Studies in Nonlinear Dynamics & Econometrics

Volume 16, *Issue* 5 | 2012 | *Article* 3

Predicting Stock Returns Using a Variable Order Markov Tree Model

Armin Shmilovici* Irad Ben-Gal†

*Ben-Gurion University of the Negev, armin@bgumail.bgu.ac.il
†Tel-Aviv University, bengal@eng.tau.ac.il
DOI: 10.1515/1558-3708.1648

Predicting Stock Returns Using a Variable Order Markov Tree Model

Armin Shmilovici and Irad Ben-Gal

Abstract

The weak form of the Efficient Market Hypothesis (EMH) states that the current market price fully reflects the information of past prices and rules out predictions based on price data alone. In an efficient market, consistent prediction of the next outcome of a financial time series is problematic because there are no reoccurring patterns that can be used for a reliable prediction.

This research offers an alternative test of the weak form of the EMH. It uses a universal prediction algorithm based on the Variable Order Markov tree model to identify re-occurring patterns in the data, constructs explanatory models, and predicts the next time-series outcome. Based on these predictions, it rejects the EMH for certain stock markets while accepting it for other markets.

The weak form of the EMH is tested for four international stock exchanges: the German DAX index; the American Dow-Jones30 index; the Austrian ATX index and the Danish KFX index. The universal prediction algorithm is used with sliding windows of 50, 75, and 100 consecutive daily returns for periods of up to 12 trading years. Statistically significant predictions are detected for 17% to 81% of the ATX, KFX and DJ30 stock series for about 3% to 30% of the trading days. A summary prediction analysis indicates that for a confidence level of 99% the more volatile German (DAX) and American (DJ30) markets are indeed efficient. The algorithm detects periods of potential market inefficiency in the ATX and KFX markets that may be exploited for obtaining excess returns.

1. Introduction

The problem of predicting future values in a time series is related to different applications in various fields. In *econometrics* a relevant problem is predicting future values of a financial time-series, such as daily stock returns, or alternatively, evaluating the efficiency of different markets. In *information theory* a related problem is compressing a data sequence, where a better predictability results in a better *compression* rate of the data. In this paper we apply an information theory and data compression model to an econometrics problem. In particular, we use the *Variable Order Markov tree model* (abbreviated, henceforth, as the *VOM tree*) as a tool to predict the outcome of a financial time series and assess the efficiency of different stock markets.

The VOM tree, known originally as the *context tree* (Rissanen, 1984), has been developed as a *universal prediction* model, aiming to predict an *arbitrary* sequence of data symbols that follow an unknown stationary stochastic process (Cover and Thomas, 1995). The VOM tree has been shown to attain the best asymptotic convergence rate to the optimal prediction (Ziv 2001, 2002). Therefore, the algorithm is particularly effective for predicting relatively short series, such as the ones available in economic data sets.

The VOM tree is based on minimal a-priori statistical assumptions about the prediction function or about the distribution of the data values. It generalizes a wide variety of finite-memory models, such as Markov chain models. Unlike these models, the order of the VOM tree is not fixed and it is not defined a priori to its construction (the model's learning phase). Instead, the structure of the model is a function of the particular observed *patterns* that are found to be statistically significant in the dataset. Intuitively speaking, The VOM tree contains all the significant patterns that are found in a given data sequence. Therefore, when recognizing the beginning of one of these patterns, one can use it to predict the future values of that sequence: a frequently occurring pattern is expected to generate a more reliable prediction than a pattern which was rarely observed. In the framework of financial series, a high prediction rate indicates potential market inefficiency, since when the start of a previously recurred pattern is recognized, the continuation of that pattern can be predicted.

In *data compression* the predictability level of a sequence is directly associated with its randomness level (Cover, 1974). The more random the sequence is, it is said to have a higher (*stochastic*) *complexity*, a lower compression rate and the lower is the probability of a correct prediction of future values. A sequence with a low (stochastic) complexity is one in which recurring patterns can be detected – and thus can be used for compression. A measure for the predictability of a sequence is the *size* of the compressed data, often evaluated by the inverse log-likelihood of the sequence based on the used predictive model[1]. We elaborate a bit more on these measures in appendix A to associate the readers with notions of predictability and compressibility, as used in information theory to which the VOM tree belongs.

The Weak-form Efficient Market Hypothesis (abbreviated henceforth by EMH for simplicity) states that no excess returns can be *consistently* earned by using investment strategies that are based only on historical share prices. The EMH claims that these prices are the best, unbiased estimates of the value of the security that reflect all the available information. Accordingly, it rules out the possibility to *consistently* produce excess returns based on technical analysis alone. New information is discovered and quickly disseminated to reflect a change in the market price. Nonetheless, the EMH does not rule out short time-lags, in which it is possible to identify stocks that are undervalued or overvalued by using new prediction models as we aim to do here.

In this paper we apply the VOM tree to predict four different international stock exchanges: 30 stocks composing of the German DAX index, 30 stocks composing of the American Dow-Jones30 index, 20 stocks of the Austrian ATX index and 16 stocks of the Danish KFX index. The selection of these indices enables to compare financial markets of different volume and type. Our underlying hypothesis, which is supported by the empirical findings, was that in an efficient market a consistent "above random prediction" of a time series is not obtainable. However, in smaller and assumed less-efficient markets, using a new prediction model can potentially lead to "above random" financial gains, identifying periods of potential market inefficiency.

The contribution of this paper is in the use of a universal prediction model, and particularly the VOM tree, for evaluating potential periods of market efficiencies. The use of a universal prediction algorithm for time-series forecasting is well known in information theory. Yet, sporadic publications tested market efficiencies via information theory measures (Chen and Tan, 1996, Chen and Tan, 1999, and Shmilovici *et al*., 2003, 2009).

The paper is organized as follows. A literature review is given in Section 2. Section 3 introduces the VOM-tree algorithm as well as illustrative examples related to a financial time series. Section 4 details the conducted experiments by applying the VOM tree to four different markets. Section 5 outlines the main empirical results. Section 6 concludes with a short discussion.

[1] This measure is also called the log-loss. The optimal average log-loss value represents the highest compression rate 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.

2. Literature Review

Along the years the EMH has been extensively studied by numerous research papers[2]. Thorough surveys, such as the ones by Fama (1991, 1998) and Hellstrom and Holmstrom (1998), present obscure conclusions regarding the validity of the EMH theory in practical settings. Many papers that support the EMH by empirical studies, use specific predictive models that are tested against a null hypothesis that share prices are unpredictable and uncorrelated. These papers often apply statistical tests to show the insignificant predictive power of the used models with respect to the null hypothesis. Yet, the question regarding the adequacy of the assumed prediction models and, as a result, the practical validity of the EMH remains unanswered in such cases. In other words, it remains unclear whether the null hypothesis is accepted due to the validity of the EMH or simply due to the inadequacy of the proposed prediction models. This is particularly true for predictive models that are assumed a priori based on theoretical reasons without having a clear support by the gathered data. In contrast, we claim that the use of predictive models with minimum a-priori assumptions, as proposed in this paper, reduces the risk of rejecting the EMH due to model inadequacy.

Predicting stock prices is generally accepted to be a difficult task. In many cases, stock prices are assumed to follow a random-walk or a Martingale differential process most of the time (e.g., see Fama, 1998). Yet, the weak form of the EMH indicates that there might be short time-windows of market inefficiencies, where prices deviate from their regular behavior, providing opportunities for *new* forecasting techniques. This is the motivation for proposing a forecasting method in this paper. Bellgard (2002), Schwert (2003), and Timmermann and Granger (2004) demonstrated a time-lag between the introduction time of a new forecasting procedure (or a detection of a market anomaly) and the time when this procedure is no longer useful for prediction. The inefficiency time-lag is potentially longer for unique forecasting models since more time is required before its "self destruction". To summarize, successful predictions that are based on new models, such as neural networks, Bayesian networks, decision trees, Game-theoretic approach (Shafer and Vovk, 2001) and in this paper the VOM tree model, do not contradict the EMH: they can be found affective for a short time period, as it takes time for the trading community to assimilate these new exposed methods and eliminate the inefficiency (Tsibouris and Zeidenberg, 1995, Baetaens *et al.*, 1996).

The challenge of predicting sequence of values (symbols) is in the essence of information theory since the early days of the field (Shannon, 1951). Information theory uses statistical prediction to solve problems, such as data compression

[2]On 28/6/11 the SSRN Electronic Paper Collection (ssrn.com) contained 8036 records with the JEL G14 classification

(Cover and Thomas, 1995), gambling strategies (Cover, 1974) and sequential decision making (e.g., Merhav and Feder, 1993). The theory associates the predictability level of a stochastic process to its *stochastic complexity*, a measure that quantifies the amount of information stored in a stochastic sequence (e.g., Rissanen, 1984, 1989, Merhav and Feder, 1998). In other words, the information content in a sequence can be measured by how well it is *compressed*. Based on the *lossless source-coding theorem*, not all sequences are compressible, however, the longer the sequence, the lower is the probability that it is incompressible (Li and Vitanyi, 1997). *Universal compression* methods (Ziv and Lempel, 1978, Feder *et al.*, 1992) have been developed to compress an *arbitrary* sequence of symbols generated by an unknown stochastic process. These methods construct a predictive model of the process, estimating the probabilities of various sequences of symbols. They assign short codes to higher probability (frequently re-occurring) sequences and longer codes to lower-probability sequences and by that compress the data (e.g., jpeg image compression). It is known that for long sequences, the universal coding approaches the optimal compression rate, which is measured by the *entropy* of the sequence, even without having prior knowledge on the generating stochastic process. This is the reason why the compression rate of a stochastic process is closely related to its predictability level. Nonetheless, the significance of the stochastic complexity theory, which was well recognized in many fields, such as data-compression (Rissanen, 1983), machine learning (Weinberger *et al*., 1995), statistical process control (Ben-Gal *et al*., 2003), text clustering (Vert, 2001) and statistics and bioinformatics (Buhlmann and Wyner, 1999, Orlov *et al*. 2002) has not been fully assimilated in financial econometrics. This gap motivates us to apply a universal prediction model, such as the VOM tree that we are using, in an empirical study, where the stochastic processes represent stock prices over time.

The idea that an efficient market should have a high stochastic complexity and vice versa was first studied by Chen and Tan (1996). The authors tested an EMH hypothesis based on the stochastic complexity of financial time-series by using a binary Markov model of order one for prediction. Chen and Tan (1999) studied the effect of the window size on the stochastic complexity measure of different financial series by using ARMA modeling. They concluded that signals in financial series are often brief (e.g., following the news of a fore coming business deal) and therefore, using shorter window size is more advantageous for prediction purpose. The generalized VOM tree model (Ben-Gal *et al.*, 2003) was later used by Shmilovici *et al*. (2003) for the compression of financial time-series. Shmilovici *et al*. (2009) applied the VOM tree to analyze the intra-day FOREX time-series, but failed to generate profitable trading strategies of excess returns. In this paper we follow all these works with the distinction that the VOM tree models are used here for prediction of stock market data and for an empirical comparative test of the EMH among different stock markets.

3. The VOM Tree Model

In this section we describe the VOM tree model, outline its construction algorithm and exemplify how it can be used for prediction. We later use this model for the prediction of daily stock returns. The VOM tree model belongs to a set of prediction models in the field of information theory. In appendix A we sketched some of these known approaches that are used in information theory for the prediction of finite-alphabet sequences.

3.1 Prediction by the VOM tree

The Variable-Order Markov (VOM) tree can be viewed as a data structure which is used to store the (conditional) probability parameters of the different symbols conditioned on their (prefix) contexts (Ben-Gal et al., 2003). In our case, the symbols are binary, representing either a "*Gain*" ('*G*') or a "*Loss*" ('*L*') in a consecutive sequence of daily stock returns. The tree assigns a context (of past gains and losses) for each symbol in the sequence, depending on its position in the tree. It has a root node on top, from which the branches are developed downwards. In the binary case, each node has at most $|A|=2$ descendent nodes with differently labelled edges: either a past "Gain" or a past "Loss". Each node in the VOM tree contains $|A|$ conditional probability parameters of symbols (Gain or Loss) given their context (past Gains and Loses), which is represented by the *reversed* path from that node to the tree root (see Ben-Gal *et al*., 2003). Thus these conditional distributions of symbols depend on contexts of varying lengths.

