Comput Econ
DOI 10.1007/s10614-008-9153-3

Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm

Armin Shmilovici · Yoav Kahiri · Irad Ben-Gal · Shmuel Hauser

Accepted: 12 August 2008

Abstract Universal compression algorithms can detect recurring patterns in any type of temporal data—including financial data—for the purpose of compression. The universal algorithms actually find a model of the data that can be used for either compression or prediction. We present a universal Variable Order Markov (VOM) model and use it to test the weak form of the Efficient Market Hypothesis (EMH). The EMH is tested for 12 pairs of international intra-day currency exchange rates for one year series of 1, 5, 10, 15, 20, 25 and 30 min. Statistically significant compression is detected in all the time-series and the high frequency series are also predictable above random. However, the predictability of the model is not sufficient to generate a profitable trading strategy, thus, Forex market turns out to be efficient, at least most of the time.

Keywords Efficient Market Hypothesis · Universal prediction · Forex Intra-day trading · Variable Order Markov

A. Shmilovici (✉)
Department of Information Systems, Ben-Gurion University, P.O. Box 653, Beer-Sheva 84105, Israel
e-mail: armin@bgumail.bgu.ac.il

Y. Kahiri
School of Management, Ben-Gurion University, P.O. Box 653, Beer-Sheva 84105, Israel

I. Ben-Gal
Department of Industrial Engineering, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel
e-mail: bengal@eng.tau.ac.il

S. Hauser
ONO Academic College and School of Management, Ben-Gurion University, P.O. Box 653, Beer-Sheva 84105, Israel
e-mail: shauser@ono.ac.il

JEL Classification G14 · C22 · C53 · C49 · C63

Mathematics Subject Classification (2000) 62P05 · 91B84 · 62M20

1 Introduction and Motivation

In this paper we test the validity of the weak form of the Efficient Market Hypothesis (abbreviated henceforth as EMH) by trying to predict the Forex Intra-day Trading series. Our prediction is based on the Variable Order Markov (VOM) model (Rissanen 1983). The VOM model generalizes a wide variety of finite-memory models, in particular, the conventionally-used Markov Chain models (Ben-Gal et al. 2005). Unlike the Markov Chain model, the order of the VOM model is not necessarily fixed and it is not defined a priori to the learning stage, but rather the order is variable and it depends on the particular observed data-sequences that are found to be statistically significant. In other words, for certain observations the VOM model can represent a "flat" distribution of independent data, while for other observations the model can represents a conditional distribution of a higher order.

The VOM model has been developed as a *universal prediction* model aiming to predict an *arbitrary* sequence of symbols from an unknown stationary stochastic process (Rissanen 1984). Originally, the VOM model was closely related to data-compression applications, in which the prediction has been used to compress[1] an unknown sequence of discrete symbols (e.g., Ziv and Lempel 1978; Feder et al. 1992). In these cases, prediction was obtained by estimating the *conditional probabilities* of symbols given their conditioning (observed) sequences. A known measure of the compression of a sequence is its log-loss score, which is simply the inverse of the log-likelihood of the sequence given a representing model. The optimal average log-loss value represents the highest compression rate[2] of the data that, for long sequences, attain the entropy lower bound (Begleiter et al. 2004). Thus, constructing a data-compression model that minimizes the average log-loss score of a sequence is equivalent to constructing a prediction model that maximizes the likelihood of a sequence. Therefore, in this study the terms "compression" and "prediction" are often considered as equivalent terms. Note that although other universal-compression algorithms can be used for prediction, the used VOM model—a variation of Rissanen's (1983) *context tree*—has been shown to attain the best asymptotic convergence rate for a given sequence (Ziv 2001, 2002; Begleiter et al. 2004).

The significance of universal prediction Models is well recognized in applications such as time-series forecasting (Feder et al. 1992), branch prediction (Federovsky et al. 1998), error corrections of textual data (Vert 2001), Statistical Process Control (Ben-Gal et al. 2003), machine learning and bioinformatics (Shmilovici and Ben-Gal

[1] As an intuitive explanation, the compression of the data is obtained by assigning short codes to frequently re-occurring subsequences, depending on their likelihood to appear in the data.

[2] For long sequences, the universal coding approaches the optimal compression rate—the entropy of the sequence—without prior knowledge on the generating model. The crucial essence in compression is estimating the *conditional probability* for the next outcome given the past observations, so those symbols (and sub-sequences) with high conditional probabilities are assigned short codes.

2007; Zaidenraise et al. 2007). Yet, the powerful insights to be gained from these models in financial econometric series were not well studied (Shmilovici and Ben-Gal 2006).

The first contribution of this paper is the use of the Variable Order Markov (VOM) model, as a universal prediction model of time-series. Given the trend of lowering transaction costs in recent years because of technological improvements in trading, the need for enhanced models for prediction is greater then ever. The second contribution of this paper is the implication of a familiar example—the EMH—to show how universal prediction can be applied in general to financial econometrics and in particular to the Forex Intra-day Trading series.

In this empirical study, the VOM model was used to predict the outcome of the Forex Intra-day trading series. Based on these forecasts, we tested the weak form of the EMH for twelve pairs of international intra-day currency exchange rates, for one-year series that were sampled each 1, 5, 10, 15, 20, 25 and 30 min. This is the most comprehensive Intraday Forex research we know of in terms of the number of investigated data series. To the best of our knowledge, this is the first time where the compressibility-predictability relation (Feder et al. 1992) is implemented to an empirical financial study. Statistically significant compression was detected in all the time-series and predictability above a random prediction was found for all the high-frequency series. On the other hand, the significant predictability of the VOM model was not sufficient to generate a profitable trading strategies and excess return, thus, in the use of a generalized model is consistent with Timmermann and Granger (2004) argument that a good forecasting approach is to conduct a search across many prediction models employed to short data window.

The rest of the paper is organized as follows. Section 2 gives some background on the EMH and the Forex Market. Section 3 introduces the VOM model as a universal prediction model (some illustrative examples of the VOM are given in the Appendix). Section 4 reports on the experiment details and results and Section 5 concludes with a short discussion.

2 Background and Related Work

2.1 The Weak Form Efficient Market Hypothesis

According to Jensen (1978), a market is efficient with respect to an *information set* Ω_t if it is impossible to make economic profits by trading on the basis of Ω_t. Most studies in the literature of financial market returns restrict Ω_t to comprise only past and current asset prices[3] (Timmermann and Granger 2004). We follow this restriction, which is known as the *weak* form of the Efficient Market Hypothesis, and abbreviate it as the EMH. The EMH implies the absence of consistent profitable opportunities, but, it does not rule out all forms of predictability of returns for certain periods. Predictions

[3] The *semi-strong* form of the EMH expands Ω_t to include all publicly available information. Restricting the information set in such a way is designed to rule out private information which may be expensive to acquire and is harder to measure.

of returns invalidate the EMH only once they are significant enough to consistently generate economic profits that cover trade and transaction costs.

The EMH has been investigated in numerous papers (Mills 2002; Lo 2007). Thorough surveys, such as Fama (1991, 1998), Hellstrom and Holmstrom (1998) and Timmermann and Granger (2004) present conflicting conclusions regarding the validity of the EMH. Most of the arguments that justify the EMH argue that different tested models were unable to obtain a statistically-significant prediction of economic series. Albeit, the question regarding the adequacy of these tested prediction models and, as a result, the validation of the EMH, remains unclear in many of these publications.

The EMH is considered as a "backbreaker" for many forecasting methods (Timmermann and Granger 2004). Moreover, Timmermann and Granger (2004) suggest that forecasting models are "self-destructive" in an efficient market, since any advantage due to a new forecasting technique that becomes 'public knowledge' is expected to disappear in future samples. Bellgard (2002), Schwert (2003) and Sullivan et al. (1999) demonstrated a lag between the introduction time of a new forecasting procedure (or the detection of a market anomaly) and the time when this procedure is no longer useful. Thus, it may be easier to detect market inefficiencies with a *new* forecasting algorithm as we attempt to do here. The use of a generalized (universal) prediction model is also consistent with Timmermann and Granger's argument that a good forecasting approach is to conduct a search across many prediction models employed to short data window.

Claims for successful predictions of economic series based on nonlinear models, such as neural networks (e.g., Zhang 1994; Deboeck 1994; Lebaron 1999; Baetaens et al. 1996; Yu et al. 2005; Kaashoek and Van Dijk 2002) do not necessarily contradict the EMH if the trading community is not exposed to these new methods and cannot assimilate them immediately. In this paper, we test the validity of the EMH based on the Variable Order Markov (VOM) model. This model generalizes a wide variety of finite-memory models (Ben-Gal et al. 2005), thus, it is potentially more adequate for such investigation. The use of a generalized model is consistent with Timmermann and Granger (2004) argument that a good forecasting approach is to conduct a search across many prediction models employed to short data window.

