornstein–Uhlenbeck process with mean μ, standard deviation σ and $\lambda > 0$ defined by*

$$dx_t = \lambda(\mu - x_t)\,dt + \sigma\,dB_t, \tag{A1}$$

where B_t is a standard Brownian motion.

Then it can be represented as a time change of another Brownian motion W:

$$x_t = x_0 e^{-\lambda t} + \mu(1 - e^{-\lambda t}) + \frac{\sigma}{\sqrt{2\lambda}}\,e^{-\lambda t} W(e^{2\lambda t} - 1).$$

Proof: The solution of (A1) is

$$x_t = x_0 e^{-\lambda t} + \int_0^t e^{-\lambda(t-s)} \lambda\mu\,ds + \sigma \int_0^t e^{-\lambda(t-s)}\,dB_s.$$

Then

$$e^{\lambda t} x_t = x_0 + \mu\left(e^{\lambda t} - 1\right) + \sigma \int_0^t e^{\lambda s}\,dB_s.$$

Now we apply a time-change τ,

$$e^{\lambda\tau(t)} x_{\tau(t)} = x_0 + \mu\left(e^{\lambda\tau(t)} - 1\right) + \sigma \int_0^{\tau(t)} e^{\lambda s}\,dB_s. \tag{A2}$$

The last integral is

$$\int_0^{\tau(t)} e^{\lambda s}\,dB_s = \int_0^t e^{\lambda\tau(s)}\,dB_{\tau(s)} = \int_0^t e^{\lambda\tau(s)} \sqrt{\tau'(s)}\,d\widetilde{W}_s \tag{A3}$$

where $\widetilde{W}$ is an another Brownian motion,

$$\widetilde{W}_t = \int_0^t \frac{1}{\sqrt{\tau'(s)}}\,dB_s.$$

We choose τ so that

$$e^{\lambda\tau(s)}\sqrt{\tau'(s)} = 1$$

$$\tau(t) = \frac{1}{2\lambda}\log(2\lambda t + 1).$$

Take an inverse

$$\tau^{-1}(t) = \frac{1}{2\lambda}(e^{2\lambda t} - 1).$$

Now plug this into (A2)

$$
\begin{aligned}
e^{\lambda t} x_t &= x_0 + \mu \left(e^{\lambda t} - 1 \right) + \sigma \widetilde{W}(\tau^{-1}(t)) \\
\widetilde{W}(\tau^{-1}(t)) &= \widetilde{W} \left(\frac{1}{2\lambda} (e^{2\lambda t} - 1) \right) \\
&= \frac{1}{\sqrt{2\lambda}} W(e^{2\lambda t} - 1)
\end{aligned}
$$

where $W(t) = \sqrt{2\lambda}\, \widetilde{W}\left(\frac{1}{2\lambda}\right)$ is also a standard Brownian motion.

Hence, the Ornstein–Uhlenbeck process can be represented as a time-change of the Brownian motion

$$
x_t = x_0 e^{-\lambda t} + \mu(1 - e^{-\lambda t}) + \frac{\sigma}{\sqrt{2\lambda}}\, e^{-\lambda t} W(e^{2\lambda t} - 1).
$$

□

Lemma A.2: *Let $\{Y_t\}$ be the Ornstein–Uhlenbeck process with mean zero, variance one and $\lambda > 0$ on the time interval $[0, T]$*

$$
dY_t = -\lambda Y_t\, dt + dB_t. \tag{A4}
$$

We make a H*-construction on the Ornstein–Uhlenbeck process for some H as in Section 2. Then the* H*-inversion goes to infinity as time goes to infinity, that is*

$$
N_T(H, Y) \to \infty \text{ (a.s.) as } T \to \infty.
$$

Proof: Let $\varepsilon > 0$ and $Y_0 = -\varepsilon$. We take the Ornstein–Uhlenbeck process as a time-changed Brownian motion (Lemma A.1) and find the probability that the Ornstein–Uhlenbeck process is above ε, that is

$$
\mathrm{P}(Y_t > \varepsilon) \text{ as } t \to \infty
$$

The solution of (A4) is

$$
Y_t = -\varepsilon e^{-\lambda t} + \frac{1}{\sqrt{2\lambda}}\, e^{-\lambda t} W(e^{2\lambda t} - 1)
$$