Consider, for example, the VOM tree in Figure 1 which represents a typical structure based on 50 consecutive daily stock returns. The probability for a gain (loss) is given by the left (right) component in each node respectively. For example, in the tree root (origin node) the unconditional probability for a gain or a loss in this sequence is 0.6 or 0.4 respectively. If the return in the previous day was a 'loss', the same probability distribution can be applied – since there is no descendent node from the root that is labeled by a 'loss'. However, the conditional probability distribution for a gain or a loss, conditioning on a 'gain' in the previous day changes to 0.48 and 0.52 respectively (edge 1). Similarly, if it is known that the last two trade days ended with gains (edge 1 and edge 2), then the conditional probability for a daily loss is 0.75. On the other hand, if it is known that yesterday's trade ended with a gain (edge 1) and the previous day ended with a loss (edge 3), the conditional probability for a daily loss is 0.87. Note that this is the higher probability in the tree that might indicate a more reliable prediction than the Bernoulli based probability of 0.6 in the origin node. The imbalance in the VOM tree reflects the fact that some past scenarios do not affect future prediction, while others do.

In general, the tree edges are numbered and represent past gains (to the left) or past losses (to the right). The lower (deeper) edge in the tree, the longer is the conditioning context that is used. Accordingly, each node contains the conditional probabilities of a possible gain or loss given the previous returns, as indicated by branch to that node.

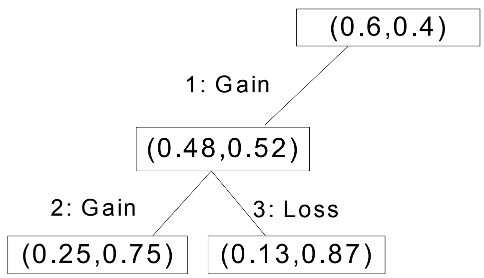

Figure 1: An example of a VOM tree

Note that the VOM 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 $|A|$ descendents). This is the main difference between this variable order model and the conventional fixed-order Markov model that can be represented by a balanced and complete tree. Thus, the VOM tree model is more general than the Markov models since it enables to consider in one hand longer sequences while eliminating other contexts that are found insignificants. This flexibility is a key feature that improves the predictability of the model. Accordingly, we define a variable model order L_j that depends on the preceding symbols to position j. In other words, the order of the Markov model becomes a function of the context at each position.

Once the VOM Tree is constructed the likelihood of a given sequence can be then computed by the product of the conditional probability components, each of which depends on variable order contexts:

$$P\left(x_1^N\right)=\prod_{j=1}^{N} P\left(X_j = x_j \middle| X_{j-L_j}^{j-1} = x_{j-L_j}^{j-1}\right),$$

where the variable order $L(x_{j-1},x_{j-2},\ldots)=L$ is itself a function of the preceding symbols. An optimal value for L_j defines the shortest context for which the transition probability of symbol x_j is practically equal to the transition probability of that symbol given the context of maximal order L (Ben-Gal et al. 2003). Note that for the fixed-order Markov chain $L(x_{j-1},x_{j-2},\ldots)=L$ for all x_j, whereas, for the suggested variable-order Markov model, $L_j \leq L$, implying that some transition probabilities of the Markov chains can be lumped together (e.g., Buhlmann and Wyner, 1999, Orlov and Potapov, 2000).

The likelihood of a sequence given the VOM tree depends on the contexts of a variable-order L_j. For example, in the VOM tree in Figure 1, the likelihood of, say, the sequence "GGGLLG" is computed as follows:

$P(GGGLLG)=P(G)\times P(G|G)\times P(G|GG)\times P(L|GGG)\times P(L|GGGL)\times P(G|GGGLL)$ = $P(G)\times P(G|G)\times P(G|GG)\times P(L|GG)\times P(L)\times P(G)=0.6\times 0.48\times 0.25\times 0.75\times 0.4\times 0.6$.

Using these likelihood computations different predictions can be made.

3.2 Construction of the VOM tree

The construction of the VOM tree contains two stages (for detailed description, see Ben-Gal *et al.*, 2003). In the first stage, the tree is grown from its origin (root) node downwards based on the training sequence X_{-N}^{0}. During this stage, counts in each node are updated to represent the conditional frequency of the symbols given their contexts. The counts denote the number of instances where symbol x_i follows the context $X_{i-1-K_{\max}}^{i-1}$ in the training sequence X_{-N}^{0}. K_{max} is the initial tree depth prior to any pruning, and it is determined by practical memory capacity constraints, as well as available training data. In our case K_{max}=10.

In the second stage, the tree branches are pruned to obtain its variable-lengths structure. The pruning is based on the Kullback-Leibler (KL) divergence for the conditional probabilities of symbols between a descendent node and its parent node. If the KL divergence measure is smaller than a pre-selected pruning threshold, the descendent node is pruned. A small KL divergence implies that there is no significant change in the symbols' distribution when using the reduced order of the model, or in other words, that the larger model order, which is represented by the descendent node, does not add much information and can be pruned without affecting the prediction probability. The pruning level is controlled by the pruning coefficient *C*.

Once the VOM tree is pruned, a pseudo-count is added to all the tree counts to compensate for unobserved subsequences with zero counts in the tree (see Ben-Gal *et al.*, 2003). Finally, the smoothed (normalized) counts in the pruned tree are used to estimate the conditional probability $P(x_i \mid X_0^{i-1})$ for prediction or compression purposes.

The outline of the context-tree algorithm, which we use in this paper, is given in Figure 2 below. The complete details of the algorithm, which has a linear complexity in the sequence length *N*, can be found in Ben-Gal et al. (2003).

The pruned VOM tree model can be represented by the joint distribution of contexts (leaves) and their symbols. The size of the final context tree model is determined by the value of the pruning coefficient *C*. Thus, if *C*=0 the tree is left unpruned and often contains too many parameters (thus often it is over-fitted). Rissanen (1983) recommended a default pruning coefficient value of *C=2* for a good compression rate. Our experience with financial time series (including those analyzed in Shmilovici *et al.*, 2003) indicates that a value of *C=0.5* usually results in better predictions.

***VOM Tree growing stage*:**

***Step* 0.** Start with the root as the initial tree, with all symbol counts equal to zero.

***Step* 1**. *Counter update*: Recursively, having constructed the current tree from the current sequence, read the next symbol x_i in the sequence. Traverse the tree along the path defined by the context X_{-k}^{0} and increment the count of the symbol x_i in all nodes until the deepest node is reached. Each symbol x_i belongs to an alphabet A with cardinality $|A|$

***Step* 2**. *Tree growth*: If the last updated count is at least 1 and the depth of the node is $k<K_{max}$ where $K_{max} \le \log(N+1)/\log(|A|)$, create a new node of depth $k = k+1$ and initialize all symbol counts to zero except for the symbol x_i whose count is set to 1. K_{max} is used to reduce the computation time and the memory requirements during run-time. For a modern desktop computers, $K_{max} =10$ can be well tolerated.

***VOM Tree pruning stage*:**

***Step* 3.** Estimate the KL divergence of the distribution of symbols between a leaf of depth *k(leaf)* and its parent node of depth *k(leaf)-1* :

$$KL(leaf) = \sum_{X_i \in A} Q\left(x_i \middle| X_{i-k(leaf)}^{i-1}\right) \log_2 \left(\frac{Q\left(x_i \middle| X_{i-k(leaf)}^{i-1}\right)}{Q\left(x_i \middle| X_{i-k(leaf)-1}^{i-1}\right)} \right)$$

Repeat for all leaves. $Q\left(x_i \mid X_{i-k}^{i-1}\right)$ is an estimate of $P\left(x_i \mid X_{i-k}^{i-1}\right)$.

***Step* 4.** Prune the leaf if $KL(leaf) \ge C(|A|+1)\log(N+1)$, where the logarithm is to the base 2, and the default value for the pruning coefficient is C=0.5. Practically, this pruning step keeps the leaf only if its symbols' distribution is sufficiently different from the symbols' distribution in its parent node.

***Step* 5.** If all leaves are left unpruned – stop. Otherwise, go back to step 3 and repeat for all the pruned leaves.

Figure 2. Outline of the construction of the VOM tree

Like other machine learning techniques, such as decision trees (that are also pruned based on the divergence of probabilities between a parent node and its descendents), the pruning process is intended to avoid over-fitting (bias) of the model to the training sequence (Ben-Gal et al., 2003). Yet, as demonstrated in Figure 1, longer contexts (deeper leaves in the tree) can indicate a potentially more reliable prediction (concentration of the probability mass in a single symbol instead of the equiprobable symbols in the origin node). Thus, if the pruning process is too aggressive (by selecting a high value for *C*), it removes contexts that correspond with the occurrences of events that generated a small number of counts, smoothes out their predictions and reduces the overall prediction level.

3.3 The Reliability threshold

In long financial series where the long-term probability of a positive (and a negative) daily return fluctuates around *0.5*, obtaining a deep VOM tree leads towards a potentially more reliable prediction. Yet, obtaining a deeper tree is not enough to guarantee a reliable prediction. A *current* daily prediction is expected to be more reliable only if the current context which is used for prediction points to a *deep leaf* in the tree. 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 (alternatively, one can say that such a tree structure reflects the fact that the market is efficient *most* of the time). In our experiments, that are presented in the next section, we decided to carry out a prediction only for instances for which the prediction reliability (probability parameter) is above a certain threshold. For example, referring to the tree in Figure 1 and defining a reliability threshold of 0.65, no prediction can be made if one knows only that yesterday's trade ended with a positive return. The reason is that in the relevant node following edge 1 the prediction reliability is 0.52, which is below the defined reliability threshold.

4. Numerical Experiments

4.1. *Assumptions and preprocessing*

The analysis in this section is based on the following two assumptions. First, that the VOM tree can be used for the prediction of recurring patterns in financial series, if these patterns do exist. Second, that there is a probability that a random series – such as a financial series from an efficient market – may contain some recurring patterns. We measure the predictability of the series by the fraction of correct predictions when using the VOM tree, and compare it to the fraction of correct predictions that is expected in a random sequence. We use this comparison to analyze the efficiencies of several markets. In particular, the performed experiments test the following null hypothesis with a 95% confidence level:

H_0: Prediction of observations in the series is random – the market is efficient.

H_1: Prediction of observations in the series is above random – the market is inefficient.

Before running the test, some preprocessing issues as well as some of the VOM tree parameters must be specified.

- The series' daily returns have to be discretized a priori since the VOM tree handles discrete data. In particular, we test the null hypothesis H_0 based on a binary prediction trend with positive returns (gains) or negative returns (losses). Note that to support a profit-gain policy, a ternary alphabet (e.g., positive, negative, stable) or a higher alphabet is probably preferable. In this paper we start with a simple binary discretization.
- The series length, *N*, has to be specified by the user. The effect of the window size on the predictability measure is itself a potential direction of research (e.g., see Chen and Tan, 1999). It is easier to detect statistically significant patterns in a longer series, yet, the VOM tree is considered efficient enough to operate on short series as well (Ziv 2002). In the following experiments we focus on sliding windows of 50, 75 and 100 consecutive trading days (about 2-5 months). That is, we try to predict the 51^{th}, the 76^{th} and the 101^{th} daily return given the respective window of past observations. Note that our implementation of the VOM tree algorithm is *not* adaptive (e.g., Federovsky *et al.*, 1998). A series that is "too long" might capture a non-stationary change in the trading system or a "trading noise". In these cases, the predictor might generate a lower reliability prediction.
- The VOM tree algorithm that is written in the MATLAB script language must be calibrated – especially the pruning coefficient *C* that determines the required number of repetitions of a pattern in order for it to be represented in the VOM tree. The value of the pruning coefficient depends on the series length, the alphabet size and the series type (see Ben-Gal *et al.*, 2003). Tuning experiments with various values of pruning coefficient in the range $C \in [0.25,...,4.0]$ revealed that the pruning coefficient that yields the highest prediction performance for the binary series of the four different stocks markets is $C=0.5$.

