

High Frequency Statistical Arbitrage via the Optimal Thermal Causal Path

V L Raju Chinthalapati

Department of Mathematics, London School of Economics
Department of Accounting & Finance, University of Greenwich

6th October 2011

Abstract

We consider the problem of identifying similarities and causality relationships in a given set of financial time series data streams. We develop further the “Optimal Thermal Causal Path” [28, 27] method, which is a non-parametric method proposed by Sornette et al. The method considers the mismatch between a given pair of time series in order to identify the expected minimum energy path lead-lag structure between the pair. Traders may find this a useful tool for directional trading, to spot arbitrage opportunities. We add a curvature energy term to the method and we propose an approximation technique to reduce the computational time. We apply the method and approximation technique on various market sectors of NYSE data and extract the highly correlated pairs of time series. We show how traders could exploit arbitrage opportunities by using the method.

Keywords Statistical Arbitrage, Time-series Classification, Optimal Thermal Causal Path.

1 Introduction

The success of a trader in the stock market depends on various factors, including how well the trader predicts the market direction. A technical trading strategy is nothing but a set of investment decisions based on the patterns in stock prices. Both academic literature and practitioner literature have identified mixed results

for the performance of the trading strategies. Lo et al. [19] study the process of identifying most popular price patterns that are commonly used by practitioners and evaluate the performance of associated trading strategies. For most traders, directional trading is an attractive trading strategy for making large profits [16]. Directional trading can be defined as opening a position (which means buying or selling) to take advantage of an expected price move in a security. It is essentially taking advantage of the forward price trends in securities. Directional trading involves a bet on the price direction of a given security. Traders will benefit from a rise in the price if they are long (buying) or from a decline in the price if they are short (selling). Trading strategies based on causality relation patterns between two time series are considered to be one of the important class of directional trading trading strategies. Limited academic literature is available on trading strategies based on causality relation. Brooks et al. [6] examine the lead-lag relationship between the FTSE 100 index and index futures price using various time series models and use the predictive ability to derive a trading strategy. Shan et al. [26] study the causality relation between news and financial instruments trading activities using data mining techniques in order to build trading models. Kleinberg et al. [18] study causal relationships in stock returns with temporal logic based methods and apply the methods to construct optimal trading rules. Directional trading strategies offer excellent returns (with risk, of course) [22]. The notable aspect of this trading strategy is that it tends to be most profitable during volatile conditions [1]. Hedge funds and proprietary trading desks of investment banks have identified directional trading as a profitable arbitrage strategy [2].

Informally, according to a directional trading strategy, we buy a security if we know the security price is going up, and we sell if we know the security price is going down. Although it seems simple, the non-trivial question is “how do we know the direction of a given security price”? One possible answer to this question is to identify a security that could “lead” the given security, meaning that the price of the first security follows (approximately) the trend in price of the second. Now, the next level of non-trivial question is “how do we find a security that leads the given security price”? It turns out that we have to identify causality relations between pairs of securities. Good causal relationships can be identified between a pair of similar time sequences. Various distance metrics can be used for similarity classification [17], but the most popular one is the

Euclidean metric. Identifying similar time series has many applications in fields like engineering, medicine, economics and finance. Numerous algorithms have been proposed for time series classification, for example [17, 23, 24]. Among all these algorithms, Discrete Time Warping (DTW) enjoys the best reputation [29].

One key application of identifying similar time series is directional trading. In this paper, we consider multiple series which evolve over time, for example, stock prices. Filtering similar time series involves: (a) extracting all pairs of similar time series and (or) (b) extracting all series similar to a specified series from a collection of time series. For a given pair of time series $X = (x(0), x(1), \dots, x(N - 1))$ and $Y = (y(0), y(1), \dots, y(M - 1))$, in DTW, we find a mapping function between X and Y that satisfies certain conditions. We consider similarity measure under Euclidean metric. In addition to minimizing the cumulative Euclidean metric between the pair of time series, the other required conditions are (1) Boundary conditions, (2) Continuity, and (3) Monotonicity. The less the cumulative Euclidean metric, the more similarity can be identified between the time series. The mapping function $\phi : X \rightarrow Y$, between the time series describes the causality relation, which can be considered as the lead-lag structure between the pair of time series.

In reality the time series data is noisy and it might have unrelated patterns that could lead to wrong conclusions on the causality relation. This motivates us to find the expected causality relationship among the given set of time series data. We use the “Optimal Thermal Causal Path” [28, 27] method in order to identify the expected causality relation. We add a curvature energy term to the method to improve causality relation accuracy. Extracting a similar pair of time series from huge data sets is a computationally intensive process. In order to reduce the computational time, we propose an upper bounding measure that is similar to the lower bound of the fast search method for dynamic time warping (FTW) [24]. We propose an approximation scheme that reduces the computational time for NYSE TAQ data sets. From NYSE TAQ data sets, we extract the highly correlated stocks and conduct extensive experiments. We show how traders could exploit arbitrage opportunities by using the method.

The organisation of the rest of the paper is as follows. Section 2 discusses related work. Section 3 covers the required background material. In section 4 we show how one can use the modified optimal thermal causal path method for

directional trading. Section 5 describes the approximation scheme to reduce the computational time. Section 6 reviews the experiments for statistical arbitrage trading.

2 Related Work

A pure arbitrage is a financial transaction practice with no net investment that makes strictly positive payoff with no possibility of a negative payoff. In reality, even in the ideal case, it is rare to find pure arbitrage. As a relaxation of pure arbitrage, we can define expectations arbitrage or statistical arbitrage which is not risk-less.

A statistical arbitrage is a financial transaction practice with no net investment that makes a strictly positive expected payoff with no possibility of a negative payoff. But, the Efficient Market Hypothesis (EMH) states that there will not be any pricing anomalies and that prices in the market fully reflect the available market information at any time. In the literature, we find a lot of empirical research that could relate to EMH [25, 9, 10]. According to EMH, we cannot predict the price movements. Random walk models for price movements were developed and their history can be traced back to Bachelier's Ph.D. thesis in 1900 [4]. Empirical research suggested that market prices can be partially predicted [20]. In the academic literature we find quite a few statistical arbitrage related scholarly compositions [14, 12, 5, 7]. Bondarenko [5] proves that if the pricing kernel is path dependent, then no statistical arbitrage opportunities exist. Burgess [7] introduces general directions in which cointegration analysis can be generalised to statistical arbitrage. Getmansky and Lo [12] explain limits of arbitrage and why some of the statistical arbitrage opportunities might not be exploitable for a small hedge fund. Hogan et al. [14] describe how to verify the existence of statistical arbitrage and counter-argue with EMH. They introduce the concept of statistical arbitrage which tests the efficiency of market without specifying an equilibrium model.

Statistical arbitrage techniques were first used at Morgan Stanley in 1980 [11]. Popular mathematical concepts [8, 19] used in statistical arbitrage are: time series analysis methods, neural networks and pattern recognition methods, particle physics concepts etc. In this paper we use particle physics concepts of free energy and energy minimisation [28, 27].

The most popular technique of statistical arbitrage is pairs trading [11]. It is essentially a mean reverting version of statistical arbitrage. In pairs trading, the arbitrageur identifies a pair of securities whose price difference (spread) is fluctuating around a constant mean (believed in mean reversion). At some time, if the price difference is deviating (increasing or decreasing) from the believed constant mean, then the arbitrageur buys the underpriced security and short sells (selling a security that has been borrowed) the overpriced one. The hope of the arbitrageur is that the payoffs difference will converge back to the believed constant and he/she will make a profit. It is difficult to predict how long the divergence trend continues. If the divergence trend continues for a long time, the arbitrageur should invest more or 'close' the positions. The arbitrageur might lose the game if he does not have enough capital to cover the expenditure for such a long time. As John Keynes says: "Markets can remain irrational longer than you can remain solvent". The limits of the pairs trading technique of arbitrage is explained by Getmansky and Lo [12]. Since contrarian investment strategies are based on mean reversion principle, the profits from pairs trading might be merely a disguised way of exploiting these previously documented profits of contrarian strategies. Gatev et al. [11] study the profitability of pairs trading in the U.S. equity market by considering daily data from 1962 to 2002. Their bootstrap results suggest that the pairs trading effect differs from previously documented mean reversion profits. In a single study, Nath [21] documents pairs trading profitability of U.S. treasury bills, notes and bonds. Andrade et al. [3] study why the prices of similar stocks diverge by considering pairs trading as a framework.

Directional trading involves a bet on the price direction of a given security. Directional trading does not come under the mean reverting based statistical arbitrage as it depends on the lead-lag relation between the pair of securities. But in this paper, we predict the direction of a given security price by understanding its causality relationship with some other security price in a statistical sense; that is, we consider the expected causality relation. Since we use the expected causality relation for betting on the price, the returns are not risk-free. Unfortunately not much academic literature is available on the profitability and performance of these kinds of directional trading strategies. For example, Jorda and Taylor [15] study the performance of directional trading strategies and verify the claim that directional forecasts beat coin toss strategies.

3 A Prelude

3.1 Time Series Classification

A time series is a sequence of measurements for a variable at different time steps. Usually these measurements are taken at uniform time intervals. For time series analysis we consider large time series data bases, for example all stock prices of the NYSE. For the given sample of a pair of time series,

$$X = (x(0), x(1), \dots, x(N - 1)) \quad (1)$$

$$Y = (y(0), y(1), \dots, y(M - 1)) \quad (2)$$

we try to figure out the similarities between them.

Similarity search helps in (a) extracting all pairs of similar time series (classification), (b) extracting all series similar to a specified series from a collection of time series (clustering) and (c) extracting causality relationship among the set of time series (association rules) [17]. Quantification of similarity can be considered as domain specific and subjective. A similar pair of time series has minimum dissimilarity, and the simplest way to compute dissimilarity is by the Euclidean distance metric. For a pair of time series X and Y of equal length ($M = N$),

$$D(X, Y) = \sqrt{\sum_{i=0}^{N-1} (x(i) - y(i))^2} \quad (3)$$

The above Euclidean distance metric gives the total dissimilarity between X and Y that is based on one-to-one alignment of the pair of time series. Even though the given two time series are similar to each other, still they may have a phase difference. Dynamic Time Warping (DTW) is the most popular similarity measure technique that could incorporate nonlinear alignments [23]. When we are trying to capture causality relationship for example, lead-lag structure between two stock prices, we need to construct nonlinear alignments so that we can identify the appropriate lead-lag relation even if there exists a phase difference.