Hence,

$$
\begin{aligned}
\mathrm{P}(Y_t > \varepsilon) &= \mathrm{P}\left(-\varepsilon e^{-\lambda t} + \frac{1}{\sqrt{2\lambda}}\, e^{-\lambda t} W(e^{2\lambda t}-1) > \varepsilon\right) \\
&= \mathrm{P}\left(W(e^{2\lambda t}-1) > \sqrt{2\lambda}\left(1+e^{\lambda t}\right)\varepsilon\right) \\
&= \mathrm{P}\left(\frac{W(e^{2\lambda t}-1)}{\sqrt{e^{2\lambda t}-1}} > \frac{\sqrt{2\lambda}(1+e^{\lambda t})\varepsilon}{\sqrt{e^{2\lambda t}-1}}\right) \\
&= 1-\Phi\left(\frac{\sqrt{2\lambda}(1+e^{\lambda t})\varepsilon}{\sqrt{e^{2\lambda t}-1}}\right) \\
&= 1-\Phi(\sqrt{2\lambda}\varepsilon) \text{ as } t\to\infty \\
&\neq 0
\end{aligned}
$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\, du.$$

In similar way we can show that the probability of the Ornstein–Uhlenbeck process with an initial value ε to be below $-\varepsilon$ does not equal zero either. So, the Ornstein–Uhlenbeck process never converges completely to its mean but fluctuates between $-\varepsilon$ and ε. Then, if we take $H \leq 2\varepsilon$

$$N_T(H, P) \to \infty \text{ (almost surely)}$$

as $T \to \infty$. □

Lemma A.3: *Limiting state probability of the recombining binomial tree approximation of the Ornstein–Uhlenbeck process* $\{y_n\}$ *being on the level* m *is*

$$\mathrm{Q}(m) = \mathrm{Q}(0)\,\frac{1}{2}\, e^{-\lambda\, m(m-1)}\left(e^{-2\lambda\, m} + 1\right)$$

where

$$\mathrm{Q}(0) = \left(1 + \sum_{i=1}^{\infty} e^{-\lambda\, i(i-1)}\left(e^{-2\lambda\, i} + 1\right)\right)^{-1}.$$

Proof: Let $\{x_t\}$ be the Ornstein–Uhlenbeck process with mean zero and $\rho > 0$

$$dx_t = -\rho x_t\, dt + \sigma dB_t$$

and let $\{y_n\}$ be a recombining binomial tree approximation of the Ornstein–Uhlenbeck process $\{x_t\}$ with the probability of moving up from the state y_n

$$\mathrm{P}^{\uparrow}(y_n) = \frac{1}{2} + \frac{1}{2}\tanh\left[\frac{\rho(-y_n)}{\sigma}\sqrt{\Delta t}\right] \tag{A5}$$

and the size of step up or down

$$H = \sigma\sqrt{\Delta t}. \tag{A6}$$

If we take $\Delta t = 1, \rho = \lambda$ and $\sigma = 1$, then the size of up or down movement equals 1 and the process $\{y_n\}$ takes integer values $y_n = m, m \in [-n, \dots, -2, -1, 0, 1, 2, \dots, n]$. The probability of moving up from the level m is

$$\mathrm{P}^{\uparrow}(m) = \frac{1}{2} + \frac{1}{2}\tanh\left(-\lambda\, m\right). \tag{A7}$$

We are interested in the limiting probability $\mathrm{P}(y_n = m)$ as $n \to \infty$, that is the process $\{y_n\}$ is on any given level m. We use a brief notation $\mathrm{Q}(m) = \mathrm{P}(y_n = m)$ as $n \to \infty$ and $\mathrm{P}^{\uparrow}(m)$ and $\mathrm{P}^{\downarrow}(m)$ for the probability of up and down movements from the level m.

The process $\{y_n\}$ is symmetrical around zero. Then

$$\mathrm{Q}(0) = \mathrm{P}^{\downarrow}(1)\,\mathrm{Q}(1) + \mathrm{P}^{\uparrow}(-1)\,\mathrm{Q}(-1) = 2\,\mathrm{P}^{\downarrow}(1)\,\mathrm{Q}(1). \tag{A8}$$

Working in the similar way and taking $\mathrm{P}^{\downarrow}(0) = \mathrm{P}^{\uparrow}(0) = 1/2$

$$\begin{aligned}
\mathrm{Q}(1) &= \mathrm{P}^{\downarrow}(2)\,\mathrm{Q}(2) + \mathrm{P}^{\uparrow}(0)\,\mathrm{Q}(0)\\
&= \mathrm{P}^{\downarrow}(2)\,\mathrm{Q}(2) + \mathrm{P}^{\uparrow}(0)\,2\,\mathrm{P}^{\downarrow}(1)\,\mathrm{Q}(1)\\
&= \mathrm{P}^{\downarrow}(2)\,\mathrm{Q}(2) + \mathrm{P}^{\downarrow}(1)\,\mathrm{Q}(1)\\
\mathrm{Q}(1)(1 - \mathrm{P}^{\downarrow}(1)) &= \mathrm{P}^{\downarrow}(2)\,\mathrm{Q}(2)\\
\mathrm{Q}(2) &= \mathrm{Q}(1)\frac{\mathrm{P}^{\uparrow}(1)}{\mathrm{P}^{\downarrow}(2)} \text{ and } \mathrm{Q}(1) = \mathrm{Q}(2)\frac{\mathrm{P}^{\downarrow}(2)}{\mathrm{P}^{\uparrow}(1)}
\end{aligned}$$