4.2 *Performance Measures: Metrics for Testing*

The first performance measure we used is the Fraction of Correct Predictions Above the Reliability Threshold (FCPART). It is defined as the empirical count of the periods with correct prediction divided by the total number of observed periods. The observed periods are those for which one can rely on the VOM tree leaf to make a prediction above a predefined reliability threshold (such as leaf 2 in Figure 1). For example, let us consider $N=2568$ predictions from sliding windows of length 50 in the ATX-INDEX series in Table 3. In only 288 (11.21%) cases the relevant leaf of the VOM tree had a probability value above a reliability threshold of 0.65 that enabled making a prediction. Of those predictions, only 159 (55.21%) were found to be correct. Therefore, FCPART = 0.5521.

The FCPART is compared against the ratio expected from a random Bernoulli process with parameter $p=0.5$. A normal approximation to the binomial distribution with mean $p=0.5$ and standard deviation S.D. $\approx \sqrt{p(1-p)/N}$ is used to compute the single-sided 95% confidence intervals (with quantile $Z_{1-0.05}=1.645$) for the fraction of correct predictions. For example, let us consider N=2568 predictions from sliding windows of length 50 in the ATX-INDEX series. The standard deviation of the random process with the same length is equal to S.D. $\approx \sqrt{0.5*0.5/2568} \approx 0.00987$. This results in a single sided 95% threshold for random fraction of correct predictions which is equal to 0.5+1.645×0.00987=51.62%. Since the actual fraction of correct predictions in this ATX-INDEX series is 50.86%, H_0 cannot be rejected in this case. However, recall that it was found that for 11.21% of the predictions with a prediction reliability higher than 0.65, the fraction of correct predictions (the FCPART) is 55.21%. Repeating the above analysis, but this time with N=2568×0.1121=288 observations, results in 95% random threshold of 0.5+1.645×0.029463=54.85%. Therefore, in this case, H_0 is rejected with a confidence level of 95%, implying that this prediction is not random. Moreover, note that a prediction success rate above 54% is often considered as satisfying for practical investment (Tsibouris and Zeidenberg, 1995, Baetaens et *al.*, 1996) in the sense of covering (on the average) the "transaction costs" of the prediction.

The second performance measure that we used is the aggregate prediction rate across a group of sequences, attempting to detect random flagging of predictable sequences. When testing multiple hypotheses simultaneously, one has to pay attention to the first type statistical errors. For example, consider the 20 stock series that compose the ATX index: even if all series are random walk series with a confidence level of 95%, still there would be 20×0.05=1 stock on the average for which the random walk hypothesis will be rejected. Therefore, one cannot assure that such a series indicates "above random predictability”. In this context, a multiple testing correction, such as the traditional Bonferroni’s correction or the False Discovery Rate approach (Benjamini and Hochberg, 1995) can be implemented in future research.

4.3 *Testing the DJ, DAX, ATX, and KFX indices*

As indicated in Shmilovici *et al.* (2003), series from different financial markets can demonstrate quite a different behavior. In general, it is expected that high volume markets will behave differently than low volume markets. In this study, we collected data from four different stock markets: one with a large daily trade volume (the American market), one with a medium daily trade volume (the German market) and two with a low daily trade volume (the Austrian and the

Danish markets). During the period[3] between 2/1/90 – 30/06/03 we considered the daily closing values[4] of: i) the 30 stocks that comprise of the German DAX stock index; ii) the 30 stocks that comprise of the American DJ30 stock index; iii) 20 of the stocks that comprise of the Austrian ATX stock index; and iv) 16 of the stocks that comprise of the Danish KFX stock index[5]. The predictions were performed in the following manner:

- The daily returns were computed and each series was coded into a binary series according to the sign of the daily return (zero returns were coded identically to the negative returns due to transaction costs).
- A VOM tree model was constructed for each running-length window of size 50 (and, respectively, of sizes 75 and 100).
- The VOM tree model was then used to predict the trend on the 51th day (and, respectively, on the 76th and the 101th days).
- The predicted trends were compared to the actual trends and statistics were collected in tables 1,2,3,4 for each stock and for each window length. The "TP rate" statistic was collected for all the predictions,
- The same procedure was repeated for the predictions above the following reliability thresholds: 0.60, 0.65 and 0.70. A low reliability threshold such as 0.60 can provide a sufficient number of trading days to implement a trading strategy, however, it is not expected to provide a sufficiently high FCPART (see appendix B). On the other hand, a high reliability threshold, such as 0.70, can provide a high FCPART but not a sufficient number of trading days to implement a profitable trading strategy and reject H_0 with a 95% confidence level.

Figure 3 presents typical histograms of the number of trading days (for which predictions were generated) for two stocks based on different reliability thresholds. The histograms represent two stock series from the KFX market with running windows of length 100. The top figure – generated from the Coloplast B stock trend series – demonstrates a significant number of trading days with prediction reliability above the 0.60 threshold, namely, 39.22%; The bottom figure – generated from the Novo Nordisk B stock trend series – demonstrates a limited number of trading days with prediction reliability above the 0.60 threshold, namely, 12.76%.

[3]Series shorter than two years were rejected. Data in some series is incomplete. We ignored the effect of the few missing values.
[4]Collected from Yahoo! Finance Investing, World Stock Exchanges http://uk.biz.yahoo.com/uk_world.html
[5]Index stock can be very volatile. For example since companies close down. We chose stocks for which most of the data was available.

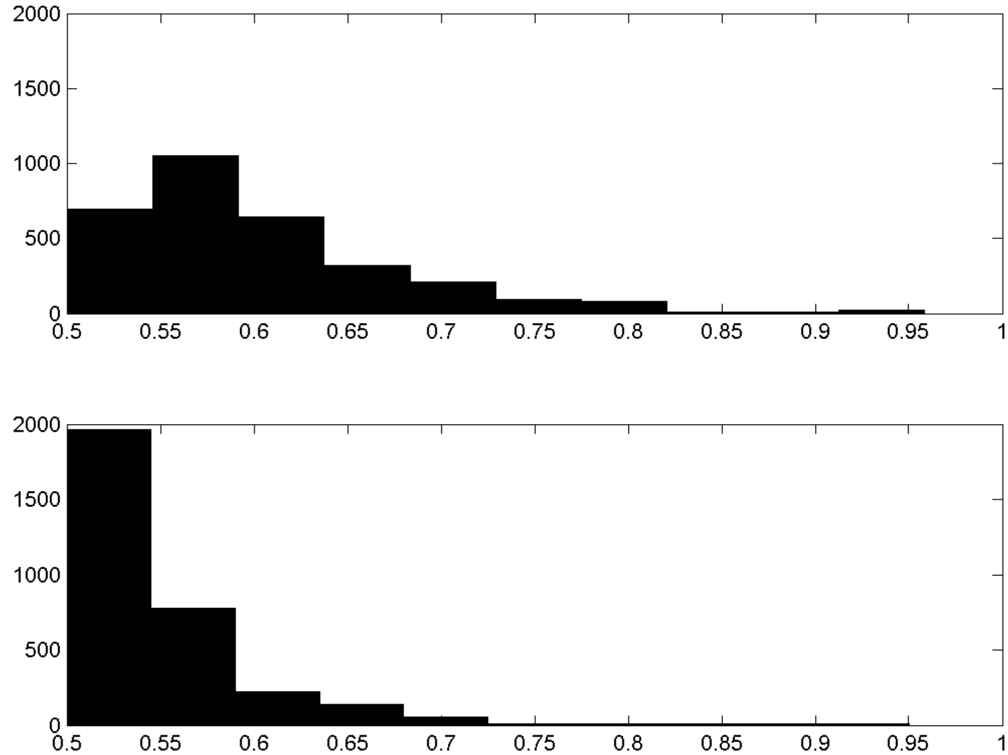

Figure 3: Histograms of the number of trading days (predicted days) for different reliability thresholds. The Series of length 100 were taken from the KFX stock market. Top - the Coloplast B stock trend series; Bottom - the Novo Nordisk B stock trend series

Tables 1, 2, 3 and 4 present, respectively, the results for the stocks composing the KFX, ATX, DAX and DJ30 indices. The *first column* for each stock (in each table, respectively) shows the stock name. The *second column* for each stock, describes the initial day from which the data was collected and the number of predictions made from running windows of length 50 (in order to obtain the number of available predictions for running windows of length 75 and 100, one should reduce this number by 25 and by 50 respectively). The *third column* describes the FCPART for running windows of length 50 (respectively in columns 6 and 9 for sliding windows of lengths 75 and 100). The *third column* presents the FCPART for predictions with a reliability threshold of 0.60, and the percentage of the prediction days that satisfy this reliability condition (respectively the 6th and 8th columns for sliding windows of lengths 75 and 100). The 4th and 5th columns present respectively the FCPART for those predictions with reliability thresholds of 0.65 and 0.70, and the percentage of the prediction days that satisfy these conditions (respectively, in columns 7, 8 and 10, 11 for running windows of lengths 75 and 100). Table entries in **bold** represent series for which H_0 is rejected with a confidence level of 95%, thus, leading to a conclusion that the prediction is not random and rejecting H_0.

Table 1: Predictability measures for 16 of the stocks composing the KFX top 20 index of Kopenhagen, Denmark. (Bold numbers are above the 95% confidence)

Stock Name\ window size		**50**			**75**			**100**		
Name	Start Period (#prediction days for series length 50)	% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction		
		0.60	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**
KFX TOP20 INDEX	26-Jan-93 (2548)	51.93 21.31	**56.33** **8.99**	48.21 2.20	**55.53** **18.99**	54.67 5.95	37.78 1.78	**56.13** **14.05**	51.0 4.0	45.83 0.96
CARLSBERG B	2-Jan-90 (3051)	**55.87** **29.86**	**56.18** **16.45**	**57.99** **7.18**	**58.47** **26.73**	**60.89** **12.59**	**66.46** **5.22**	**58.76** **23.19**	**61.81** **9.60**	**63.16** **3.80**
COLOPLAST B	2-Jan-90 (3186)	**61.12** **38.51**	**64.89** **21.81**	**67.34** **13.94**	**60.81** **37.30**	**64.33** **19.87**	**67.42** **12.62**	**61.54** **39.22**	**62.56** **18.65**	**64.02** **10.46**
DANSKE BANK	2-Jan-90 (3317)	**53.70** **27.28**	**54.89** **14.17**	**57.89** **5.73**	**54.95** **20.57**	**55.79** **10.24**	**59.71** **4.22**	**55.39** **16.74**	**61.09** **7.32**	**61.32** **3.64**
DANISCO	2-Jan-90 (3316)	**53.74** **26.60**	**56.67** **15.59**	50.53 5.67	**57.52** **24.25**	**59.18** **11.09**	55.93 3.59	**57.34** **20.45**	**59.04** **7.62**	46.48 2.17
DSV	2-Jan-90 (2797)	**63.09** **51.91**	**66.47** **30.50**	**74.38** **17.16**	**63.97** **49.46**	**64.81** **28.50**	**74.23** **16.38**	**63.46** **49.51**	**70.55** **22.50**	**75.16** **17.15**
GROUP 4 FALCK	7-Apr-95 (1991)	50.96 18.33	52.05 7.33	52.54 2.96	46.21 4.48	34.09 4.48	35.48 1.58	48.37 7.88	39.29 1.44	45.45 0.57
GN STORE NORD	2-Jan-90 (3214)	**55.89** **34.85**	**55.59** **18.36**	55.14 5.76	**55.78** **31.98**	**57.23** **14.74**	54.70 2.67	**55.85** **29.99**	**61.24** **9.70**	54.29 2.21
ISS	2-Jan-90 (3314)	**53.40** **28.42**	52.20 13.70	52.67 4.53	**54.51** **23.59**	**54.73** **10.28**	55.75 3.98	**54.28** **19.70**	52.36 5.85	54.44 2.76
JYSKE BANK	2-Jan-90 (3299)	**58.10** **46.59**	**60.25** **26.46**	**63.03** **10.82**	**60.11** **45.65**	**62.39** **20.71**	**65.17** **10.17**	**60.39** **42.35**	**63.60** **15.73**	**62.83** **8.28**
H. LUNDBECK	18-Jun-99 (949)	48.85 18.34	49.14 12.22	42.31 5.48	51.72 15.69	46.48 7.68	47.37 4.11	44.68 5.23	**100.0** **0.56**	100.0 0.33
MOELLER MAERSK A	2-Jan-90 (2197)	**59.81** **42.92**	**61.71** **30.31**	**61.64** **13.88**	**61.05** **35.82**	**62.89** **26.43**	**62.70** **14.69**	**62.35** **35.63**	**65.31** **22.96**	**61.81** **12.20**
MOELLER MAERSK B	2-Jan-90 (3153)	**54.72** **30.26**	**55.46** **15.10**	56.52 4.38	**54.36** **23.47**	**54.82** **10.61**	54.43 2.53	53.24 19.92	55.00 4.51	55.56 1.74
NORDEA	2-May-00 (725)	51.96 24.69	42.11 7.86	44.44 2.48	46.84 22.57	39.13 3.29	45.45 1.57	53.03 19.56	66.67 1.78	80.00 0.74
NEG MICON	14-Nov-95 (1792)	**56.55** **39.17**	**58.60** **20.76**	54.20 7.31	**55.04** **36.50**	54.48 16.41	46.99 4.70	**54.91** **32.72**	55.13 8.96	57.14 2.81
NOVO NORDISK B	2-Jan-90 (3255)	49.91 17.60	51.65 10.23	53.72 5.78	50.42 14.74	53.44 8.11	50.41 3.75	**55.01** **12.76**	48.66 5.83	52.11 2.22
TDC	11-May-94 (2227)	50.58 23.35	47.46 10.60	50.47 4.80	53.27 18.76	50.89 5.09	58.93 2.54	**55.50** **17.55**	51.22 3.77	52.38 1.93

