Case-based Algorithm for Pattern Recognition and Extrapolation (APRE Method)

Branko Pecar, PhD

University of Gloucestershire
bpecar@glos.ac.uk

Abstract. The method developed and described in this paper departs from the traditional time series analysis approach. The starting premise is that a time series can be broken down into a number of characteristic cases, each of which potentially holds the key for indicating the value of the subsequent observation. The case that constitutes the beginning of the forecasting horizon (the reference case) is compared with all the past cases and the best-case match is identified. The differences between the best historical case and the reference case are used for predicting the first value in the forecasting horizon. The method is compared with some of the most prominent forecasting methods (exponential smoothing and Box-Jenkins ARIMA modelling). To validate behavioural properties of the original series versus the extrapolated values (forecasted series), measurements using the correlation (fractal) dimension are undertaken.

Keywords. Case based approach, time series analysis, pattern recognition, extrapolation, APRE, correlation dimension

1. Introduction

The subject area of time series analysis and extrapolation has been successfully tackled by a number of different approaches. Decomposition [14], exponential smoothing [5] [10] [21], stochastic modelling [4], state space modelling [9] [12] and Bayesian models [11], are some of the traditional approaches. More contemporary methods have roots in neural networks [16] and fuzzy logic [2] [20] [22], i.e. artificial intelligence domain, or genetic algorithms [7]. Pattern recognition as a subset of artificial intelligence also contributed to this subject area. The best example for this approach is Singh's PMRS[1] [18], which has origins in the paper published by Sket-Motnikar, Pisanski & Cepar [19]. Although the method we describe in this paper has resemblance to the pattern-imitation method by Sket-Motnikar, Pisanski & Cepar and Singh's PMRS, the APRE method described in this paper has been developed completely independently.

The common thread that flows through all the above approaches is some sort of a rule approach. An assumption is made that a time series belongs to one of the general processes, or probability distribution is assumed, or errors are measured and used as a rule for correction, etc. This paper will attempt to depart from this generic approach to time series analysis and introduce a different treatment of time series.

2. Possible Alternatives to Rules Approaches

A time series is a snapshot, contained in a time window, of the historical behavioural pattern of the observed variable. The clues about the direction and the dynamics of the future behaviour of the variable are hidden in the time series. If a time series represents a historical behavioural pattern, than it inevitably consists of miniature instances that define this pattern. Predictably, these instances could be called cases. Every instance, i.e. every case, holds the key for the future behaviour of the time series. In other words, if we know the case that currently characterises the variable and there is a precedent case, we can predict what the next move is likely to be.

This line of thinking inevitably leads towards the case-based reasoning (CBR) approach to problem solving. In this paradigm, specific knowledge about the behaviour of the variable is implicitly embedded in individual cases. This approach can be successfully deployed for time series extrapolation purposes, but

[1] Pattern Modelling and Recognition System.

before we define specific interpretation of cases in the time series context, we need briefly to remind ourselves of some of the starting premises of CBR.

CBR assumes that the library of past cases holds the expertise about the system behaviour, rather than encoding this behaviour by a series of rules. If we can identify (match) past cases with the current case, we have the foundation for predicting the future outcomes. CBR usually follows the process of **retrieving** similar cases, **reusing** the retrieved cases, **revising** the solution and **retaining** the solution. It is irrelevant whether a specific approach to CBR is based on trivial syntactic similarities, or more complex semantic ones. The process is the same.

3. Definition of Cases in the Time Series Context

In CBR terminology a case is a problem situation [1]. If we split time series into smaller pattern sequences, than each and every one of these patterns could be treated as a case. For example, we could break the series down into a sequence of three rolling observation patterns. In other words, the last interval in a series of, say, three observations, could contain x_n, x_{n-1} and x_{n-2}, the one before the last one x_{n-1}, x_{n-2} and x_{n-3}, and so forth until we reach x_3, x_2 and x_1.