Now we employ the same approach for the next level m

$$\begin{aligned}
\mathrm{Q}(2) &= \mathrm{P}^{\downarrow}(3)\,\mathrm{Q}(3) + \mathrm{P}^{\uparrow}(1)\,\mathrm{Q}(1)\\
&= \mathrm{P}^{\downarrow}(3)\,\mathrm{Q}(3) + \mathrm{P}^{\uparrow}(1)\,\mathrm{Q}(2)\frac{\mathrm{P}^{\downarrow}(2)}{\mathrm{P}^{\uparrow}(1)}\\
&= \mathrm{P}^{\downarrow}(3)\,\mathrm{Q}(3) + \mathrm{P}^{\downarrow}(2)\,\mathrm{Q}(2)\\
\mathrm{Q}(2)(1 - \mathrm{P}^{\downarrow}(2)) &= \mathrm{P}^{\downarrow}(3)\,\mathrm{Q}(3)\\
\mathrm{Q}(3) &= \mathrm{Q}(2)\frac{\mathrm{P}^{\uparrow}(2)}{\mathrm{P}^{\downarrow}(3)}
\end{aligned}$$

We can repeat this exercise for the following levels and get a recursive relation for the limiting probability of being on level m is

$$\mathrm{Q}(m) = \mathrm{Q}(m-1)\frac{\mathrm{P}^{\uparrow}(m-1)}{\mathrm{P}^{\downarrow}(m)} \tag{A9}$$

To prove the claim (A9) for the general case we assume that it is true for $m \leq k$ and check if it holds for $m = k+1$. It is clear that the probability of being on level k is

$$\mathrm{Q}(k) = \mathrm{Q}(k-1)\mathrm{P}^{\uparrow}(k-1) + \mathrm{Q}(k+1)\mathrm{P}^{\downarrow}(k+1).$$

By the formula (A9) we get

$$\begin{aligned}\mathrm{Q}(k) &= \mathrm{Q}(k-1)\frac{\mathrm{P}^{\uparrow}(k-1)}{\mathrm{P}^{\downarrow}(k)}\\ &\Rightarrow \mathrm{Q}(k-1)\mathrm{P}^{\uparrow}(k-1) = \mathrm{Q}(k)\mathrm{P}^{\downarrow}(k).\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{Q}(k) &= \mathrm{Q}(k)\mathrm{P}^{\downarrow}(k) + \mathrm{Q}(k+1)\mathrm{P}^{\downarrow}(k+1)\\ \mathrm{Q}(k)[1-\mathrm{P}^{\downarrow}(k)] &= \mathrm{Q}(k+1)\mathrm{P}^{\downarrow}(k+1)\\ \mathrm{Q}(k)\mathrm{P}^{\uparrow}(k) &= \mathrm{Q}(k+1)\mathrm{P}^{\downarrow}(k+1)\\ \mathrm{Q}(k+1) &= \mathrm{Q}(k)\frac{\mathrm{P}^{\uparrow}(k)}{\mathrm{P}^{\downarrow}(k+1)}.\end{aligned}$$

So, claim (A9) is true for $m = k+1$, then we can conclude that (A9) holds for all m by the Principle of Mathematical Induction.

The probability $\mathrm{Q}(m) = \mathrm{Q}(-m)$ as the process $\{y_n\}$ is symmetrical. Then for $m > 0$

$$\mathrm{Q}(m) = \mathrm{Q}(0)\prod_{j=0}^{m-1}\frac{\mathrm{P}^{\uparrow}(j)}{\mathrm{P}^{\downarrow}(j+1)}$$

We know the probability of up or down movement, then

$$\begin{aligned}\frac{\mathrm{P}^{\uparrow}(i)}{\mathrm{P}^{\downarrow}(i+1)} &= \frac{\frac{1}{2}+\frac{1}{2}\tanh(-\lambda i)}{\frac{1}{2}-\frac{1}{2}\tanh(-\lambda(i+1))}\\ &= \frac{1+\dfrac{e^{-2\lambda i}-1}{e^{-2\lambda i}+1}}{1-\dfrac{e^{-2\lambda(i+1)}-1}{e^{-2\lambda(i+1)}+1}} = e^{-2\lambda i}\,\frac{e^{-2\lambda(i+1)}+1}{e^{-2\lambda i}+1}\end{aligned}$$