Table 2: Predictability measures for 20 of the stocks composing the ATX top 22 index of Vienna. (Bold numbers are above the 95% confidence)

Stock Name\ window size		**50**			**75**			**100**		
Name	Start Period (#prediction days for series length 50)	% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction		
		0.60	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**
ATX-INDEX VIENNA	11-Nov-92 (2568)	52.58 21.92	**55.21** **11.21**	48.18 4.28	52.51 16.48	52.17 8.14	52.17 2.71	53.27 12.15	56.63 3.30	58.82 1.35
BWT AG	13-May-92 (2489)	**56.52** **33.27**	**54.83** **17.88**	**57.14** **7.31**	**55.52** **25.37**	**59.58** **15.46**	**56.67** **7.31**	**57.09** **20.54**	**56.64** **11.73**	50.00 3.85
BBAG STAMM	27-Apr-92 (2697)	52.42 26.03	53.28 13.57	52.02 6.41	50.58 22.42	52.43 9.99	53.91 4.30	52.11 17.04	45.69 4.38	52.17 2.61
SEMPERIT HDG	20-Sep-95 (1857)	**58.64** **33.98**	**59.63** **17.34**	**63.38** **7.65**	**57.12** **29.15**	**58.13** **15.78**	**59.09** **8.41**	**57.75** **28.56**	**58.71** **14.61**	54.41 7.53
VA TECH- NOLOGIE	25-May-94 (2187)	**55.25** **28.30**	54.22 15.18	50.45 5.08	**54.41** **24.14**	52.86 10.50	48.48 3.05	**55.80** **18.95**	53.38 6.22	54.24 2.76
AUSTRIAN AIRL	2-Jan-90 (3274)	**57.75** **29.57**	**59.14** **17.04**	**54.50** **5.77**	**57.18** **25.30**	**60.37** **11.73**	**60.40** **4.59**	**55.36** **22.58**	**60.31** **7.97**	56.76 3.44
ERSTE BANK ST	22-Nov-93 (2260)	**67.11** **40.22**	**70.77** **31.15**	**72.79** **20.49**	**67.53** **36.11**	**70.40** **27.96**	**73.05** **21.74**	**70.95** **32.71**	**72.96** **27.78**	**72.69** **21.54**
OMV AG	21-May-91 (2935)	**53.44** **21.81**	51.97 13.83	49.08 5.55	**54.90** **15.09**	50.00 8.04	54.22 2.85	**57.28** **18.58**	55.19 6.34	**60.27** **2.53**
TELEKOM AUST	21-Nov-00 (583)	48.94 16.12	50.00 5.83	47.83 3.95	36.96 8.24	42.31 4.66	45.45 1.97	52.38 15.76	52.63 10.69	49.15 2.44
VERBUND	19-Mar-90 (3222)	**55.12** **34.85**	**55.26** **16.51**	55.15 5.12	**55.66** **27.37**	**55.01** **13.42**	54.00 4.69	**54.12** **23.36**	55.26 7.19	56.67 3.78
EVN	6-May-91 (2944)	**54.50** **29.42**	**54.03** **15.76**	54.79 6.39	53.52 23.36	50.80 10.65	54.62 4.45	**55.41** **20.77**	**55.61** **7.71**	**52.78** **2.49**
PALFINGER	4-Jun-99 (949)	**56.78** **41.94**	**56.72** **28.24**	57.45 9.91	**55.89** **39.50**	54.82 17.97	53.73 7.25	**55.59** **32.81**	54.46 11.23	**71.88** **3.56**
UNIQA VERS.	10-Dec-90 (3001)	**63.07** **63.88**	**64.06** **46.45**	**65.52** **22.33**	**62.84** **65.29**	**64.04** **43.92**	**63.96** **22.38**	**62.69** **68.11**	**62.97** **37.61**	**61.68** **18.57**
FLUGHAFE N WIEN	15-Jun-92 (2669)	50.28 19.75	50.92 10.23	51.00 3.75	53.14 16.87	54.19 6.77	**66.04** **2.00**	50.90 14.93	51.61 4.73	50.00 1.45
BOEHLER UDDEHOLM	10-Apr-95 (1967)	50.00 12.20	50.82 6.20	53.41 4.47	47.71 7.88	53.06 5.05	54.43 4.07	56.62 7.09	**59.57** **4.90**	**63.49** **3.29**
GENERALI HDG	2-Jan-91 (3027)	**56.38** **32.87**	**57.03** **21.61**	**55.22** **8.85**	**55.33** **29.08**	**55.60** **16.06**	**59.22** **5.96**	**57.25** **25.93**	**56.47** **9.34**	56.60 3.56
RHI	2-Jan-90 (3271)	**55.63** **30.66**	53.42 14.31	53.11 5.41	**56.90** **26.80**	**57.07** **11.34**	56.48 3.33	**55.60** **23.01**	**55.46** **7.11**	56.70 3.01
VOESTALPIN E	9-Oct-95 (1844)	52.83 20.12	47.55 7.75	46.43 3.04	50.68 16.05	56.44 5.55	51.22 2.25	52.41 10.42	52.46 3.40	45.83 1.34
WIENERBER GER	2-Jan-90 (3273)	**53.63** **24.84**	53.83 11.18	**57.58** **4.03**	**53.31** **2044**	52.40 7.70	47.67 2.65	**54.43** **19.95**	49.49 6.14	50.70 2.20
BRAU UNION	15-Jul-93 (2397)	**55.45** **30.25**	**56.55** **18.15**	**59.00** **8.34**	**53.43** **29.51**	51.47 14.33	55.83 5.06	**57.49** **24.46**	**57.47** **7.41**	57.81 2.73
MAYR MELNHOF	22-Apr-94 (2205)	50.98 20.82	52.24 11.11	46.85 5.03	52.01 15.96	51.91 6.01	56.25 2.94	55.10 11.37	**61.05** **4.41**	59.65 2.65

Each line in Table 5 presents the total percentage of the bold cells in tables 1, 2, 3 and 4 respectively for which H_0 is rejected. For example, when considering the KFX stocks with running windows of length 50 and a reliability threshold of 0.70, one can note that 6 out of 16 stocks (37.5%) were flagged as predictable stocks "above random". Using the binomial distribution with parameters $\hat{p} = 0.05$ and N=16 results in a single sided 99% threshold of 3 "randomly predictable stocks". Since in this case the number of predictable stocks is 6, H_0 is rejected for the KFX stock series, as it is found for all the other ATX and KFX stock series. On the other hand, in the DAX stocks with running windows of length 50 and a reliability threshold of 0.60 only 2 out of 30 stocks were flagged as "predictable above random". This result stands within the 99% single-sided confidence interval of the binomial distribution with parameters $\hat{p} = 0.05$ and N=30, which is equal to 5 "randomly predictable" stocks. Thus, one cannot reject H_0 for those particular stocks, as indicated for most DAX and DJ30 series (beside the DJ30 series for a window size of 75 days).

Table 5: Aggregate results for tables 1, 2, 3 and 4 respectively. For the bold cells, the EMH is rejected.

Index Name\ window size		**50**				**75**				**100**			
Name of index from which the aggregates were taken	# of stock series per index	#Pred	#stocks predictable with 95% confidence % of total			#Pred	#stocks predictable with 95% confidence % of total			#Pred	#stocks predictable with 95% confidence % of total		
			0.60	**0.65**	**0.70**		**0.60**	**0.65**	**0.70**		**0.60**	**0.65**	**0.70**
ATX-INDEX VIENNA	20	**14/20 70**	**14/20 70**	**10/20 50**	**8/20 40**	**13/20 65**	**13/20 65**	**8/20 40**	**7/20 35**	**14/20 70**	**14/20 70**	**11/20 55**	**6/20 30**
KFX-INDEX KOPENHAGEN	16	**12/16 75**	**11/16 68.7**	**10/16 62.5**	**6/16 37.5**	**12/16 75**	**11/16 68.7**	**10/16 62.5**	**6/16 37.5**	**13/16 81.2**	**12/16 75**	**9/16 56.2**	**6/16 37.5**
DAX-INDEX FRANKFURT	30	1/30 3.33	2/30 6.67	0/30 0	1/30 3.33	0/30 0	0/30 0	1/30 3.33	1/30 3.33	1/30 3.33	1/30 3.33	0/30 0	1/30 3.33
DJ30-INDEX NEW-YORK	30	8/30 26.67	4/30 13.33	4/30 13.33	2/30 6.66	6/30 20.00	**5/30 16.67**	**6/30 20.00**	**5/30 16.67**	8/30 26.67	4/30 13.33	2/30 6.66	2/30 6.66

4.4 Testing Volume and Predictability relations for the ATX Stocks

In the previous experiments, a significant predictability level was detected in two of the international stock markets: **all** the Vienna ATX (Austria) and the Kopenhagen KFX (Denmark) stock series demonstrate significant predictability for **all** the considered reliability thresholds and for **all** sliding windows lengths. In these two markets the null hypothesis was rejected, implying that the market is inefficient. In contrast, and as expected in larger and potentially more efficient markets[6], the implementation of the same test procedures to the Frankfurt DAX[7] (Germany) and the DJ30 (New-York) stock series resulted in a much lower predictability rate and, thus, to the acceptance of the null hypothesis, implying that the market is efficient. The purpose of this section is to test if the relation between volume and predictability holds not only across markets, but also within a market over time.

Risk is often related to the predictability level: the less predictable the future value of an asset is, the more risky it is. Modern financial theory asserts that investors expect higher returns from riskier assets; therefore, the factors that influence the risk of a stock may also influence its predictability. Fama and French (1993) construct a three-factor asset pricing model for stocks that includes the conventional market correlation factor (*beta*) and two additional risk factors related to stock size and book to market equity. The model implies that the expected return on a portfolio in excess of the risk free rate is explained by the sensitivity of its return to three factors: (*i*) the excess return on a broad market portfolio; (*ii*) the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks (SMB); and (*iii*) the difference between the return on a portfolio of high-book-to-market stocks and the return on a portfolio of low-book-to-market stocks (HML). The size effect and the book to equity effect are detected in many international stock markets[8] (Fama and French, 1998; Maroney and Protopapadakis, 2002). The purpose of this section is to test the Fama and French three factors model for the predictability of the stocks composing the ATX top 22 index of Vienna.

Stock data[9] was collected for the stocks[10] composing the ATX top 22 index of Vienna. Table 5 presents stock value[11] (in Billion Euro) as well as the average and standard deviation of the daily turnover for each stock (in Million Euro). The *beta* (correlation with the ATX index) was computed for each stock. Based on their values[12], the top and bottom 30% of the stocks were classified as Big or Small (respectively).