In DTW, we find the best possible alignment warp [17] ϕ between X and Y . For that we define a local distance measure $e(x(i), y(j)) = |x(i) - y(j)|$ and do a systematic comparison between all values of the given time series X

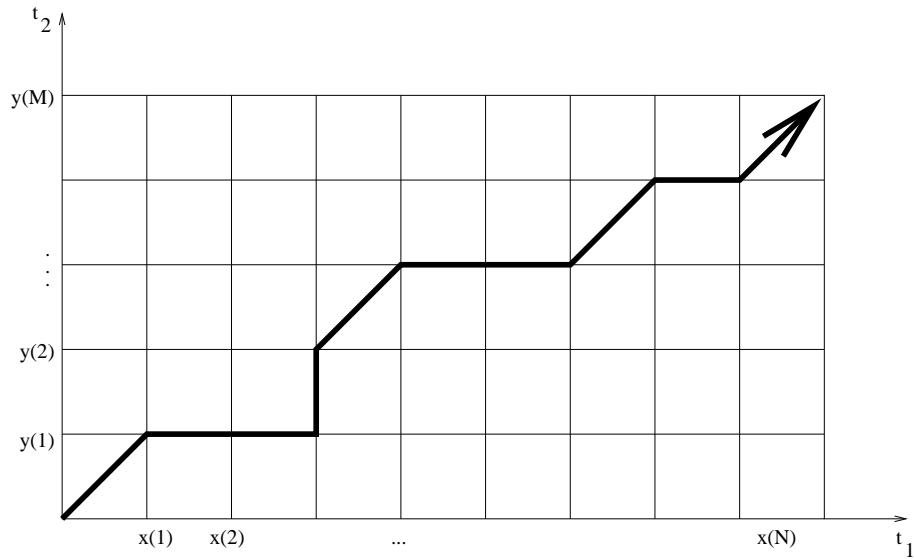


Figure 1: DTW

and Y . Figure 1 shows a two dimensional grid where each node $(x(i), y(j))$ is assigned a value that is equivalent to $e(x(i), y(j))$. As an analogy, $e(x(i), y(j))$ is considered as the local energy mismatch at the nodes of the grid for the given pair of time series. Extending the analogy further, one can consider the grid as the energy matrix for the given pair of time series. The causality relationship between X and Y can be explained by using the mapping from X to Y (the “warping path” between X and Y). The dissimilarity $E(X, Y)$, for the given mapping function ϕ is the cumulative sum of the local energies

$$E^\phi(X, Y) = \sum_i e(x(i), y(\phi(i))). \quad (4)$$

Since we need to minimise dissimilarity between X and Y , we simply search for the mapping achieving minimum energy, among all possible mappings in the energy grid. Let

$$E(X, Y) = \min_{\phi(t_1), t_1=0,1,\dots,N-1} \sum_{t_1=0}^{N-1} |x(t_1) - y(\phi(t_1))|, \quad (5)$$

We impose some constraints on ϕ for this optimisation problem: (1) end constraints: $\phi(0) = 0$ and $\phi(N - 1) = M - 1$ (where M is a parameter and

$M \leq N$, considered to be fixed for the moment) and (2) monotonicity and smoothing constraints: $0 \leq \phi(t_1 + 1) - \phi(t_1) \leq 1$. Note that the mapping ϕ can be a multi-valued function and in that case it is difficult to interpret. In order to ensure that ϕ is a well-defined function that express the dependence relation between the two time series, we may follow the convention [28] to map t_1 to the largest value t_2 of the vertical segment corresponding to t_1 [28, 27]. For example, in Figure 1, for instance, we have $\phi(3) = 2$. Note that, subject only to the second constraint, we have some flexibility in imposing the end constraints. The smoothing constraint, implies

(1) A path that starts at $(0, 0)$ will end at $(N - 1, \phi(N - 1))$, where $\phi(N - 1) \leq N - 1$

(2) A path that starts at $(k, \phi(k))$ will end up at $(N - 1, r)$,

where $r \leq \phi(k) + N - 1 - k$.

In reality, the lead-lag relation between two time series could interchange. That means, the leading and lagging roles can interchange between the given time series X and Y . The solution to the above optimisation problem (5) is based on the following dynamic programming for a given set of arbitrary end points. Let $E(t_1, t_2)$ be the minimum accumulated energy between the starting grid point $(0, 0)$ and the ending grid point (t_1, t_2) .

$$E(t_1, t_2) = e(x(t_1), y(t_2)) + \min[E(t_1 - 1, t_2), E(t_1, t_2 - 1), E(t_1 - 1, t_2 - 1)], \quad (6)$$

where $t_2 = \phi(t_1)$.

3.2 Optimal Thermal Causal Path

Unfortunately, time series are not noise free, and the above method may extract unrelated structures between X and Y . Another problem is that most of the financial time series are not strictly stationary and lead and lag roles would dynamically change between the given pair of stocks. One can use Granger causality methods [13], but these tests require substantial amount of data. Sornette et al. [28, 27] choose an interesting approach to predict causality relations by using statistical physics techniques.

In the above optimisation method (5), Sornette et al [28, 27] consider weighted average over many potential mappings (which have more energy than the opti-

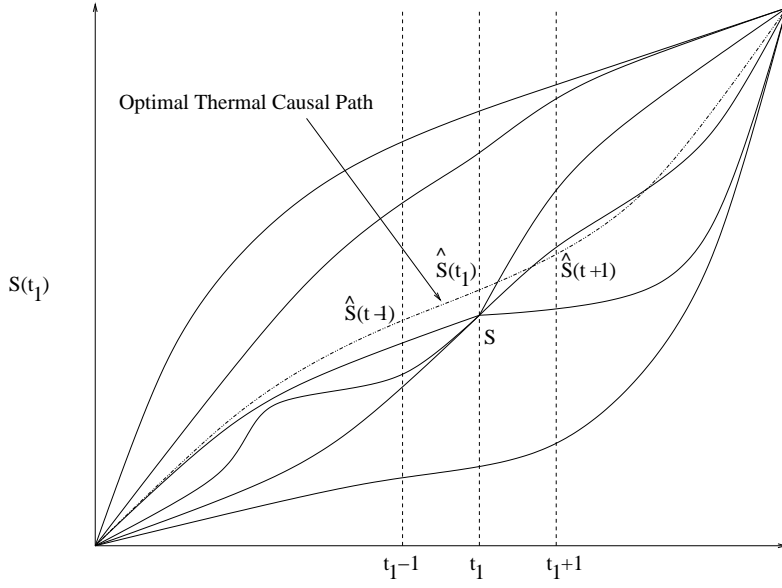


Figure 2: Optimal thermal causal path

mal path) around the optimal path. The weight for a mapping is proportional to $e^{-\frac{E\phi}{T'}}$, where T' describes the allowed deviation from the minimum energy path. It turns out that for each of the nodes (t_1, s) on the grid, finding the probability $P(t_1, s)$ that a path passes through s at time t_1 . Here,

$$P(t_1, s) \propto \sum_l e^{\frac{E_{sl}}{T'}}$$

where E_{sl} represents accumulated energy of a path l between $(0, 0)$ and (t_1, s) .

$$P(t_1, s) = \hat{\omega} \sum_l e^{\frac{-E_{sl}}{T'}}$$

where $\hat{\omega}$ is proportionality constant. Then the position of the optimal thermal causal path at t_1 is

$$\hat{s}(t_1) = \sum_{s=0}^{M-1} sP(t_1, s). \quad (7)$$

The sequence $(\hat{s}(0), \hat{s}(1), \dots, \hat{s}(N-1))$ represents the optimal thermal path trajectory on the grid as shown in Figure 2.

Let $G(t_1, s) = \sum_l e^{-\frac{E_{sl}}{T'}}$ and $G(t_1) = \sum_s G(t_1, s)$. Since $\sum_s P(t_1, s) = 1$ and

$$\hat{\omega} = \frac{1}{\sum_s G(t_1, s)}, \quad (8)$$

it can be verified that

$$P(t_1, s) = \frac{G(t_1, s)}{G(t_1)} \quad (9)$$

and

$$\hat{s}(t_1) = \sum_{s=0}^{M-1} s \frac{G(t_1, s)}{G(t_1)}. \quad (10)$$

In statistical mechanics, $G(t_1, s)$ and $G(t_1)$ are known as partition function and total partition function respectively. Since all paths those are reaching $(t_1 + 1, s + 1)$ must pass through one of the points (t_1, s) , $(t_1, s + 1)$ and $(t_1 + 1, s)$, it can be verified that the partition function satisfies the following recursive relation

$$\begin{aligned} G(t_1 + 1, s + 1) &= \sum_{\text{all paths } l \text{ between } (0,0) \text{ and } (t_1, s)} \exp\left(\frac{-(E_l + e(x(t_1), y(s + 1)))}{T'}\right) \\ &+ \sum_{\text{all paths } l \text{ between } (0,0) \text{ and } (t_1, s+1)} \exp\left(\frac{-(E_l + e(x(t_1), y(s + 1)))}{T'}\right) \\ &+ \sum_{\text{all paths } l \text{ between } (0,0) \text{ and } (t_1+1, s)} \exp\left(\frac{-(E_l + e(x(t_1 + 1), y(s + 1)))}{T'}\right) \\ &= [G(t_1, s) + G(t_1, s + 1) + G(t_1 + 1, s)] \exp\left(\frac{-(e(x(t_1 + 1), y(s + 1)))}{T'}\right) \end{aligned}$$

where E_l represents the energy of a path l between the given end points.

Synthetic Example: Let us consider a synthetic example and study the lead-lag relation. Figure 3 shows the daily price dynamics of stock X over a period of time. Assume that there exists a stock Y that is lagging behind X by L (some constant) units of time (of course, in reality it is almost impossible to find such a pair of stocks!). Then the price dynamics of Y can be expressed as $Y(t) = X(t - L)$ and the lead-lag relation should be $t_2 = t_1 - L$. For this trivial synthetic example, if we apply the optimal thermal causal path method in order to get the lead-lag relation between X and Y then the thermal path must be (since there is no noise in the time series) $t_2 = \hat{s}(t_1) = t_1 - L$. While considering the lead-lag relation between two time series, it would be convenient to translate the co-ordinates from (t_1, t_2) to (\tilde{x}, t) , where $\tilde{x} = t_2 - t_1$ and $t = t_2 + t_1$. Since $\tilde{x} = t_2 - t_1 = L$, the transformation is convenient in order to clearly understand the lead (or lag) between the two time series. Now, the optimal thermal causal path equation (10) can be expressed [28, 27] as

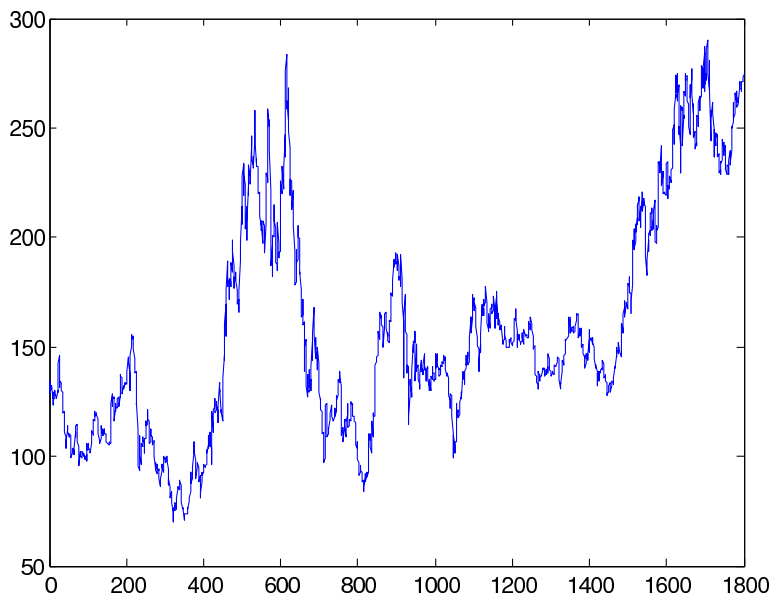


Figure 3: Stock X's Price Dynamics

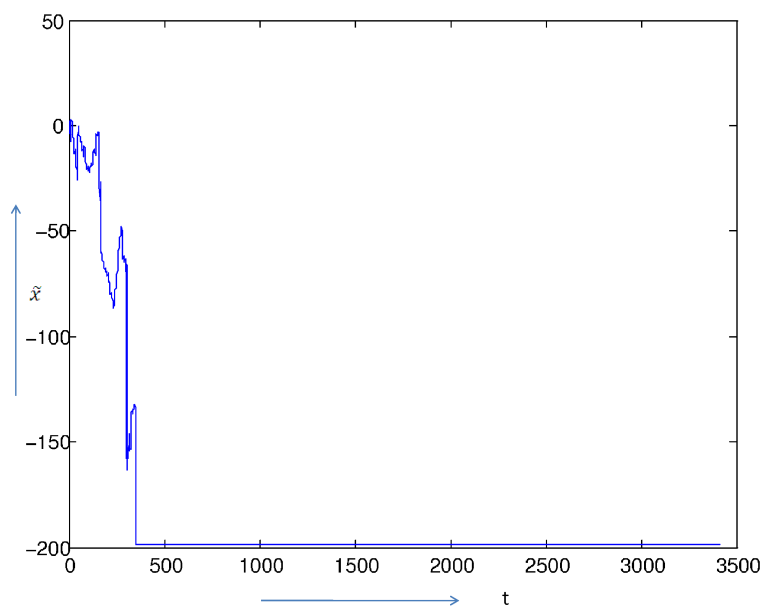


Figure 4: Lead-lag Relation $\tilde{x}(t)$ between X and Y

$$\tilde{x}(t) = \sum_{\tilde{x}} \tilde{x} \frac{G(\tilde{x}, t)}{G(t)}, \quad (11)$$