2.2 Predicting the Forex Intra-day Trading Series

The currency exchange market—also referred as the Forex market—is the world's largest market (Millman 1995), having a daily trading volume in excess of one trillion dollars. There is no single unified foreign exchange market. Trading is executed via telephone and computer links between dealers in different centers. The main trading centers are located in London, New York, and Tokyo, however, many secondary dealers, such as local banks throughout the world, participate in this on-going trading of 24 h a day (except for weekends). Currencies are traded against one another.[4]

[4] Each pair of currencies thus constitutes an individual product and is traditionally denoted as "XXXYYY" (Neely 1997). For instance, EURUSD is the price of the euro expressed in US dollars, as in 1 euro = 1.2045 dollar.

The bid/ask spread[5] is the difference between the price at which a bank or market maker will sell ("ask", or "offer") and the price at which a market-maker will buy ("bid") from a wholesale customer. Competition has greatly increased with pip spreads shrinking on the majors currencies to as little as 1–1.5 pips (basis points).

The *daily* efficiency of Forex markets has been examined extensively (Chung and Hong 2007 and its references). However, there is not much research on high-frequency financial series, such as the *intraday* Forex trading, since the plentiful trade records (tick by tick) are not uniformly distributed (Dacorogna et al. 2001; Tsay 2002). Classical linear models such as the Purchasing Power Parity model (Bahmani-Oskooee et al. 2006; Choong et al. 2003; Taylor and Taylor 2004) fail to explain the temporal currency fluctuations. It is demonstrated that nonlinear forecasting methods, as the one we propose here, are often better than linear ones for predicting the Forex trading series (Kamruzzaman and Sarker 2004; Boero and Marrocu 2002). Dempster et al. (2001) investigated three currency pairs and concluded that the highest frequency available should be used for forecasting. Goldschmidt and Bellgrad (1999) compared several forecasting models for a single currency pair.

Papageorgiou (1997) conducted a similar research to the one presented here by using a Markov model (which is a particular instance of the VOM) to predict a future ternary-state ('*Increase*', '*Decrease*', and '*Stability*') of a single currency series (CHFUSD). Shmilovici et al. (2003) and Shmilovici and Ben-Gal (2006) applied the VOM model to a simple *binary* series (having either an '*Increase*' or a '*Decrease*' states) to test the compressibility of *daily* stock series. In this paper, we extend the above works by applying the VOM model to numerous ternary-state series in the *intra-day* Forex market.

3 Introduction to the VOM

The VOM model was first suggested by Rissanen (1983) as the "context tree" for data compression purpose. Variants of the model were used in applications such as genetic text modeling (Orlov et al. 2002), classification of protein families (Bejerano and Yona 2001), and statistical process control (Ben-Gal et al. 2003). Ziv (2001) proves that in contrast to other models the convergence of the context tree model to the 'true distribution' model is fast and does not require an infinite sequence length. The VOM model which we used here is a variant of the *Prediction by Partial Match* (PPM) tree, which was found in Begleiter et al. (2004) to outperform other variants of the VOM model. Our version of the VOM model is different in its parameterization, growth, smoothing procedure and pruning stages from the previous versions of the model. These differences might be significant, especially when applying the model to small datasets (Buhlmann and Wyner 1999; Ziv 2001; Begleiter et al. 2004).

[5] This spread is minimal for actively traded pairs of currencies, usually in the order of only 1–3 pips. For example, the bid/ask quote of EURUSD might be 1.2200/1.2203. The minimum trading size for most deals is usually $1,000,000. These spreads might not apply to retail customers at banks, which will routinely mark up the difference to say 1.2100/1.2300 for transfers. Spot prices at market makers vary, but on EUR/USD are usually no more than 5 pips wide (i.e., 0.0005).

The VOM model represents a collection of statistically-significant patterns in time series by a parsed tree. Each node in the tree contains the conditional distribution of symbols given a pattern that is represented by the branch (path) from the root to the node. The branches are labeled by the ternary symbols (‘*Increase*’, ‘*Decrease*’, and ‘*Stability*’ in our study) representing the reoccurring patterns in the series. The branches are not necessarily equal in length—a main difference between this context-specific model and conventional fixed-order Markov models. Therefore, the algorithm is particularly effective for predicting relatively short series, such as the ones available in economic datasets. The algorithm reduces the number a-priori statistical assumptions that are usually required by conventional models on the model structure and the distribution of the data (for example, it does not require to fix the order of the model as required by conventional Markov chains). While the advantages of such a class of models become evident in machine learning, it has not been well analyzed in econometrics.

Next, we shortly present the VOM model that we used in our experiments. We follow the explanations and style in Begleiter et al. (2004) and Ben-Gal et al. (2003) that contain further details on the model and its construction.

Let $\sum$ be a finite alphabet of size $|\sum|$. For example, $\sum = \{$*Stability*, *Increase, Decrease*$\}$ and $|\sum| = 3$. Consider a sequence $\sigma_1^n = \sigma_1\sigma_2 \ldots \sigma_n$ where $\sigma_i \in \Sigma$ is the symbol at position i, with $1 \leq i \leq n$ in the sequence and $\sigma_i\sigma_{i+1}$ is the concatenation of σ_i and σ_{i+1}. Based on a training set σ_1^n, the VOM construction algorithm learns a model $\hat{P}$ that provides a likelihood assignment for any future symbol given its past (previously observed) symbols. Specifically, the VOM generates a conditional probability distribution $\hat{P}(\sigma|s)$ for a symbol $\sigma \in \Sigma$ given a context $s \in \Sigma^*$, where the * sign represents a context of any length, including the empty context (for unconditional distribution of symbols in the root of the tree). VOM models attempt to estimate conditional distributions of the form $\hat{P}(\sigma|s)$, where the context length $|s| \leq D$ varies depending on the available statistics. In contrast, conventional Markov models attempt to estimate these conditional distributions by assuming a fixed contexts’ length $|s| = D$ and, hence, can be considered as special cases of the VOM models. Effectively, for a given training sequence, the VOM models are found to obtain better model parameterization than the fixed-order Markov models (Ben-Gal et al. 2005).

Most VOM learning algorithms include three phases: counting, smoothing, and context selection (Begleiter et al. 2004). In the counting phase, the algorithm constructs an initial context tree T of maximal depth D, which defines an upper bound on the dependence order[6] (i.e., the contexts’ length). The tree has a root node, from which the branches are developed. A branch from the root to a node represents a context that appears in the training set in a reversed order. Thus, an extension of a branch by adding a node represents an extension of a context by an earlier observed symbol. Each node has at most $|\sum|$ children. The tree is not necessarily balanced (i.e., not all the branches need to be of the same length) nor complete (i.e., not all the nodes need to have $|\sum|$ children). The algorithm constructs the tree as follows. It incrementally parses the sequence, one symbol at a time. Each parsed symbol σ_i and its D-sized context, σ_{i-D}^{i-1}, define a potential path in T, which is constructed if it does not yet exist.

6 We use $D \leq \log(n+1)/\log(|\Sigma|)$, where n denotes the lengths of the training sequences. In our experiments $D = 8$.

Note that after parsing the first D symbols, each newly constructed path is of length D. Each node contains $|\sum|$ counters of symbols given the context. The algorithm updates the contexts by the following rule: traverse the tree along the path defined by the context σ_{i-D}^{i-1} and increment the count of the symbol σ_i in all the nodes until the deepest node is reached. The count $N_\sigma(s)$ denotes the number of occurrences where symbol $\sigma \in \Sigma$ follows context s in the training sequence. These counts are used to calculate the probability estimates of the predictive model. Appendix A illustrates an example of the VOM learning algorithm.