[6] In a small volume market, large buy (sell) orders are typically partitioned over several consecutive days to avoid price swings. This might explain some of the predictability results we obtained.

[7] Gurgul et al. (2007) investigated the dynamic relations between price and volume in the DAX stocks

[8] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#International

[9] Data taken from http://en.wienerborse.at/prices_statistics/statistics/yearly/index.html for the year 2000.

[10] The index is updated semi-yearly. Only 20 stocks were included for prolonged periods of time.

[11] In the year 2000 in Billion Euro. Older data is not in available in the Euro currency.

[12] This classification is typical in implementing the Fama and French model. However, the Fama and French model is typically implemented on multiple stocks on a monthly or yearly basis. We had only 20 stocks and therefore implemented it on the full period which could be over 10 years.

For the first experiment, we computed the time series of the Small-minus-Big portfolio (the difference between the daily average[13] of the small stocks and the daily average of the big stocks). Applying the VOM tree to the Small-minus-Big time series, we failed to detect 'above random predictability' cases. No correlation was detected between any of the stock time series and the Small-minus-Big time series (beta smaller than 0.01). Note that the Fama and French model attempts to detect excess returns. The VOM tree attempts to detect above random predictability. There is no direct relation between predictability and excess returns.

Table 6: Predictability measures for 20 of the stocks composing the ATX top 22 index of Vienna. The data used for regression analysis based on the length 75 series.

STOCK NAME	# of trading days	Volume Billion Euro (2000)	Size Category 1=Small 2=Big	Average Daily Turnover Million Euro	StDev of Daily Turnover Million Euro	Correlation with ATX Index (Beta)	%Correct predictions for 0.65 threshold	%Trading days above 0.65 threshold
BWT AG	2489	0.58	1	1.91	3.86	0.04	59.58	15.46
BBAG STAMM	2697	0.40		1.24	2.89	0.04	52.43	9.09
SEMPERIT HDG	1157	0.21	1	2.59	5.28	0.05	58.13	15.78
VA TECHNOLOGIE	2187	0.48	1	22.96	36.20	0.47	52.86	10.50
AUSTRIAN AIRL	3274	0.42	1	0.69	1.88	0.05	60.37	11.73
ERSTE BANK ST	2260	2.42	2	7.00	16.10	0.07	70.40	27.96
OMV AG	2935	2.23	2	21.89	40.74	0.28	50.00	8.04
TELEKOM AUST	583	1.05		3.69	8.64	0.29	42.31	4.66
VERBUND	3222	1.63	2	9.38	21.06	0.00	55.01	13.42
EVN	2944	1.11	2	12.60	23.29	-0.02	50.80	10.65
PALFINGER	949	0.26		0.24	0.46	0.10	54.82	17.97
UNIQA VERS	3001	0.75		0.09	0.29	0.44	64.04	43.92
FLUGHAFEN WIEN	2669	0.63		7.76	14.51	0.06	54.19	6.77
BOEHLER UDDEHOLM	1967	0.39	1	12.17	20.23	0.25	53.06	5.05
GENERALI HDG	3027	1.41	2	2.36	5.21	0.05	55.60	16.06
RHI	3271	0.42	1	3.00	7.14	0.07	57.07	11.34
VOESTALPINE	1844	1.28		12.31	18.42	-0.01	56.44	5.55
WIENERBERGER	3273	1.33	2	8.05	16.31	0.01	52.40	7.70
BRAU UNION	2397	0.42		1.77	3.50	0.02	51.47	14.33
MAYER MELNHOF	2205	0.56		5.78	10.71	-0.02	51.91	6.01

In the second experiment, we used the SPSS statistical package to run a stepwise regression on the data in Table 6 to predict two variables: *i*) the FCPART for a 0.65 reliability threshold and *ii*) the percentage of trading days above 0.65 threshold. The regression for the FCPART turned out to be statistically significant (F-statistic with P_{value} =0.007; Adjusted R^2 =0.674; P_{value}<0.05 for the regression

Stock values and rankings may change significantly over time. Therefore, we also used the 2003 data to screen out stock that changed their category significantly.

[13] The portfolios were adjusted daily such that the shares retain an equal monetary value. Some stocks had shorter time series than others. Therefore, we adjusted the weights of the existing shares in the portfolio to compensate for the smaller number of shares.

coefficients' t-statistic): The regression model contains a constant factor, and the variables *stock's volume*, *stocks size category*, and *standard deviation of the stocks' daily turnover*. Note that this experiment corroborates Fama and French's hypothesis that small stocks behave differently than large stocks.

5. Results

Following the above tests, a significant predictability is detected in two of the international stock markets. In particular, **all** the Vienna ATX (Austria) and the Kopenhagen KFX (Denmark) stock series demonstrate significant predictability (above the random prediction reference) for **all** reliability thresholds and for **all** windows lengths. Overall percentages indicate that 30% to 81% of those stock series (see Table 5) lead to an “above random” FCPART and to the rejection of the null hypothesis. Thus, these two markets were found to be inefficient in the considered period. By contrast, the implementation of the same test procedures to the Frankfurt DAX[14] (Germany) and the DJ30 (New-York) stock series result in a much lower predictability rate and, thus, to the acceptance of the null hypothesis that the market is efficient. As expected in larger and potentially more efficient markets[15], the DJ30 (New-York) stock series demonstrate some predictability level (on the limit of a random prediction reference) for windows of length 75 and around 20% of the stocks. A regression experiment for the stocks in the ATX index corroborate Fama and French's hypothesis that small stocks behave differently than large stocks in certain time periods

Within the predictable stock series, the FCPART can be considered as a measure for the stock’s efficiency. While the FCPART can go up to 63.07% (see Uniqa Vers in Table 3), the typical proportion of predictable trading days for the ATX and KFX series is around 30% for the 0.60 reliability threshold, while dropping to about 10% for the 0.70 reliability threshold. The equivalent figures for the DJ30 series are lower – about 20% for the 0.60 reliability threshold, while dropping to about 4% for the 0.70 reliability threshold. This result implies that even the stock series that demonstrated above random predictability are efficient around 70% to 90% of the time (80% to 96% for the DJ30). Once again, it corresponds with the EMH theoretical convention that expects the markets to be efficient most of the time.

The reliability threshold has to be practically determined when implementing the prediction scheme in an actual investment strategy. As expected, the FCPART is lower than the threshold and tends to increase with the reliability threshold (see

[14] Gurgul et al. (2007) investigated the dynamic relations between price and volume in the DAX stocks

[15] In a small volume market, large buy(sell) orders are typically partitioned over several consecutive days to avoid price swings. This might explain some of the predictability.

appendix B). However, an increase of the reliability threshold reduces both the number of inefficient stocks and the percentage of predictable days. A threshold of 0.60 seems as a reasonable choice for an investment strategy, since an increase of the threshold beyond that value reduces the number of available trading days and often results in a lower prediction rate (see Table 5).

In an attempt to validate the conclusions of section 5 in regard to the market efficiency, and in conjunction to the field of information theory, we also computed the *compressibility* level beyond random measure[16] for each stock series and for each running window length, in a manner similar to that of Shmilovici *et al.* (2003). The conclusions regarding the market efficiency remained unchanged. Note that unlike the used predictability measures, the compressibility measure is impractical in terms of a potential investment strategy but rather reflects the randomness level of the data.

For further validation purpose, we also inspected the actual daily returns for some of the most predictable windows. As expected, it reveals that prolonged periods of positive (negative) daily returns correspond to predictability beyond the level of random prediction.

6. Discussion

From its early beginnings, the EMH has woven together *two* theoretical threads: *i*) the hypothesis that prices incorporate all relevant information; and *ii*) the hypothesis that there are no steady profitable trading strategies (Lo, 2007). The experiments in sections 4.3 lead to the rejection of the EMH based on the first thread – detecting statistically-significant information patterns in the daily stock time-series.

There is evidence that the daily stock time-series exhibit mean reversion toward an equilibrium level and that the degree of mean reversion is stronger when the deviation from the equilibrium is larger. Moreover, such return reversals for the market as a whole may be quite consistent with the efficient functioning of the market since they could result, in part, from the tendency of interest rates to be mean reverting, or that transaction costs produce a band of inaction in which the big traders allow the daily stock prices to float freely. Consequently, the adjustment process takes place *only* when the rates approach *the upper or lower limit of the inaction band* (Chung and Hong 2007). Therefore, the found predictability cases could be attributed to the intervention of the big traders at specific (yet unknown) threshold values. The market efficiency 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

[16] Results are not presented here but are available from the authors upon request.

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). The VOM tree model that we used in this paper is capable of capturing patterns of reversion to equilibrium (Singer and Ben-Gal, 2007), as long as the reversion is within the window of observation (50 to 100 consecutive trading days in our experiments). This capability to predict beyond random can be explained as follows: when the deviation from equilibrium is small, and the stock market can move either way, the VOM tree can captures only a small predictability level above random (such as seen in leaf 1 in Figure 1). On the other hand, in periods of large deviation from equilibrium, when the stock market tends to move towards the equilibrium, the VOM tree can captures a strong predictability above random (such as seen in leaf 2 in Figure 1).

The regression experiment in section 4.4 corroborates the results in section 4.3: the stock's size category is negatively correlated – therefore the stock is less predictable for the "Big" category than for the "Small" category. The standard deviation of the stocks daily trading volume (a common measure of the stock's risk) is negatively correlated with the stock's predictability (FCPART), as expected from conventional financial theory.

A “real” test for the market efficiency is in the ability to suggest a trading strategy that demonstrates excess returns, e.g., consistent returns above the "buy and hold" strategy. This is a difficult task considering the market’s infrequent inefficiencies and the limited reliability of the forecast. Some attractive attributes of the VOM tree lie in its ability to predict the sign[17] of the price change at each period, estimate the reliability of these sign predictions and detect periods of the market’s inefficiency. Unlike the conventional test for market efficiency that is largely a one-shot game, the VOM tree can be used to measure the fraction of the time that the market is efficient. Econometricians learned similar ideas from the co-integration analysis, while the latter does not automatically provide a measure on the time of disequilibrium.

The main limitation of the VOM tree is that it ignores the *actual values* of the expected return. That is, the current version of the algorithm is based on a binary alphabet, thus, it is limited to the forecasting of either positive or negative returns without differentiating among the expected returns. Such a limited prediction of sign sequences addresses a "weaker" form of the market efficiency. As a first practical step, a binary predictor should at least indicate if the predicted return is above the trade commissions (e.g., the “bid and ask” spreads). Taking further steps in this direction and integrating the VOM tree in a strategic trading tool that might generate excess returns is a matter of active research. Another limitation is

[17] For a ternary discretization of the price change, the VOM model can predict more than the sign change

the undefined discretization process of the series that is required prior to the implementation of the algorithm. Kahiri *et al.* (2004), Shmilovici *et al.* (2009) used the VOM tree to predict ternary trends (i.e., increase symbol, stable symbol and decrease symbol) in the FOREX market. The trading commissions determined the discretization levels. 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".

As seen in Table 5, given a reliability threshold the percentages of predictable stock series are fairly equal for the window lengths. Thus, it seems that the considered window lengths have no apparent effects on the percentage of predictable series. Theoretically, the predictability should grow with the window length (see appendix B). However, the considered lengths might be too small and, moreover, a theory of universal prediction for noisy sequences is not yet available in this direction (Hutter, 2001). Longer sequences are possibly more sensitive to noise and temporal market effects, and may need an adaptive version of the VOM tree algorithm, such as the one proposed in Federovsky *et al.*, (1998). No attempt was made in this paper to optimize the VOM tree (e.g., by optimizing the pruning coefficient C and other structure parameters) for each window length and stock and it is left for future research. However, recall that tuning experiments revealed that the VOM tree is fairly robust to the choice of C in the range $C \in [0.25,...,4.0]$.

Another practical deficiency of the VOM tree – the limited number of prediction days with a high reliability – can be ameliorated by independently implementing the VOM tree for each series in a *portfolio* of stocks. The theory of universal portfolios (Cover 1991, Blum and Kalai 1997) 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.