where $G(\tilde{x}, t)$ follows the recursion relation

$$G(\tilde{x}, t) = [G(\tilde{x} - 1, t - 1) + G(\tilde{x} + 1, t - 1) + G(\tilde{x}, t - 2)] \exp\left(-\frac{e(\tilde{x}, t)}{T'}\right). \quad (12)$$

coming back to the synthetic example, suppose for example that Y is lagging behind X by 200 units of time. When we use the optimal thermal causal path method and find the lead-lag relation \tilde{x} between X and Y then it is expected that we will find the lead-lag function $\tilde{x} = -200$. Figure 4 shows the optimal thermal causal path $\tilde{x}(t)$ between X and Y . Note that the number of time units on the abscissa of Figure 4 is more than that of Figure 3. This is due to the co-ordinate transformation $\tilde{x} = t_2 - t_1$ and $t = t_2 + t_1$. Note that, the causal path that is shown in Figure 4 is the result of imposing one of the end constraints that the path starts at $(0, 0)$. Since the other end constraint is not imposed, the optimal thermal causal path method is free to choose the actual lag ($\tilde{x}(t) = -200$) for all values of t . Figure 5 describes the causal path when both the end constraints are imposed.

4 A Method for Directional Trading

Here, we propose a trading strategy for directional trading. A successful trader can learn market direction and adopt trading strategies that could generate profits consistently. There are variety of techniques to pin down market direction for directional trading. Here we are proposing a method for directional trading that is based on time series classification. Once we know the price direction of a security, it is relatively easy to open positions, either long or short, in order to get profits from the investments.

Let us consider that we have a large financial time series data base and there is a pair of similar time series X and Y . Assume that we use time series classification techniques as explained in Section 3 and extract the lead-lag relation between X and Y by using the history of the time series up to time t . Figure 6 depicts a lead-lag relation between X and Y . In this case X is leading Y by L time units. Since Y is lagging behind X by L time steps, the next L price

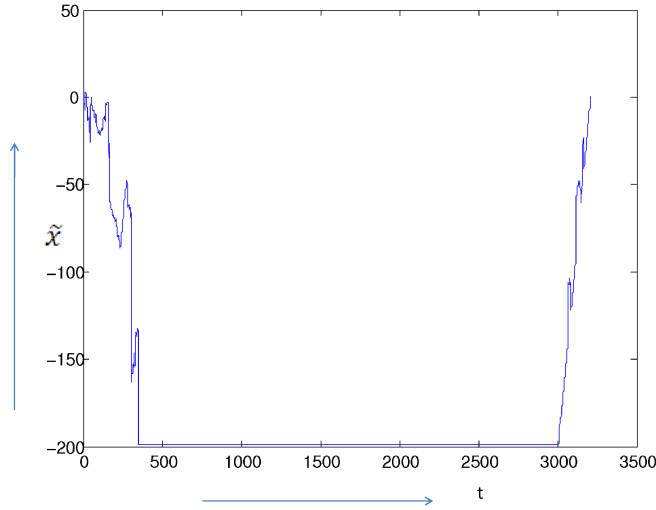


Figure 5: Lead-lag relation $x(t)$ between X and Y with both the end constraints

movements $(Y(t), \dots, Y(t + L))$ for Y can be anticipated by considering the price movements $(X(t - L), \dots, X(t))$ in X.

More realistic models:

In real life it might be too ambitious to look for a pair of stocks that has a constant lead-lag relation which can then be exploited by using directional trading. But, it may be possible to find a pair of stocks that has a reasonably constant lead-lag relation with minimum fluctuations. As an example, let us construct $Y(t_2) = X(t_2 - L) + \eta$, where $L = 200$ is artificial lag and η is noise.

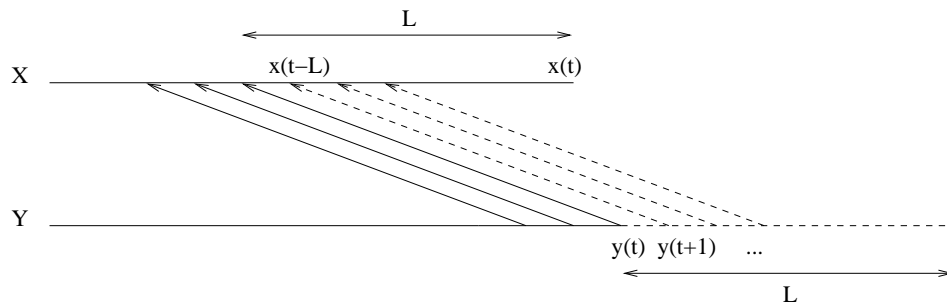


Figure 6: Lead-lag

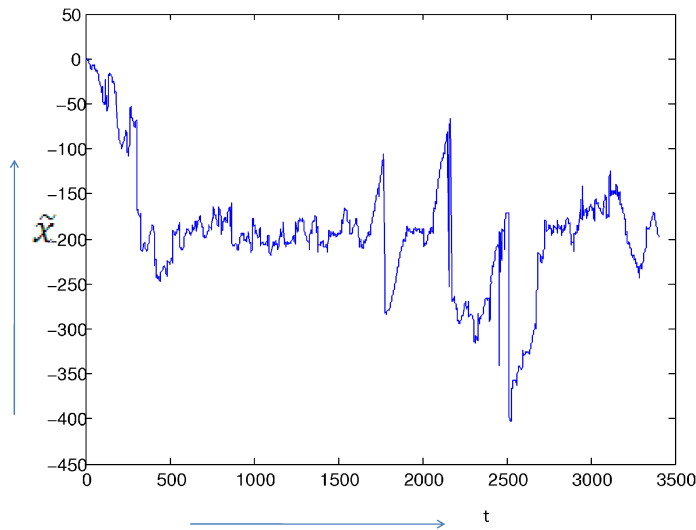


Figure 7: Lead-lag Relation between X and Y with 10% Noise

Figure 7 depicts a lead-lag relation between X and Y (only imposing one end constraint $(0, 0)$) with 10% zero mean Gaussian noise.

Since in reality it is difficult to find a lead-lag relation similar to one that is shown in Figure 4, it is not possible to construct a trading strategy that can bet on lagging series Y by just looking into the history of X on a one-to-one basis as shown in Figure 6. There are at least two different ways in which things can be more complicated: (1) noise (but fixed lag) considered earlier (depicted in Figure 7) and (2) fluctuations in lag (other possible ways are dependence on more than one lag and dependence on lags from more than one series, but we are not considering them here). We may identify a lead-lag relation that has small fluctuations ($L \pm r$, where $r \leq \tau$ and τ is a constant). In this case, in contrast to the case described in previous page, the ‘noise’ is, in a sense, attached to the lag. It is reasonable to expect that any substantial price changes that occur in X during the time interval $(t - L - \tau, t - L + \tau)$ might lead to similar substantial price movements in Y during the time interval $(t - \tau, t + \tau)$.

Let $M_t = (m_0^t, m_1^t, \dots, m_K^t)$ be K -dimensional price movement vector that describes anticipated substantial price movements of Y for the next K time steps from time $t - \tau$. Here $m_i^t \in \{-1, 0, +1\}$ represents substantial fall (-1) ,

m_i^t	Strategy for Y at time $t + i - \tau$
-1	Short sell (borrow and sell) one unit of Y
0	Do nothing
+1	Buy one unit of Y

Table 1: Trading strategy for the given m

no substantial change (0) and substantial up (+1) in the price movements of Y . Since X is leading Y , any substantial price change occurring in X during the time interval $(t + i - L - \tau, t + i - L + \tau)$ might lead to similar substantial price movements in Y during the interval $(t + i - \tau, t + i + \tau)$.

In order to construct the price movement vector, we identify the substantial price movements in X whenever the percentage change in the stock price crosses a given threshold value B_{ps} which is defined in terms of basis points. Let the percentage change in the price of X during the time period $(t - L + i - \tau, t + i - L + \tau)$ is P_c^X . The construction of m_i^t can be explained as

$$m_i^t = \begin{cases} -1 & \text{if } P_c^X < -B_{ps} \\ +1 & \text{if } P_c^X > B_{ps} \\ 0 & \text{otherwise.} \end{cases}$$

Based on the price movement vector, the trader may anticipate the price direction of Y during the time interval $(t - \tau, t + K + \tau)$ by looking at M_t , and constructing his trading strategy as shown in Table 1.

Since the trading strategy is constructed based on the *expected* lead-lag structure between the two time sequences, it can be classified as a statistical arbitrage trading strategy. The success of the above trading strategy which is derived from the optimal thermal causal path based lead-lag relation can be verified by looking at the hit ratio. The success of the trading strategy is quantified by (1) the number of trading opportunities or total number of trades (buy and sell signals) in the given out-of-sample time interval and (2) the number of successful anticipation of the Y 's price direction and magnitude beyond the threshold value (B_{ps}) in the out-of-sample time interval.

$$\text{Hit ratio} = \frac{\text{no. of successful anticipation}}{\text{total no. of trades}}. \quad (13)$$

We investigate this experimentally later.