The purpose of the second phase of the VOM construction is to use the counts as a basis for generating the predictor $\hat{P}(\sigma|s)$. The following equation is used to smooth the probability to account also for events of zero frequency (Ben-Gal et al. 2005)

$$\hat{P}(\sigma|s) = \frac{\frac{1}{2} + N_\sigma(s)}{\frac{|\Sigma|}{2} + \sum_{\sigma' \in \Sigma} N_{\sigma'}(s)}.$$

The purpose of the third phase of the algorithm is to reduce the model size in order to avoid an over-fitting of the model to the training sequence and enhance memory usage and computation time. Given a long training set, the second phase of the algorithm might results in a context tree having contexts that occurred only a small number of times, and thus, are not statistically significant. If $s = \sigma_k\sigma_{k-1}\ldots\sigma_1$ marks a leaf node, then its "parent node" is its longest suffix $s' = \sigma_{k-1}\ldots\sigma_1$. The algorithm prunes any leaves that do not contribute "additional information" in predicting σ relative to its "parent" node. This additional information is measured by the *Kullback-Leibler* (*KL*) divergence of the distribution of symbols between all leaves of depth k and their parent node of depth $k-1$:

$$KL(\text{leaf}(s)) = \sum_{\sigma' \in \Sigma} \hat{P}(\sigma|s)\log_2\left(\frac{\hat{P}(\sigma|s)}{\hat{P}(\sigma|s')}\right).$$

A leaf is pruned if $KL(\text{leaf}(s)) \le C(|\Sigma|+1)\log_2(n+1)$.[7] Practically, this pruning step keeps the leaf only if its symbols 'distribution is sufficiently different from the symbols' distribution in its parent node. The pruning process can continue recursively to deeper nodes in the context tree.[8]

Once the VOM tree is constructed it can be used to derive the likelihood scores of test sequences, $\hat{P}(\sigma_1^T) = \prod_{i=1}^{T}\hat{P}(\sigma_i|\sigma_1\ldots\sigma_{i-1})$, were σ_0 denotes the empty set. Sequences with similar statistical properties to sequences from the training set (i.e., sequences that belong to the same class of the training dataset) are expected to obtain a higher likelihood score, or equivalently a lower log-loss which is the logarithm of the inverse likelihood, $-\log_2\hat{P}(\sigma_i|\sigma_1\ldots\sigma_{i-1})$. The log-loss is known to be the ideal compression or "code length" of σ_i, in bits per symbol, with respect to the conditional distribution $\hat{P}(\sigma|\sigma_1\ldots\sigma_{i-1})$ (Begleiter et al. 2004). That is, a good compression model that minimizes the log-loss can be used as a good prediction model that maximizes the likelihood and vice-versa (Feder and Merhav 1994).

7 Rissanen (1983) recommends a default of $C = 2$. In our experiments $C = 0.5$ produced better results.

8 Further details about the truncation process and partial leaves are given in Ben-Gal et al. (2005).

The VOM model is used to predict a symbol as follows: Given a sequence s, one can predict the next symbol $\hat{\sigma}$ in the series as the symbol that maximize the likelihood $\hat{\sigma} = \arg\max_{\sigma'}\{\hat{P}(\sigma'|s)\}$, where $\hat{P}(\sigma|s)$ denotes the estimated likelihood of the VOM model.

As noted above, the existence of recurring patterns in a sequence enables data compression. Each branch in the tree represents a recurring pattern (sub-sequence) called a "context". The entire series can then be coded by these contexts. If the length (in bits) of the coded sequence is shorter than the length of the original sequence, then compression is obtained. Reoccurring patterns in the data enhance its predictability—sequences that are highly compressible are easy to predict and, conversely, incompressible sequences are difficult to predict.[9]

The imbalance in the context tree (i.e., having shorter and longer branches) reflects the fact that some patterns do not affect future predictions, while others do. In general, the *deeper* the leaf in the tree, the higher is the "reliability" of its prediction.[10] Since practically most contexts do not point to a deep leaf (and since most trees in a noisy and random sequence are rather *flat*), the predictions are expected to be reliable only for a *fraction* of the time (in this study it means that the market is efficient most of the time). In our experiments, we investigated what happens if we carry out a prediction only for those scenarios in which the likelihood (expected probability) is above a certain threshold. The results are given in the next section.

4 Numerical Experiments

The purpose of this section is to present part of the experiments that tested the EMH for the Forex market by using the VOM model (Kahiri 2004). We start by presenting the used datasets and the pre-processing procedures and follow by describing three sets of experiments: the series compression experiments, the series prediction experiments, and the simulated investment experiments.

4.1 Used Data and Preprocessing Procedures

The data used for this research is the "tick by tick" bid prices of 12 currency pairs for the year 2002.[11] Table 1 lists the currencies and the number of available minutes for each currency.[12] The following and experimental procedures were used:

---

[9] Note that different sequences can have the same compressibility, so the compressibility of a sequence does not uniquely determine its predictability.

[10] Error bounds for several universal predictors are introduced in Feder and Federovski (1999) for binary series and in Hutter (2001) for non-binary series. Merhav and Feder (1998) present further results, such as the relation between the number of leaves in the context tree and the information content in the sequence.

[11] Data received from www.forexite.com

[12] Apparently some currencies are not traded 24 h per day. We ignore the "gaps" in the trade. The number of gaps is small compared to the size of the data, and it can only "weaken" the conclusions.

Table 1 Currency data from 2002 used for this research

Currency pair	No. of minutes
EURUSD	250,608
GBPUSD	241,368
USDJPY	269,154
AUDUSD	150,546
CHFJPY	248,997
EURCAD	251,800
EURCHF	276,609
EURGBP	194,433
EURJPY	278,678
GBPCHF	305,573
GBPJPY	297,378
USDCHF	269,154

- The series were sampled by using 1, 5, 10, 15, 20, 25, or 30 min intervals.
- The series were discretized a priori since the proposed VOM model handles only discrete data.[13] The difference series were quantized to ternary-symbols series, such that an increase (decrease) of 3 pips (i.e., 0.0003) or more was coded by "1" (or respectively by "3"). Any change less than 2 pips[14] was considered insignificant and coded by the stability symbol "2". Obviously, the high-frequency sampling series (1 and 5 min) obtained a different distribution of symbols than the low-frequency sampling series (25 and 30 min). For example, in the latter series the stability symbol "2" is much more frequent.
- Temporal patterns can be of short duration as one can not assume stationarity in data dependence. Accordingly, we constructed a VOM model[15] for each temporal ('sliding') window and used three window lengths[16]: 50, 75, and 100 symbols. Thus, each experiment was conducted $12 \times 7 \times 3 = 252$ times,[17] averaging the statistics of thousands of VOM models per each experiment.
- The implementation procedure was written in the MATLAB script language. An important user-defined parameter in the construction algorithm of the VOM model

13 For example, the EURUSD 5-min series starts as follows (0.8891, 0.8890, 0.8890, 0.8891, 0.8888, 0.8889, 0.8883, 0.8883, 0.8882, 0.8888, 0.8886, 0.8884, 0.8885, 0.8887, 0.8889, 0.8892). Using a ternary quantization, the coded series looks like (N/A, 2, 2, 2, 3, 2, 3, 2, 2, 1, 2, 2, 2, 2, 2, 1).

14 The commission paid by a retail trader is later considered to be 2 pips, thus, the above pips ranges are meaningful from a prediction point of view. It could be smaller for the more volatile currencies.

15 Each VOM model was then used to analyze the data in that temporal window and predict the next data point following that window. A "too short" window may not capture sufficient repetitions of a pattern for the construction of a deep tree, while a "too long" window may capture meaningless transient events that do not affect to the conditional distribution of the predicted data.

16 A window covered between 50 and 100 trading minutes for the 1-min series; and 25–50 consecutive trading hours for the 30 min series. The sliding windows differ by one sample each, and a unique VOM model was constructed for each window.

17 Currency-pair × sampling frequency × sequence length.

is the *truncation coefficient*[18] (denoted by C) that determines the size of the model. Further details of the algorithm are described in Ben-Gal et al. (2003).

4.2 Experiment 1: Compressibility Above Random

In this section we describe the first set experiments that focus on the compressibility of the series, with respect to a compressibility benchmark that is derived from a random series.

4.2.1 Methodological Note

As well known, a random series contains almost no recurring patterns, and therefore, it can not be well compressed[19] (Cover and Thomas 1991; Feder et al. 1992). Accordingly, we used a "random compression" value as a benchmark (lower bound) for the compressibility of series. In particular, we measured the compressibility of different series by the ratios of compressed sequence-windows with respect to a random series. We used this compressibility measure to indirectly analyze the validity of the EMH. The experiments test the following null hypothesis with a 90% confidence level:

H_0: Compressibility of the series is random—accept the EMH.
H_1: Compressibility of the series is above random—reject the EMH.