It follows that

$$\begin{aligned}\prod_{j=0}^{m-1}\frac{\mathrm{P}^{\uparrow}(j)}{\mathrm{P}^{\downarrow}(j+1)} &= e^{-2\lambda 0}\,\frac{e^{-2\lambda 1}+1}{e^{-2\lambda 0}+1}\,e^{-2\lambda 1}\,\frac{e^{-2\lambda 2}+1}{e^{-2\lambda 1}+1}\,e^{-2\lambda 2}\,\frac{e^{-2\lambda 3}+1}{e^{-2\lambda 2}+1}\cdots e^{-2\lambda(m-1)}\,\frac{e^{-2\lambda m}+1}{e^{-2\lambda(m-1)}+1}\\ &= \frac{1}{2}\exp\left(-\lambda\, m(m-1)\right)\left(\exp\left(-2\lambda\, m\right)+1\right).\end{aligned}$$

Also due to the symmetry of the Ornstein–Uhlenbeck process the limiting probability of being on level $m = 0$ can be calculated as

$$
\begin{aligned}
\mathrm{Q}(0) &= 1 - 2(\mathrm{Q}(1) + \mathrm{Q}(2) + \mathrm{Q}(3) + \dots) \\
&= 1 - 2\,\mathrm{Q}(0) \sum_{m=1}^{\infty} \left(\prod_{j=0}^{m-1} \frac{\mathrm{P}^{\uparrow}(j)}{\mathrm{P}^{\downarrow}(j+1)} \right) \\
&= 1 - \mathrm{Q}(0) \sum_{m=1}^{\infty} \exp\left(-\lambda\, m(m-1)\right) \left(\exp\left(-2\lambda\, m\right) + 1\right) \\
&= \frac{1}{1 + \sum_{m=1}^{\infty} \exp\left(-\lambda\, m(m-1)\right) \left(\exp\left(-2\lambda\, m\right) + 1\right)}
\end{aligned}
$$

Hence, the limiting probability that the Ornstein–Uhlenbeck process $\{y_n\}$ is on level m is

$$
\mathrm{Q}(m) = \mathrm{Q}(0)\, \frac{1}{2}\, e^{-\lambda\, m(m-1)} \left(e^{-2\lambda\, m} + 1\right)
$$

where

$$
\mathrm{Q}(0) = \left(1 + \sum_{i=1}^{\infty} e^{-\lambda\, i(i-1)} \left(e^{-2\lambda\, i} + 1\right)\right)^{-1}.
$$

$\square$

Remark: If we consider a more general case with an arbitrary value of the step up or down H, that is a classical renko chart, to get a recombining binomial tree approximation, the value Δt in (A5) should be scaled

$$
\Delta t = \left(\frac{H}{\sigma}\right)^2.
$$

Alternatively, one can keep $\Delta t = 1$ and uses (A7) with the unit increments and scaled coefficient of mean-reversion

$$
\lambda = \rho \frac{H}{\sigma}.
$$

Lemma A.4: *The Ornstein–Uhlenbeck process satisfies the strong mixing condition (α-mixing).*

Proof: For any two sequences $\{\xi\} = \mathcal{U}'$ and $\{\eta\} = \mathcal{U}''$ with finite second moments we have the following index (Kolmogorov and Rozanov 1960):

$$
\rho(\mathcal{U}', \mathcal{U}'') = \sup_{\xi,\eta} \frac{|\mathbf{E}[(\xi - \mathbf{E}[\xi])(\eta - \mathbf{E}[\eta])]|}{\sqrt{\mathbf{E}[(\xi - \mathbf{E}[\xi])^2]\mathbf{E}[(\eta - \mathbf{E}[\eta])^2]}}
$$

If $\mathcal{U}'$ and $\mathcal{U}''$ are respectively the collections of all random variables which are measurable with respect to the σ-algebras $\mathcal{M}'$ and $\mathcal{M}''$, then

$$\rho(\mathcal{M}', \mathcal{M}'') = \rho(\mathcal{U}', \mathcal{U}'')$$

is the maximal correlation coefficient between the σ-algebras $\mathcal{M}'$ and $\mathcal{M}''$.

Let $x(t)$ be the Ornstein–Uhlenbeck process

$$dx(t) = -\lambda\, x(t)\, dt + \sigma\, dB_t$$

then its stationary solution is

$$x(t) = \sigma \int_{-\infty}^{t} e^{-\lambda(t-u)}\, dB_u.$$

For the process $x(t)$ we have the following two measures of dependence $\alpha(\tau) = \alpha(\mathcal{M}_{-\infty}^{t}, \mathcal{M}_{t+\tau}^{\infty})$ and $\rho(\tau) = \rho(\mathcal{M}_{-\infty}^{t}, \mathcal{M}_{t+\tau}^{\infty})$, where $\mathcal{M}_{s}^{t}$ is the σ-algebra of events which is determined by $x(u), s \le u \le t$ (Kolmogorov and Rozanov 1960).