4.1 Modification using Curvature Energy

We now consider a modification of the method, by using a ‘curvature’ term. Each node is associated with an energy. Sornette et al. [28, 27] mentioned a generalisation of the optimisation problem (5) by adding a path curvature energy, but they did not pursue it. For a well-behaved lead-lag relationship, the path, $\phi : t_1 \mapsto t_2$, in addition to being continuous ought also to be smooth. The constraint: $0 \leq \phi(t_1 + 1) - \phi(t_1)$, makes sure that the mapping be continuous and smooth in the continuous time limit. Further smoothness can be achieved by adding an analogue of a differentiability condition, which can be done by introducing a curvature energy term in the optimisation problem. Curvature energy is expressed by the second derivative in the continuous case as:

$$E_{curv} = \left| \frac{d^2\phi(t_1)}{dt_1^2} \right|^2. \quad (14)$$

The discrete version of curvature energy can be considered as:

$$\frac{d^2\phi(t_1)}{dt_1^2} \approx \left| \frac{d\phi(t_1)}{dt_1} - \frac{d\phi(t_1 - 1)}{dt_1} \right|.$$

Since $\frac{d\phi(t_1)}{dt_1} \approx \phi(t_1 + 1) - \phi(t_1)$ and $\frac{d\phi(t_1 - 1)}{dt_1} \approx \phi(t_1) - \phi(t_1 - 1)$, we have

$$\frac{d^2\phi(t_1)}{dt_1^2} \approx |\phi(t_1 + 1) - 2\phi(t_1) + \phi(t_1 - 1)|$$

and so we have

$$E_{curv} \approx |\phi(t_1 + 1) - 2\phi(t_1) + \phi(t_1 - 1)|^2. \quad (15)$$

While considering the curvature energy, the optimisation problem can be written as

$$\min E = \min_{\phi} (1 - \beta) \sum_{t_1=0}^{N-1} |X(t_1) - Y(\phi(t_1))| + \beta \sum_{t_1=1}^{N-2} |\phi(t_1 + 1) - 2\phi(t_1) + \phi(t_1 - 1)|^2 \quad (16)$$

subject to $0 \leq \phi(t_1 + 1) - \phi(t_1) \leq 1$. Here, β is an appropriate weight for curvature energy.

5 Approximate Solution

Let C and E_C denote a typical path and its energy respectively, between the given end nodes. For the given two time series $X = \{x(0), x(1), \dots, x(N-1)\}$ and $Y = \{y(0), y(1), \dots, y(M-1)\}$, let

$$E_l(X, Y) = \min_C E_C(X, Y) \quad (17)$$

$$E_u(X, Y) = \max_C E_C(X, Y) \quad (18)$$

be lower and upper bounds respectively for $E(X, Y)$. Obviously,

$$\min_C E_C \leq E \leq \max_C E_C$$

are the trivial bounds. As we discussed in section 3, one can use the DTW method and find out $E_l(X, Y)$ and $E_u(X, Y)$ by constructing a two dimensional grid, for the X and Y time series. $E_l(X, Y)$ (or $E_u(X, Y)$) represents the minimum (or maximum) possible energy that is accumulated while travelling from $(0, 0)$ to $(N-1, M-1)$ subject to the constraints. $E_l(X, Y)$ and $E_u(X, Y)$ can be evaluated by using the dynamic programming approach (6) and considering,

$$E_l(X, Y) = e_l(N, M) \quad (19)$$

where e_l is defined recursively by

$$\begin{aligned} e_l(i, j) &= \alpha|X(i) - Y(j)| + \beta(l_i - 2l_{i-1} + l_{i-2})^2 \\ &\quad + \min\{e_l(i, j-1), e_l(i-1, j), e_l(i-1, j-1)\}. \end{aligned}$$

Here, $l_i = j$, and l_{i-1}, l_{i-2} are the vertical components of the two nodes prior to the node (i, j) on the minimum energy path passing through (i, j) . Also,

$$E_u(X, Y) = e_u(N, M), \quad (20)$$

where e_u is defined by:

$$\begin{aligned} e_u(i, j) &= \alpha|X(i) - Y(j)| + \beta(l_i - 2l_{i-1} + l_{i-2})^2 \\ &\quad + \max\{e_u(i, j-1), e_u(i-1, j), e_u(i-1, j-1)\}. \end{aligned}$$

Here, the curvature energy curvature is weighted by weight factor β and $\alpha = 1 - \beta$.

5.1 Energy Approximations

We use Yoshikawa et al.'s [24] method for constructing approximate segments for a given time series. Let

$$X = \{x(0), x(1), \dots, x(N-1)\}$$

be the given time series and N and M be multiples of T ($N = nT$ and $M = mT$, but this can be relaxed for more general case), by defining i^{th} approximate segment $x^A(i)$ of X as a pair of numbers

$$x^A(i) = \{L^x(i), U^x(i)\},$$

where

$$L^x(i) = \min\{x(a), \dots, x(b)\}$$

$$U^x(i) = \max\{x(a), \dots, x(b)\},$$

and where $b = Ti$ and $a = (i-1)T + 1$ if $2 \leq i \leq n$. The numbers $L^x(i)$ and $U^x(i)$ are lower and upper bounds on the values of $x(i)$ for i in the range $a = (i-1)T + 1$ to $b = Ti$.

This means, we approximate X with an unordered set $X^A = \{x^A(1), x^A(2), \dots, x^A(n)\}$ by dividing the grid into $\frac{N}{T}$ ranges and constructing $\frac{N}{T}$ sets $\{L^x(i), U^x(i)\}$. Similarly, Y can be approximated by $Y^A = \{y^A(1), y^A(2), \dots, y^A(m)\}$. Figure 8 depicts the coarse grid (shown in solid circles) that is constructed by considering $T = 3$ from the given pair of time series X and Y . Note that for most of the applications it is appropriate to consider $n = m$. We approximate $E_l(X, Y)$ and $E_u(X, Y)$ by considering approximate segments X^A and Y^A for X and Y respectively. Let $E_u(X^A, Y^A)$ represent the maximum possible energy that is accumulated while travelling from $(0, 0)$ to $(n-1, m-1)$ subject to the constraints. At any node in the coarse grid, for $1 \leq i \leq n$ and $1 \leq j \leq m$, the upper bound for approximated energy is defined as

$$e_u^a(i, j) = e_{cell}^a(i, j) + \max\{e_u^a(i, j-1), e_u^a(i-1, j), e_u^a(i-1, j-1)\}, \quad (21)$$

where $e_{cell}^a(i, j)$ is the maximum energy that is accumulated while the path is passing through (within) the cell in the grid just before the node (i, j) . $e_{cell}^a(i, j)$

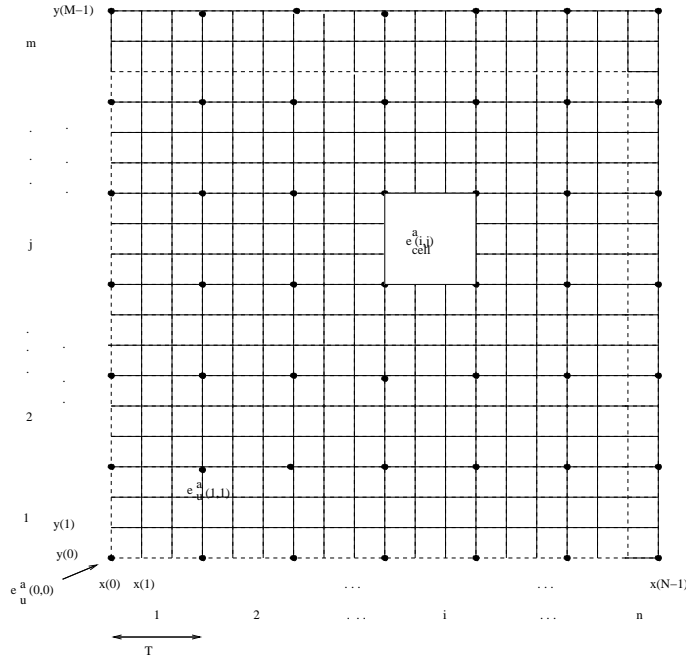


Figure 8: Grid and coarse grid

can be computed by considering the associated approximate segments $x^A(i)$ and $y^A(j)$ of the cell (note that every cell $(.,.)$ in the coarse grid is associated with two segments $x^A(.)$ and $y^A(.)$). Since the maximum curvature energy that can be considered for the given cell is T^2 (see the definition of the curvature energy in the previous section 4.1), subject to the constraints

$$e_u^a(0,0) = |x_0 - y_0|, \quad e_u^a(i,0) = e_u^a(0,j) = 0,$$

it can be verified that

$$e_{cell}^a(i,j) = \alpha TD(L^x(i), U^x(i), L^y(j), U^y(j)) + \beta T^2 \quad (22)$$

where

$$D(L^x(i), U^x(i), L^y(j), U^y(j)) = \max\{|U^x(i) - L^y(j)|, |U^y(j) - L^x(i)|\}.$$

Here α and β are the weights for internal and curvature energies respectively in $e_{cell}^a(i,j)$. Using the similar analysis to that proposed by Yoshikawa et al. for lower bounding distance measure with Segmentation (LBS) [24], we can verify the following theorem.

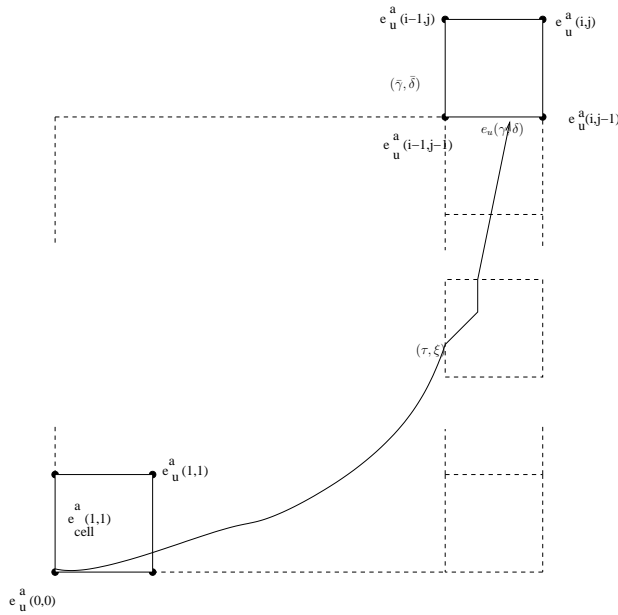


Figure 9: Proof

Theorem 1 For a pair of equal length ($N = M$) time series X and Y , $E_u(X, Y) \leq E_u(X^A, Y^A)$.

Proof By induction on coarse grid size n .

Base case: Verify $e_u(T, T) \leq e_u^a(1, 1)$

Let $(\rho, \psi) \in \{(T, 0), (T, 1), \dots, (T, T)\} \cup \{(0, T), (1, T), \dots, (T, T)\}$. We have (by definition)

$$e_u(\rho, \psi) = \alpha|X(\rho) - Y(\psi)| + \beta(l_\rho - 2l_{\rho-1} + l_{\rho-2})^2 + \max\{e_u(\rho, \psi-1), e_u(\rho-1, \psi), e_u(\rho-1, \psi-1)\}.$$

Since, by definition,

$$D(L^x(i), U^x(i), L^y(j), U^y(j)) = \max\{|U^x(i) - L^y(j)|, |U^y(j) - L^x(i)|\} \quad (23)$$

and the accumulated curvature energy portion in $e_u(\rho, \psi)$ is:

$$\sum_i^\psi (l_i - 2l_{i-1} + l_{i-2})^2 \leq T^2,$$

it follows that $\forall(\rho, \psi)$,

$$e_u^a(1, 1) \geq e_u(\rho, \psi), \text{ so } e_u^a(1, 1) \geq e_u(T, T).$$

Inductive Step: We now verify that

$$e_u^a(i-1, j-1) \geq e_u(T(i-1), T(j-1)) \Rightarrow e_u^a(i, j) \geq e_u(Ti, Tj).$$

Start with

$$e_u^a(i-1, j-1) \geq e_u(T(i-1), T(j-1)),$$

let $e_u(\gamma, \delta)$ be the maximum energy between the nodes $(0, 0)$ and (γ, δ) , where $(i-1)T \leq \gamma \leq Ti$ and $\delta = T(j-1)$. Assume that this maximum energy path travels through (τ, ξ) , $\{\tau = T(i-1), 0 \leq \xi \leq T(j-1)\}$. Then

$$e_u^a(i-1, j-1) \geq e_u(T(i-1), T(j-1)) \geq e_u(\lfloor \tau/T \rfloor + T, \lfloor \xi/T \rfloor + T) \geq e_u(\tau, \xi).$$

Since the energy accumulated at each and every cell that belongs to the last column of cells in Figure 9 must be less than $e_{cell}^a(., .)$ we have $e_u^a(i, j-1) \geq e_u(\gamma, \delta)$. For the given $\bar{\gamma} = T(i-1)$ and $T(j-1) \leq \bar{\delta} \leq Tj$, similar argument can be made and verified that $e_u^a(i-1, j) \geq e_u(\bar{\gamma}, \bar{\delta})$. Now, it is trivial that $e_u^a(i, j) \geq e_u(Ti, Tj)$.