Note that even for "almost random" sequences, such as a financial series in efficient market, there is some probability that recurring patterns will occur randomly and some (short) sequences will be compressible. In order to minimize the effect of this Type-I statistical error, we followed a simple empirical procedure:

- An independent (uncorrelated) series can be generated from any given marginal distribution (of ternary symbols in our case). This distribution represents[20] in this study the uncorrelated distribution of each currency-pair and frequency sampling. This distribution can be used to derive the benchmark threshold of compressibility for the series.

[18] A selection of a "too small" truncation coefficient produces an over-fitting model, while a selection of a "too large" truncation coefficient produces an under-fitting model. In this work we did not optimize the value of the truncation coefficient for each sequence. Instead, we roughly tested the following three truncation coefficients: 0.25, 0.5, 1.0 for a sample of four currencies (USDJPY, GBPUSD, EURUSD, EURCHF) and computed a measure for the efficiency of the VOM model to correctly predict the symbols "1","3" with respect to the basic trinomial distribution of these symbols in the same series (e.g., if the unconditional frequency of the symbol "1" in the data was 20%, then predicting it correctly in 35% of the cases was considered an improvement over unconditional prediction). It was found that the value of the truncation coefficient that produced VOM models with the best prediction efficiency was $C = 0.5$. Accordingly, this value was used in all the experiments.

[19] Consider for example the large size of a compressed JPEG file for a picture having pixels with a random color distribution.

[20] To avoid excessive computations, we computed a single distribution for each frequency sampling. We use the same distribution for all the currency pairs, which was computed by averaging the distribution over four currency pairs (EURUSD, USDJPY, EURCHF, GBPUSD). For example, the resulting distribution of symbols for a 20-min sampling series was (0.324, 0.344, 0.332).

- Based on the above generated distributions, 150,000 uncorrelated strings of length 50, 75 and 100 symbols were generated for each of the seven sampling intervals.
- A VOM model was then constructed for each uncorrelated sequence, and was used to obtain representing distribution of compressibility[21] values. A percentile of this distribution can be used as a benchmark value to determine if a sequence is indeed uncorrelated.
- We used the 10th percentile of the (empirical) compressibility distribution (i.e., the compression value which is found to be smaller than 90% of the empirical compression values) to define a rejection region[22] and a type I error, as conventionally done in hypothesis testing.
- For each sequence of the 252 combinations of currency-pair, sampling frequency and sequence length, we counted[23] the number of sequences with a compression value below the 10th percentile. This count was used as a statistic in the hypothesis testing and determine if the VOM model compress the sequence above the random threshold.

4.2.2 Results

In 248 cases out of the 252 combinations of currency-pair, sampling frequency and sequence length, the compressibility of the respective VOM model was above the critical value that is expected from a random sequence. Thus, in this statistical hypothesis testing, which is based on compressibility statistic of a random series, one can reject the EMH.

4.2.3 Discussion Note

Previous statistically oriented research works (e.g., Chung and Hong 2007 and related references) managed to detect anomalies in *daily* Forex series. Therefore, it is not surprising that anomalies were detected also in *intraday* series. Yet, it was unexpected to find so many series with non-random patterns (for example in comparison to interday stock series in the study of Shmilovici and Ben-Gal 2006). One possible explanation for this phenomenon is that large buy/sell transactions (that are common in the Forex market) are typically executed as multiple small transactions over a short period of

[21] For example, the mean representation of a random ternary sequence of length 75 with 20-min sampling distribution is approximately $75 \cdot \log_2 3 \approx 119$ bits. If the log-likelihood (to the base 2) of a *specific* random sequence, as computed by its VOM model, is 112 bits, then this specific sequence attains a compression value of $112/119 \approx 0.941$.

[22] For example, for the window of length 75 with a 20-min sampling, 90% of the compression values for the random sequences were above 0.998 that was defined as the critical value for the hypothesis testing. If a specific ternary coded Forex sequence of length 75 (taken from a 20-min sampling series) obtains a compression smaller than the *compression threshold* of 0.998 (e.g., the series obtained a compression of 0.991), then, it is assumed that there is less than 10% probability that the sequence is random.

[23] For example, 99.97% of the sequences for the EURCAD series, using windows of length 75 and a 20-min sampling, were compressed below the compressibility threshold 0.998. This is far more than the expected 10%.

Table 2 Confusion matrix for EURUSD series for a 30-min sampling frequency and window length of 100 symbols

Observed values	Stability	Decrease	Increase
Predicted values			
Stability	0.0712	0.0597	0.0573
Decrease	0.1920	0.2804	0.2903
Increase	0.0129	0.0158	0.0204
Marginal	0.2761	0.3559	0.3680

time, therefore, generating non-random temporal patterns that influence significantly the compressibility statistic.

4.3 Experiment 2: Predictability Above Random

The purpose of this sub-section is to describe how we test the capability of the VOM to predict[24] the next outcome in a financial sequence.

4.3.1 Methodological Note

The statistics of predictions can be well represented in the case of ternary-symbol series by a 3×3 confusion matrix. Let us consider, for example, the confusion matrix in Table 2 for the prediction of the EURUSD series for a 30-min sampling and a window length of 100 symbols. Note that in 7.12% of the sequences, a "stability" symbol is predicted, when the observation was "stability", as defined above. Accordingly, the prediction accuracy of "stability" can be defined by the estimated conditional probability for observing "stability" given a prediction of "stability". This accuracy can be evaluated by applying Bayes rule as follows,

$$\Pr\left(\text{observe ``stability''}\,|\text{predict ``stability''}\right) = \frac{0.0712}{0.0712 + 0.0597 + 0.0573} \approx 37.83\%$$

of the cases. Since stability occurred in 27.61% of the cases, the VOM-based prediction can improve a random guess based on a uniform distribution (with probability of 33.33%) or a naive guess based on the past percentages of "stability" symbols in the data (with probability of 27.61%).

The experiments test the following null hypothesis with a 90% confidence level:

H_0: Prediction of observations in the series is random—accept the EMH.
H_1: Prediction of observations in the series is above random (e.g., better than a naïve prediction, which is based on historical distribution)—reject the EMH.

[24] For example, for a series of length 100, a VOM model can be constructed and used to predict the (discretized) data point at position 101. This prediction can be then compared to the observed data point at position 101 to find if the prediction is correct or erroneous and update the confusion matrix accordingly.

The test statistics was defined as the proportion of correct predictions (called also the "success rate"), which is compared to the expected proportion of correct predictions based on a random Bernoulli process.[25]

4.3.2 Results

- The predictability test was conducted for 252 × 3 times. The EMH for the 1-min sampling series was rejected for *all* the experiments and *all* the prediction possibilities. For the 5-min sampling series the EMH was rejected for *most* of the experiments, and for the 30-min sampling series the EMH was rejected for about half of the experiments. Thus, as expected, it seems that the higher frequency series are more predictable by the VOM model than the lower frequency series.
- A higher success rate (proportion of correct predictions) was obtained when predicting "stability" than the success rate when predicting an increase or a decrease value in the series.

4.4 The Kappa Statistic Test

4.4.1 Methodological Note

Another measure that was used for testing the degree of agreement between forecasted data and observed data is the *Kappa*[26] statistic (Sim and Wright 2005). We used the implementation of Annette[27] (1997) to compute the value of *Kappa* and its confidence interval. The experiments test the following null hypothesis with a 95% confidence level:

H_0: $Kappa = 0$ (prediction is random)—accept the EMH.
H_1: $Kappa > 0$ (prediction is better than random)—reject the EMH.

[25] The distribution of the "success rate" of a symbol in a random series can be approximated by a Gaussian distribution with mean p and standard deviation $S.D. \approx \sqrt{p(1-p)/N}$, where p denotes the probability of the symbol in the series. We use this approximation to compute the single-sided 90% confidence interval, or the acceptance region (with a quantile $Z_{0.90} = 1.285$) for the fraction of correct predictions. For example, consider the EURUSD series with a windows size of 100 data points. It results in a sample size of N=250,608−100=250,508 predictions from windows of length 100. Note from Table 2 that the marginal probability of *increase* is 0.368. The standard deviation of a random process with the same length is equal to $S.D. \approx \sqrt{0.368 \times 0.632/250,508} \approx 0.00096$. This results in a single sided 90% threshold for random success rate which is equal to $0.368 + 1.285 \times 0.00096 = 36.9\%$. Now, note that the observed proportion of correct predictions for this EURUSD series is (see Table 2) is $0.0204/(0.0129+0.0158+0.0204) \approx 41.5\%$, thus, outside the acceptance region—leading to the rejection of the EMH in this case.