For the Ornstein–Uhlenbeck process the maximal correlation coefficient between σ-algebras $\mathcal{M}_{-\infty}^{t}$ and $\mathcal{M}_{t+\tau}^{\infty}$ equals the module of the correlation coefficient between two closest points from the above σ-algebras—$x(t)$ and $x(t+\tau)$ and it depends only on τ

$$\rho(\tau) = \rho(\mathcal{M}_{-\infty}^{t}, \mathcal{M}_{t+\tau}^{\infty}) = |\rho(x(t), x(t+\tau))| = e^{-\lambda\tau} \tag{A10}$$

The correlation coefficient between any linear combination of any other random variables from σ-algebras $\mathcal{M}_{-\infty}^{t}$ and $\mathcal{M}_{t+\tau}^{\infty}$ is less than (A10). We can see it in the following example.

Example: Let $x(t)$ be the Ornstein–Uhlenbeck process and $s \le t < z$, $z = t + n$. Find the correlation between $x(t) + x(s)$ and $x(z)$.

The covariance between $x(t) + x(s)$ and $x(z)$ is

$$\begin{aligned}
&\mathbf{Cov}\,(x(t) + x(s), x(z)) = \\
&= \mathbf{E}\left[\sigma\left(\int_{-\infty}^{t} e^{-\lambda(t-u)}\, dB_u + \int_{-\infty}^{s} e^{-\lambda(s-u)}\, dB_u\right)\sigma\left(\int_{-\infty}^{z} e^{-\lambda(z-u)}\, dB_u\right)\right] \\
&= \sigma^2 \mathbf{E}\left[\int_{-\infty}^{t} e^{-\lambda(t-u)}\, dB_u \int_{-\infty}^{t+n} e^{-\lambda(t+n-u)}\, dB_u + \int_{-\infty}^{s} e^{-\lambda(s-u)}\, dB_u \int_{-\infty}^{t+n} e^{-\lambda(t+n-u)}\, dB_u\right] \\
&= \sigma^2\left(e^{-\lambda(t+t+n)} \int_{-\infty}^{t} e^{2\lambda u}\, du + e^{-\lambda(s+t+n)} \int_{-\infty}^{s} e^{2\lambda u}\, du\right) \\
&= \frac{\sigma^2}{2\lambda}\left(e^{-\lambda n} + e^{-\lambda(t-s+n)}\right) = \frac{\sigma^2}{2\lambda} e^{-\lambda n}\left(1 + e^{-\lambda(t-s)}\right)
\end{aligned}$$

The variance of $x(z)$ is $\sigma^2/2\lambda$ and variance of $x(t) + x(s)$ is

$$\mathbf{Var}(x(t) + x(s)) = \mathbf{E}\left[\left(\sigma \int_{-\infty}^{t} e^{-\lambda(t-u)}\, dB_u + \sigma \int_{-\infty}^{s} e^{-\lambda(s-u)}\, dB_u\right)^2\right] = \frac{\sigma^2}{\lambda}\left(1 + e^{-\lambda(t-s)}\right)$$

Then the correlation between $x(t) + x(s)$ and $x(z)$ is

$$\begin{aligned}\mathbf{Corr}\,(x(t)+x(s),x(z)) &= \frac{\mathbf{Cov}\,(x(t)+x(s),x(z))}{\sqrt{\mathbf{Var}(x(t)+x(s))\,\mathbf{Var}(x(z))}}\\ &= e^{-\lambda n}\sqrt{\frac{1+e^{-\lambda(t-s)}}{2}} \le e^{-\lambda n} \text{ as } t \ge s\end{aligned}$$

The maximal correlation coefficient $\rho(\tau)$ between σ-algebras $\mathcal{M}^t_{-\infty}$ and $\mathcal{M}^{\infty}_{t+\tau}$ goes to zero as $\tau \to \infty$. This equivalents to $\alpha(\tau) \to 0$ as $\tau \to \infty$ (Kolmogorov and Rozanov 1960, Bradley 2005), where $\alpha(\tau)$ is a measure of dependence (Rosenblatt 1956)

$$\alpha(\tau) = \alpha(\mathcal{M}', \mathcal{M}'') = \sup_{A' \in \mathcal{M}', A'' \in \mathcal{M}''} |\mathrm{P}(A' \bigcap A'') - \mathrm{P}(A')\mathrm{P}(A'')|.$$

Hence, the Ornstein–Uhlenbeck process $\{x_t\}$ possesses the property of strong mixing. □

Theorem A.5 : H-*volatility of the Ornstein–Uhlenbeck process.*
Let $P(t)$ be an Ornstein–Uhlenbeck process with mean zero and $\rho > 0$

$$dP(t) = -\rho P(t)\,dt + \sigma dB_t$$

Then for any positive H satisfying (1), the H-*volatility is less than* $2H$

$$\lim_{T\to\infty} \xi_T(H,P) < 2H \tag{A11}$$

Proof: Let $\{(\tau^a_n, \tau^b_n), n = 0, 1, ..., N\}$ be a time sequence defined on the Ornstein–Uhlenbeck process $P(t)$ as in Section 2.

