Quantitative Finance, Vol. 00, No. 00, November 2012, 1–20

A Continuous Time Bayesian Network Classifier for Intraday FX Prediction

S. VILLA† and F. STELLA*†

†Department of Informatics, Systems and Communication
Faculty of Mathematics, Physics and Natural Sciences
University of Milano - Bicocca, Italy

(v2.0 released December 2013)

Prediction of foreign exchange (FX) rates is addressed as a binary classification problem in which a continuous time Bayesian network classifier is developed and used to solve it. An exact algorithm for inference on continuous time Bayesian network classifiers is introduced. The performance of an instance of these classifiers is analyzed and compared to that of dynamic Bayesian network by using real tick by tick foreign exchange rates. Performance analysis and comparison, based on different metrics such as accuracy, precision, recall, and Brier score, evince a predictive power of these models for foreign exchange rates at high frequencies. The achieved results also show that the proposed continuous time Bayesian network classifier is more effective and more efficient than dynamic Bayesian network classifier. In particular, it allows to perform high frequency prediction of foreign exchange rates in cases where dynamic Bayesian networks based models are computationally intractable.

Keywords: continuous time Bayesian networks, continuous time classifiers, FX prediction.

1. Introduction

High frequency financial data, made available by electronic order driven markets, has started a revolution on data processing and statistical modeling. This revolution forces quantitative finance analysts to cope with challenging theoretical and computational problems. Indeed, high frequency financial data requires to take into account micro structure effects which cannot be appreciated when using longer time intervals (i.e., hours, days, or months).

High frequency transaction data from the foreign exchange market is the largest and most liquid financial market. Its rapid growth over the last few years has been facilitated by the wider use of electronic trading at both levels, i.e., broker-dealer markets and customer markets deployed through on line trading platforms. The foreign exchange market is affected by many correlated economical, political and even psychological factors that make its forecasting a hard

*Corresponding author. Email: stella@disco.unimib.it

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

task. Researchers and practitioners have been striving for an explanation of the movement of exchange rates. It has long been known that structural models fail to predict exchange rates at short term horizons, see for example Chinn and Meese (1995), and Kilian and Taylor (2003).

The basic question to be answered is whether it is possible or not to forecast the behavior of an economic variable over a short time interval. The efficient market hypothesis (EMH), a fundamental pillar of modern economics introduced by Fama (1965), states that forecasting is not possible. The weak form of the this hypothesis asserts that the price of an asset reflects all of the information that can be obtained from its past prices. Accordingly to the EMH, the asset price follows a random walk and thus the best prediction of the future asset price is given by its current price. To state it differently, the movement of the price of an asset is unpredictable. The EMH makes the assumption that all news is promptly incorporated into the asset price, and since the news is unpredictable the asset price is unpredictable too.

However, a common myth of traders is that there is a certain predictability of market dynamics. Many efforts have been devoted to proving or disproving the EMH with several works rejecting it with specific reference to intraday data. Zhang (1999) exploited conditional entropy to show that even the most competitive markets are not strictly efficient; Baviera *et al.* (2000) rejected the random walk hypothesis for high frequency FX returns and proposed a Markovian model which reproduces the available information of the financial series; Renner *et al.* (2001) found evidence of Markov properties for high frequency exchange rate data; Baviera *et al.* (2002) found anti-persistent Markov behavior of log-price fluctuations in high frequency FX rates, which in principle allows the possibility of a statistical forecast; Ohira *et al.* (2002) found that there is a rather particular conditional probability structure in high frequency FX data; Tanaka-Yamawaki (2003) observed stability of a Markovian structure in high frequency FX data; recently, Shmilovici *et al.* (2009) tested the EMH using a variable order Markov model for twelve pairs of international intraday currency exchange rates for one year series with different time granularity. The authors found that intraday currency exchange rates are predictable above the random guess. These empirical evidences reveal that some predictability of high frequency FX rates exists.

The specialized literature describes FX forecasting techniques based on different assumptions and methods. Traditional parametric forecasting methods such as auto-regressive integrated moving average (ARIMA) and exponential smoothing models have been widely used for time series forecasting, see Box *et al.* (2008). ARIMA models assume a linear relationship between past and current value of underlying variables as well as for error terms. Time series models are highly nonlinear. Mean and variance of the time series can change over time. In order to overcome this difficulty, an auto-regressive conditional heteroskedastic (ARCH) model was introduced by Engle (1982) and successively generalized by Bollerslev (1986). Different implementations and extensions of GARCH models have been proposed to overcome some difficulties, i.e., nonlinearities and long term memory. Vilasuso (2002) showed that the fractionally integrated generalized ARCH (FIGARCH) model is better than both GARCH and IGARCH models to capture the salient features of the exchange rate volatility. However, FIGARCH models are unable to capture subtle nonlinear patterns hidden in the FX rate time series.

Recent developments in artificial intelligence and machine learning techniques, together with a continuously increasing availability of computational power, have allowed nonparametric models to be applied to FX forecasting. Most widely used techniques in the FX rates forecasting are artificial neural networks, with specific reference to multilayer perceptron, radial basis function networks, and recurrent neural networks, see Yu *et al.* (2007). Kuan and Liu (1995) compared the performance of multilayer perceptron and recurrent neural networks when predicting daily FX rate. Their work provides evidence that supports the usefulness of neural network models to predict FX rates. The same conclusion was drawn by Yao and Tan (2000) where time series data and technical indicators were fed into neural networks to capture the underlying rules of the movement in currency exchange rates. Other types of artificial neural networks were studied, Leung *et al.* (2000) compared the forecasting accuracy of multilayer perceptron to that of generalized regression neural networks. The authors concluded that the latter performs statistically better than other models for different currencies. Kamruzzaman and Sarker (2003) investigated three artificial neural networks for Australian Dollar rate prediction using moving average indicators and they concluded that these models outperform the ARIMA model on five performance metrics. Although artificial neural network models achieve success in FX rates forecasting due to their nonlinearity, robustness and adaptivity, they still have some disadvantages in some practical applications. They suffer from several problems and limitations such as over-fitting, lack of interpretability, difficulty to handle qualitative factors, and local minimum problems often incurred when adopting the empirical risk minimization principle to solve the learning task.

To overcome the latter problem support vector machines have been introduced. These models minimize structural risks as opposed to empirical risks, used by artificial neural networks. Kamruzzaman *et al.* (2003) showed support vector machines to be a powerful tool for forecasting six different foreign currency exchange rates against Australian Dollar and achieved better performance than artificial neural networks and ARIMA based models.

Genetic algorithms approaches have been applied to the FX market. Dunis and Zhou (1998) developed a FX trading system that uses genetic algorithms to optimize parameters for technical trading indicators by using FX data, while Dempster and Jones (2001) introduced a framework for automatic trading system construction and adaptation based on genetic algorithms.

In the recent literature, hybrid models have also been proposed, such as Ince and Trafalis (2006). In their paper the authors combined parametric and nonparametric techniques to achieve a better performance for daily exchange rate forecasting. Gradojevic (2007) combined artificial neural networks and fuzzy logic controllers to obtain the optimal daily currency trading rules. Kablan and Ng (2011) introduced an adaptive neuro-fuzzy inference system for financial trading, which learns to predict FX price movements of tick by tick data.

In this paper, the high frequency FX rate prediction problem is formulated as binary classification in which the class is expected to occur in the future, the time flows continuously, and the classification decision must be made at any point in time using trained continuous time naïve Bayes classifiers, see Stella and Amer (2012). The performance of this continuous time classifier is analyzed and compared to that of a dynamic Bayesian networks by exploiting tick by tick FX rates for three currency pairs. Performance analysis and comparison are based on different metrics such as accuracy, precision, recall, and Brier score.

The paper is laid out as follows. Section 2 introduces the theory underpinning the continuous time Bayesian network classifiers and the inference framework. Section 3 presents the used market data together with its basic analysis. Section 4 describes an instance of the continuous time Bayesian network classifier, together with the results of the performed numerical experiments. Finally, Section 5 is devoted to comments and conclusions.

2. Background

One of the most salient features of high frequency transaction data is that transactions do not occur at regularly spaced time intervals. Different models have been developed which are suitable for high frequency data, e.g., Engle and Russell (1998) proposed the auto-regressive conditional duration model that treats the waiting time between events as a stochastic process and proposes a class of point processes with dependent arrival rates. Unfortunately, this model and its extensions do not incorporate information coming from the price process, i.e., they cannot represent a set of random variables together with their conditional dependency structure.

Bayesian networks introduced by Pearl (1985) are probabilistic graphical models which allow the efficient description and management of highly dimensional joint probability distributions. They are specifically designed for cases when the random variables can have considerable dimension and/or it is difficult to come up with traditional parametric models of the joint probability distribution. Bayesian networks are used to model dependencies between variables (as a directed acyclic graph), but they do not model temporal dependencies, see Jensen and Nielsen (2007).

Dynamic Bayesian networks (DBNs) introduced by Dean and Kanazawa (1989) extend Bayesian networks to model dynamical systems. These networks do not model time explicitly but discretize it into fixed length Δt intervals (time slices). The variables can be denoted as the state of the DBN because they include the temporal dimension. The states of a system described as a DBN satisfy the Markovian condition, i.e., the state of a system at time $t + \Delta t$ depends only on its immediate past, i.e., its state at time t. DBNs allow connections within time slices (called intra-slice) and between time slices (called inter-slice). These temporal connections incorporate the conditional probabilities between variables from two different time slices, while their conditional probability tables represent the transition matrices from time t to $t+\Delta t$. Using a DBN model is it possible to answer probabilistic queries concerning the state of the system at a particular position in the sequence, such as filtering, smoothing, prediction, decoding, and classification, see Murphy (2002) for detailed analysis.