We conclude by noting that a predictability measure of a series (e.g., the rate of correct predictions), as 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 process. Unlike the proposals for testing the Martingale Difference Hypothesis (MDH)[18] (Domínguez and Lobato, 2003) that require unreasonably large amounts of data, the proposed predictability measure is computable even for relatively short sequences.

Website: For available VOM tree web server see http://www.eng.tau.ac.il/~bengal/

[18] The MDH states that the best prediction (in least mean square sense) of the future values of a time series given the current information set is just the unconditional expectation. Hence, past information does not help to improve the forecast of future values of a MDS.

Table 3: Predictability measures for the stocks composing the German DAX 30 index. (Bold numbers are above the 95% confidence)

Stock Name\ window size		**50**			**75**			**100**		
Name	Start Period (#prediction days for series length 50)	% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction		
		0.60	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**
XETRA DAX PF	26-Nov-90 (3107)	**56.32** **19.09**	**55.56** **10.43**	48.60 3.44	**55.49** **17.13**	**55.77** **8.44**	55.42 2.69	**57.24** **14.0**	**58.43** **5.43**	54.35 1.50
ADIDAS SALOMON	17-Nov-95 (1855)	52.19 17.25	50.31 8.68	45.31 3.45	50.20 13.83	45.54 5.52	42.50 2.19	49.50 11.19	44.74 2.11	37.50 0.89
HENKEL KGAA VZ	8-Apr-91 (2783)	48.04 13.76	55.80 6.50	**58.59** **4.60**	51.11 8.16	54.49 6.06	55.81 3.12	50.51 7.24	48.12 4.10	46.51 1.57
ALLIANZ AG	5-Apr-91 (3014)	54.28 18.21	55.56 9.56	55.15 4.51	49.20 12.58	48.55 5.79	50.00 2.28	47.86 8.67	38.10 2.83	38.30 1.59
HYPOVEREINSBANK	5-Apr-91 (2990)	51.04 19.33	51.60 9.40	49.61 4.25	49.31 14.64	47.74 6.71	42.19 2.16	48.13 10.88	14.55 3.95	45.28 1.80
ALTANA	19-Jan-96 (1669)	49.83 17.32	53.06 8.81	50.00 4.41	53.30 12.90	56.25 5.84	48.94 2.86	51.04 11.86	52.31 4.01	46.88 1.98
INFINEON TECH N	13-Mar-00 (767)	54.05 14.12	51.14 11.47	57.89 2.48	48.67 15.23	54.84 4.18	60.00 2.02	57.95 12.27	57.14 2.93	53.85 1.81
BASF AG	5-Apr-91 (3015)	51.31 17.78	54.18 9.12	52.05 4.84	49.01 13.51	48.53 6.82	42.86 2.34	52.83 10.73	56.74 4.76	59.62 1.75
LINDE	5-Apr-91 (2905)	48.18 17.01	48.26 6.92	46.67 2.58	45.05 7.71	39.73 2.53	43.33 1.04	48.80 4.38	57.14 1.47	55.88 1.19
BAY MOT WERKE	5-Apr-91 (3014)	51.52 18.61	51.37 9.69	55.56 5.37	49.85 11.07	51.95 5.15	48.98 3.28	50.80 8.43	52.14 3.95	48.68 2.56
MAN AG	5-Apr-91 (2963)	49.93 23.05	49.68 10.39	43.55 4.18	48.13 15.49	49.59 4.19	45.00 1.36	51.00 12.05	56.67 3.09	53.33 1.03
BAYER AG	5-Apr-91 (3015)	50.94 19.40	50.67 7.46	50.83 3.98	49.75 13.44	51.28 5.22	53.54 3.31	50.75 8.97	51.69 3.98	56.16 2.46
METRO AG	22-Jul-96 (1688)	50.16 18.78	47.48 8.23	41.18 3.02	50.20 14.85	44.44 7.04	35.48 1.86	51.08 11.36	31.03 1.77	38.10 1.28
COMMERZBANK AG	5-Apr-91 (3015)	52.04 17.84	49.17 8.03	43.04 2.62	51.96 11.07	53.74 4.98	41.67 1.20	52.58 9.81	46.96 3.88	34.48 0.98
MLP	15-Oct-98 (959)	49.20 26.07	45.33 7.82	50.00 4.80	46.77 13.28	44.12 3.64	52.63 2.03	46.67 9.90	53.85 4.29	52.94 3.74
DAIMLERCHRYSLER	5-Apr-91 (3009)	50.00 17.08	50.55 9.14	52.32 5.02	50.54 12.47	54.55 6.64	51.52 3.332	54.05 12.50	50.67 5.07	46.48 2.40
MUENCH. RUECK N	26-Jul-94 (1873)	48.26 16.92	47.50 6.41	53.85 41.16	47.80 9.85	47.06 4.60	49.18 3.30	47.27 6.03	45.00 2.19	45.95 2.03
DEUTSCHE BANK N	5-Apr-91 (3010)	49.79 16.08	47.62 8.37	48.78 5.45	47.44 10.45	50.66 5.09	40.00 2.51	50.70 7.26	51.69 3.99	56.90 1.96
RWE ST A	5-Apr-91 (3015)	**54.53** **17.58**	54.66 8.19	53.77 3.52	49.71 11.71	51.48 5.65	52.31 2.17	48.35 9.21	54.24 3.98	56.72 2.26

DEUTSCHE POST NA	20-Nov-00 (591)	45.45 22.34	52.31 11.00	42.86 5.92	49.04 18.37	54.05 6.54	53.85 2.30	41.18 9.43	33.33 2.22	20.00 0.92
SAP AG	13-Sep-94 (2015)	48.34 13.45	50.00 5.16	56.14 2.83	46.43 7.04	37.50 2.41	37.50 1.61	38.71 3.16	37.50 1.22	23.53 0.87
DT BOERSE N	5-Feb-01 (539)	48.72 14.47	41.86 7.98	38.10 3.90	55.00 3.89	57.14 1.36	57.14 1.36	20.00 1.02	0.00 0.41	0.00 0.41
SCHERING AG	5-Apr-91 (2949)	48.80 17.02	52.34 8.85	50.00 4.88	49.44 12.11	44.94 5.40	55.56 2.46	48.09 11.76	50.89 3.86	50.67 2.59
DT LUFTHANSA AG	5-Apr-91 (2937)	51.6 19.14	48.58 7.22	49.51 3.51	48.62 13.7	47.06 5.25	46.51 2.95	44.36 9.53	44.30 2.74	44.44 1.87
SIEMENS N	5-Apr-91 (3210)	50.34 13.86	50.79 5.95	48.00 3.12	52.82 9.45	55.94 4.49	56.63 2.61	52.38 7.31	52.14 3.70	55.71 2.22
DT TELEKOM N	18-Nov-96 (1599)	53.01 16.64	53.85 9.76	57.32 5.13	55.56 13.72	**58.59** **6.29**	**65.63** **4.07**	**57.14** **10.39**	57.32 5.29	**60.29** **4.39**
THYSSEN KRUPP	5-Apr-91 (3013)	50.0 17.13	54.05 8.60	50.98 5.08	50.41 12.15	49.75 6.59	45.63 3.45	48.81 9.96	50.47 3.61	52.94 1.72
E.ON AG	5-Apr-91 (3015)	47.21 16.65	50.2 8.46	46.51 4.28	49.72 11.91	50.0 7.22	54.0 3.34	50.72 11.70	49.71 5.90	50.52 3.27
TUI AG	5-Apr-91 (2999)	50.68 24.34	46.68 11.77	45.59 4.53	51.42 16.54	51.31 6.42	55.26 2.56	51.47 11.53	53.06 3.32	60.42 1.63
FRESENIUS MEDI	4-Oct-96 (1635)	**55.18** **18.29**	55.32 8.62	57.33 4.59	51.59 1565	55.24 6.52	56.10 2.55	51.91 11.55	53.23 3.91	51.43 2.21
VOLKSWAGEN AG	5-Apr-91 (3215)	52.98 22.95	51.30 9.58	55.15 4.23	50.89 15.83	53.81 6.18	54.31 3.64	52.45 11.63	54.00 4.74	53.25 2.43

Table 4: Predictability measures for the stocks composing the American Dow-Jones 30 index. (Bold numbers are above the 95% confidence)

Stock Name\ window size		**50**			**75**			**100**		
Name	Start Period (#prediction days for series length 50)	% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction			% predictability above threshold % samples above threshold used for prediction		
		0.60	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**	**0.60**	**0.65**	**0.70**
DOW-JONES30 (USA)	2/1/90-31/12/01 (2977)	50.89 16.96	52.57 5.88	56.38 3.16	53.18 11.72	52.67 5.08	49.32 2.47	46.55 7.93	48.35 3.11	52.24 2.29
ALCOA INC	2/1/90-31/12/01 (2977)	47.57 17.30	45.62 7.29	49.57 3.93	45.12 10.06	46.85 4.84	45.45 2.24	47.98 5.91	51.06 3.21	51.16 1.47
AMERICAN EXPRESS CO	2/1/90-31/12/01 (2977)	**54.36** **18.11**	51.40 9.61	52.91 5.78	51.37 13.58	48.92 6.30	51.49 3.42	51.51 11.34	50.43 3.93	49.15 2.02
BOEING CO	2/1/90-31/12/01 (2977)	51.59 22.20	54.39 9.57	51.92 3.49	52.99 15.85	**56.25** **6.50**	**60.87** **3.12**	**56.47** **11.62**	55.92 5.19	52.70 2.53
CITIGROUP	2/1/90-31/12/01 (2977)	53.041 5.45	53.49 8.67	50.00 4.77	52.16 12.53	53.81 6.67	52.04 3.32	53.23 15.34	54.32 5.53	57.89 2.60
CATERPILLAR INC	2/1/90-31/12/01 (2977)	51.98 19.52	51.92 11.39	49.61 4.33	54.18 13.38	53.91 7.79	56.18 3.01	51.31 13.05	51.45 4.71	55.38 2.22
DU PONT CO	2/1/90-31/12/01 (2977)	51.59 21.16	52.30 10.21	51.94 4.33	52.49 16.33	49.73 6.33	42.65 2.30	52.94 11.62	55.65 3.93	50.00 1.50