Conclusion: By induction, we can conclude that

$$E_u^a(n, m) \geq e_u(Tn = N, Tm = M),$$

so $E_u(X^A, Y^A) \geq E_u(X, Y)$. ■

One can try for better bounds for the optimal thermal energy of the system, which can be defined as

$$E_{Th} = \frac{\sum_C E_C e^{-E_C/T}}{\sum_C e^{-E_C/T}}. \quad (24)$$

Let $f(i, j)$ be the number of paths between $(0, 0)$ and (i, j) from the *Grid*, defined as $Grid = \{(i, j) | i \in \{0, 1, \dots, N-1\} \text{ and } j \in \{0, 1, \dots, M-1\}\}$. Then

$$f(i, j) = \sum_{k=0}^{\min(i, j)} \binom{i+j-k}{k, i-k, j-k}. \quad (25)$$

Here

$$\binom{i+j-k}{k, i-k, j-k} = \frac{(i+j-k)!}{k!(i-k)!(j-k)!}.$$

To see this, we note that any path from $(0, 0)$ to (i, j) must consist of a sequence of three different types of individual ‘moves’ or ‘segments’: (i) *vertical*

moves of the type $(r, s) \rightarrow (r, s+1)$; (ii) *horizontal* moves $(r, s) \rightarrow (r+1, s)$; and (iii) *diagonal* moves $(r, s) \rightarrow (r+1, s+1)$. Let k denote the number of diagonal moves made in a path from $(0, 0)$ to (i, j) . The total horizontal distance covered by the path is i and the vertical distance is j . Noting that each diagonal move moves the path horizontally and also vertically forward by a unit, we see that we must have both $k \leq i$ and $k \leq j$, so $k \leq \min\{i, j\}$. Furthermore, if V and H are the number of vertical and horizontal moves in the path (respectively), then we must have: $j = k + V$ and $i = k + H$, so $V = j - k$ and $H = i - k$. The total number of moves in the path is $k + V + H = i + j - k$. The number of paths from $(0, 0)$ to (i, j) which make k diagonal moves is therefore the number of sequences of length $i + j - k$ which contain k diagonal moves, $i - k$ horizontal moves and $j - k$ vertical moves. There are $\binom{i + j - k}{k}$ possible positions in the sequence for the diagonal moves and, given these, there are $\binom{i + j - 2k}{i - k}$ possible positions for the horizontal moves. The positions of the vertical moves are then determined. Therefore, the total number of paths with k diagonal moves is

$$\begin{aligned} \binom{i + j - k}{k} \binom{i + j - 2k}{i - k} &= \frac{(i + j - k)!}{k!(i + j - 2k)!} \frac{(i + j - 2k)!}{(i - k)!(j - k)!} \\ &= \frac{(i + j - k)!}{k!(i - k)!(j - k)!}. \end{aligned}$$

The total number of paths from $(0, 0)$ to (i, j) is therefore

$$\sum_{k=0}^{\min\{i, j\}} \frac{(i + j - k)!}{k!(i - k)!(j - k)!}.$$

The proportion of paths $p_f(i, j)$ containing the node (i, j) out of total number of paths between $(0, 0)$ and (N, M) is defined as

$$p_f(i, j) = \frac{f(i, j)f(N - i, M - j)}{f(N, M)} \quad (26)$$

Theorem 2 $E_{Th} \leq \sum_{(i, j) \in Grid} p_f(i, j)e(i, j) \leq \max_C E_C$.

Proof The expected energy of the system with respect to the uniform distribution can be expressed as

$$\frac{\sum_C E_C}{\sum_C 1} = \frac{\sum_C E_C}{\text{Total \# of paths}}$$

$$\begin{aligned}
&= \frac{\sum_C \sum_{(i,j) \in C} e(i,j)}{\text{Total \# of paths}} \\
&= \sum_{(i,j)} \left(\frac{\sum_{\{C | (i,j) \in C\}} e(i,j)}{\text{Total \# of paths}} \right) \\
&= \sum_{(i,j)} p_f(i,j) e(i,j)
\end{aligned}$$

For evaluating E_{Th} , the assigned probability to a path is inversely proportional to the energy of the path, so E_{Th} is no more than the expected energy that is calculated by assigning equal probabilities to all paths. The second inequality in Theorem 2 is straightforward. ■

5.2 Filtering Process

In our application, we try to identify the stock pairs that have a good causality relationship by searching a given huge database. Computing the thermal causal path for all the possible pairs of stocks is a computationally very intensive process. We consider the following three phases for a filtering process to speed this up. From the previous analysis,

$$\min_C E_C \leq E_{Th} \leq \sum_{(i,j) \in Grid} p_f(i,j) e(i,j) \leq \max_C E_C.$$

For the given approximate segments, one can use Yoshikawa et al.'s LBS [24] and Theorem 2 in order to approximate $\min_C E_C$ and $\max_C E_C$ respectively.

First Phase:

Step-a: Let (X_k, Y_k) represent the k^{th} pair of time series in the database. For the given level of coarse segments A ,

$$E_l(X_k^A, Y_k^A) \leq \min_C E_C^k \text{ and } \max_C E_C^k \leq E_u(X_k^A, Y_k^A).$$

It is clearly sensible to eliminate the k^{th} pair if

$$E_l(X_k^A, Y_k^A) \geq \min_q E_u(X_q^A, Y_q^A).$$

This is because if there is a pair (X_q^A, Y_q^A) such that $E_u(X_q^A, Y_q^A)$ is less than $E_l(X_k^A, Y_k^A)$, certainly $E(X_q^A, Y_q^A) < E(X_k^A, Y_k^A)$ and (X_q, Y_q) would be a better choice than (X_k, Y_k) .

Step-b: Consider finer approximation segments and repeat step-a. After a few iterations of the first phase filtration, there may not be much advantage (in eliminating such non-similar pairs). We go to the next phase with further refinement of the grid.

Second Phase:

Step-c: We eliminate the k^{th} pair if

$$E_l^k(X_k, Y_k) \geq \min_{\text{over all pairs}} \left\{ \sum_{(i,j) \in \text{Grid}} p_f(i,j) e(i,j) \right\}.$$

Step-d: Consider finer approximation segments and repeat step-c.

Algorithms 1 and 2 below explain the first and second phases.

Third Phase:

After using the above two phases of filtration, we expect a significant reduction in the number of pairs of time series remaining to be considered for best thermal causal relationship. In this last phase of filtration, we extract the pairs based on the minimum thermal energy and the causality relation between the pairs by using the optimal thermal causal path method. The extracted pairs can be used for directional trading.

Algorithm 1 Algorithm for Filtering (First Phase)

- 1: Filter(dataset) {all pairs $(X_k, Y_k) \in \text{dataset}$ as input}
 - 2: **for** i=1 to n **do**
 - 3: construct dataset[A^i] {constructing dataset that has coarse time series pairs of approximations of level A^i }
 - 4: **for** all pairs $(X_k^{A^i}, Y_k^{A^i})$ **do**
 - 5: UPPERENERGY[k] = $E_u(X_k^{A^i}, Y_k^{A^i})$
 - 6: **end for**
 - 7: Compute MIN[i] = $\min_k \{UPPERENERGY[k]\}$
 - 8: **for** all pairs $(X_k^{A^i}, Y_k^{A^i})$ **do**
 - 9: **if** $E_l(X_k^{A^i}, Y_k^{A^i}) \geq MIN[i]$ **then**
 - 10: dataset[A^i] = dataset[A^i] \ $(X_k^{A^i}, Y_k^{A^i})$
 - 11: **end if**
 - 12: **end for**
 - 13: **end for**
-

Algorithm 2 Algorithm for Filtering (Second Phase)

```
1: Filter(dataset) {all pairs  $(X_k, Y_k) \in \text{dataset}$  as input}
2: for i=1 to m do
3:   construct dataset[ $A^i$ ] {constructing dataset that has coarse time series
   pairs of approximations of level  $A^i$ }
4:   for all pairs  $(X_k^{A^i}, Y_k^{A^i})$  do
5:     UPPERENERGY[ $k$ ] =  $\sum_{(i,j) \in \text{Grid for pair } k} p_f(i,j)e(i,j)$ 
6:   end for
7:   Compute MIN[ $i$ ] =  $\min_k \{UPPERENERGY[k]\}$ 
8:   for all pairs  $(X_k^{A^i}, Y_k^{A^i})$  do
9:     if  $E_l(X_k^{A^i}, Y_k^{A^i}) \geq MIN[i]$  then
10:      dataset[ $A^i$ ] =  $\text{dataset}[A^i] \setminus (X_k^{A^i}, Y_k^{A^i})$ 
11:    end if
12:  end for
13: end for
```

6 Experiments with NYSE Data

For the experiments discussed below, we consider 97 stocks from NYSE. Wharton Research Data Services (WRDS) was used for collecting NYSE tick-by-tick TAQ data on 1st JAN 2006 from various sectors like (1) consumer goods (automobile, food, and household), (2) basic materials, (3) consumer services, (4) utilities, (5) technology, and (6) oil and gas. While cleaning the data in addition to eliminate outliers, we eliminate the redundant data so that we get one tick for every second. From 97 stocks, we construct around 300 pairs of stocks. While considering a pair of stocks we make sure that both the observations are taken at the same time and they are 10 seconds (or 20 seconds) apart. For extracting the best set of pairs we use first three hours of high frequency data to calculate the similarity (in terms of optimal thermal energy) measure for each and every pair of time series. We assign ranks to the pairs that are inversely proportional to their optimal thermal energy. Since calculating optimal thermal energy of each and every pair is computationally intensive, we use the first and second phases of the filtering process and eliminate the unnecessary pairs.