Dynamic Bayesian networks allow to model the dependency structure between variables by using transition probabilities between states at times t and $t + \Delta t$. However, "since DBNs slice time into fixed increments, one must always propagate the joint distribution over the variables at the same rate", as stated by Nodelman *et al.* (2002). Therefore, if the system consists of processes which evolve at different time granularities and/or the observations are irregularly spaced in time, the inference process may become computationally intractable.

2.1. Continuous time Bayesian networks

Continuous time Bayesian networks overcome dynamic Bayesian networks limitations by explicitly representing temporal dynamics and thus allow you to recover the probability distribution over time when specific events occur. Continuous time Bayesian networks have been applied in different fields, such as to model the presence of people on their computers, Nodelman and Horvitz (2003), for dynamical systems reliability modeling and analysis, Boudali and Dugan (2006), for network intrusion detection, Xu and Shelton (2008), to model social networks, Fan and Shelton (2009), and to model cardiogenic heart failure, Gatti *et al.* (2011).

A continuous time Bayesian network is a graphical model whose nodes are associated with random variables whose state evolves in continuous time. The evolution of each variable depends on the state of its parents in the graph. A continuous time Bayesian network consists of two main components: (i) an initial probability distribution and (ii) the dynamics which rule the evolution over time of the probability distribution associated with the continuous time Bayesian network.

Definition 2.1: Continuous time Bayesian network (CTBN), Nodelman *et al.* (2002). Let $\boldsymbol{X}$ be a set of random variables $X_1, X_2, \ldots, X_N$. Each X_n has a finite domain of values $Val(X_n) = \{x_1, x_2, \ldots, x_I\}$. A continuous time Bayesian network $\aleph$ over $\boldsymbol{X}$ consists of two components: the first is an initial distribution $P^0_{\boldsymbol{X}}$, specified as a Bayesian network $\mathcal{B}$ over $\boldsymbol{X}$, the second is a continuous time transition model specified as:

- a directed (possibly cyclic) graph $\mathcal{G}$ whose nodes are $X_1, X_2, \ldots, X_N$;
- a conditional intensity matrix, $\boldsymbol{Q}^{Pa(X_n)}_{X_n}$, for each variable $X_n \in \boldsymbol{X}$, where $Pa(X_n)$ denotes the set of parents of X_n in $\mathcal{G}$.

Given the random variable X_n, the conditional intensity matrix (CIM) $\boldsymbol{Q}_{X_n}^{Pa(X_n)}$ consists of a set of intensity matrices, one intensity matrix:

$$\boldsymbol{Q}_{X_n}^{pa(X_n)} = \begin{bmatrix} -q_{x_1}^{pa(X_n)} & q_{x_1 x_2}^{pa(X_n)} & . & q_{x_1 x_I}^{pa(X_n)} \\ q_{x_2 x_1}^{pa(X_n)} & -q_{x_2}^{pa(X_n)} & . & q_{x_2 x_I}^{pa(X_n)} \\ . & . & . & . \\ q_{x_I x_1}^{pa(X_n)} & q_{x_I x_2}^{pa(X_n)} & . & -q_{x_I}^{pa(X_n)} \end{bmatrix},$$

for each instantiation $pa(X_n)$ of the parents $Pa(X_n)$ of node X_n, where I is the cardinality of X_n, $q_{x_i}^{pa(X_n)} = \sum_{x_j \neq x_i} q_{x_i x_j}^{pa(X_n)}$ can be interpreted as the instantaneous probability to leave x_i for a specific instantiation $pa(X_n)$ of $Pa(X_n)$, and $q_{x_i x_j}^{pa(X_n)}$ can be interpreted as the instantaneous probability to transition from x_i to x_j for a specific instantiation $pa(X_n)$ of $Pa(X_n)$.

Continuous time Bayesian networks allow point evidence and continuous evidence, while dynamic Bayesian networks allow point evidence only. Point evidence is an observation of the value x of a variable X_n at a particular instant in time, i.e., $X_n(t) = x$, while continuous evidence provides the value of a variable throughout an entire interval, which we take to be a half-closed interval $[t_{i-1}, t_i)$. Continuous time Bayesian networks are instantiated with a J-evidence-stream.

Definition 2.2: J-time-stream, Stella and Amer (2012). A J-time-stream, over the left-closed time interval $[0, T)$, is a partitioning into J left-closed intervals $[0, t_1); [t_1, t_2); \ldots; [t_{J-1}, T)$.

Definition 2.3: J-evidence-stream, Stella and Amer (2012). Given a CTBN $\aleph$, consisting of N nodes, and a J-time-stream $[0, t_1); [t_1, t_2); \ldots; [t_{J-1}, T)$, a J-evidence-stream is the set of joint instantiations $\mathbf{X} = \mathbf{x}$ for any subset of random variables X_n, $n = 1, 2, \ldots, N$ associated with each of the J time segments. A J-evidence-stream will be referred to as $(\mathbf{X}^1 = \mathbf{x}^1, \mathbf{X}^2 = \mathbf{x}^2, \ldots, \mathbf{X}^J = \mathbf{x}^J)$ or for short as $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J)$.

A J-evidence-stream is said to be fully observed in the case where the state of all the variables X_n is known along the whole time interval $[0, T)$. A J-evidence-stream which is not fully observed is said to be partially observed. Given a continuous time Bayesian network and a J-evidence-stream, inference is concerned with the computation of the posterior for those variables whose state is unknown. Exact inference is intractable in the general case as the state space of the system grows exponentially with the number of variables. For this reason, several approximate algorithms have been proposed: Nodelman *et al.* (2005a) provided the expectation propagation algorithm which allows point and continuous evidence; Saria *et al.* (2007) introduced an expectation propagation algorithm which utilizes clusters contain distributions that can overlap in both space and time; El-Hay *et al.* (2008) proposed a Gibbs sampling procedure; and Fan *et al.* (2010) developed importance sampling algorithm.

The problem of learning a continuous time Bayesian network from data has been addressed in Nodelman *et al.* (2003) and Nodelman *et al.* (2005b) as the problem of finding the structure $\mathcal{G}$ which maximizes a Bayesian score. While it is known that learning the optimal structure of a Bayesian network is NP-complete, see Chickering (1996), the same does not hold true in the context of continuous time Bayesian networks learning where all edges are across time and thus represent the effect of the current value of one variable on the next value of the other variables. Therefore, no acyclicity constraints arise, and it is possible to optimize the parent set for each variable of the network independently, see Nodelman (2007).

2.2. Continuous time Bayesian network classifiers

The continuous time Bayesian network model has been exploited to define a class of continuous time classifiers, see Stella and Amer (2012). In particular, a continuous time Bayesian network classifier is a supervised classification model which explicitly represent the evolution in continuous time of set of random variables X_n, $n = 1, 2, \ldots, N$.

Definition 2.4: Continuous time Bayesian network classifier (CTBNC), Stella and Amer (2012). A continuous time Bayesian network classifier is a pair $\mathcal{C} = \{\aleph, P(Y)\}$ where $\aleph$ is a CTBN model with attribute nodes $X_1, X_2, \ldots, X_N$, class node Y with marginal probability $P(Y)$ on states $Val(Y) = \{y_1, y_2, \ldots, y_K\}$, and $\mathcal{G}$ is the graph, such that the following conditions hold:

- $\mathcal{G}$ is connected;
- $Pa(Y) = \emptyset$, the class variable Y is associated with a root node;
- Y is fully specified by $P(Y)$ and does not depend on time.

In this paper, we focus on a specific instance from the class of continuous time Bayesian network classifiers where the class variable Y, which is the root node, is the only parent for the attributes X_n, $n = 1, 2, \ldots, N$ as shown in Figure 1(a). This instance is called continuous time naïve Bayes classifier and it is formally defined as follows.

Definition 2.5: Continuous time naïve Bayes classifier (CTBNC-NB), Stella and Amer (2012). A continuous time naïve Bayes classifier is a is a continuous time Bayesian network classifier $\mathcal{C} = \{\aleph, P(Y)\}$ such that $Pa(X_n) = Y$, $n = 1, 2, \ldots, N$.

As benchmark purposes, we have used a dynamic naïve Bayes classifier (DBNC-NB) in which at each time we have a naïve Bayes classifier (intra-slice topology) and the attributes nodes are connected from one time to the next (inter-slice topology), see Murphy (2002) and Martinez and Sucar (2008). An example of dynamic naïve Bayes classifier is depicted in Figure 1(b) where the dashed lines identify the temporal relationships from time t to $t + 1$.

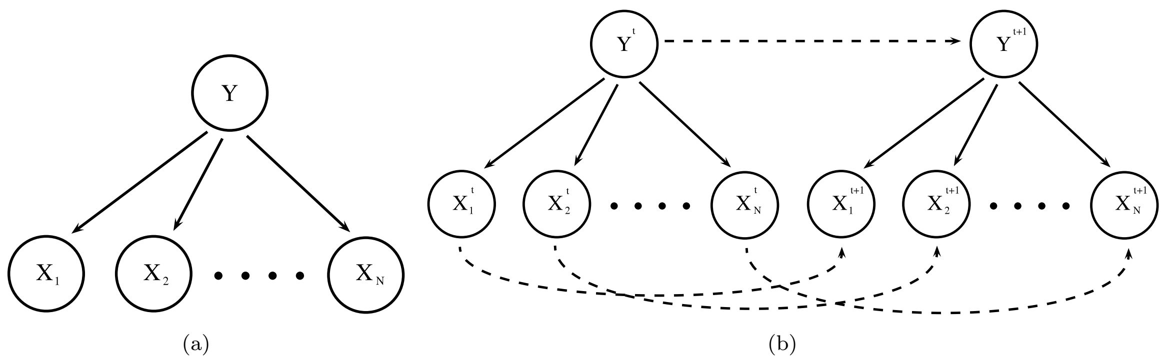


Figure 1. An instance of a continuous time naïve Bayes classifier (a), and an instance of a dynamic naïve Bayes classifier (b) in which the dashed lines identify the temporal relationships from time t to $t + 1$.