The H-inversion is a number of times H-process changes its direction and equals the number of stopping times τ^b_n when that change of direction manifests itself. Then by Lemma A.2

$$N = N_T(H,P) \to \infty \text{ (almost surely)} \tag{A12}$$

as $T \to \infty$.

We define the distance between the two sequential local extremums

$$\begin{aligned} c_n &= |P(\tau^a_n) - P(\tau^a_{n-1})|\\ &= (P(\tau^a_n) - P(\tau^a_{n-1}))\cdot \mathrm{sign}(P(\tau^a_n) - P(\tau^a_{n-1}))\\ &= (P(\tau^a_n) - P(\tau^b_n) + P(\tau^b_n) - P(\tau^a_{n-1}) + P(\tau^b_{n-1}) - P(\tau^b_{n-1}))\\ &\quad \cdot \mathrm{sign}(P(\tau^a_n) - P(\tau^a_{n-1}))\\ &= \left[\left(P(\tau^a_n) - P(\tau^b_n)\right) - \left(P(\tau^a_{n-1}) - P(\tau^b_{n-1})\right) + \left(P(\tau^b_n) - P(\tau^b_{n-1})\right)\right]\\ &\quad \cdot \mathrm{sign}\left(P(\tau^a_n) - P(\tau^a_{n-1})\right). \end{aligned}$$

The distance between $P(\tau^a_n)$ and $P(\tau^b_n)$ is equal to H by the rules of renko and kagi constructions, but we need to know the sign for that distance. There are two possible cases:

1. $P(\tau_n^a)$ is a local maximum and $P(\tau_{n-1}^a)$ is a local minimum, then

$$\left[\left(P(\tau_n^a)-P(\tau_n^b)\right)-\left(P(\tau_{n-1}^a)-P(\tau_{n-1}^b)\right)\right]\cdot\operatorname{sign}\left(P(\tau_n^a)-P(\tau_{n-1}^a)\right)$$
$$=[H-(-H)]\cdot 1=2H$$

2. $P(\tau_n^a)$ is a local minimum and $P(\tau_{n-1}^a)$ is a local maximum, then

$$\left[\left(P(\tau_n^a)-P(\tau_n^b)\right)-\left(P(\tau_{n-1}^a)-P(\tau_{n-1}^b)\right)\right]\cdot\operatorname{sign}\left(P(\tau_n^a)-P(\tau_{n-1}^a)\right)$$
$$=[-H-H]\cdot(-1)=2H$$

It follows that

$$\begin{aligned}c_n&=|P(\tau_n^a)-P(\tau_{n-1}^a)|\\&=2H+(P(\tau_n^b)-P(\tau_{n-1}^b))\cdot\operatorname{sign}(P(\tau_n^a)-P(\tau_{n-1}^a))\end{aligned}\tag{A13}$$

The value of $\operatorname{sign}(P(\tau_n^a)-P(\tau_{n-1}^a))$ is completely defined by the process $\{P(t),t\in[\tau_{n-1}^b,\tau_n^b]\}$ and known at the stopping time τ_n^b of the Ornstein–Uhlenbeck process, but $c_n=|P(\tau_n^a)-P(\tau_{n-1}^a)|$ are not independent. However, it is *'nearly'* independent. The sequence $\{c_n\}$ is stationary, as the distribution of the random vector $(c_n,c_{n+1},...,c_{n+k})$ does not depend on n, and α-mixing with $\alpha_n=0$ for large n by Lemma A.4. Hence by the Central Limit Theorem for Dependant Variables (Billingsley (1995), Theorem 27.4)

$$\begin{aligned}\lim_{T\to\infty}\xi_T(H,P)&=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}c_n\\&=\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}|P(\tau_n^a)-P(\tau_{n-1}^a)|\\&\to\mathbf{E}[|P(\tau_1^a)-P(\tau_0^a)|]\ \text{(a.s.) as } T\to\infty\end{aligned}\tag{A14}$$

Now we have to separate the proofs for renko and kagi constructions.