We managed to eliminate 46 pairs after performing (4 iterations each) first and second phases of the filtering process. It is observed that after few iterations

of filtering, it is not much further advantage with the two phases of the filtration. One of the reasons for that is the weight factor (β) for the curvature energy component. A priori, it is difficult to determine the numerical value for β . 78 pairs are filtered out from the first and second phases and it is estimated that around 20% of the computing time is reduced due to the filtering process. Note that the reduction in the computational time would depend on the starting set of equity pairs for the analysis. In the third phase of filtration, we use the optimal thermal causal path method and calculate the optimal thermal energy for each pair of time series. At the end of the third phase, we identify top ten ranked pairs (based on their optimal thermal energy) from different sectors. In order to get the stable lead-lag relation, we understand from the experiments that not much significant advantage is received from the inclusion of curvature energy. After selecting the ten pairs, we carefully observe the changes in the lead-lag relation of each and every selected pair by updating lead-lag (optimal thermal causal path) once in every 20 minutes. When we identify a stable lead-lag relation between a pair of time series, we take either long or short positions on the lagging stock according to the directional trading that we discussed earlier. In the following from Figure 10 to Figure 19, we show the normalised values of stock prices of the selected pairs of stocks and the associated lead-lag relations those are used for the directional trading. The normalised values of a time series are calculated with respect to its standard deviation. For example, the top portion of Figure 10 shows the normalised stock prices for the pair CLX and PHM and the bottom portion of Figure 10 shows the lead-lag relation for the pair during the time period when we have the directional trading opportunity. Note that in the figures we have used (t_1, t_2) and (\tilde{x}, t) co-ordinates for normalised values and lead-lag relation respectively. The trader may believe that a stable lead-lag relation (that may have small fluctuations around the mean of \tilde{x}) is useful for directional trading. For the pair CLS and PHM, the time period between 2900 and 3200 is considered for directional trading as the lead-lag has minute fluctuations around the mean value (in this case $\tilde{x}(t) = 93$). by fixing $\tilde{x}(t) = 93$ for $2900 \leq t \leq 3200$, one can find the associated sequence of the pairs (t_1, t_2) from the co-ordinate transformation rule. The obtained sequence of the pairs can be used for constructing long/short positions for directional trading as each order pair in the sequence would explain the association between $x(t_1)$ and $y(t_2)$. We discuss the performance of directional trading for the selected

Pair of Stocks	BPS	Total no. of Trades	Successful Trades	Hit Ratio
CLX, PHM	50	100	61	61 %
CLX, PHM	80	92	57	61.9 %
CLX, PHM	90	89	55	61.8 %
CAG, DF	50	300	188	62.6 %
CPB, CQB	50	200	120	60 %
K, KFT	50	200	115	57.5 %
ABX, GG	50	200	133	66.5 %
ABX, GG	60	199	131	65.8 %
AAI, IGT	50	277	163	58.8 %
AAI, IGT	60	273	157	57.5 %
AMR, YUM	50	100	63	63 %
AMR, YUM	60	100	61	61 %
AMR, YUM	80	100	61	61 %
AMD, GLW	50	100	58	58 %
AMD, GLW	60	100	58	58 %
AMD, GLW	80	100	58	58 %
GLW, MOT	50	200	132	66 %
GLW, MOT	60	200	132	66 %
GLW, MOT	80	199	132	66.3 %
GLW, TXN	50	147	80	54.4 %
GLW, TXN	60	145	79	54.4 %
GLW, TXN	80	95	58	61 %

Table 2: Performance of Directional Trading

pairs of stocks in Table 2. For each pair of stocks, we find different number of trades for the given threshold values. Last two columns of Table 2 shows the performance of optimal thermal causal path based directional trading in terms of number of successful trades and hitratio.

7 Conclusions

We considered the “Optimal Thermal Causal Path” method as a tool to identify the arbitrage opportunities in financial markets. We propose an approximation

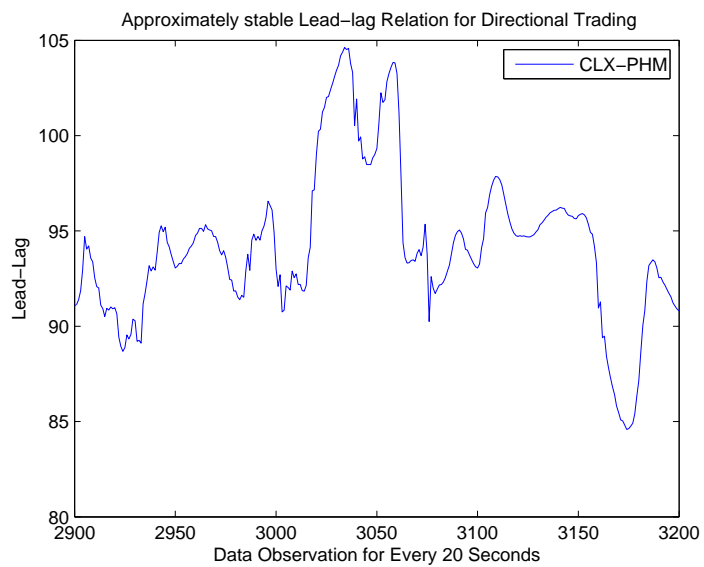
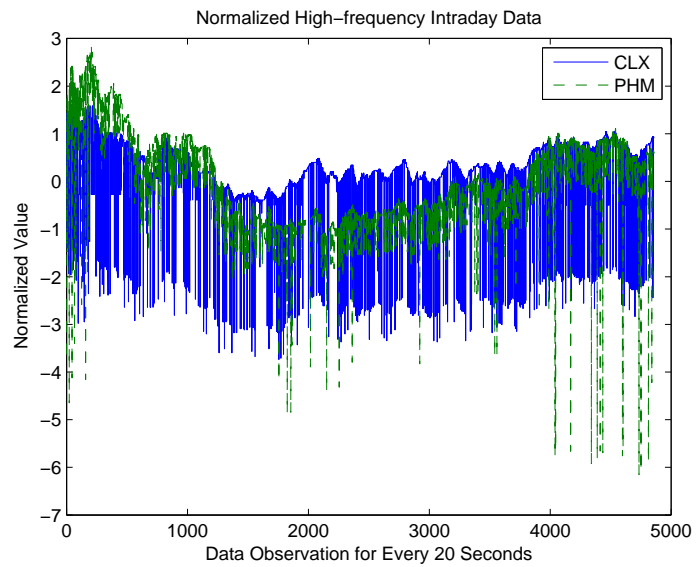


Figure 10: Lead-lag between CLX and PHM

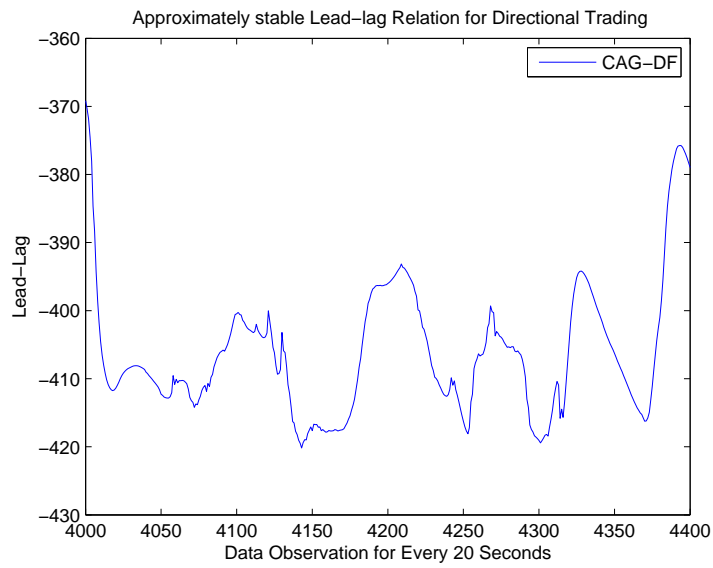
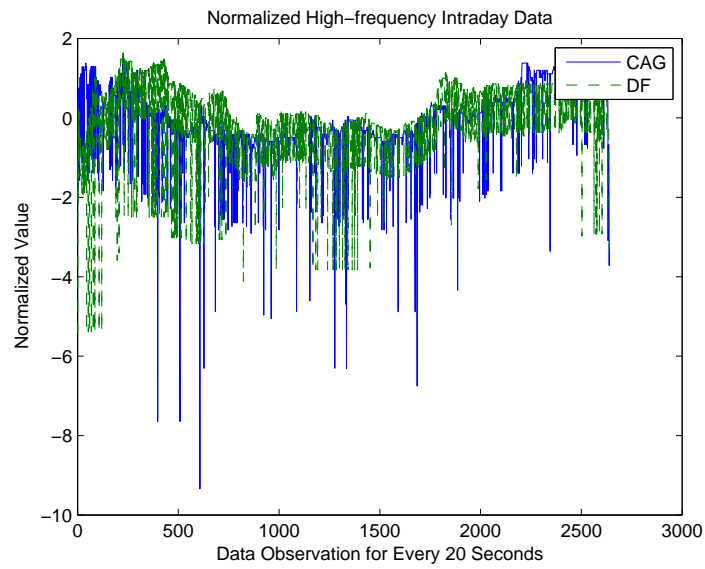


Figure 11: Lead-lag between CAG and DF

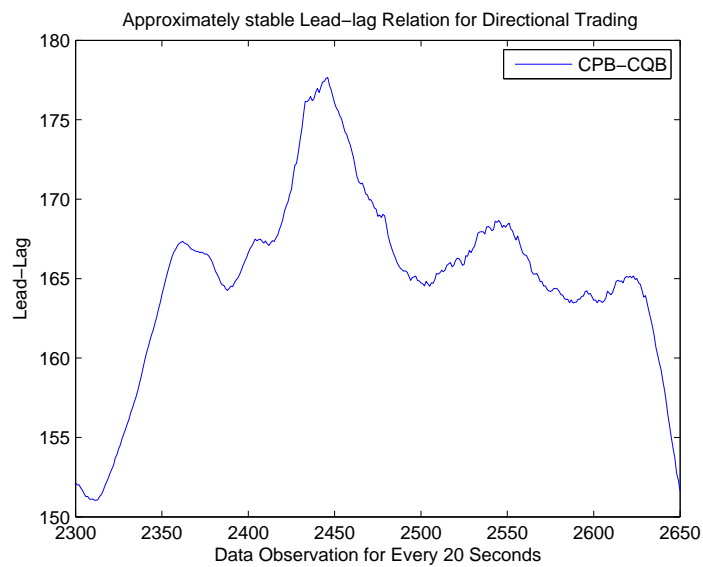
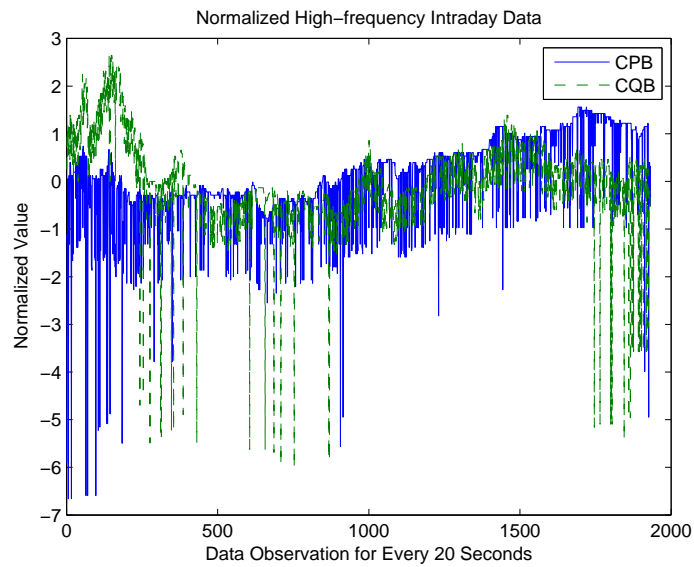