Given a dataset of fully observed J-evidence-streams, the learning task of a continuous time naïve Bayes classifier consists of estimating the prior probability associated with the class node Y, and estimating, for each attribute node X_n, $n = 1, 2, \ldots, N$, the conditional intensity matrix $\boldsymbol{Q}_{X_n}^{pa(X_n)}$ for each instantiation $pa(X_n)$ of its parents $Pa(X_n)$. The learning algorithm for continuous time Bayesian network classifiers is described in detail in Stella and Amer (2012). Once the learning problem has been solved, the model can be used to classify fully observed J-evidence-streams.

2.3. Inference framework

A continuous time Bayesian network classifier classifies a fully observed J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J)$ by the class value which maximizes the posterior probability:

$$P(Y|\left(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J\right)) = \frac{P(\left(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J\right)|Y)P(Y)}{P((\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J))}, \tag{1}$$

where $P(Y)$ represents the prior probability of the class Y and $P(\left(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J\right)|Y)$ is the likelihood of the fully observed J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J)$ given the class Y, while $P(\left(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J\right))$ is the probability of the fully observed J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^J)$.

The likelihood can be written as follows:

$$P(\left(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^J\right)|Y) = \prod_{j=1}^{J} P(\mathbf{x}^j|Y)P(\mathbf{x}^{j+1}|\mathbf{x}^j, Y), \tag{2}$$

where $P(\mathbf{x}^j|Y)$ represents the probability that $\mathbf{X}$ stays in state $\mathbf{x}^j$ during the time interval $[t_{j-1}, t_j)$ given the class variable Y, while the term $P(\mathbf{x}^{j+1}|\mathbf{x}^j, Y)$ represents the probability that $\mathbf{X}$ transitions from state $\mathbf{x}^j$ to state $\mathbf{x}^{j+1}$ at time t_j given the class variable Y (to ensure consistency we set $P(\mathbf{x}^{J+1}|\mathbf{x}^J, Y) = 1$).

In the case where a continuous time Bayesian network classifier is used, we can write:

$$P(\mathbf{x}^j|Y)P(\mathbf{x}^{j+1}|\mathbf{x}^j, Y) = \prod_{n=1}^{N} P(x_n^j|Y)P(x_n^{j+1}|x_n^j, Y). \tag{3}$$

Continuous time Bayesian networks allow single transition only, i.e., at any time t_j a single variable X_m can change its state while the states of the remaining variables X_n with $n \neq m$ remain unchanged. Therefore, for X_n, $n \neq m$, we have $P(x_n^{j+1}|x_n^j, Y) = 1$, while for X_m we have:

$$P(x_m^{j+1}|x_m^j, Y) = \frac{q_{x_m^j x_m^{j+1}}^{pa(X_m)}}{q_{x_m^j}^{pa(X_m)}}, \tag{4}$$

where:

- $q_{x_m^j}^{pa(X_m)}$ is the parameter associated with state x_m^j, in which the variable X_m was during the j^{th} time segment, given the state $pa(X_m)$ of its parents $Pa(X_m)$ during the j^{th} time segments;
- $q_{x_m^j x_m^{j+1}}^{pa(X_m)}$ is the parameter associated with the transition from state x_m^j, in which the variable X_m was during the j^{th} time segment, to state x_m^{j+1}, in which the variable X_m will be during the $(j+1)^{th}$ time segment, given the state $pa(X_m)$ of its parents $Pa(X_m)$ during the j^{th} and the $(j+1)^{th}$ time segments.

It is worthwhile to mention that $P(x_n^j|Y)$ can be further specified for the variable X_m which transitions at time t_j $(x_m^j \neq x_m^{j+1})$ and for the variables X_n $n \neq m$ that do not change their state at time t_j $(x_n^j = x_n^{j+1})$.

More precisely, we can write the following:

$$P(x_n^j|Y) = \begin{cases} \exp\left(-q_{x_n^j}^{pa(X_n)}(t_j - t_{j-1})\right) & \text{if } x_n^j = x_n^{j+1}, \\ q_{x_n^j}^{pa(X_n)} \exp\left(-q_{x_n^j}^{pa(X_n)}(t_j - t_{j-1})\right) & \text{if } x_n^j \neq x_n^{j+1}. \end{cases} \tag{5}$$

Equations (4) and (5), when combined with (3), allow us to write:

$$P(\mathbf{x}^j|Y)P(\mathbf{x}^{j+1}|\mathbf{x}^j, Y) = q_{x_m^j x_m^{j+1}}^{pa(X_m)} \prod_{n=1}^{N} \exp\left(-q_{x_n^j}^{pa(X_n)}(t_j - t_{j-1})\right), \tag{6}$$

which combined with Equations (1) and (2) allow to conclude that a continuous time Bayesian network classifier classifies a fully observed J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$ by selecting the class $y^* \in Val(Y)$ as follows:

$$y^* = \arg\max_{y \in Val(Y)} \left\{ P(y) \prod_{j=1}^{J} q_{x_{m_j}^j x_{m_j}^{j+1}}^{pa(X_{m_j})} \prod_{n=1}^{N} \exp\left(-q_{x_n^j}^{pa(X_n)}(t_j - t_{j-1})\right) \right\}, \tag{7}$$

where m_j is the index of the variable which transitions at time t_j.

Given a continuous time Bayesian network classifier $\mathcal{C} = \{\aleph, P(Y)\}$ consisting of N attribute nodes $X_1, X_2, \dots, X_N$, a class node Y such that $Val(Y) = \{y_1, y_2, \dots, y_K\}$, and a fully observed J-evidence-stream $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J)$, Algorithm **1** computes the posterior probability and returns the most probable class. The for statement reported from line 1 to 3 computes the a priori probability for each class, while the core of the algorithm, from line 4 to 13, computes the logarithm of Equation (7). The three for statements range over the K classes (from line 4 to 13), the J-evidence-streams (from line 5 to 12), and the N attributes (from line 6 to 11). The most probable class y^* is computed by the argmax function in line 14.

Algorithm 1 CTBNC Inference

Require: a CTBNC $\mathcal{C} = \{\aleph, P(Y)\}$ consisting of N attribute nodes, a class node Y such that $Val(Y) = \{y_1, y_2, \dots, y_K\}$, and a fully observed J-evidence-stream $\left(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J\right)$.
Ensure: the maximum aposteriori classification y^* for $\left(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J\right)$.
1: **for** $k \leftarrow 1$ to K **do**
2:    $logp(y_k) \leftarrow \log P(y_k)$
3: **end for**
4: **for** $k \leftarrow 1$ to K **do**
5:    **for** $j \leftarrow 1$ to J **do**
6:       **for** $n \leftarrow 1$ to N **do**
7:          $logp(y_k) \leftarrow logp(y_k) - q_{x_n^j}^{pa(X_n)}(t_j - t_{j-1})$
8:          **if** $x_n^j \neq x_n^{j+1}$ **then**
9:             $logp(y_k) \leftarrow logp(y_k) + \log\left(q_{x_n^j x_n^{j+1}}^{pa(X_n)}\right)$
10:         **end if**
11:      **end for**
12:   **end for**
13: **end for**
14: $y^* \leftarrow \arg\max_{y_k \in Val(Y)} \{logp(y_k)\}$

3. Market data

It has been recognized that the analysis of the FX market is extremely complex, see Dacorogna *et al.* (2001). Indeed, it is a 24 hour globally traded market, it is active 5 days a week in different countries with calendars having different public holidays. It is a decentralized market where transactions are not recorded by a central institution and is not subject to formal regulations, see BIS (2011). Furthermore, there is no comprehensive data source, data vendors capture the transaction data from the global market or with their own platforms. These data are then fed in real time to their customers.

Table 1. Datasets used for the experiments and summary statistics for the log-returns.

Dataset	**Pair**	**Records**	**Tick (secs)**	**Mean** ($\times 10^{-9}$)	**Std** ($\times 10^{-5}$)	**Skew**	**Kurt** ($\times 10^{2}$)
TFX	EUR/USD	80,817,600	0.0010	-0.3810	1.5614	-32.3505	445.1200
	GBP/USD	36,878,945	0.0010	-0.0486	1.7621	-1.0388	39.3435
	EUR/CHF	45,171,474	0.0010	-0.5635	3.3986	0.5946	271.4766
DKY	EUR/USD	25,744,055	0.0010	-1.3756	2.1651	3.0588	36.2991
	GBP/USD	24,814,328	0.0010	-0.0821	1.8048	-0.4253	9.3222
	EUR/CHF	22,220,534	0.0010	-1.4233	2.9947	1.8330	77.2407
GCL	EUR/USD	5,356,005	1.0000	-5.5655	6.7719	0.4124	1.6812
	GBP/USD	4,892,874	1.0000	-0.3810	5.8300	-0.0385	0.5520
	EUR/CHF	4,451,148	1.0000	-5.6720	8.7358	18.1648	80.8507
BMK	GRW	22,463,844	1.0000	1.8904	12.7484	0.0018	0.0412

The lack of a common, multi-source, and publicly available tick by tick FX rate database required us to design and develop an integrated database. It contains market data provided by three FX trading platforms publicly available over the Internet, namely: TrueFX, Dukascopy, and GainCapital. We call the datasets built with these sources as TFX, DKY, and GCL respectively.