[26] The value of *Kappa* is between 0 and 1 and used to measure the degree of compatibility in a confusion matrix. A value of 0 indicates no compatibility, while a value of 1.0 indicates perfect compatibility. It is generally regarded as a more robust measure than the simple success rate since the *kappa* measure takes into account the probability to obtain an agreement as a result of random events.

[27] We used the MedCalc software from www.madlogic.com/madcalc.html.

Table 3 The *Kappa* coefficient for several currency series

Currency	1 Min			10 Min	
	Window	Kappa	95% CI	Kappa	95% CI
EURUSD	100	0.010	−0.015 to 0.04	**0.056**	0.046 to 0.065
	75	0.014	−0.015 to 0.043	**0.052**	0.042 to 0.062
	50	0.021	−0.007 to 0.05	**0.054**	0.045 to 0.064
GBPUSD	100	0.007	−0.019 to 0.033	**0.039**	0.03 to 0.049
	75	0.011	−0.011 to 0.036	**0.027**	0.018 to 0.036
	50	0.017	−0.008 to 0.043	**0.034**	0.024 to 0.043
AUDUSD	100	0.011	−0.038 to 0.059	**0.044**	0.03 to 0.058
	75	0.012	−0.036 to 0.06	**0.040**	0.026 to 0.054
	50	0.021	−0.026 to 0.068	**0.044**	0.031 to 0.058
CHFJPY	100	0.003	−0.03 to 0.037	**0.019**	0.009 to 0.029
	75	0.007	−0.026 to 0.041	**0.015**	0.006 to 0.025
	50	0.015	−0.018 to 0.047	**0.015**	0.006 to 0.025

Positive numbers indicate predictability. Bold numbers indicates statistically significant compatibility (at 95%)

4.4.2 Results

Table 3 illustrates the results of the kappa test for some of the considered series. Note the following facts:

- In *all* the series, the obtained value of *Kappa* was very *small*. The bold styled numbers in Table 3 indicate that the 95% confidence interval for the 10 min series is strictly positive (i.e., do not contain the value of zero), thus, indicating that the compatibility between observed and predicted values in the confusion matrix is statistically significant at a confidence level of 95%.
- For the *all* the 10-min sampling series, and for *some* of the 30-min sampling series, the *Kappa* coefficient is statistically significant.

We can conclude that even when a compatibility is found between the predicted data and the observed data, this difference is very small. Note that unlike the previous hypothesis testing, this test lumps together the three prediction states.

4.4.3 Discussion Note

The above hypothesis testing indicate that the VOM model can be used for prediction with a success rate which is higher than the naïve, historic-based prediction. This is especially true, when using the higher frequency sampling series. This fact is in agreement with previous results that indicate that the dependence between values in financial time series decrease in the lag size between the observations.

Chung and Hong (2007) used a model-free evaluation procedure for daily directional Forex predictability for six currencies and also found significant evidence on directional predictability. It is an empirically-challenging task to find whether the direction of Forex change is predictable. Technical trading rules are built on the fundamental assumption that patterns in the Forex market are regular and repeatable.

Technical trading rules widely used by Forex dealers are heavily based on forecasts of the direction of change (Pring 1991; Taylor and Allen 1992). Our results demonstrate that although prediction above random is potentially available, the Forex series are "almost" random; therefore, it may be very difficult to generate profit on the basis of predictability rules alone.

4.5 Testing the Reliability of the VOM Model Prediction

4.5.1 Methodological Note

As indicated in the section above, the VOM-based prediction provides an improvement over a random or a naïve (historic based) guess. Note that, unfortunately, partially due to the ternary structure of the series—the predictions were correct in less than 50% of the cases. Yet, one might indicate that the VOM model can provide not only a "directional" prediction, but also an estimate for the "reliability" of a prediction.[28] This reliability is based on the *likelihood* to obtain a particular symbol given a context of past observations. For example, if the likelihood to obtain an "increase" symbol is 70% in one leaf and 55% in another leaf, than an "increase" prediction in the first case is considered as more reliable. Accordingly, one might suggest to limit the predictions only to cases were the likelihood (the reliability) is higher than a certain threshold.

Since the data is of high frequency, one may be satisfied with making predictions (and trades) only for a fraction of the time, as long as the likelihood[29] of the predictions is higher than a predefined value, say 50%.

The purpose of this subsection is to test the following hypothesis regarding the effects of the prediction likelihood[30] on the test statistics.

H_0: The prediction likelihood does not impact the prediction success.
H_1: The use of prediction of higher likelihood improves the success rate.

To test this hypothesis, we recomputed a confusion matrix (such as the one in Table 2) but now we considered *only* those predictions with likelihood higher than 0.5 (and respectively higher than 0.7 and 0.8). In practical investments we are mostly interested in predicting an increase (decrease). Thus, we reformulated the above hypothesis testing (with a 90% confidence interval, ignoring predictions of "stability"),

[28] Typically, deeper leaves in the VOM tree (such as branch '311' in Fig. A.1) correspond to predictions with higher reliability since in the construction stage leaving a deeper leaf requires a higher statistical significance. However, most VOM models do not have deep leaves when using conventional values for the truncation coefficient C, and even when a deep leaf is detected, it is used for prediction only for a fraction of the time.

[29] As a typical example, consider the 10-min sampling series. Note that only 2.1% of the predictions in these series obtained a "reliability" higher than 0.7, and only 0.5% of the predictions obtained a "reliability" higher than 0.8.

[30] Prediction reliability (likelihood) higher than 0.5, 0.7 and 0.8.

such that the difference in the proportion of success rate with and without considering the likelihood is significant.[31]

4.5.2 Results

- The null hypothesis H_0 was rejected only in one case of the 30-min sampling series (having a window length of 75). In all other cases, H_0 was not rejected.
- Computing the confusion matrix and the *Kappa* statistic for the samples with high prediction likelihood *reduced* the number of cases where the *Kappa* test indicated a statistically significant forecasting capability.

4.5.3 Discussion Note

The results indicate that using the likelihood as an indicative value of the prediction is not sufficient to improve the accuracy of the VOM-based predictions. Note that this is a surprising result. One possible explanation is that the number of observations with high prediction likelihood is too small for rejecting H_0. A somewhat similar notion to our prediction likelihood can be found in the Game-Theoretic EMH test proposed by Wu and Shafer (2007). The authors prove that "above random predictability" is not sufficient to reject the EMH and a very high degree of reliability is required for a profitable trading strategy

4.6 Testing a Trading Strategy

In the above sections we found that there are some Forex series that are predictable "above random" when using the proposed VOM model. Recall that the weak form EMH implies the absence of consistent profitable opportunities, but it does not rule out all forms of predictability. Predictable patterns invalidate the EMH only if they produce excess returns that are consistently large enough to cover for the transaction costs that are usually associated with such trading. Accordingly, the purpose of the next sub-section is to test weather the VOM model, in its current form, is powerful enough for providing a profitable trading strategy that can cover the associated transaction costs.

4.6.1 Methodological Note

Note that for the experiments in the previous sub-sections we discretized the Forex series to three levels (three symbols), while the difference between an increase (or a decrease) and stability was determined to be *at least* 3 pips. In practice, it may happen that the prediction of a large increase (i.e., much larger than 3 pips) is more accurate than the prediction of a small increase (e.g., exactly 3 pips). Therefore, even for series

31 In other words, that the probability of correctly predicting an increase with a likelihood threshold of 0.50 is higher than the probability of correctly predicting while ignoring the likelihood. The acceptance region is based on the normal approximation for a random proportion of correct predictions.

Table 4 Summary of the trading strategy (in pips) for some currency series and sampling

	Minutes		
	1	5	10
EURUSD			
Profit	1610/17	9080/4	17810/40
Loss	1563/11	9250/20	17760/40
Commission	3586/34	10610/22	15670/22
Net profit	−3539/−28	−10780/−38	−15620/−22
GBPUSD			
Profit	1947/8	15440/39	32540/51
Loss	1951/4	16270/12	33810/18
Commission	3976/20	15600/32	24840/30
Net profit	−3980/−16	−16430/−5	−26110/3

were the predictions were found to be statistically insignificant, the VOM model may still be useful in devising a trading strategy.

The purpose of this sub-section is to test the following hypothesis:

H_0: No profitable trading strategy is found when using the VOM model—accept the EMH.

H_1: A profitable trading strategy can be found when using the VOM model—reject the EMH.

To test this hypothesis, we conducted a simulation of "opening" and "closing" positions of the following trading strategy[32]: If the VOM model prediction indicates an increase (decrease[33]), open a position and close it immediately one time unit later. We ignored interest rates, and both the profit and the net profit were computed as cumulative pips over 2002. The process was repeated for all 252 Forex series, considering prediction reliabilities (likelihood values) of 0.5, 0.7 and 0.8.