WALT DISNEY CO	2/1/90-31/12/01 (2977)	46.83 15.35	41.92 6.65	40.40 3.33	53.47 11.21	51.91 4.44	51.52 2.24	46.94 8.37	43.42 2.60	45.10 1.74
EASTMAN KODAK	2/1/90-31/12/01 (2977)	46.10 14.65	47.30 8.10	48.06 4.33	53.05 13.35	49.21 6.40	56.32 2.95	49.14 11.89	52.63 3.89	50.79 2.15
GENERAL ELECTRIC CO	2/1/90-31/12/01 (2973)	52.18 22.34	50.57 11.82	51.18 5.71	52.24 15.89	**56.02** **7.32**	**58.40** **4.23**	53.91 12.68	**59.59** **4.99**	**64.63** **2.80**
GENERAL MOTORS	2/1/90-31/12/01 (2977)	50.84 20.09	51.71 7.86	50.00 3.49	47.38 12.94	46.09 3.90	42.03 2.34	50.66 7.76	52.17 2.36	44.44 1.54
HOME DEPOT INC	2/1/90-31/12/01 (2975)	48.00 10.92	50.34 5.01	50.00 3.56	50.31 5.39	48.89 3.05	44.00 1.69	52.00 5.12	43.08 2.22	39.02 1.40
HONEYWELL INTERNATION.	2/1/90-31/12/01 (2977)	52.85 18.88	**56.42** **8.63**	**59.54** **4.40**	49.18 14.46	50.00 7.72	50.88 3.86	49.52 10.76	46.67 4.10	49.38 2.77
HEWLETT-PACKARD	2/1/90-31/12/01 (2977)	**54.84** **15.62**	54.87 7.59	55.56 4.84	**55.42** **11.25**	**58.12** **6.47**	**61.90** **4.27**	**55.08** **10.42**	**59.14** **6.35**	**61.47** **3.72**
INTLLIGENT BUS MACHINE (IBM)	2/1/90-31/12/01 (2977)	48.21 20.69	46.15 8.30	47.37 3.83	49.87 13.04	50.32 5.32	47.37 1.93	47.72 8.23	35.00 1.37	36.36 0.75
INTEL CORP	2/1/90-31/12/01 (2977)	49.52 17.37	51.74 8.70	47.22 3.63	48.89 12.20	46.06 5.49	40.35 1.93	48.49 10.22	46.67 3.07	54.55 1.88
INTERNATION AL PAPER CO	2/1/90-31/12/01 (2977)	47.18 11.89	50.52 6.52	52.48 4.74	49.36 7.96	54.49 5.66	53.51 3.86	51.32 7.79	51.32 5.19	51.43 2.39
JOHNSON& JOHNSON	2/1/90-31/12/01 (2975)	48.27 11.62	48.97 4.87	54.43 2.65	53.04 8.37	53.04 3.90	50.59 2.88	49.32 4.99	53.01 2.84	50.00 1.84
JP MORGAN	2/1/90-31/12/01 (2975)	51.85 18.14	53.71 7.69	53.60 4.20	**55.67** **13.14**	**57.69** **7.93**	39.13 2.34	51.98 11.24	54.55 6.01	40.35 1.95
COCA COLA CO	2/1/90-31/12/01 (2977)	43.15 13.23	44.37 5.07	45.67 4.03	49.77 7.22	51.00 3.39	55.71 2.37	49.18 4.17	50.65 2.63	54.00 1.71
MCDONALDS CORP	2/1/90-31/12/01 (2977)	47.83 20.86	48.50 8.94	47.22 4.84	49.06 14.43	45.99 6.33	37.08 3.01	47.87 10.42	42.40 4.27	42.31 1.78
3M COMPANY	2/1/90-31/12/01 (2976)	41.60 8.80	41.12 3.59	39.74 2.62	40.48 4.27	39.47 2.57	40.74 1.83	51.47 2.32	50.00 1.37	50.00 1.16
ALTRIA GROUP	2/1/90-31/12/01 (2977)	52.30 21.90	**54.90** **10.28**	55.47 4.60	47.54 15.18	48.99 6.71	43.08 2.20	49.09 11.21	50.54 3.18	51.85 1.84
MERCK & CO	2/1/90-31/12/01 (2977)	52.03 18.21	52.56 7.86	55.06 2.99	**55.36** **11.69**	55.00 4.07	59.42 2.34	54.85 9.16	47.89 2.43	50.00 1.78
MICROSOFT CP	2/1/90-31/12/01 (2977)	46.40 13.54	43.58 6.01	37.25 3.43	46.58 7.42	39.22 3.46	29.82 1.93	51.70 5.02	56.82 1.50	58.33 1.23
PROCTER & GAMBLE	2/1/90-31/12/01 (2977)	47.95 13.10	53.89 6.48	51.94 4.33	48.45 5.45	47.66 3.62	48.00 2.54	41.86 2.94	45.28 1.81	44.83 0.99
SBC COMMS	2/1/90-31/12/01 (2977)	48.18 18.47	45.37 7.26	46.99 2.79	51.36 13.72	50.34 5.05	55.17 1.96	49.53 10.97	43.66 2.43	50.00 1.16
AT&T CORP	2/1/90-31/12/01 (2977)	**54.61** **25.53**	**55.74** **14.04**	53.37 5.98	**56.82** **22.83**	**58.36** **9.93**	**60.80** **4.23**	**55.19** **21.05**	51.94 7.04	55.34 3.52
UNITED TECH CP	2/1/90-31/12/01 (2976)	50.45 22.64	48.81 9.84	48.04 3.43	49.29 16.70	45.29 5.76	46.94 1.66	48.79 12.68	48.42 3.25	59.65 1.95
WAL-MART STORES	2/1/90-31/12/01 (2976)	**54.53** **23.72**	**56.65** **11.62**	**59.02** **6.15**	**56.11** **19.14**	**57.09** **9.55**	**62.09** **5.18**	**57.25** **17.66**	54.24 6.05	55.65 3.93
EXXON MOBIL	2/1/90-31/12/01 (2976)	46.42 17.37	44.36 8.94	45.80 4.40	49.03 8.71	50.40 4.23	48.33 2.03	55.06 6.08	57.45 3.21	54.55 1.50

Appendix A: Universal Prediction and Estimation

This section illustrates prediction related problems, as modeled and addressed in the field of information theory. It sketches some known approaches for the prediction of finite-alphabet sequences that are assumed to be generated by a stochastic process. We use such an approach for the prediction of daily stock returns.

Following the notation in Ziv (2001, 2002), consider finite-alphabet sequences of symbols $X_{-N}^{m} \equiv X_{-N},\ldots,X_0,\ldots,X_m$ generated by a stationary source with unknown properties, where each symbol X_i belongs to an alphabet A with cardinality $|A|$.

The symbol prediction problem (Merhav and Feder, 1998): the optimal prediction of X_1 for any observed suffix X_{-N}^{0} is achieved by choosing the symbol $X_1 \in A$ that maximizes the conditional probability $P(X_1 \mid X_{-N}^{0})$. The conditional probability is unknown and has to be estimated from the training sequence X_{-N}^{0}. Consider the class of universal predictors of X_1 conditioned on the *context* $X_{-K_0}^{0}$, which is a sub-sequence of X_{-N}^{0}. The length of the training sequence $K_0 = K_0(X_{-N}^{0})$ is itself an integer function of the training sequence X_{-N}^{0} and determines the required (varying) memory length ($0 \le K_0 \le N$) for the prediction based on the observed context (this is the reason that this prediction is also called 'context specific'). For example, the Bernoulli case, where the symbols are independent of past observed symbols, is indicated by $K_0=-1$, while a Markov model of order one is represented by $K_0=0$.

A simple universal prediction algorithm follows. Find all the different instances of the context $X_{-K_0}^{0}$ in the sequence X_{-N}^{0}. K_0 is chosen as the largest integer such that $X_{-K_0}^{0}$ appears at least n times in the sequence X_{-N}^{0}. Note that n is limits the number of contexts considered in the predictor training phase. Thus, it avoids over-fitting the training data to small number of contexts and limits the storage capacity consumed by the algorithm. X_1 is then predicted as the majority vote symbol $\hat{X}_1 \in A$ that follows the observed instances of $X_{-K_0}^{0}$.

Example: Let $X_{-6}^{0} = 0101100$ and predict X_1 – the next symbol in the sequence – by using the above-mentioned approach. For this short sequence, let us define $n=3$. Note that the relevant contexts for $n=3$ (reading symbols from the last position in the sequence) are "0", "00" and "100". However, since the subsequence "00" (or the longer subsequence "100") does not appear anywhere before the last suffix, the context is defined as "0", i.e., $K_0(X_{-6}^0)=0$. Since the symbol "0" is followed twice by "1" and once by "0", X_1 is predicted to be the majority vote symbol "1".

***The probability estimation problem**:* given X_{-N}^0, estimate $P(X_1 | X_{-N}^0)$. To estimate $P(X_1 | X_{-N}^0)$, one assigns a conditional probability measure $Q(X_1 | X_{-N}^0)$, such that it is "close" in some sense to the true probability distribution $P(X_1 | X_{-N}^0)$, (see Ziv, 2001). For a *universal probability estimation* algorithm the redundancy – i.e. the difference in the complexity measure resulting from using $Q(\cdot)$ instead of $P(\cdot)$– is bounded uniformly, with respect to all distributions in a given class (Rissanen, 1984). A simple universal estimation algorithm defines K_0 as a function of X_{-N}^0 and requires $X_{-K_0}^1$ to appear at least n times in X_{-N}^0. Then, it estimates $P(X_1 | X_{-N}^0)$ by the frequency of the *realization* of the subsequence $\left(X_{-K_0}^0, X_1\right)$ over all the observed *realizations* of subsequences $\left(X_{-K_0}^0, X_i\right)$, $X_i \in A$ in X_{-N}^0.

Example: Given $X_{-6}^0 = 0101100$ and n=1, let us estimate $P(X_1 = 1 | X_{-6}^0)$. For this short sequence, $K_0(X_{-6}^0)=0$ because the subsequence 01 is the longer subsequence that appears in X_{-6}^0 while 001 (or longer subsequences) does not appear in X_{-N}^0. Since the subsequences 01 and 00 appear, respectively, twice and once in X_{-6}^0, the estimation is $Q(1 | X_{-K_0}^0) = Q(1|0) = \frac{\#(01)}{\#(00)+\#(01)} = \frac{2}{1+2} = \frac{2}{3}$, where $\#(\cdot)$ denotes the frequency of its argument in the sequence. Note that the *probability estimation* scheme provides more information regarding the quality of the prediction when compared to the *symbol prediction* scheme. In the paper we use such information to decide in which situations to predict the daily stock returns**.**

The problem of universal compression: when solving this problem one has to minimize the relative entropy or the Kullback Leibler (KL) divergence $E_Q \log \frac{Q(X_1 | X_{-N}^0)}{P(X_1 | X_{-N}^0)}$ between the real unknown distribution $P(\cdot)$ and the estimated one $Q(\cdot)$ (e.g., see Cover and Thomas, 1995). Ziv (2001, 2002) presents non-asymptotic lower bounds for the expected compression rate of *any* universal algorithm that is sequential and has limited training data. Several universal algorithms that are proposed in the literature can achieve these tight bounds. The advantage of context-tree algorithms, such as the Context Tree Weighting (CTW) (Willems *et al.*, 1995), the Prediction by Partial Matching (PPM) (Federovsky *et al.*, 1998) and the VOM tree model we use in this paper is that they can approach Ziv's bounds with *the most efficient learning rate* (Weinberger *et al.,* 1995). Practically, this result means that even with the use of a relatively short sequence, the context-tree model can converge to the (unknown) true model which generated the sequence.

Federovsky *et al.*, (1998) demonstrate the capabilities of the CTW and PPM algorithms for branch prediction in programs. Begleiter *et al.*, (2004) investigate the capabilities of six prominent prediction algorithms on various types of sequences. They find that the CTW and the PPM outperform all other algorithms in sequence prediction tasks. In this paper, we apply a variant of the PPM universal prediction algorithm to estimate $Q(X_1 \mid X_{-N}^0)$. Using this algorithm, called the VOM tree algorithm outlined in section 4.2, we compute the probability estimates for *every context* in the daily stock training sequence. We then use those estimates for a *universal prediction* of X_1, where $\hat{X}_1 \equiv \arg\max_{X_1 \in A} \{Q(X_1 \mid X_{-N}^0)\}$.

Compressibility and Predictability in the VOM tree

As noted before, the existence of recurring patterns in a sequence enables data compression. Each branch in the tree represents a recurring sub-sequence called a "context". The entire sequence can be coded by these sub-sequences in the tree. If the length (in bits) of the coded sequence is shorter than the length of the original sequence, then compression is obtained[19]. The higher is the imbalance among branches in the tree, the higher is the compression rate that can be obtained. At the same time, the recurring patterns in the data introduce the possibility of prediction, in the sense that sequences that are highly compressible are easy to predict and, conversely, incompressible sequences are difficult to predict.

Although prediction and compression are closely related, there is no one-to-one correspondence between the predictability and the compressibility of a sequence[20] (Feder *et al*., 1992). The crucial essence in compression is estimating the *conditional probability* for the next outcome given the past observations, so those symbols with high conditional probabilities are assigned short codes. The estimated probability can be used also for prediction purposes. A prediction algorithm, for which the redundancy is bounded uniformly with respect to all

[19] With an arithmetic encoder, it is guaranteed that the redundancy does not exceed two bits per sequence (Willems *et al.*, 1997).

[20] Some permutations of a sequence may have the same compressibility rate

distributions in some given class, is called a *universal prediction* algorithm with respect to that class (Rissanen, 1984).

Feder *et al.*, (1992) present upper and lower bounds for the relation between the compressibility function rate, $\rho(X)$, and the predictability function rate, $\pi(X)$, of a binary sequence *X*. They show that $\rho(X)/2 \geq \pi(X) \geq h^{-1}(\rho(X))$, where $h(X)$ is the binary entropy function. Different sequences can have the same compressibility, so the compressibility of a sequence does not uniquely determine its predictability. The upper and lower bounds intersect when $(\rho = 0, \pi = 0)$ and when $(\rho = 1, \pi = 1/2)$ (Figure 3). At the extreme points the bounds imply the intuitively appealing idea that a sequence is perfectly predictable if and only if it is totally redundant and, conversely, a sequence is totally unpredictable if and only if it is incompressible. Merhav and Feder (1998) present further results, such as the relation between the number of leaves in the VOM tree and the information content in the sequence.