Each dataset is composed by tick by tick bid-ask prices, spanning from January 1st, 2011 to December 30th, 2011 (260 trading days), of three commonly traded currency pairs: EUR/USD, GBP/USD, and EUR/CHF. The database amounts to 250,346,963 bid-ask observations. Time granularity of TFX and DKY datasets equals 1 millisecond, while the time granularity of the GCL equals 1 second. Furthermore, one year of tick by tick data, generated from a Gaussian random walk (GRW) model was added to the database for comparison purposes (BMK), as in Glattfelder *et al.* (2011) we set $x_0 = 1.336723$, $\Delta t = 1$ second, $\mu = 0$, and $\sigma = 1/6769.6$.

The data were submitted to basic data analysis procedures. Although FX data consists of bid-ask quotes, statistical characteristics of mid-prices have been analyzed. Moreover, only those transactions where the bid price or the ask price changed from the previous tick have been used.

A preliminary empirical study of these data confirmed the existence of significant trends in the FX market as reported in Dacorogna *et al.* (2001). Temporal dependence is a typical characteristic of high frequency data. Periodic patterns can be found in volatility, trade frequency, volume, and spreads as depicted in Figures 2(a) and 2(b). Serial correlation patterns of the log-returns can be also discovered as displayed in Figure 2(c), see Cont (2001) for details.

The most important characteristic of high frequency data is that they are irregularly spaced over time. The duration between successive trades reflects the intensity of the trading activity, see Hautsch (2004) for review. Figure 2(e) shows the empirical cumulative distribution of the time lag between consecutive transactions for TFX EUR/USD. The characteristic of irregularly spaced observations plays a central role when developing continuous time models, as opposed to discrete time models based on fixed time steps.

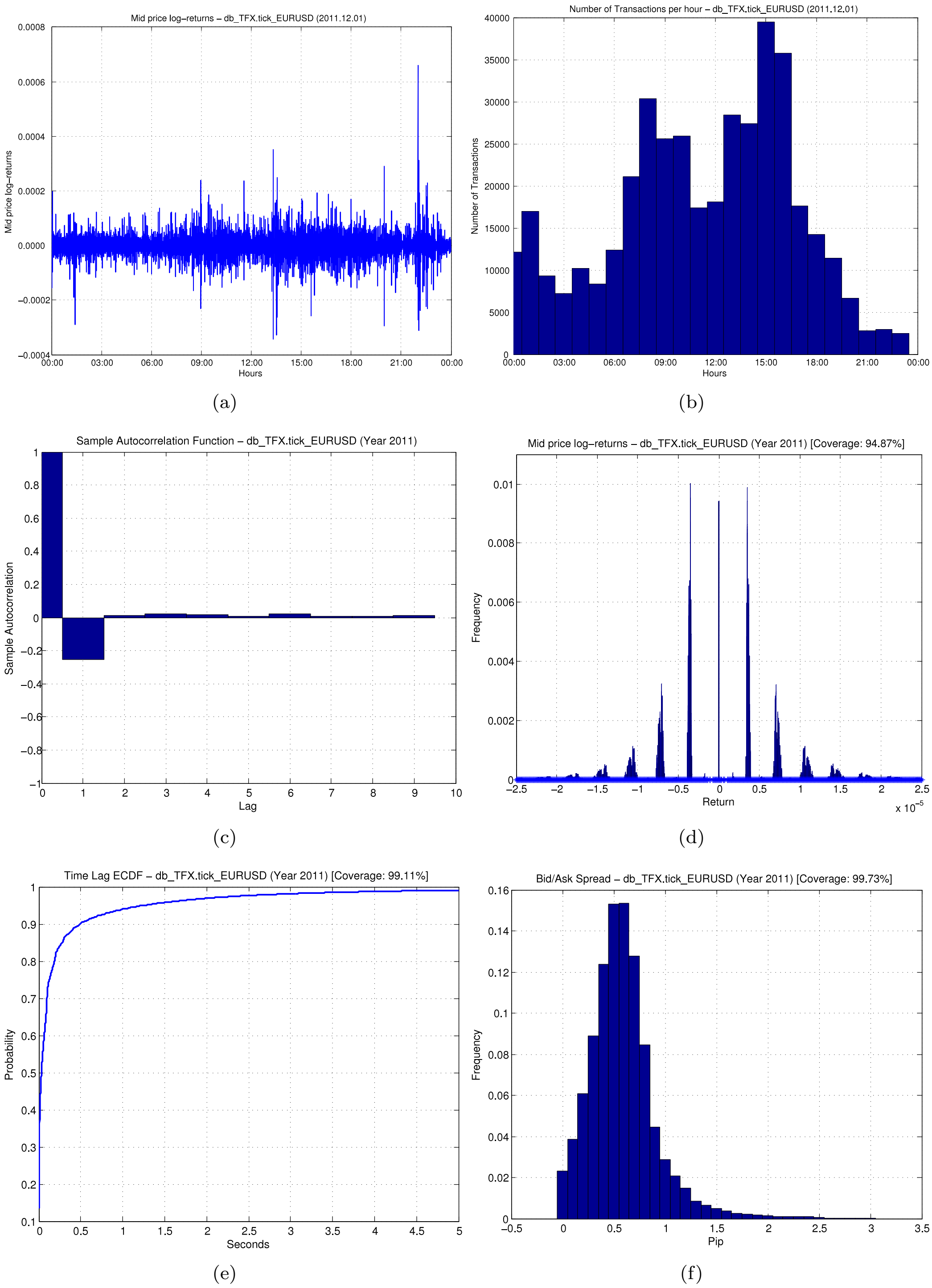


Figure 2. Data analysis of the TFX EUR/USD spot rate dataset for year 2011: (a) tick by tick log-returns as of December 1st, 2011; (b) number of transactions per hour as of December 1st, 2011; (c) sample autocorrelation function of the log-returns; (d) tick by tick log-returns distribution (94.87% of the dataset is displayed); (e) empirical cumulative distribution function of the time lag between two transactions (99.11% of the dataset is displayed); (f) bid/ask spread in terms of tenth of a *pip* (price interest point, i.e., 10^{-4} of an exchange rate, 99.73% of the dataset is displayed).

4. Numerical experiments

Numerical experiments were planned and performed by following standard tasks and guidelines of data mining, see Bellazzi and Zupan (2008) for review. Continuous time naïve Bayes classifier is compared to dynamic naïve Bayes classifier by using the following performance measures: learning and inference time, accuracy, precision, recall, receiver operating characteristic (ROC) curve, and Brier score. Comparison concerns three currency pairs: EUR/USD, GBP/USD, and EUR/CHF, of the following datasets: TFX, DKY, and GCL.

Prediction of foreign exchange rates is addressed as a binary classification problem. This task consists of predicting if a trajectory, a sequence of FX mid price values, will result in an up movement of the FX spot rate by a fixed amount of υ pips or not. Therefore, the class is a binary variable Y, where $Y = 1$ if the FX spot rate shows an up movement of υ pips, and $Y = 2$ otherwise. Table 2 reports the number and the duration (in seconds) of up movements and down movements of $\upsilon = 10$ pips used for the experiments.

Table 2. Number and duration (in secs) of an up ($Y = 1$) and down ($Y = 2$) movements of $\upsilon = 10$ pips.

		Y=1			**Y=2**		
Dataset	**Pair**	**Number**	**Mean** ($\times 10^3$)	**Std** ($\times 10^3$)	**Number**	**Mean** ($\times 10^3$)	**Std** ($\times 10^3$)
TFX	EUR/USD	12,740	0.8593	1.5544	12,702	0.8518	1.5300
	GBP/USD	10,916	0.9933	1.7694	10,883	1.0014	1.7921
	EUR/CHF	11,939	0.9323	2.2367	11,935	0.8784	2.1394
DKY	EUR/USD	12,384	0.9064	1.6154	12,422	0.8921	1.5266
	GBP/USD	10,521	1.0682	1.8198	10,533	1.0532	1.7362
	EUR/CHF	11,116	1.0442	2.4836	11,157	0.9516	2.1523
GCL	EUR/USD	12,622	0.8864	1.6076	12,659	0.8784	1.5487
	GBP/USD	10,781	1.0459	1.8038	10,785	1.0223	1.6706
	EUR/CHF	11,096	1.0327	2.3953	11,134	0.9639	2.2121
BMK	GRW	206,652	0.0543	0.0442	206,579	0.0544	0.0441

4.1. Model

The attributes used by the classifiers to solve the binary classification problem have been selected by exploiting basic concepts of digital signal processing with specific reference to moving averages, see Proakis and Manolakis (2006) for review. The moving average technique, for smoothing short-term price variation, has been used. Moving average involves the computation of the mean value of the past mid price as follows:

$$MA_t^M = \frac{1}{M} \sum_{i=0}^{M-1} p_{t-i}, \tag{8}$$

where M represents the length of the moving average while p_{t-i} is the mid price at time $t - i$. Although it is possible to associate different weights to different past prices, a simple moving average, assigning the same weight to different past prices, was used. Different moving averages were used to compute a set of moving average triggers. A moving average trigger is defined as:

$$X_{M_1,M_2} = \begin{cases} 1 & \text{if } MA_t^{M_1} > MA_t^{M_2}, \\ 2 & \text{if } MA_t^{M_1} \leq MA_t^{M_2}. \end{cases} \tag{9}$$

To visualize the concept, an up movement of the mid price for the EUR/USD spot rate is depicted in Figure 3(a), while the smoothed movement obtained through two moving averages with $M_1 = 40$ and $M_2 = 20$ together with their relative trigger $X_{40,20}$ are shown in Figure 3(b). The time evolutions of the triggers constitute the input of the classifier as shown in Figure 4(a), while the output is the class probability computed by the classifier as depicted in Figure 4(b).