Table 4 below illustrates an example of some of the simulation results for series of length 100. In each box, the number to the left of the "/" sign is the number of pips when the strategy "ignores" the prediction likelihood. The number to the right of the "/" sign is the number of pips when the strategy was conditioned on a prediction likelihood of 0.8 or higher. For each currency, we calculate separately the profit transactions, the loss transactions (in pips), and the commission (twice the number of transactions). The net profit[34] was calculated as follows:

$$Net\_Profit = Profit - Loss - Commission$$

[32] To simplify the computation, we assumed a constant bid-ask spread and used only the bid prices. For example, if the current bid price was 1.0000, and the VOM model predicted an increase, then buy at 1.0000 and sell it one time unit after that. If in the next time unit the price was 1.0051, than the gained profit is 0.0051. The transaction costs were assumed to be 2 pips per transaction. Accordingly, the net profit was 0.0049.

[33] Short selling is permitted.

[34] For example, for the 1 min EURUSD series with a likelihood threshold of 0.8, the trading strategy generated 17 transactions with a cumulative commission of 34 pips. The sum of the profitable transactions was 17 pips and the sum of the loss transactions was 11 pips. When ignoring the transaction that generated neither a loss nor a profit, the obtained net profit was $17 - 11 - 34 = 28$ pips.

4.6.2 Results

Several patterns can be detected from Table 4 (and from other results that are given in Kahiri 2004) as follows:

- The number of cases where the profit was larger than the loss is less than 50%.
- The trading strategy demonstrated a tendency for higher profit for the USDJPY and the EURGBP. Yet, this tendency is not statistically significant.
- Even when the profit was higher than the loss (e.g., in the EURUSD 1-min series), automatic prediction (without considering the likelihood threshold) generated too many transactions—that resulted in a net loss.
- Considering the prediction likelihood in the trading strategy reduced significantly the number of transactions as expected. For a prediction likelihood threshold of 0.8, only 5 cases (out of 252) resulted in a net profit (e.g., GBPUSD, 10 min series, in Table 4). Unfortunately, even in these cases, the net profit was very small. Recall that 1 pips is equal to 0.01%, the net profit generates a few percents at most per year—less than the risk-free interest rate.

Effectively, in none of the tested cases, H_0 could be rejected.

4.6.3 Discussion Note

The statistically significant predictability of the VOM model, as found in Section 4.4 is not sufficient to generate a reliable profitable trading strategy. This result is consistent with previous results in the literature that considered simpler models. For example, in a similar research, Papageorgiou (1997) used a conventional Markov chain model (which is a particular instance of the VOM model) to predict a future ternary state (‘*Increase*’, ‘*Decrease*’, and ‘*Stability*’), for a *single* currency (CHFUSD). The author found that prediction based on a shorter observation window is better, yet, he also failed to construct a profitable trading strategy that consistently covers the trading fees.

5 Discussion

This paper combined a practical econometric problem—forecasting a financial time series, and a theoretical econometric problem—testing the Efficient Market Hypothesis. The paper introduced a universal data compression algorithm, which is based on the VOM model, and applied it to test the validity of the EMH for the intraday Forex market. The VOM model detected statistically-significant patterns in the data (12 series of currency pairs) that result in above-random compressibility for most of the Forex series. The VOM model was also used to forecast future trends in the currencies series and produced many examples of forecasts that were found to be statistically significant—this observation was particularly valid for the low-frequency sampling series. On the other hand, a trading strategy based on these forecasts failed to obtain consistent excess returns, even when we ignored the trading fees. In other words, our experiments demonstrate that the obtained *theoretical* market's inefficiencies, as reflected by hypothesis testing with respect to a random predictability, do not necessarily lead to *practical* market inefficiencies and profitable *arbitrage opportunities*.

Based on our experiments, it seems that the intraday Forex market is efficient, at least most of the time. Recalling that the Forex market is considered as the most volatile financial market having continuous trading activities around the globe—the above findings are not surprising.

From its early beginnings, the EMH has woven together *two* theoretical threads: the hypothesis that prices incorporate all relevant information, and the hypothesis that there are no profitable trading strategies (Lo 2007). The experiments in Sections 4.2, 4.3, 4.4 and 4.5 lead to the rejection of the EMH based on the first thread—detecting statistically-significant information patterns in the Forex time-series. The experiments in Sect. 4.6 are based on the second thread and confirm the EMH.

There is evidence that Forex rates exhibit mean reversion toward an equilibrium level and that the degree of mean reversion is stronger when the deviation from the equilibrium is larger. It is conjectured that transaction costs produce a band of inaction in which the big traders allow the Forex to float freely. Consequently, the adjustment process takes place *only* when the perceived misalignment is large enough to cover such costs or the *rates approach the upper or lower limit of the inaction band* (Taylor and Taylor 2004; Chung and Hong 2007). Therefore, the found predictability could be attributed to the intervention of the big traders at specific (yet unknown) threshold values. The EMH is confirmed whenever apparently profitable trading strategies are ruled out by market friction (Malkiel 2003), in other words, some statistically significant anomalies are not economically significant. If the level of transaction costs needed to generate profits from an anomaly (therefore, eliminating it) is far below the level that actually exists in the market, it could explain why a reasonably efficient market allows the anomaly to exist (Wu and Shafer 2007).

5.1 Limitations in the Current Research

The main limitation of the VOM model is that it ignores the *actual values* of the expected returns. That is, the used version of the algorithm is based on a ternary alphabet, thus, it is limited to the forecasting of either "positive", "negative", or "stable" returns disregarding the different amounts of the expected returns. This limitation addresses a "weaker" form of the EMH. Tino et al. (2000) discussed the relation between the discretization strategy, the sliding window length, and the size of the model. They concluded that "discretization should be viewed as a form of knowledge discovery revealing the critical values in the continuous domain". There are other "algorithmic learning" aspects of the VOM model that could be further optimized (such as the truncation coefficient C, or the window length N) that might increase the model validity.

Another practical deficiency of the universal prediction algorithm—the limited number of prediction instances with a high likelihood—can be ameliorated by implementing the universal prediction algorithm for each series in a *portfolio* of financial assets. The theory of universal portfolios (Cover 1991) analyses an investment strategy when a prediction is available for each series in the portfolio. Preliminary results reported in Alon-Brimer (2002) indicate that such a strategy is, in fact, feasible.

Note that the conventional test for the EMH is largely a "one-shot" game—in the sense that only a *single* model is selected and then tested. Here, for *each* running window we effectively selected the *best* model out of all the possible VOM models. The flexibility of the VOM model that can represent both independent data as well as non-linear trends enhance further such a selection.

Furthermore, a *predictability measure* of a time series (such as the rate of correct predictions or the *Kappa* statistic that were used in this paper), can be regarded as a generic econometric feature that is applicable to the analysis of *any* time series to measure its "closeness" to a random stochastic process. In a financial series a predictability measure is considered a more "direct" measure of randomness than a complexity measure (Chen and Tan 1996, 1999). Econometricians learned similar ideas from the co-integration analysis, while the latter does not automatically provide a measure on the time in disequilibrium.