Let us now use a linguistic equivalent to describe the series dynamics and say that every observation in the pattern that constitutes the case, in relation to the previous observation, can go up (P for positive move), down (N for negative move) or stay on the same level (Z for a zero move)[2]. This implies that we can identify intervals that consist of a series of three-observation patterns, something like PPN, PNN, NNZ, NZZ, ZZP, ZPP, etc. Each of these patterns constitutes a case. If we analyse all the identical cases, we'll probably discover that identical cases are usually followed by a similar move. This gives us the foundation to change the paradigm and think about a time series as a case history.

However, because we are dealing with numerical values that form patterns, although multiple cases keep on repeating themselves through the series and consist of identical patterns, they exhibit different magnitudes. The magnitude (or, the interval length) can be calculated as the cumulative value of all the observations in an interval. In certain cases cumulative values could apply to differences between observations in an interval, or even moving averages of the differences. These options have not been pursued in this paper and this remains one of the possibilities for further research.

Returning to the issue of the case formation, we have to say that there is no rigorous method to define what a typical case is, in other words, whether consists of only 3 rolling observations, 4, 5 or more. In accordance with the CBR approach the best way is to suspend judgement and allow the coexistence of different cases. In practical terms this implies forming a library of 3, 4, 5, etc. pattern observations and treating every group as a case category. For the sake of convenience, to speed up the computing time and restrict the storage requirements, we restricted ourselves to a maximum of 12 observations in a pattern. As this corresponds with the number of months in a year, the assumption is that cases are also capable of detecting seasonal variations.

4. Algorithmic Approach to Case Matching in Time Series

Following this logic, we could approach the task of pattern recognition and extrapolation of the time series with the following algorithm:

1. Break the series down into r-interval patterns consisting of sequential rolling observations.
2. Store all these interval-patterns as historical cases.
3. Take the last interval in the series with an arbitrary number of observations in the interval. This interval will constitute the current, i.e. the reference case.
4. Note the reference case's pattern and the measure the magnitude (e.g. the sum of all observations in the case).
5. Retrieve all the cases from the past with the identical pattern to the reference case.

[2] P, N and Z are equivalent to Singh's binary patterns [18].

6. Measure the magnitude of the movements (the length) of all these cases.
7. The retrieved matching case whose length shows the smallest distance from the length of the reference case can be used for predictions.

The prediction can be implemented in a number of ways. The simplest way is to select the observation that **follows** the closest matching historical case and add it to the reference case. Depending on the stationary character of the series, other methods can be used. Again, these possibilities are left unexplored in this paper.

The above concept of identifying unique patterns (or data strings) is comparable to Algorithmic Information Content (AIC) [6]. AIC measures the compressibility of a data string, in other words, a string (or a series) of '011011011…', could be compressed by '011' that is repeated *i* times. A random series will exhibit very high AIC, whilst low AIC indicates a very predictable curve of a given shape. This implies that a low AIC series can be modelled using time as an independent factor (curve fitting, for example). A high AIC series can only be modelled by a probabilistic model, where every observation follows some probability distribution function. However, medium AIC is most difficult to model, and most real-life series fall into this category. Most of the medium AIC series can be reasonably successfully modelled by a self-referential model (autoregressive, for example). However, identification of the correct model can be very difficult and could be considered the domain of art, rather that science.

5. Forecasting Using the Case Based Approach to Time Series Analysis

Rather than identifying self-referential models, which would take us back to rule based approaches, we'll propose the notation that would enable us to deploy an algorithm for case based reasoning, as described above. We'll call this new approach an **A**lgorithmic approach to time series **P**attern **R**ecognition and **E**xtrapolation, and we'll refer to it as the APRE method.