A preliminary set of experiments was performed to select the set of moving average triggers to be used by the classifiers to solve the binary classification problem. To clarify the preliminary experiments, we let $\mathcal{M} = \{M_1, \ldots, M_m\}$ to be a set of moving averages of length $M_1, \ldots, M_m$, while $\mathbf{X}$ is a set of trigger variables $X_{i,j}$ with $i, j \in \mathcal{M}$ and $i > j$. For example, if $\mathcal{M} = \{80, 60, 40, 20\}$ then $\mathbf{X} = \{X_{80,60}, X_{80,40}, X_{80,20}, X_{60,40}, X_{60,20}, X_{40,20}\}$. Preliminary experiments are associated with configurations that are defined as pairs $\mathcal{C} = \{\upsilon, \mathcal{M}\}$.

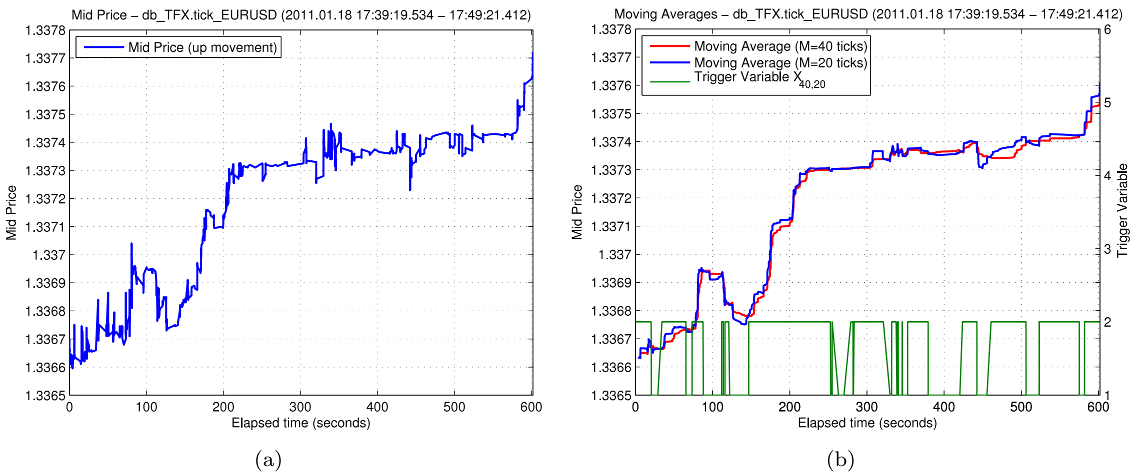


Figure 3. (a) mid price up movement of $\upsilon = 10$ pips of the TFX EUR/USD spot rate, (b) the smoothed movement obtained by two moving averages (left axis) and their trigger (right axis).

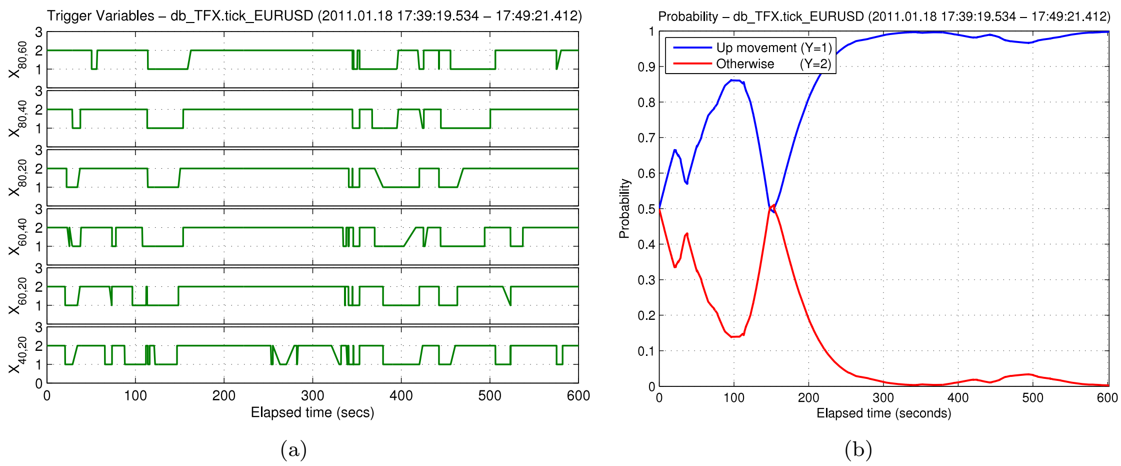
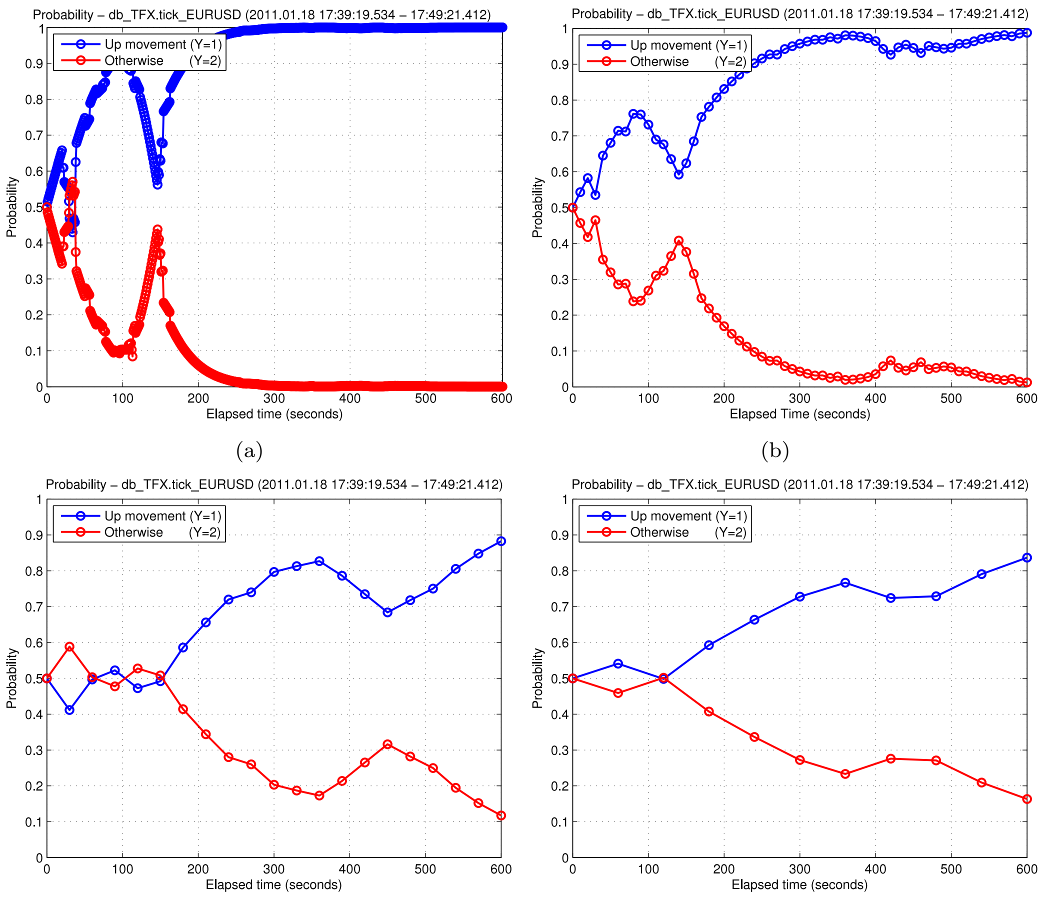


Figure 4. (a) time evolutions of six trigger variables that constitute the input for the classifier, (b) time evolution of the class probability computed by continuous time Bayesian network classifier.

4.2. Benchmarking

The dynamic naïve Bayes classifier was used as benchmark where the parameter learning has been performed with the expectation maximization algorithm as described by Murphy (2002), while inference is performed by using the BK algorithm introduced by Boyen and Koller (1998).

Figures 5(a), 5(b), 5(c), and 5(d) show the evolution of the posterior probability associated with the class variable Y 'up movement of $\upsilon = 10$ pips' for the TFX EUR/USD spot rate computed by the dynamic classifier. Note that the continuous time classifier allows to obtain the posterior probability of the class variable at each point in time, conversely the same does not occur for the dynamic classifier which provides probability values at discrete time points (Δt) computed via the approximate inference algorithm. It is worthwhile to notice that, for large values of the time slice Δt, the dynamic classifier simply cannot capture the transition dynamics accurately enough to converge to competitive performance.



(a) (b) (c) (d)

Figure 5. Class probability evolutions of a mid price up movement of $\upsilon = 10$ pips using trained dynamic classifiers with (a) $\Delta t = 1$ sec, (b) $\Delta t = 10$ secs, (c) $\Delta t = 30$ secs, and (d) $\Delta t = 60$ secs.

There are some differences between the two models underpinning the classifiers, namely DBN and CTBN. DBN allows to model non-Markovian behavior between time slices, but you have to cope with the exponential blowup in performing inference over time. The core datum for a DBN is the observed state depending on Δt, whereas the core datum for a CTBN is the observed transition. A DBN can be interpreted as a model of a process as the learned model can be sensitive to the chosen time interval, while a CTBN models the process, see Nodelman (2007).