Appendix A—Examples for the VOM Algorithm

We illustrate the VOM learning algorithm by the following toy example. Consider $\Sigma \equiv \{1, 2, 3\}$, and a training sequence σ_1^{150} composed of 30 consecutive repetitions of the pattern "11123". Figure A.1 presents the resulting VOM tree with a maximum context length (branch depth) of $D = 3$. Only nodes that were traversed at leased once (i.e., at least one of the counts is non-zero) are shown. Each node is labeled by

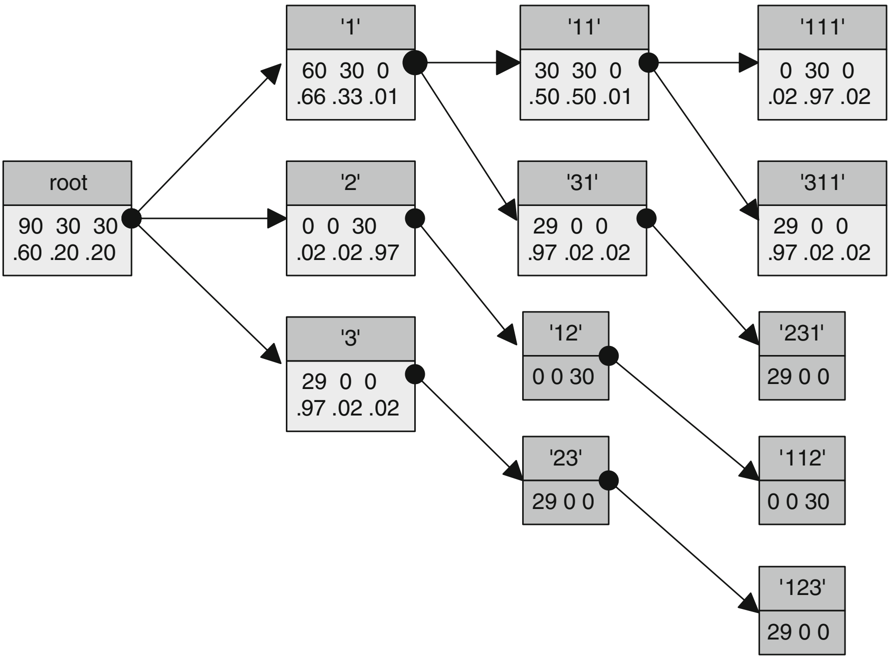

Fig. A.1 A VOM model (tree) generated from 30 consecutive repetitions of the pattern "11123". The first line of numbers below the node's (shaded) label present the three counts ordered with respect to symbols {1,2,3} conditioned on the context that is represented by the label. The second line of numbers below the node label presents the three predictors. Truncated nodes are shaded the context which leads to it. The first line of numbers below each shaded node label present the three counts $N_\sigma(s)$. For example, for the node labeled as "11", the counts are $N_1(11) = 30$, $N_2(11) = 30$ and $N_3(11) = 0$. This means that from the total of $\sum_{\sigma'} N_{\sigma'}(s) = 60$ times that the substring "11" appeared in the training sequence, it was succeeded 30 times by the symbol "1" and 30 times by the symbol "2".

The following equation is used to smooth the probability to account for events of zero frequency,

$$\hat{P}(\sigma|s) = \frac{\frac{1}{2} + N_\sigma(s)}{\frac{|\Sigma|}{2} + \sum\limits_{\sigma' \in \Sigma} N_{\sigma'}(s)}.$$

Based on this equation the second line of numbers below the node label presents the three predictors, $\hat{P}(\sigma|s)$, i.e., the conditional probability of symbols given the context, for example,

$$\hat{P}(1|11) = \frac{\frac{1}{2} + 30}{\frac{3}{2} + 30 + 30 + 0} \cong .50.$$

For illustration purpose Fig. A.1 also shows the shaded nodes that are truncated by the VOM algorithms. For example, leaf node labeled "112" has exactly the same count distribution as its parent node "12", having *KL*(leaf(112)) = 0. Thus, the longer context "112" does not add a new information regarding the conditional distribution of symbols, $\hat{P}(\sigma|112) \approx \hat{P}(\sigma|12)$, and should be truncated by the constructing algorithm of the VOM model. The truncation process is repeated recursively, and the node labeled "12" is truncated too from the same reasons with respect to its parent node "2".

Consider for example the VOM model in Fig. A.1 and a test sequence $\sigma_1^5 = 23112$. The likelihood of this sequence is computed as follows: P(23112) = P(2) × P(3|2) × P(1|23) × P(1|231) × P(2|2311) ≅ P(2) × P(3|2) × P(1|3) × P(1|31) × P(2|311) = 0.20 × 0.97 × 0.97 × 0.97 × 0.02 ≅ 0.00365 (represented respectively, by nodes: "root", "2", "3", "31", and "311" in Fig. A.1). The number of bits required to represent this sequence $\sigma_1^5 = 23112$ is approximately $-\log_2(0.00365) \cong 4.78$ bits. Note that a binary coding of 2 bits per symbol[35] would require 10 bits to code this five symbols sequence. Thus, the VOM model succeeds to compress σ_1^5. Moreover, the longer the sequence, the higher is the probability to obtain a better compression.

For illustration, consider the VOM tree in Fig. A.1 and the prediction of the next symbol in the sequence s = "1121". Given this tree, the longest context from this sequence is '1' (node '1' since there is no '21' node) with the symbol '1' obtaining the maximal likelihood, P(1|1) = 0.66. Therefore, the symbol '1' is selected as the most probable prediction. Suppose that the symbols {"1", "2", "3"} indicate, respectively, a {*Stability, Increase, Decrease*} in consecutive value of a certain foreign exchange ratio. Then, the following prediction rules applies: If there is no information regarding the previous daily return, the a-priori prediction is for a *Stability* (zero return) having

[35] The most efficient representation of uniformly distributed symbols requires $5\log_2(3) \approx 7.92$ bits. In this example, the VOM provides a better compression since it captured well the reoccurring patterns.

the highest probability of 0.6 (see the tree root). A *Stability* is further predicted if the previous observation was either a *Stability* ('1') or a *Decrease* ('3'), or one of combinations represented by the nodes "11", "31", and "311". A *Decrease* is predicted if the previous observation was an *Increase* (node "2"). *Increase* is predicted if the last three consecutive symbols indicated *Stability* (node "111"). The tie in node "11" can be broken arbitrarily.

Using the tree in Fig. A.1 and defining a reliability threshold of 0.70, no prediction is carried out if we only know that the previous observation was *Stability*. The reason is that in node "1" the prediction reliability is 0.66, which is below the defined threshold that is required for making a prediction.

References

Alon-Brimer, Y. (2002). *Measuring the efficiency of the Israeli stock market by the context tree model*. M.Sc. dissertation (in Hebrew), Ben-Gurion University, Israel.

Annette, M. G. (1997). Kappa statistics for multiple raters using categorical classifications. In *Proceedings of the Twenty-Second Annual SAS Users Group International Conference* (available online), San Diego, CA, March.

Baetaens, D. J. E., van den Berg, W. M., & Vaudrey, H. (1996). Market inefficiencies, technical trading and neural networks. In C. Dunis (Ed.), *Forecasting financial markets, financial economics and quantitative analysis* (pp. 245–260). Chichester, England: Wiley.

Bahmani-Oskooee, M., Kutan, A. M., & Zhou, S. (2006). Do real exchange rates follow a non-linear mean reverting process in developing countries. EMG Working Paper Series, WP-EMG-02-2006.

Begleiter, R., El-Yaniv, R., & Yona, G. (2004). On prediction using variable order Markov models. *Journal of Artificial Intelligence, 22*, 385–421.

Bejerano, G., & Yona, G. (2001). Variations on probabilistic suffix trees: Statistical modeling and prediction of protein families. *Bioinformatics, 17*, 23–43. doi:10.1093/bioinformatics/17.1.23.

Bellgard, C. D. (2002). *Diffusion of forecasting technology: Nonlinearity and the information advantage effect*. Ph.D. dissertation, The University of Western Australia.

Ben-Gal, I., Morag, G., & Shmilovici, A. (2003). CSPC: A monitoring procedure for state dependent processes. *Technometrics, 45*(4), 293–311. doi:10.1198/004017003000000122.

Ben-Gal, I., Shani, A., Gohr, A., Grau, J., Arbiv, S., Shmilovici, A., et al. (2005). Identification of transcription factor binding sites with variable-order Bayesian networks. *Bioinformatics, 21*(11), 2657–2666. doi:10.1093/bioinformatics/bti410.

Boero, G., & Marrocu, E. (2002). The performance of non linear exchange rate models: A forecasting comparison. *Journal of Forecasting, 21*, 513–542. doi:10.1002/for.837.

Buhlmann, P., & Wyner, A. J. (1999). Variable length Markov chains. *Annals of Statistics, 27*(2), 480–513. doi:10.1214/aos/1018031204.

Chen, S.-H., & Tan, C.-W. (1996). Measuring randomness by Rissanen's stochastic complexity: Applications to the financial data. In D. L. Dowe, K. B. Korb, & J. J. Oliver (Eds.), *Information, statistics and induction in science*. Singapore: World Scientific.

Chen, S.-H., & Tan, C.-W. (1999). Estimating the complexity function of financial time series: An estimation based on predictive stochastic complexity. In *Proceedings of the 5th International Conference of the Society for Computational Economics*, Boston, June 24–26. Also published in *Journal of Management and Economics, 3*(3) (an electronic journal).

Choong, C. K., Poon, W. C., Habibullah, M. S., & Yusop, Z. (2003). *The validity of PPP theory in asean five: Another look on cointegration and panel data analysis*. Faculty of Accountancy and Management, University Tunku Abdul Rahman.