First we prove (A11) for renko construction. We consider a sequence of random variables $\{d_k,k=1,2,...\}$ such that

$$d_k=\begin{cases}1, & p_k\\-1, & 1-p_k\end{cases}\tag{A15}$$

Define the process

$$\gamma_n=\sum_{k=1}^{n}d_k,\quad n=1,2,...\tag{A16}$$

It is clear that the process $\{\gamma_n\}$ is a recombining binomial tree approximation of the Ornstein–Uhlenbeck process (Hoek and Elliott 2006) which has the following general formula for the probability of moving up

$$p_n = \frac{1}{2} + \frac{1}{2}\tanh\left[\frac{\rho(\mu - P(n))}{\sigma}\sqrt{\Delta t}\right]$$

For the process $\{\gamma_n\}$ we take the probability in (A15) as

$$p_n = \frac{1}{2} + \frac{1}{2}\tanh\left(-\lambda\gamma_n\right).$$

Under the probability p_n the process $\{\gamma_n\}$ defined by (A15) and (A16) is a recombining binomial tree approximation of the Ornstein–Uhlenbeck process (19) with $\lambda = \rho\frac{H}{\sigma}$ and $\mu = 0$.

It follows from the definition of the stopping times τ_i for the renko process in (2) that

$$\frac{P(\tau_i)}{H} \overset{Law}{=} \gamma_n \tag{A17}$$

$$\frac{P(\tau_i) - P(\tau_{i-1})}{H} \overset{Law}{=} d_n \tag{A18}$$

We define a random variable

$$\nu = \min\{n \geq 1 : \gamma_n = n - 2\} \tag{A19}$$

or equivalently

$$\nu = \min\{n \geq 1 : \max_{t\in[0,n]}(\gamma_t) - \gamma_n = 1\} \tag{A20}$$

which is a time of the first downfall of $\{\gamma_n\}$. We assume that γ_{n-1} is a local maximum discovered at time $\nu = n$. The case with local minimum works in similar way due to the symmetry of the Ornstein–Uhlenbeck process.

From (A13) we have

$$\begin{aligned} |P(\tau_n^a) - P(\tau_{n-1}^a)| &\overset{Law}{=} (2H + \gamma_\nu H) \\ &= (2H + (\nu - 2)H) \\ &= \nu H \end{aligned} \tag{A21}$$

$$\mathbf{E}[|P(\tau_1^a) - P(\tau_0^a)|] = H\mathbf{E}[\nu] \tag{A22}$$

As the variable ν is a time of the first downfall after the number of raises then its probability follows a geometric distribution with probability of ‘success’ $p_n = \frac{1}{2} - \frac{1}{2}\tanh\left(-\lambda\gamma_n\right)$. The expected value of ν is

$$\mathbf{E}[\nu] = \sum_{n=1}^{\infty} n\left(\frac{1}{2} + \frac{1}{2}\tanh(-\lambda\gamma_n)\right)^{n-1}\left(\frac{1}{2} - \frac{1}{2}\tanh(-\lambda\gamma_n)\right)^{1} \tag{A23}$$

where the current value of the process $\gamma_n = \gamma_0 + n - 1$.

An initial value of the process γ_0 can take any integer value from the minimal to maximal value of the process $\{\gamma_n\}$.

Then

$$\mathbf{E}[\nu] = \sum_{k=-\infty}^{\infty} \mathrm{P}(k) \sum_{n=0}^{\infty} \frac{(n+1)}{2^{n+1}} \left(1+\tanh(-\lambda(k+n))\right)^n \left(1-\tanh(-\lambda(k+n))\right)$$
$$= \sum_{n=0}^{\infty} \frac{(n+1)}{2^{n+1}} \sum_{k=-\infty}^{\infty} \mathrm{P}(k) \left(1+\tanh(-\lambda(k+n))\right)^n \left(1-\tanh(-\lambda(k+n))\right) \qquad \text{(A24)}$$

where k takes integer values from $(-\infty, \infty)$ and $\mathrm{P}(k)$ is the probability that the initial value equals k

$$\mathrm{P}(k) = \mathrm{P}(\gamma_0 = k).$$

The density function of γ_0 is provided by Lemma A.3.

$$\mathrm{P}(k) = \mathrm{P}(0)\, \frac{1}{2}\, e^{-\lambda\, k(k-1)} \left(e^{-2\lambda\, k} + 1\right)$$

where

$$\mathrm{P}(0) = \left(1 + \sum_{i=1}^{\infty} e^{-\lambda\, i(i-1)} \left(e^{-2\lambda\, i} + 1\right)\right)^{-1}.$$