The computational effort required for learning and inference was studied with respect to the number of trajectories belonging to the TFX EUR/USD dataset. To avoid implementation bias, the same development environment, the same data, configuration, and machine power for both classifiers were used, see Keogh and Kasetty (2002). As reported in Table 3, the continuous time classifier is extremely efficient when compared to the dynamic classifier. For example, to train the continuous time classifier with 1,000 trajectories, it requires about 2 mins against 17 mins of dynamic classifier with $\Delta t = 60$ secs. Moreover, the time required for inference is 182 times less then that required by its discrete counterpart with $\Delta t = 10$ secs.

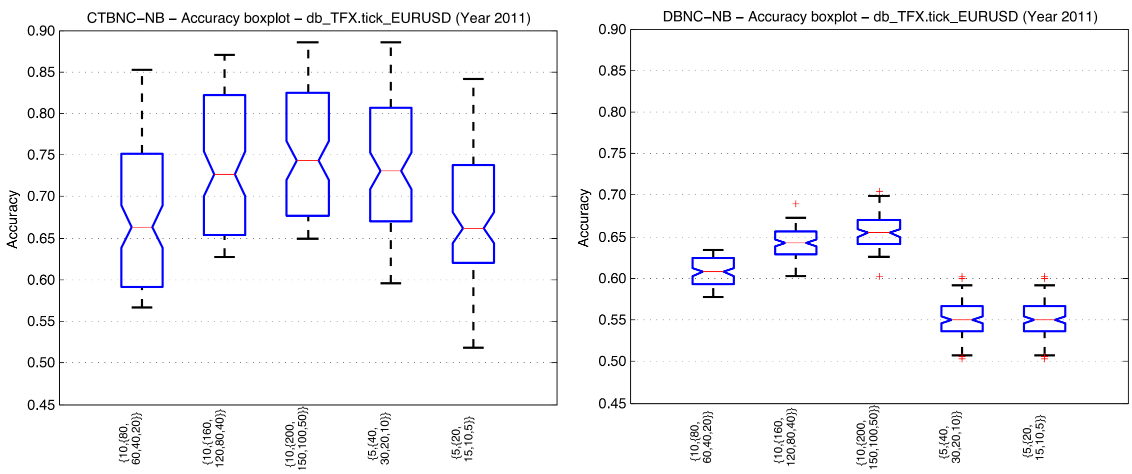
Table 3. Learning and inference time comparison (in secs) between continuous and dynamic classifiers with different Δt on the first 1,000 trajectories with step 100 of TFX EUR/USD dataset.

	Learning			Inference		
Model	**Slope**	**R^2**	**Speedup**	**Slope**	**R^2**	**Speedup**
CTBNC-NB	0.1185	0.9637	1.0000	3.6872	0.9906	1.0000
DBNC-NB $\Delta t = 60$s	0.9979	0.9910	8.4242	13.6014	0.9331	3.6888
DBNC-NB $\Delta t = 30$s	2.1015	0.9950	17.7412	57.5217	0.9011	15.6003
DBNC-NB $\Delta t = 10$s	6.6612	0.9885	56.2363	672.9401	0.9028	182.5066

4.3. Configurations analysis

Selecting the optimal configuration $\mathcal{C}^* \in \mathcal{C} = \{\upsilon, \mathcal{M}\}$, i.e., the configuration which achieves the best prediction accuracy, is prohibitive by the huge amount of possible configurations. Therefore, the performance achieved by moving average triggers, which exploit the moving averages most commonly used by traders, have been analyzed and compared.

Performance values achieved on the TFX EUR/USD dataset by continuous time and dynamic classifiers, for five configurations, are depicted in Figures 6(a) and 6(b). Accuracy values are computed by training a classifier on 1,000 contiguous trajectories and by subsequently testing it on the following 1,000 contiguous trajectories, i.e., by using the rolling window schema. It is worthwhile to mention that the accuracy achieved by the continuous time classifier is well above the one achieved by its discrete counterpart ($\Delta t = 60$ secs) even if the dispersion is higher.



(a) (b)

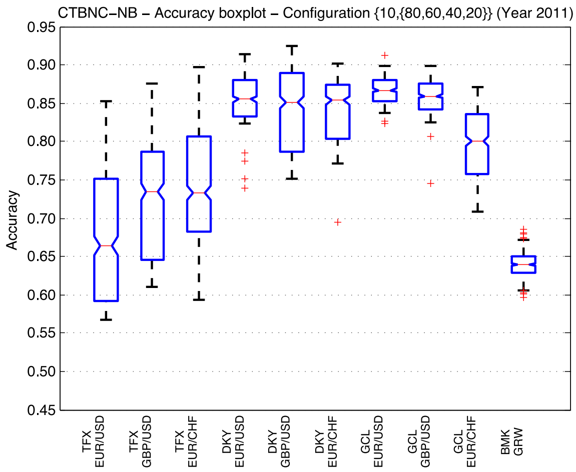
Figure 6. Forecasting accuracy comparison across five configurations for the TFX EUR/USD spot rate for (a) continuous time classifier, and (b) dynamic classifier with $\Delta t = 60$ secs.

4.4. Performance analysis

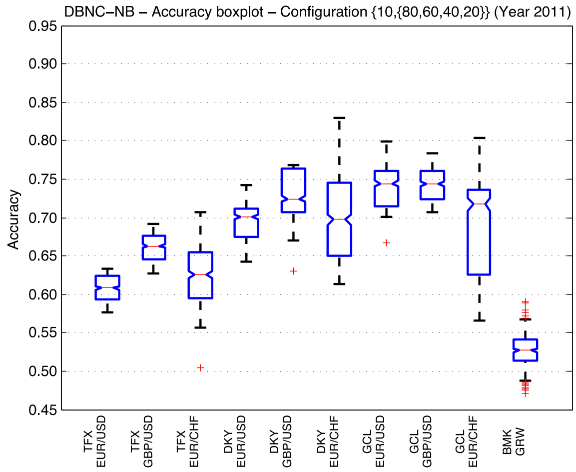
The results obtained from configuration analysis suggested the configuration $\mathcal{C} = \{\upsilon, \mathcal{M}\}$ with $\upsilon = 10$ and $\mathcal{M} = \{80, 60, 40, 20\}$ to be further studied for performance analysis using different datasets and currency pairs. This analysis includes the basic prediction accuracy assessment, the calculation of specific performance measures related to the confusion matrices (such as precision, recall, and F-measure), the Brier score evaluation, the ROC curve construction, the area under ROC curve (AUC) computation, and the out of sample error analysis.

The first type of performance analysis is related to prediction accuracy. It has been accomplished by using the rolling windows schema used for configuration analysis section. Multiple rounds have been performed by using different contiguous partitions of the dataset. Validation results, averaged over multiple rounds, are summarized in Figures 7(a) and 7(b) by means of boxplots. The empirical findings which emerge from this analysis can be summarized as follows:

(i) the accuracy values achieved by continuous time and dynamic classifiers are well above the performance value achieved by random guess;
(ii) the accuracy values achieved by the continuous time classifier are greater than the accuracy values achieved by the dynamic classifier on all datasets;
(iii) the dispersion of the accuracy achieved by the continuous time classifier is greater than the dispersion of the accuracy achieved by the dynamic classifier. This could be due to the smoothing effect induced by the time discretization process associated with DBNs.



(a)



(b)

Figure 7. Forecasting accuracy comparison using the configuration $\{10, \{80, 60, 40, 20\}\}$ over all datasets for (a) continuous time classifier, and (b) dynamic classifier with $\Delta t = 60$ secs.

Performance analysis has been extended through the inspection of the confusion matrices for both classes. The results obtained for class $Y = 1$ and for class $Y = 2$ are summarized respectively in Table 4 and Table 5. This analysis allowed to discover the following findings:

(i) the performance values of precision, recall, and F-measure achieved by the continuous time classifier are better than those achieved by the dynamic classifier for all datasets;
(ii) Brier score values, that measure the average squared deviation between predicted probabilities and outcomes, are very small: 0.17 on average for the continuous time classifier and 0.21 on average for the dynamic classifier;
(iii) AUC values are all above the random guess (i.e., 0.5) for both classifiers.

Table 4. Micro averaging performances by rolling window schema with $\{10, \{80, 60, 40, 20\}\}$ for $Y = 1$.

Classifier	Dataset	Pair	Accuracy	Precision	Recall	F-measure	Brier	AUC
CTBNC-NB	TFX	EUR/USD	0.6815	0.6838	0.6774	0.6806	0.2799	0.7266
		GBP/USD	0.7239	0.7163	0.7411	0.7285	0.2383	0.7649
		EUR/CHF	0.7421	0.7682	0.6928	0.7286	0.2184	0.7775
	DKY	EUR/USD	0.8479	0.8489	0.8464	0.8477	0.1285	0.8927
		GBP/USD	0.8460	0.8597	0.8259	0.8424	0.1290	0.8921
		EUR/CHF	0.8394	0.8464	0.8280	0.8371	0.1333	0.8863
	GCL	EUR/USD	0.8663	0.8671	0.8649	0.8660	0.1071	0.9226
		GBP/USD	0.8537	0.8477	0.8616	0.8546	0.1185	0.9043
		EUR/CHF	0.7945	0.7952	0.7915	0.7934	0.1540	0.8546
	BMK	GRW	0.6393	0.6394	0.6397	0.6396	0.2415	0.7431
DBNC-NB	TFX	EUR/USD	0.6153	0.6294	0.5642	0.5950	0.2546	0.6659
		GBP/USD	0.6661	0.6825	0.6205	0.6500	0.2280	0.7348
		EUR/CHF	0.6294	0.6701	0.5083	0.5781	0.2437	0.6881
	DKY	EUR/USD	0.7040	0.7298	0.6480	0.6865	0.2013	0.7807
		GBP/USD	0.7319	0.7547	0.6847	0.7180	0.1860	0.8097
		EUR/CHF	0.7054	0.7549	0.6052	0.6718	0.1964	0.7851
	GCL	EUR/USD	0.7461	0.7744	0.6936	0.7318	0.1712	0.8305
		GBP/USD	0.7448	0.7671	0.7015	0.7329	0.1721	0.8297
		EUR/CHF	0.6945	0.7313	0.6118	0.6663	0.1983	0.7730
	BMK	GRW	0.5468	0.6425	0.2114	0.3181	0.2495	0.5830

Table 5. Micro averaging performances by rolling window schema with $\{10, \{80, 60, 40, 20\}\}$ for $Y = 2$.