The variables are defined as follows:

n = number of observations in the series
x_n = the n^{th} observation in the series
i = total number of cases containing r elements
r = number of elements in every case
t_i= specific case
k_q= magnitude, or, the length of every case (sum of all observations)
d_{q-r+1}= case distance (differences between case magnitudes)

For j=1, … , n-1 we first find differences:

$$m_j = x_j - x_{j-1} \tag{1}$$

The letters P, N or Z are assigned according to the value of difference:

$$m_j = \begin{cases} \text{"P"} & \text{for} \quad m_j > 0 \\ \text{"N"} & \quad\quad\;\; m_j < 0 \\ \text{"Z"} & \quad\quad\;\; m_j = 0 \end{cases} \tag{2}$$

For i = 1, 2, …, (n - r + 1) and $p = i$, …, (i + r - 1), let $t_{i,p}$ represent cases with different number of observation differences:

$$t_{i,p} = \{m_p\} \tag{3}$$

For $q = r, \ldots, n$ we calculated the magnitude for every case:

$$k_q = \sum_{j=q-r+1}^{q} x_j \tag{4}$$

As mentioned earlier, in some instances perhaps a better method of calculating k_q is:

$$k_q = \sum_{j=q-r+1}^{q} \frac{x_j}{r} \tag{4a}$$

The above effectively means that we have taken moving averages as a magnitude value for every case.

We now select an arbitrary point in the series and declare that the case whose last observation in the pattern coincides with this particular point is the reference case:

$$t_{c,p} = \{m_p\} \tag{5}$$

Where index c represents the beginning of the ex-post forecasting horizon.

Once we know the value of k_c for the reference case, we measure the distances between this case and all the historical ones for $q = r+1, \ldots, n$:

$$d_{q-r+1} = k_c - \mathrm{k}_{q-r+1} \tag{6}$$

However, we only do this for cases that exhibit the same movement pattern as the reference case, i.e. for $c = 1, \ldots, (n - r)$ where:

$$t_{c,p} = t_{n-r,p} \tag{7}$$

Where the value of d_{q-r+1} = min, we can take the value of the observation x_{j+1}, succeeding the case $t_{c,p}$, as the optimum for this pattern extrapolation and use it to get the forecast for x_{n+1}. In other words, for d_{q-r+1} = min,

$$x_{n+1} = x_{j+1} \tag{8}$$

Newly generated observation (forecast) becomes part of the new reference case. This enables us to extrapolate beyond one observation in the future. We continue reusing the cases from the historical period to render ex-ante forecasting.

If we took case magnitudes to be the moving averages of the series, then the forecast could take multiplicative form, where:

For d_{q-r+1} = min, $d_{fq-r+1} = d^*$, and the forecast is:

$$x_{n+1} = x_n \times d^* \tag{8a}$$

However, one has to be very cautious with the approach given in (8a) as it can produce dynamics that are far too excessive and can resemble forecasts obtained using a large smoothing constant value in exponential smoothing methods. As indicated earlier, this option has not been pursued in this paper.

6. An Example of the APRE Method

A detailed VB code that was compiled to automate the APRE method[3] enabled a comparative evaluation of this method vs. several other well-established time series analysis methods. The methods selected for comparison were Single, Double and Triple Exponential Smoothing, Holt's Two-Parameter Smoothing Method, and the Box-Jenkins ARIMA method[4]. The series selected for analysis was picked virtually randomly from the stock exchange. To make it more interesting, a decision was made to look at the value of

[3] The author welcomes any interest in the code (embedded in an Excel spreadsheet) and can be contacted via email.
[4] Further in the text the methods are abbreviated as SES, DES, TES, Holt, ARIMA and APRE.

well-known company's shares and Microsoft was selected. A five-year period from 9 March 1998 to 14 October 2002 was chosen. The data represented the closing weekly value of Microsoft shares.

The values of the smoothing constants for the comparison methods were set to: 0.5 for alpha and 0.05 for Holt's gamma. The autocorrelation and partial autocorrelation analysis showed that Microsoft shares follow the first order autoregressive process, and ARMA (1,0,0) model was used for the Box-Jenkins method. For the APRE method, the cases stored cover pattern intervals from 2 to a maximum of 12 moving observations per case.

The series consists of 240 observations, but the first 120 values were used as a basis for the initial case reasoning. Although most of the comparison methods used do not require a training period, to maintain consistency the ex-post forecasts were produced for all the methods starting with the period 121. For all the methods except SES[5], a forecasting horizon of 10 future observations was arbitrarily selected. However, the errors were considered for only ex-post forecasts from the 121st to the 240th observation. Three types of errors were measured, and they were: ME (Mean Error), MAD (Mean Absolute Deviation) and MSE (Mean Square Error). The results rendered are given in Fig. 1.