Figure 12: Lead-lag between CPB and CQB

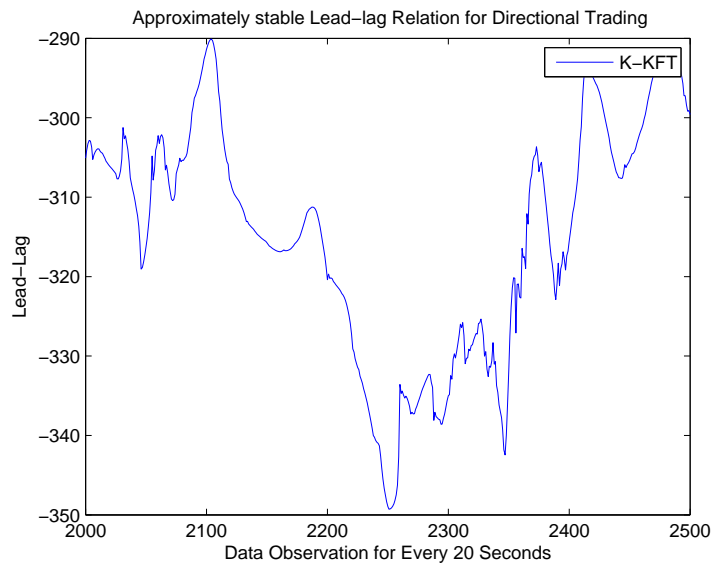
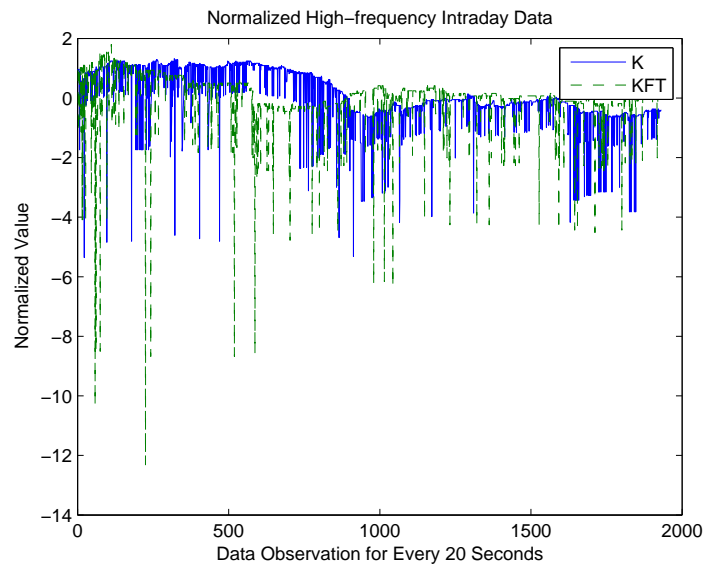


Figure 13: Lead-lag between K and KFT

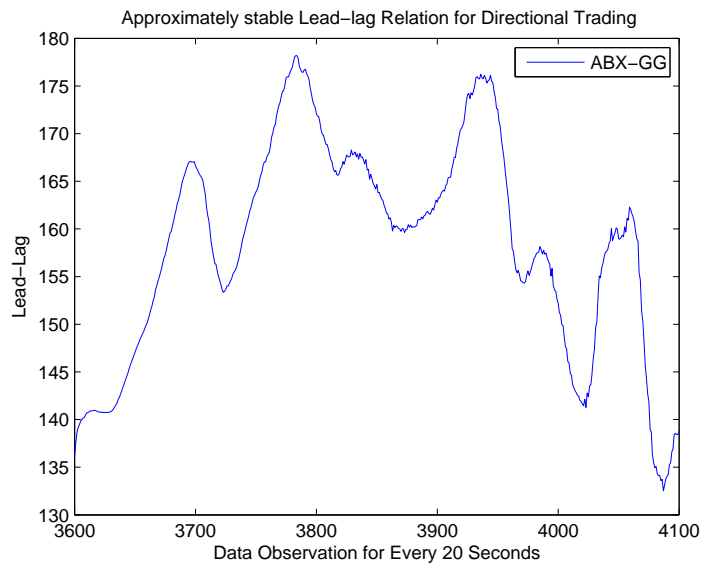
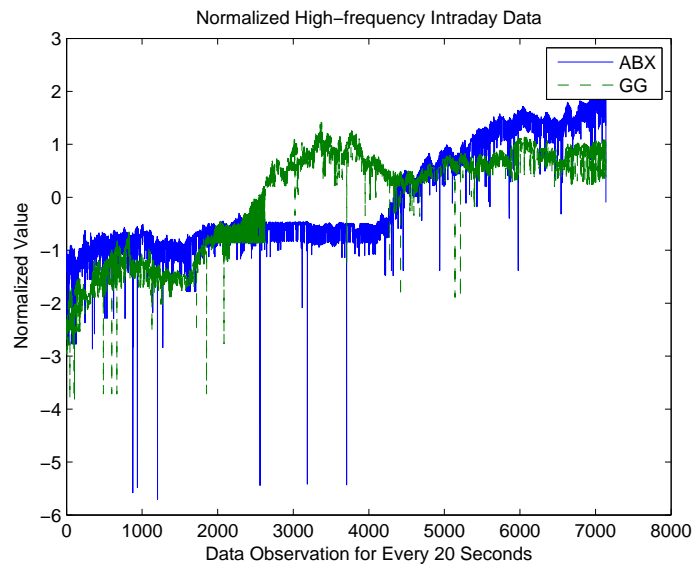


Figure 14: Lead-lag between ABX and GG

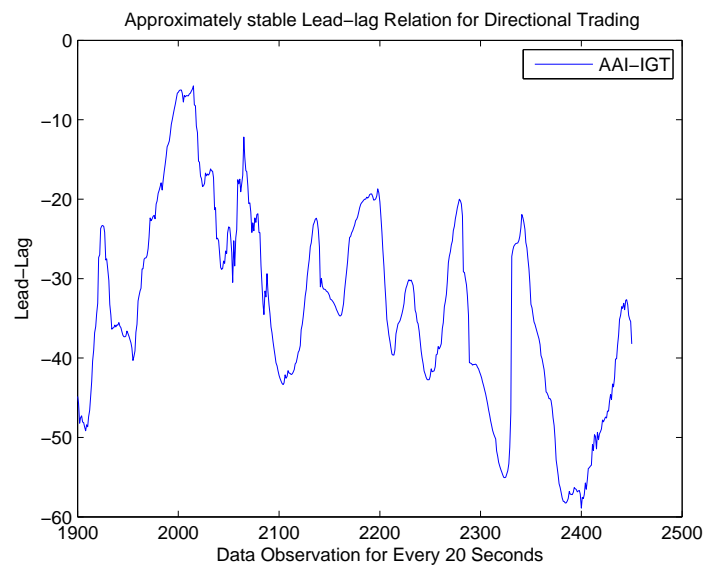
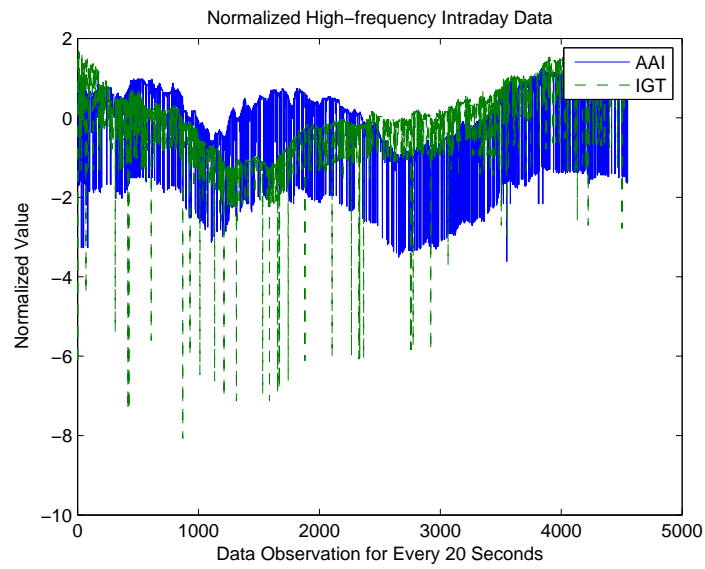


Figure 15: Lead-lag between AAI and IGT

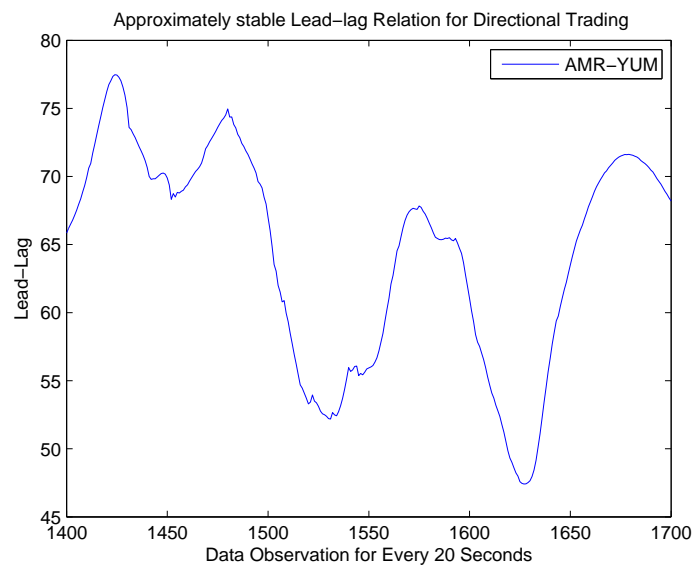
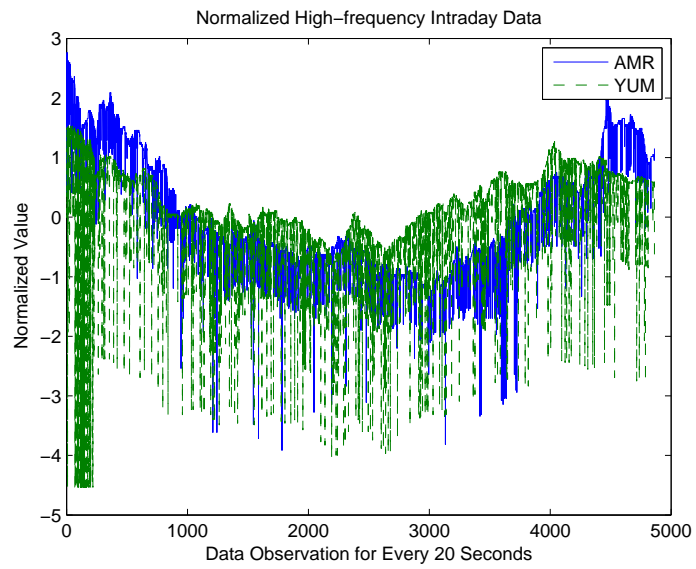