Classifier	Dataset	Pair	Accuracy	Precision	Recall	F-measure	Brier	AUC
CTBNC-NB	TFX	EUR/USD	0.6815	0.6793	0.6856	0.6824	0.2799	0.7266
		GBP/USD	0.7239	0.7320	0.7067	0.7191	0.2383	0.7649
		EUR/CHF	0.7421	0.7207	0.7914	0.7544	0.2184	0.7775
	DKY	EUR/USD	0.8479	0.8469	0.8493	0.8481	0.1285	0.8927
		GBP/USD	0.8460	0.8334	0.8660	0.8494	0.1290	0.8921
		EUR/CHF	0.8394	0.8328	0.8508	0.8417	0.1333	0.8863
	GCL	EUR/USD	0.8663	0.8656	0.8678	0.8667	0.1071	0.9226
		GBP/USD	0.8537	0.8599	0.8458	0.8528	0.1185	0.9043
		EUR/CHF	0.7945	0.7938	0.7974	0.7956	0.1540	0.8546
	BMK	GRW	0.6393	0.6393	0.6389	0.6391	0.2415	0.7431
DBNC-NB	TFX	EUR/USD	0.6153	0.6038	0.6666	0.6337	0.2546	0.6659
		GBP/USD	0.6661	0.6523	0.7116	0.6807	0.2280	0.7348
		EUR/CHF	0.6294	0.6045	0.7502	0.6695	0.2437	0.6881
	DKY	EUR/USD	0.7040	0.6835	0.7601	0.7197	0.2013	0.7807
		GBP/USD	0.7319	0.7131	0.7788	0.7445	0.1860	0.8097
		EUR/CHF	0.7054	0.6725	0.8049	0.7327	0.1964	0.7851
	GCL	EUR/USD	0.7461	0.7232	0.7985	0.7590	0.1712	0.8305
		GBP/USD	0.7448	0.7260	0.7879	0.7557	0.1721	0.8297
		EUR/CHF	0.6945	0.6681	0.7766	0.7183	0.1983	0.7730
	BMK	GRW	0.5468	0.5279	0.8823	0.6606	0.2495	0.5830

ROC curves associated with continuous time and dynamic classifier have been constructed when predicting the class $Y = 1$ for the EUR/USD datasets, these curves are depicted in Figures 8(a) and 8(b) respectively. The pictures show the fraction of true up movements out of the total number of up movements (true positive rate) against the fraction of false up movements out of the total number of down movements (false positive rate) for different probability thresholds. ROC curves have been computed by vertical averaging, i.e., by taking vertical samples of the

ROC curves for fixed false positive rates and by averaging the corresponding true positive rates, see Fawcett (2006). ROC curves remark the above random predictions for both classifiers with some differences within classifiers and datasets as highlighted by the achieved AUC values.

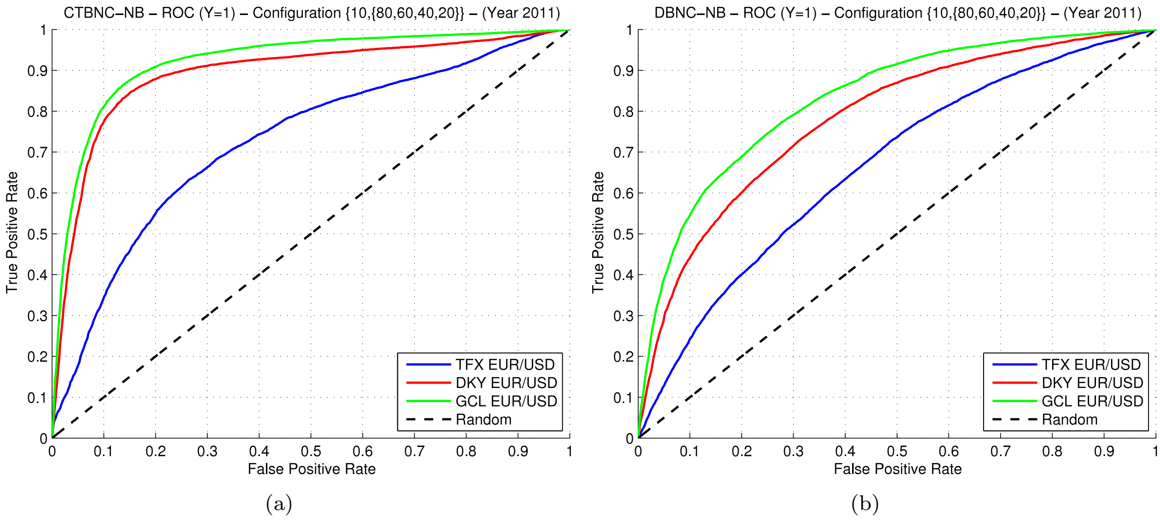


Figure 8. ROC curves using the configuration $\{10, \{80, 60, 40, 20\}\}$ for EUR/USD datasets associated with (a) continuous time, and (b) dynamic classifier with $\Delta t = 60$ secs.

Finally, a numerical experiment similar to the one described in Ding *et al.* (2008) has been performed to compare the behaviors of the out of sample error by increasing the size of the training data, i.e., the number of training trajectories. Therefore, we have used as dataset the first 10,000 trajectories and increasingly large training data sets by 500 trajectories, while the remaining are used as test set. We have measured the classification error using both continuous time and dynamic classifier, Figures 9(a) and 9(b) show the results for the EUR/USD datasets. From this empirical analysis, it is possible to note a decreasing trend of the classification error for the continuous time classifier, while the error rate for the dynamic classifier is almost flat.

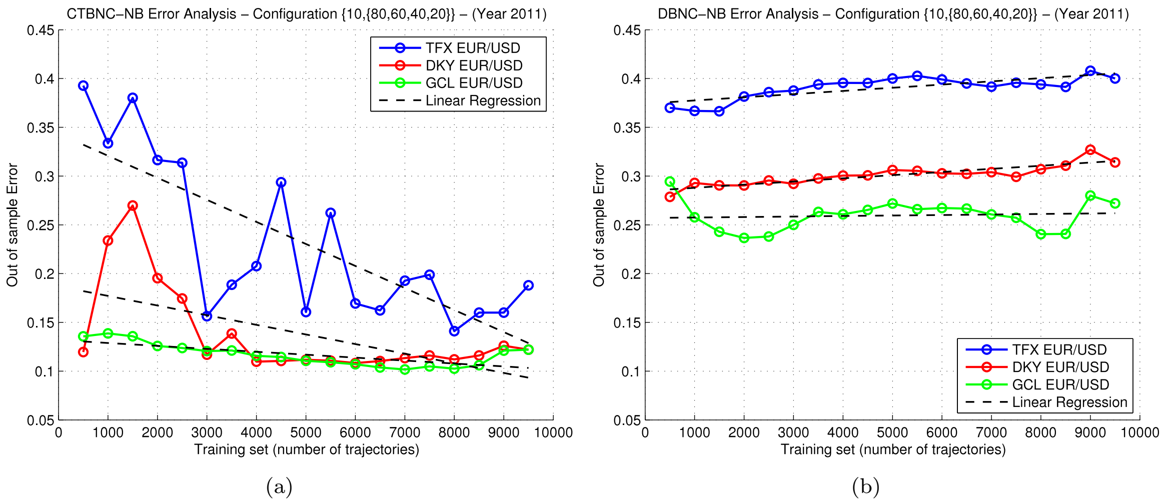


Figure 9. Error analysis by increasing the training data using the configuration $\{10, \{80, 60, 40, 20\}\}$ for EUR/USD datasets associated with (a) continuous time, and (b) dynamic classifier with $\Delta t = 60$ secs.

5. Conclusions and future directions

In recent years, we have seen a rapid growth of high frequency data as result of the advancements in computational and mass storage technology. This has led to an increase of the complexity of financial markets and a demand of new approaches to data processing and statistical analysis.

In this paper, the FX forecasting is formulated as a continuous time classification problem to natively deal with unevenly spaced time data, i.e., the main characteristic of the financial high frequency data. An exact algorithm for inference on continuous time Bayesian network classifiers is presented. A multi-source database of real tick by tick financial data for the major currencies has been designed, implemented, and analyzed. This database has been used to train a continuous time Bayesian network classifier and compare with its discrete time counterpart based on dynamic Bayesian network. Different empirical tests show that continuous time Bayesian network classifiers are more effective and more efficient than dynamic Bayesian network classifiers.

One direction of the future research will focus on improving the performance of continuous time Bayesian network classifiers parameterized with the richer class of phase distributions, see Nodelman *et al.* (2005b), the other one will focus on dynamic decision strategies to support financial decision-making in continuous time. It is worthwhile to mention that the high values of accuracy reached don't imply that trading profitability has been achieved. For this purpose, we highlight at least two relevant aspects. Firstly, you need to consider the issue of early classification and study the trade off between earliness and accuracy, as reported in Xing *et al.* (2009) and Xing *et al.* (2010). Secondly, you need to design a dynamic decision strategy which generates trading signals based on the probability evolutions, see for example Bertsimas *et al.* (2003).