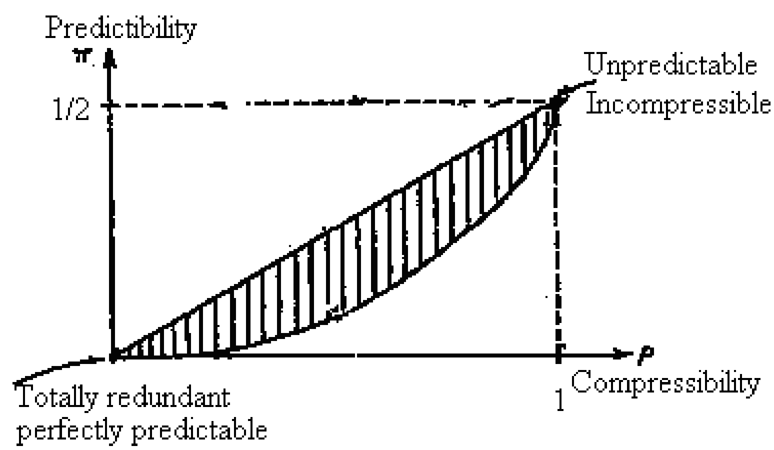

Figure A: Upper and lower bounds for relations between compressibility and predictability of a binary sequence (taken from Feder et al., 1992).

Error bounds for several universal predictors are introduced in Feder and Federovsky (1999) for binary series and in Hutter (2001) for non-binary series.

Appendix B: Bounds to the Prediction Rate

Here, we exemplify how a universal error bound can be used to find the expected prediction rate of a VOM tree. The used parameters are typical to the financial series in our experiments.

For example, in the Bernoulli case (Feder and Federovsky, 1999), the expected prediction rate (fraction of correct predictions) $\prod_N$ which is obtained from training a saturated counter predictor on a data window of length *N* is equal to:

$$\prod_N = p-(p-q)\cdot\left(\frac{q}{p}\right)^{\frac{M}{2}} - \frac{M}{4N} + (smaller\_terms)$$

where $p > \frac{1}{2}$ is the probability of a successful prediction in a single Bernoulli experiment, $q = 1-p$ and *M* denotes the number of states of the predictor. In the context tree, *M* denotes the number of independent contexts – related to the number of leaves in the pruned context tree. *M*=*O*(log *N*) – see Ben-Gal *et al.*, (2003).

Let us now plug some typical values, similar to the ones used in the section 5, such as $p = 0.65$, $N = 50$, $M = 3$. Accordingly, the expected prediction rate is $\prod_N = 0.65 - 0.1185 - 0.015 \approx 0.5165$. Several comments can be stated:

i) $\prod_N < p$, thus, the prediction rate of the entire sequence is smaller than the prediction rate in a single Bernoulli experiment.

ii) Having a longer training sequence, *N*, often increases the number of contexts *M* and thus results in an increased $\prod_N$.

iii) There exists a certain context tree with *M* contexts that maximizes $\prod_N$. Note that *M* is affected by the user-defined pruning coefficient *C*.

In practical terms, the "prediction loss" $(\prod_N - p)$ is expected whenever the prediction model is either "too simple" to accurately represent the sequence (e.g., a model with small number of state parameters), or when the training sequence length *N* is too short for training the prediction mechanism to its full capacity. The prediction loss is further increased when the sequence is contaminated by noise. In the experiments in section 5, $\prod_N$ is estimated from short and noisy sequences, and only the lower bound on *p* is determined. Effectively, this means that many experiments are needed to reliably conclude about the quality of the predictor.

References

Alon-Brimer Y. (2002): "Measuring the Efficiency of the Israeli Stock Market by the Context Tree Model", M.A. dissertation (in Hebrew), School of Management, Ben-Gurion University, Israel.

Baetaens D.J.E, van den Berg W.M, Vaudrey H. (1996): "Market inefficiencies, Technical Trading and Neural Networks", in Dunis C. (Editor), *Forecasting Financial Markets, Financial Economics and Quantitative Analysis*, 245-260, John Wiley & Sons, Chichester, England.

Begleiter R., El-Yaniv R., Yona G. (2004): "On Prediction Using Variable Order Markov Models", *Journal of Artificial Intelligence*, 22:385-421.

Bellgard C.D. (2002): "Diffusion of Forecasting Technology: nonlinearity and the Information Advantage Effect", Ph.D. dissertation, The University of Western Australia.

Blum A., Kalai A. (1997): "Universal Portfolios With and Without Transaction Costs", *Proceedings of the Tenth Annual Conference on Computational Learning Theory (COLT 1997)*, July 6-9, 1997, Nashville, Tennessee, COLT 1997: 309-313.

Ben-Gal I., Morag G., Shmilovici A. (2003): "CSPC: A Monitoring Procedure for State Dependent Processes", *Technometrics*, 45(4):293-311.

Benjamini Y., Hochberg Y. (1995): "Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society section B*, MET, 57 (1):289-300.

Buhlmann P., and Wyner A. J. (1999): "Variable length Markov chains," *Ann. Statist.* 27(2), 480-513.

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 and J. J. Oliver (eds.), Information, Statistics and Induction in Science, World Scientific, Singapore.

Chen S.-H., Tan C.-W. (1999): "Estimating the Complexity Function of Financial Time Series: An Estimation Based on Predictive Stochastic Complexity", Proc. of the 5th Int. Conference of the Society for Computational Economics, Boston, June 24-26. Also published in *Journal of Management and Economics*, 3 (3) (an electronic journal).

Chung J., Hong Y., (2007): "Model-free Evaluation of Directional Predictability in Foreign Exchange Markets", *J. of Applied Econometrics*, 22:855-889.

Cover T.M. (1974): "Universal Gambling Schemes and the Complexity Measures of Kolmogorov and Chaitin", Technical Report no. 12, dept. of statistics, Stanford university, October.

Cover T.M. (1991): "Universal Portfolios", Mathematical Finance, 1(1):1-29.

Cover T.M., Thomas J.A. (1995): *Elements of Information Theory*, Wiley, New-York.

Domínguez M.A., Lobato I.N. (2003): "Testing the Martingale Difference Hypothesis", *Econometric Reviews*, 22(2):351-377.

Fama E.F. (1991): "Efficient Capital Markets: II", *Journal of Finance*, 46:1575-1611

Fama E. F., French K. R. (1993): "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics*, 33 (1): 3–56.

Fama E.F. (1998): "Market Efficiency, Long Term Returns, and Behavioral Finance", *Journal of Financial Economics*, 49:283-306.

Fama E. F., French K. R. (1998): "Value versus growth: the international evidence, Journal of Finance" 53: 1975-1979.Feder M., Federovski E., 1999, "Universal Prediction with Finite Memory", *IT Workshop on Detection, Estimation, Classification and Imaging*, Santa Fe, NM, USA, Feb.24-26.

Feder M., Merhav N., Gutman M. (1992): "Universal Prediction of Individual Sequences", *IEEE Transactions on Information Theory*, 38(4):1258-1270.

Federovsky E., Feder M., Weiss S. (1998): "Branch Prediction Based on Universal Data Compression Algorithm", *ACM SIGARCH Computer Architecture News*, 26(3):62-72.

Gurgul H., Majdosz P., Mestel R. (2007): "Price-Volume Relations of DAX Companies", *Financial Markets and Portfolio Management*, 21(3):353-379.

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", *Proceedings of the 12th Eurpean Conference on Machine Learning ECML-2001*, 239-250.

Jensen M. (1978): "Some anomalous evidence regarding market efficiency", *Journal of Financial Economics*, 6:95-101.

Kahiri Y., Shmilovici A., Hauser S. (2004): "Forecasting with a Universal Data Compression Algorithm: The Forex Market Case", *The International Conference on Computational Methods in Science and Engineering (ICCMSE)*, 19-23 November, Athens, Greece, 4p.

Lo A. W. (2007): "Efficient Markets Hypothesis", in Blume L. and Durlauf S. (eds.), The New Palgrave: *A Dictionary of Economics*, 2nd ed., Palgrave MacMillan, London.

Malkiel B.G. (2003): "The Efficient Market Hypothesis and its Critics", *Journal of Economic Perspectives*, 17:59-82.

Maroney N, Protopapadakis A. (2002): "The Book-to-Market and Size Effects in a General Asset Pricing Model: Evidence from Seven National Markets", *European Finance Review*, 6:189-221.

Merhav N. Feder M. (1993): "Universal Schemes for Sequential Decisions from Individual Sequences", *IEEE Transactions on Information Theory*, 39, 1280-1292.

Merhav N. Feder M., (1998): "Universal Prediction", IEEE Transactions on Information Theory, IT-44, 2124-2147.

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.

Rissanen J. (1983): "A Universal Data Compression System", *IEEE Transactions on Information Theory*, 29(5):656-664.

Rissanen J. (1984): "Universal Coding, Information, Prediction, and Estimation", IEEE *Transactions on Information Theory*, 30(4):629-636.

Rissanen J. (1989): *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore.

Schwert G. (2003): "Anomalies and Market Efficiency", Chapter 15 in Constantinides G.M., Harris. M., Stulz R. (Eds.), *Handbook of the Economics of Finance*, Elsevier, 939-972.

Shafer G., Vovk V. (2001): Probability and Finance: It's Only a Game!, Wiley.

Shannon C. E., 1951, "Prediction and Entropy of Printed English" *Bell Systems Technical Journal*, 30:51-64.

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.

Shmilovici A., Kahiri Y., Ben-Gal I., Hauser S. (2009): "Measuring the Efficiency of the Intraday Forex Market with a Universal Data Compression Algorithm", *Computational Economics*. 33 (2):131-154.

Singer G., Ben-Gal I. (2007): "The Funnel Experiment: A Markov-Based SPC approach", *Quality and Reliability Engineering International*, 23:899–913.

Timmermann A., Granger C.W.J. (2004): "Efficient Market Hypothesis and Forecasting", *International Journal of Forecasting*, 20(1):15-27.

Tino P., Schittenkopf C., Dorffner G. (2000): "A Symbolics Dynamics Approach to Volatility Prediction", in Abu-Mustafa Y.S., and LeLeBaron B. (Eds.), *Proceedings of the 6th international Conference on Computational Finance*, Jan. 6-8 1999, Leonard N. Stern School of Bussiness, MIT Press, Cambridge, MA, pp. 137-151.

Tsibouris G., Zeidenberg M. (1995): "Testing the Efficient Market Hypothesis with Gradient Descent Algorithms", in Refenes A.P. (Editor), *Neural Networks in the Capital Markets*, chapter 8, 127-136, John Wiley & Sons, Chichester, England.

Vert J.P. (2001): "Adaptive context trees and text clustering", *IEEE Transactions on Information Theory*, 47(5), 1884-1901.

Weinberger M., Rissanen J., Feder M. (1995: “A Universal Finite Memory Source”, *IEEE Transactions on Information Theory*, 41(3):643-653.

Willems, F.M.J. Shtarkov Y.M. and Tjalkens T.J. (1995): "The context-tree weighting method: Basic properties", *IEEE Trans. on Information Theory*, 41(3):653-664.

Wu W., Shafer G. (2007): "Testing Lead-Lag Effects under Game-Theoretic Efficient Market Hypotheses", Available Online http://www.probabilityandfinance.com/articles/23.pdf

Yahoo! Finance Investing, World Stock Exchanges http://uk.biz.yahoo.com/uk_world.html

Zaidenraise K.O.S, Shmilovici A. Ben-Gal I. (2007): "Gene-Finding with the VOM Model", *J. of Computational Methods in Science and Engineering*, 7(1): 45-54.

Ziv J. (2001): "A Universal Prediction Lemma and Applications to Universal Data Compression and Prediction", *IEEE Transactions on Information Theory*, 47(4):1528- 1532.

Ziv J. (2002): "An efficient universal prediction algorithm for unknown sources with limited training data", Available at: 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.