Chung, J., & Hong, Y. (2007). Model-free evaluation of directional predictability in foreign exchange markets. *Journal of Applied Econometrics, 22*, 855–889. doi:10.1002/jae.965.

Cover, T. M. (1991). Universal portfolios. *Mathematical Finance, 1*(1), 1–29. doi:10.1111/j.1467-9965.1991.tb00002.x.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New-York: Wiley.

Dacorogna, M., Gencay, R., Muller, U., Olsen, R., & Pictet, O. (2001). *An introduction to high-frequency finance*. New York: Academic Press.

Deboeck, G. (1994). *Trading on the edge: Neural, genetic, and fuzzy systems for chaotic financial markets*. New York: Wiley.

Dempster, M. A. H., Payne, T. W., Romahi, Y., & Thompson, G. W. P. (2001). Computational learning techniques for intraday FX trading using popular technical indicators. *IEEE Transactions on Neural Networks, 12*(4), 744–754. doi:10.1109/72.935088.

Fama, E. F. (1991). Efficient capital markets: II. *The Journal of Finance, 46*, 1575–1611. doi:10.2307/2328565.

Fama, E. F. (1998). Market efficiency, long term returns, and behavioral finance. *Journal of Financial Economics, 49*, 283–306. doi:10.1016/S0304-405X(98)00026-9.

Feder, M., & Federovski, E. (1999), Universal prediction with finite memory. IT Workshop on Detection, Estimation, Classification and Imaging, Santa Fe, NM, USA, 24–26 February.

Feder, M., & Merhav, N. (1994). Relations between entropy and error probability. *IEEE Transactions on Information Theory, 38*(1), 259–266. doi:10.1109/18.272494.

Feder, M., Merhav, N., & Gutman, M. (1992). Universal prediction of individual sequences. *IEEE Transactions on Information Theory, 38*(4), 1258–1270. doi:10.1109/18.144706.

Federovsky, E., Feder, M., & Weiss, S. (1998). Branch prediction based on universal data compression algorithm. *ACM SIGARCH Computer Architecture News, 26*(3), 62–72.

Goldschmidt, P., & Bellgrad, C. (1999). *Forecasting foreign exchange rates: Random walk hypothesis, linearity and data freqency*. Department of Information Management and Marketing, The University of Western Australia.

Hellstrom, T., & Holmstrom, K. (1998). Predicting the stock market. Technical report Ima-TOM-1997-07, Umea University, Sweden. http://citeseer.ist.psu.edu/20821.html.

Hutter, M. (2001). Convergence and error bounds for universal prediction of nonbinary sequences. In: *Proceedings of the 12th European Conference on Machine Learning ECML-2001* (pp. 239–250).

Jensen, M. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics, 6*, 95–101. doi:10.1016/0304-405X(78)90025-9.

Kaashoek, J. F., & Van Dijk, H. K. (2002). Neural network pruning applied to real exchange rate analysis. *Journal of Forecasting, 21*, 559–577. doi:10.1002/for.835.

Kahiri, Y. (2004). *Measuring the efficiency of the Forex market via the context tree model*. M.Sc. dissertation (in Hebrew), Ben-Gurion University, Israel.

Kamruzzaman, J., & Sarker, R. A. (2004). ANN based forecasting of foreign currency exchange rates. *Neural Information Processing—Letters and Reviews, 3*(2), 49–58.

Lebaron, B. (1999). Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics, 49*, 125–143.

Lo, A. W. (2007). Efficient markets hypothesis. In L. Blume & S. Durlauf (Eds.), *The new palgrave: A dictionary of economics* (2nd ed.). London: Palgrave MacMillan.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives, 17*, 59–82. doi:10.1257/089533003321164958.

Merhav, N., & Feder, M. (1998). Universal prediction. *IEEE Transactions on Information Theory, IT-44*, 2124–2147. doi:10.1109/18.720534.

Millman, G. J. (1995). *Around the world on a trillion dollars a day*. New York: Bantam Press.

Mills, T. (Ed.). (2002). *Forecasting financial markets*. UK, Cheltenham: Edward Elgar.

Neely, C. J. (1997). *Technical analysis in the foreign exchange market: A Layman's guide*. Federal Reserve Bank of St. Louis.

Orlov, Y. L., Filippov, V. P., Potapov, V. N., & Kolchanov, N. A. (2002). Construction of stochastic context trees for genetic texts. *In Silico Biology, 2*(3), 233–247.

Papageorgiou, C. P. (1997). High frequency time series analysis and prediction using Markov models. In *Proceedings of the conference on Computational Intelligence for Financial Engineering IEEE/IAFE* (pp. 182–188), 23–25 March, New York City, NY.

Pring, M. J. (1991). *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points*. New-York: McGraw-Hill.

Rissanen, J. (1983). A universal data compression system. *IEEE Transactions on Information Theory, 29*(5), 656–664. doi:10.1109/TIT.1983.1056741.

Rissanen, J. (1984). Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory, 30*(4), 629–636. doi:10.1109/TIT.1984.1056936.

Schwert, G. (2003). Anomalies and market efficiency (Chap. 15). In G. M. Constantinides, M. Harris, & R. Stulz (Eds.), *Handbook of economics of finance* (pp. 939–960). Amsterdam: Elsevier.

Shmilovici, A., Alon-Brimer, Y., & Hauser, S. (2003). Using a stochastic complexity measure to check the efficient market hypothesis. *Computational Economics, 22*(3), 273–284. doi:10.1023/A:1026198216929.

Shmilovici, A., & Ben-Gal, I. (2006). Assessing the efficient market hypothesis in stock exchange markets via a universal prediction statistic. Unpublished technical report.

Shmilovici, A., & Ben-Gal, I. (2007). Using a VOM model for reconstructing potential coding regions in EST sequences. *Computational Statistics, 22*(1), 49–69. doi:10.1007/s00180-007-0021-8.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*, 257–268.

Sullivan, R., Timmerman, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *The Journal of Finance, 54*, 1647–1692. doi:10.1111/0022-1082.00163.

Taylor, M. P., & Allen, H. (1992). The use of technical analysis in the foreign exchange market. *Journal of International Money and Finance, 11*, 304–314. doi:10.1016/0261-5606(92)90048-3.

Taylor, A. M., & Taylor, M. P. (2004). The purchasing power parity debate. *The Journal of Economic Perspectives, 18*, 135–158. doi:10.1257/0895330042632744.

Timmermann, A., & Granger, C. W. J. (2004). Efficient market hypothesis and forecasting. *International Journal of Forecasting, 20*(1), 15–27. doi:10.1016/S0169-2070(03)00012-8.

Tino, P., Schittenkopf, C., & Dorffner, G. (2000). A symbolics dynamics approach to volatility prediction. In Y. S. Abu-Mustafa & B. LeBaron (Eds.), *Proceedings of the 6th International Conference on Computational Finance* (pp. 137–151), January 6–8 1999. Cambridge, MA: Leonard N. Stern School of Bussiness, MIT Press.

Tsay, R. (2002). *Analysis of financial time series*. New York: Wiley.

Vert, J. P. (2001). Adaptive context trees and text clustering. *IEEE Transactions on Information Theory, 47*(5), 1884–1901. doi:10.1109/18.930925.

Wu, W., & Shafer, G. (2007). Testing lead-Lag effects under game-theoretic efficient market hypotheses. http://www.probabilityandfinance.com/articles/23.pdf.

Yu, L., Wang, S., & Lai, K. K. (2005). Adaptive smoothing neural networks in foreign exchange rate forecasting. In V. S. Sunderam et al. (Eds.), *ICCS, LNCS 3516* (pp. 523–530). New York: Springer.

Zaidenraise, K. O. S., Shmilovici, A., & Ben-Gal, I. (2007). Gene-finding with the VOM model. *Journal of Computational Methods in Science and Engineering, 7*(1), 45–54.

Zhang, X. (1994). Non-linear predictive models for intra-day foreign exchange trading. *Intelligent Systems in Accounting, Finance and Management, 3*, 293–302.

Ziv, J. (2001). A universal prediction lemma and applications to universal data compression and prediction. *IEEE Transactions on Information Theory, 47*(4), 1528–1532. doi:10.1109/18.923732.

Ziv, J. (2002). An efficient universal prediction algorithm for unknown sources with limited training data. www.msri.org/publications/ln/msri/2002/infotheory/ziv/1/.

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable rate coding. *IEEE Transactions on Information Theory, 24*, 530–536. doi:10.1109/TIT.1978.1055934.