Acknowledgements

The authors are grateful to Michele Fabiano for his valuable contribution to development of the software infrastructure. The authors acknowledge the many helpful suggestions of anonymous referees which helped to improve the paper clarity and quality.

References

Baviera, R., Pasquini, M., Serva, M., Vergni, D. and Vulpiani, A., Antipersistent Markov Behavior in Foreign Exchange Markets. *Physica A*, 2002, **312**, 565–576.

Baviera, R., Vergni, D. and Vulpiani, A., Markovian Approximation in Foreign Exchange Markets. *Physica A*, 2000, **280**, 566–581.

Bellazzi, R. and Zupan, B., Predictive Data Mining in Clinical Medicine: Current Issues and Guidelines. *International Journal of Medical Informatics*, 2008, **77**, 81–97.

Bertsimas, D., Mersereau, A.J. and Patel, N.R., Dynamic Classification of Online Customers. In *Proceedings of the 3rd SIAM International Conference on Data Mining (SMD)*, 2003.

BIS, High-frequency Trading in the Foreign Exchange Market. Bank for International Settlements 2011.

Bollerslev, T., Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 1986, **31**, 307–327.

Boudali, H. and Dugan, J.B., A Continuous Time Bayesian Network Reliability Modeling and Analysis Framework. *IEEE Transactions on Reliability*, 2006, **55**, 86–97.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., *Time Series Analysis: Forecasting and Control* 2008, John Wiley & Sons.

Boyen, X. and Koller, D., Tractable Inference for Complex Stochastic Processes. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI)*, 1998.

Chickering, D.M., Learning Bayesian Networks is NP-Complete. In *Learning from Data: Artificial Intelligence and Statistics V*, pp. 121–130, 1996, Springer-Verlag.

Chinn, M. and Meese, R., Banking on Currency Forecasts: How Predictable is Change in Money?. *Journal of International Economics*, 1995, **38**, 161–178.

Cont, R., Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance*, 2001, **1**, 223–236.

Dacorogna, M.M., Gencay, R., Muller, U., Olsen, R.B. and Olsen, O.V., *An Introduction to High Frequency Finance* 2001, Academic Press, New York.

Dean, T. and Kanazawa, K., A Model for Reasoning about Persistence and Causation. *Computational Intelligence*, 1989, **5**, 142–150.

Dempster, M.A.H. and Jones, C.M., A Real-time Adaptive Trading System Using Genetic Programming. *Quantitative Finance*, 2001, **1**, 397–413.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X. and Keogh, E., Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. In *Proceedings of the Very Large Data Bases Endowment (PVLDB)*, Vol. 1, pp. 1542–1552, 2008.

Dunis, C. and Zhou, B., *Nonlinear Modelling of High Frequency Financial Time Series* 1998, New York, John Wiley.

El-Hay, T., Friedman, N. and Kupferman, R., Gibbs Sampling in Factorized Continuous Time Markov Processes. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 169–178, 2008.

Engle, R. and Russell, J., Autoregressive Conditional Duration: A New Model for Irregularly Spaced Data. *Econometrica*, 1998, **66**, 1127–1162.

Engle, R., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. *Econometrica*, 1982, **50**, 987–1008.

Fama, E.F., The Behaviour of Stock Market Prices. *Journal of Business*, 1965, **38**, 34–105.

Fan, Y. and Shelton, C.R., Learning Continuous-Time Social Network Dynamics. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 161–168, 2009.

Fan, Y., Xu, J. and Shelton, C.R., Importance Sampling for Continuous Time Bayesian Networks. *Journal of Machine Learning Research*, 2010, **11**, 2115–2140.

Fawcett, T., An Introduction to ROC Analysis. *Pattern Recognition Letters*, 2006, **27**, 861–874.

Gatti, E., Luciani, D. and Stella, F., A Continuous Time Bayesian Network Model for Cardiogenic Heart Failure. *Flexible Services and Manufacturing Journal*, 2011, pp. 1–20.

Glattfelder, J.B., Dupuis, A. and Olsen, R., Patterns in High-frequency FX Data: Discovery of 12 Empirical Scaling Laws. *Quantitative Finance*, 2011, **11**, 599–614.

Gradojevic, N., Non-linear, Hybrid Exchange Rate Modeling and Trading Profitability in the Foreign Exchange Market. *Journal of Economic Dynamics & Control*, 2007, **31**, 557–574.

Hautsch, N., *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Lecture Notes in Economics and Mathematical Systems 2004, Springer-Verlag.

Ince, H. and Trafalis, T.B., A Hybrid Model for Exchange Rate Prediction. *Decision Support Systems*, 2006, **42**, 1054–1062.

Jensen, F.V. and Nielsen, T.D., *Bayesian Networks and Decision Graphs* 2007, Springer-Verlag.

Kablan, A. and Ng, W.L., Intraday High-frequency FX Trading with Adaptive Neuro-fuzzy Inference Systems. *International Journal of Financial Markets and Derivatives*, 2011, **1/2**, 68–87.

Kamruzzaman, J. and Sarker, R.A., Forecasting of Currency Exchange Rates Using ANN: A Case Study. In *Proceedings of the International Conference on Neural Networks and Signal Processing*, pp. 793–797, 2003.

Kamruzzaman, J., Sarker, R.A. and Ahmad, I., SVM Based Models for Predicting Foreign Currency Exchange Rates. In *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM)*, pp. 557–560, 2003.

Keogh, E. and Kasetty, S., On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. In *Proceedings of the ACM SIGKDD Conference*, pp. 102–111, 2002.

Kilian, L. and Taylor, M., Why Is It so Difficult to Beat Random Walk Forecast of Exchange Rates?. *Journal of International Economics*, 2003, **60**, 85–107.

Kuan, C.M. and Liu, T., Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks. *Journal of Applied Econometrics*, 1995, **10**, 347–364.

Leung, M.T., Chen, A.S. and Daouk, H., Forecasting Exchange Rates Using General Regression Neural Networks. *Computers & Operations Research*, 2000, **27**, 1093–1110.

Martinez, M. and Sucar, L.E., Learning Dynamic Naive Bayesian Classifiers. In *Proceedings of the 21st International Florida Artificial Intelligence Research Society (FLARIS) Conference*, 2008.

Murphy, K., Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, UC, Berkeley, 2002.

Nodelman, U., Continuous Time Bayesian Networks. PhD thesis, Stanford University, 2007.

Nodelman, U., Koller, D. and Shelton, C.R., Expectation Propagation for Continuous Time Bayesian Networks. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 431–440, 2005a.

Nodelman, U., Shelton, C.R. and Koller, D., Continuous Time Bayesian Networks. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 378–387, 2002.

Nodelman, U., Shelton, C.R. and Koller, D., Learning Continuous Time Bayesian Networks. In *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 451–458, 2003.

Nodelman, U., Shelton, C.R. and Koller, D., Expectation Maximization and Complex Duration Distributions for Continuous Time Bayesian Networks. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 421–430, 2005b.

Nodelman, U. and Horvitz, E., Continuous Time Bayesian Networks for Inferring Users Presence and Activities with Extensions for Modeling and Evaluation. 2003, Technical report, MSR-TR-2003-97, Microsoft Research.

Ohira, T., Sazuka, N., Marumo, K., Shimizu, T., Takayasu, M. and Takayasu, H., Predictability of Currency Market Exchange. *Physica A*, 2002, **308**, 368–374.

Pearl, J., Bayesian Networks: A Model of Self-Activated Memory for Evidential Reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, pp. 329–334, 1985.

Proakis, J.G. and Manolakis, D.K., *Digital Signal Processing: Principles, Algorithms and Applications* 2006, Prentice Hall.

Renner, C., Peinke, J. and Friedrich, R., Evidence of Markov Properties of High Frequency Exchange Rate Data. *Physica A*, 2001, **298**, 499–520.

Saria, S., Nodelman, U. and Koller, D., Reasoning at the Right Time Granularity. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, 2007.

Shmilovici, A., Kahiri, Y., Ben-Gal, I. and Hauser, S., Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm. *Computational Economics*, 2009, **33**, 131–154.

Stella, F. and Amer, Y., Continuous Time Bayesian Network Classifiers. *Journal of Biomedical Informatics*, 2012, **45**, 1108–1119.

Tanaka-Yamawaki, M., Stability of Markovian Structure Observed in High Frequency Foreign Exchange Data. *Annals of the Institute of Statistical Mathematics*, 2003, **55**, 437–446.

Vilasuso, J.R., Forecasting Exchange Rate Volatility. *Economics Letters*, 2002, **76**, 59–64.

Xing, Z., Pei, J. and Keogh, E., A Brief Survey on Sequence Classification. *ACM SIGKDD Explorations Newsletter*, 2010, **12**, 40–48.

Xing, Z., Pei, J. and Yu, P.S., Early Prediction on Time Series: A Nearest Neighbor Approach. In *Proceedings of the 21st International Joint Conference on Artifical intelligence (IJCAI)*, pp. 1297–1302, 2009.

Xu, J. and Shelton, C.R., Continuous Time Bayesian Networks for Host Level Network Intrusion Detection. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, pp. 613–627, 2008.

Yao, J. and Tan, C.L., A Case Study on Using Neural Networks to Perform Technical Forecasting of Forex. *Neurocomputing*, 2000, **34**, 79–98.

Yu, L., Wang, S. and Lai, K.K., *Foreign Exchange Rate Forecasting with Artificial Neural Networks*, International Series in Operations Research & Management Science 2007, Springer-Verlag.

Zhang, Y., Toward a Theory of Marginally Efficient Markets. *Physica A*, 1999, **269**, 30–44.