Figure 16: Lead-lag between AMR and YUM

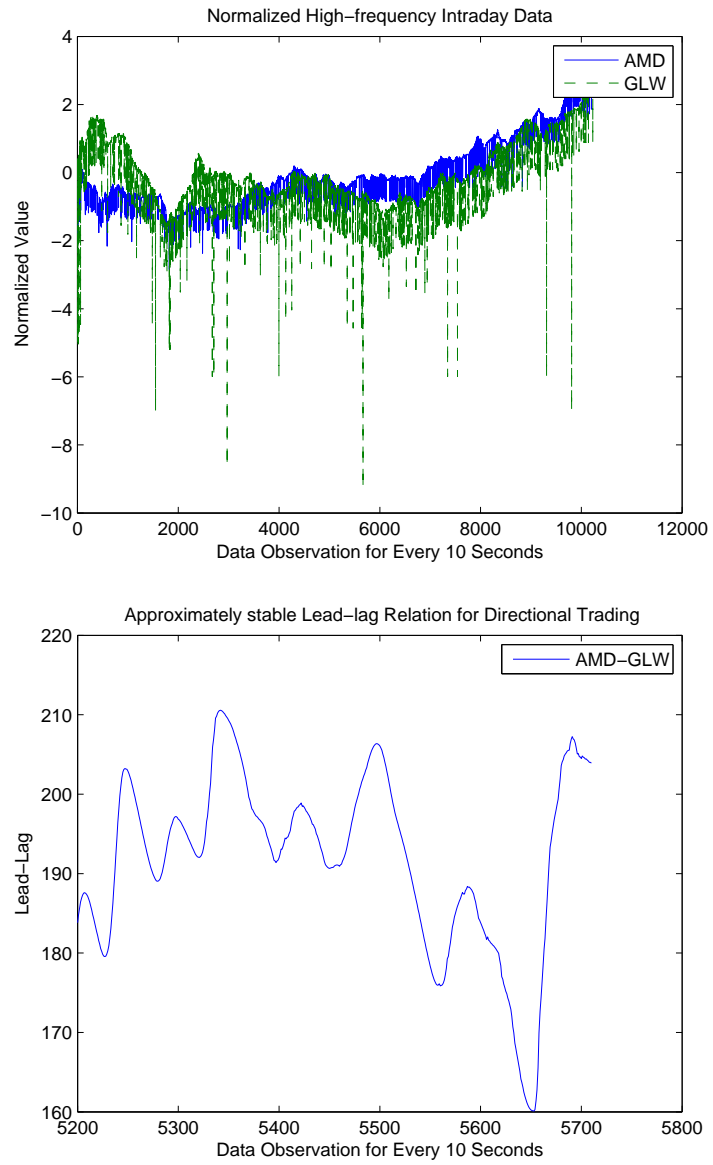


Figure 17: Lead-lag between AMD and GLW

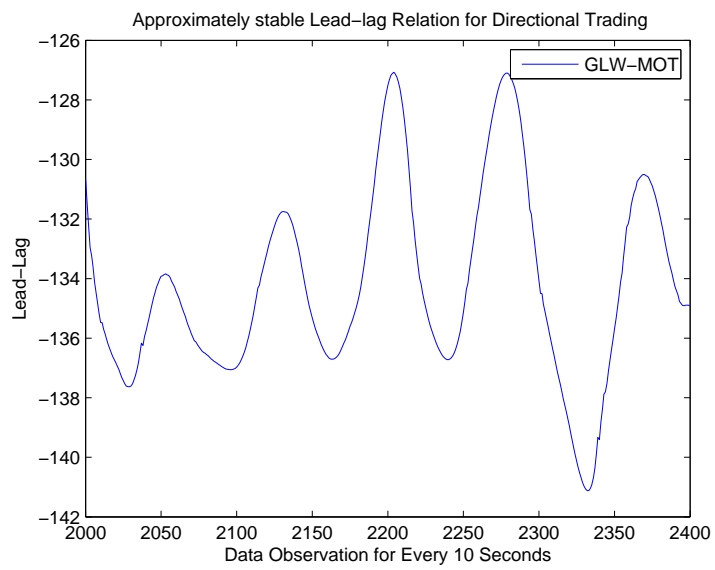
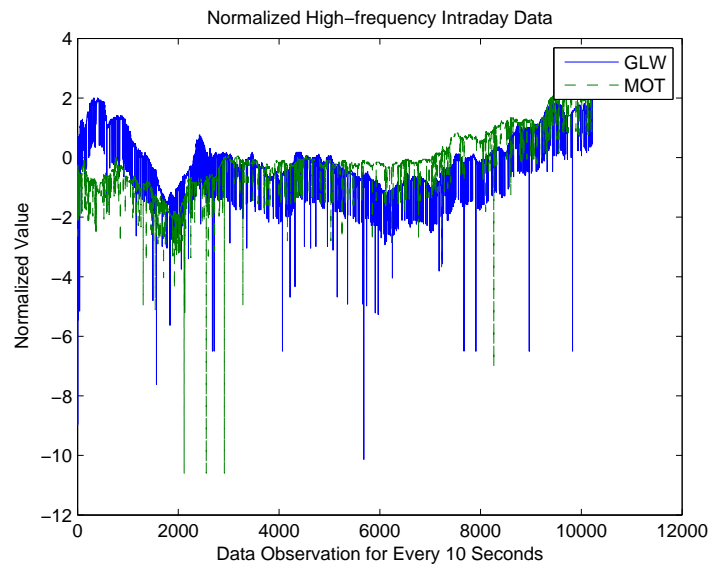


Figure 18: Lead-lag between GLW and MOT

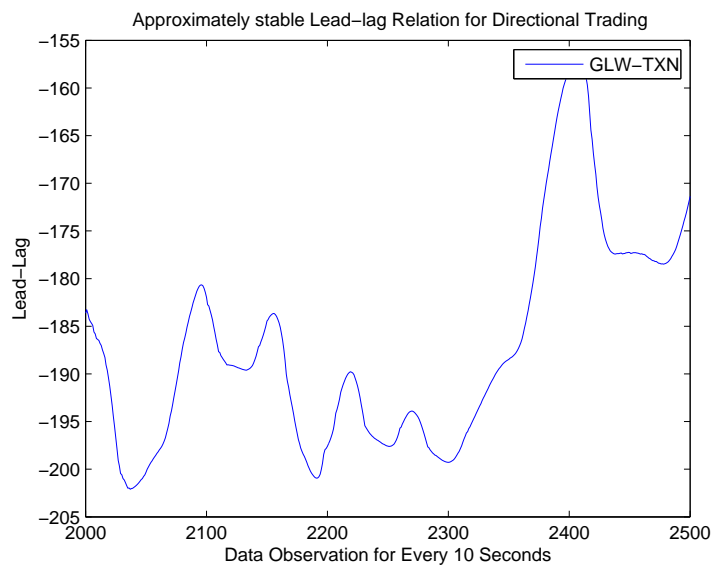
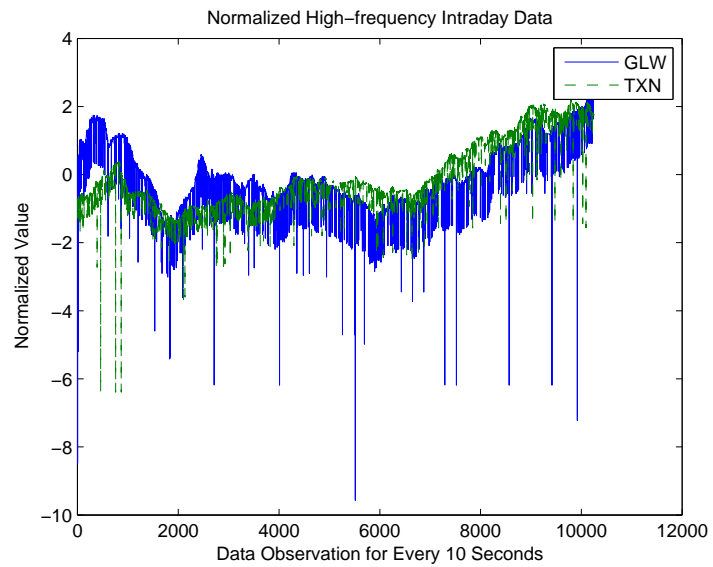


Figure 19: Lead-lag between GLW and TXN

technique that could reduce the computational time, and we implement it on various market sectors of NYSE data, extracting the lead-lag relations for highly correlated pairs of time series. We show how traders can exploit arbitrage opportunities by considering hit-ratio as a criteria. One can extend this work by doing rigorous back-testing and verify the technique's credibility for arbitrage trading.

References

- [1] *www.arbitrage-trading.com*, 2009.
- [2] I. Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. John Wiley & Sons, Inc, 2010.
- [3] C. S. Andrade, V. Pietro, and S. M. Seasholes. Understanding the profitability of pairs trading. *UC Berkeley Working Paper*, 2005.
- [4] L. Bachelier. Theorie de la speculation. *Ann. Sci. Ecole Norm. Sup.*, 17:21–86, 1900.
- [5] O. Bondarenko. Statistical arbitrage and securities prices. *Review of Financial Studies*, 16(3):875–919, 2003.
- [6] C. Brooks, A. G. Rew, and S. Ritson. A trading strategy based on the lead-lag relationship between the sopt index and futures contract for the FTSE 100. *International Journal of Forecasting*, 17:31–44, 2001.
- [7] A. N. Burgess. Statistical arbitrage models of the FTSE 100. *6th International Conference on Computational Finance*, 1999.
- [8] K. Chung-Ming and H. White. Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13:1–91, 1994.
- [9] F. F. Fama. Efficient capital markets: II. *Journal of Finance*, 1991.
- [10] F. F. Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 1998.
- [11] G. Gatev, N. Goetzmann, and K. Rouwenhorst. Pairs trading: Performance of a relative value arbitrage rule. *Review of Financial Studies*, 19(3):797–827, 2006.

- [12] M. Getmansky. Limits of arbitrage: Understanding how hedge funds fail. *Proceedings of the 23rd International Conference of the System Dynamic Society*, 2005.
- [13] C. W. J. Granger. *Essays in Econometrics: The Collected Papers of Clive W. J. Granger*. Cambridge University Press, 2001.
- [14] S. Hogan, R. Jarrow, M. Teo, and M. Warchka. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics*, 73(3):525–565, 2004.
- [15] O. Jorda and A. M. Taylor. Performance evaluation of zero net-investment strategies. *NBER Working Paper*, (17150), 2011.
- [16] P. J. Kaufman. *New Trading Systems and Methods*. John Wiley & Sons, Inc, 2005.
- [17] E. Keogh and M. Pazzani. Derivative dynamic time warping. *First SIAM International Conference on Data Mining*, 2001.
- [18] S. Kleinberg, P. Kolm, and B. Mishra. Investigating causal relationships in stock returns with temporal logic based methods. *Working Paper, New York University*, 2010.
- [19] A. W. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The Journal of Finance*, LV(4):1705–1765, 2000.
- [20] B. G. Malkiel. The efficient market hypothesis and its critics. *CEPS Working Paper*, 91, 2003.
- [21] P. Nath. High frequency pairs trading with U.S. treasury securities: Risks and rewards for hedge funds. *Working Paper, London Business School*, 2003.
- [22] A. Pole. *Statistical Arbitrage*. John Wiley & Sons, Inc, 2007.
- [23] C. A. Ratanamahatana and E. Keogh. Making time-series classification more accurate using learned constraints. *SIAM International Conference on Data Mining*, 2004.

- [24] Y. Sakurai, M. Yoshikawa, and C. Faloutsos. Fast similarity search under the time warping distance. *24th ACM SIGMOD International Conference on Principles of Database Systems (PODS'05)*, 2005.
- [25] G. Sanford and J. E. Stiglitz. On the impossibility of informationally efficient markets. *American Economic Review*, 70:393–408, 1980.
- [26] W. Shan, Z. Zhe, Y. Kang, H. Wang, and X. Chen. An ontology for causal relationships between news and financial instruments. *Expert Systems with Applications*, 35(3):569–580, 2008.
- [27] D. Sornette and W. Zhou. Non-parametric determination of real-time lag structure between two time series: the optimal thermal causal path method. *Quantitative Finance*, 28:195–224, 2006.
- [28] D. Sornette and W. Zhou. Non-parametric determination of real-time lag structure between two time series: the optimal thermal causal path method with applications to economics data. *Journal of macroeconomics*, 28:195–224, 2006.
- [29] L. Wei and E. J. Keogh. Semi-supervised time series classification. In *In Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 748–753, 2006.