Consider the second summation in (A24)

$$\sum_{k=-\infty}^{\infty} \mathrm{P}(k) \left(1+\tanh(-\lambda(k+n))\right)^n \left(1-\tanh(-\lambda(k+n))\right) =$$
$$= \sum_{k=-\infty}^{\infty} \frac{e^{-\lambda\, k(k-1)} \left(e^{-2\lambda\, k} + 1\right)}{2\left(1 + \sum_{i=1}^{\infty} e^{-\lambda\, i(i-1)} \left(e^{-2\lambda\, i} + 1\right)\right)} \cdot$$
$$\cdot \left(1+\tanh(-\lambda(k+n))\right)^n \left(1-\tanh(-\lambda(k+n))\right) < 1 \qquad \text{(A25)}$$

It looks impossible to get a closed form solution for this equation, however numerical simulations show that (A25) is less than 1 for any $\lambda > 0$ and $n \geq 0$. Hence, (A24) takes form

$$\mathbf{E}[\nu] < \sum_{n=0}^{\infty} \frac{(n+1)}{2^{n+1}} = 2 \qquad \text{(A26)}$$

So, from (A22) and (A26) we conclude that for the renko construction on the Ornstein–Uhlenbeck process the H-volatility is less than $2H$

$$\xi_T(H, P) < 2H.$$

We now prove (A11) for the kagi construction. Let

$$\theta = \min\{u \geq 0 : \max_{t \in [0,u]} P(t) - P(u) = H\} \qquad \text{(A27)}$$

By Lemma A.1 $P(\theta)$ can be represented as a time-changed Wiener process defined by the mean-reverting property of the Ornstein–Uhlenbeck process. Then it follows

$$|P(\tau_1^b) - P(\tau_0^b)| = |P(\tau_0^b)e^{-\rho\theta} + \frac{1}{\sqrt{2\rho}}\, e^{-\rho\theta} W(e^{2\rho\theta} - 1)| \tag{A28}$$

and by (A13)

$$\mathbf{E}[|P(\tau_1^a) - P(\tau_0^a)|] = \mathbf{E}\left[2H + OU_\theta\right] \tag{A29}$$

where

$$\begin{aligned} OU_\theta &= OU(\tau_0^b, \tau_1^b) \\ &= \left(P(\tau_0^b)e^{-\rho\theta} + \frac{1}{\sqrt{2\rho}}\, e^{-\rho\theta} W(e^{2\rho\theta} - 1)\right) \operatorname{sign}(P(\tau_1^a) - P(\tau_0^a)) \end{aligned} \tag{A30}$$

As the process $P(t)$ is the Ornstein–Uhlenbeck process then the initial point $P(\tau_0^b)$ is normally distributed with mean zero and standard deviation $\sigma/\sqrt{2\rho}$. Hence

$$\begin{aligned} \frac{OU(\tau_0^b, \tau_1^b)}{\operatorname{sign}(P(\tau_1^a) - P(\tau_0^a))} &= \int_{\mathbb{R}} \left(x\, e^{-\rho\theta} + \frac{1}{\sqrt{2\rho}}\, e^{-\rho\theta} W(e^{2\rho\theta} - 1)\right) e^{-\frac{\rho x^2}{\sigma^2}} \frac{\sqrt{\rho}}{\sqrt{\pi\sigma^2}}\, dx \\ &= \frac{1}{\sqrt{2\rho}}\, e^{-\rho\theta} W(e^{2\rho\theta} - 1) \\ &= \frac{1}{\sqrt{2\rho}}\, W(1 - e^{-2\rho\theta}) \end{aligned}$$

$$W(1 - e^{-2\rho\theta}) \overset{D}{<} W(2\rho\theta) \text{ for all } \rho > 0 \tag{A31}$$

Hence the Ornstein–Uhlenbeck process is smaller in distribution than the Wiener process

$$OU_\theta < W_\theta$$

We can get the same result from the maximal inequalities for the Ornstein–Uhlenbeck process (Graversen and Peskir 2000). The Ornstein–Uhlenbeck process starting from its mean in average behaves as $\sqrt{\log(1+t)}$ while the Wiener process behaves as $\sqrt{t}$. So, the Ornstein–Uhlenbeck process is smaller in distribution than the Wiener process for any $t > 0$.

Then it follows from (A29)

$$\begin{aligned} \mathbf{E}[|P(\tau_1^a) - P(\tau_0^a)|] &= \mathbf{E}\left[2H + OU_\theta\right] \\ &< \mathbf{E}[2H + W_\theta] \\ &= H\,\mathbf{E}\left[1 + \left(1 + \frac{W_\theta}{H}\right)\right] \\ &= H\left(1 + \mathbf{E}\left[1 + \frac{W_\theta}{H}\right]\right) \\ &= H\left(1 + \int_0^\infty x\, e^{-x}\, dx\right) = 2H \end{aligned} \tag{A32}$$

So, for the kagi construction over the Ornstein–Uhlenbeck process the H-volatility is less than $2H$.

$$\xi_T(H, P) < 2H$$

$\square$