Using the above parameters the APRE method failed to outperform any of the comparison methods. However, the three error measurements selected (most often used by the majority of forecasters) only measure certain qualities of forecasts. The extrapolation capabilities beyond just one single future observation of some of these methods are questionable. Even when the forecast is extrapolated further into the future, it usually follows just the underlying trend, making it no more valuable than the ordinary trend extrapolation method. From this point of view, the APRE method exhibits qualitatively different characteristics. The APRE method has inherent freedom to oscillate as freely as the original series that it tries to mimic, nor is series homoscedasticity a requirement. The oscillations are easily extrapolated in the future, giving it a unique property to propagate the dynamics of the series and render pulsating ex-ante forecasts.

Method	ME	Rank	MAD	Rank	MSE	Rank
SES	-0.39	5	3.26	4	16.95	3
DES	-0.03	2	2.85	2	15.52	2
Holt	-0.05	3	3.25	3	17.51	4
TES	-0.02	1	3.35	5	18.86	5
ARIMA	-0.28	4	2.59	1	11.33	1
APRE	-0.49	6	3.69	6	21.75	6

Fig. 1. Error comparison and performance ranking between selected forecasting methods

Below, Fig. 2 to Fig. 4 demonstrate how well the original Microsoft shares values were approximated and show the character (appearance) of all the ex-ante forecasts.

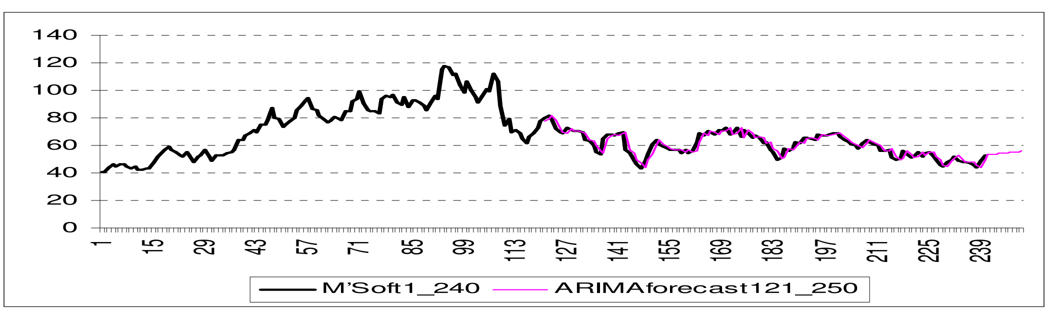


Fig. 2. Forecasts using Box-Jenkins approach to ARIMA modelling

[5] SES can extrapolate only one observation ahead.

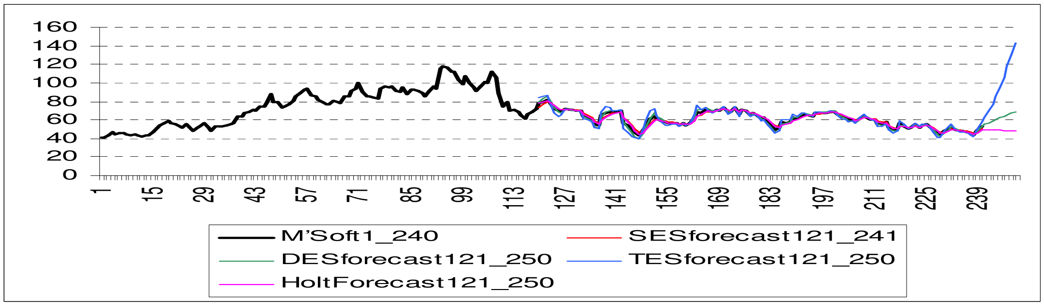


Fig. 3. Forecasts using single (SES), double (DES), triple (TES) exponential smoothing and Holt's exponential smoothing method

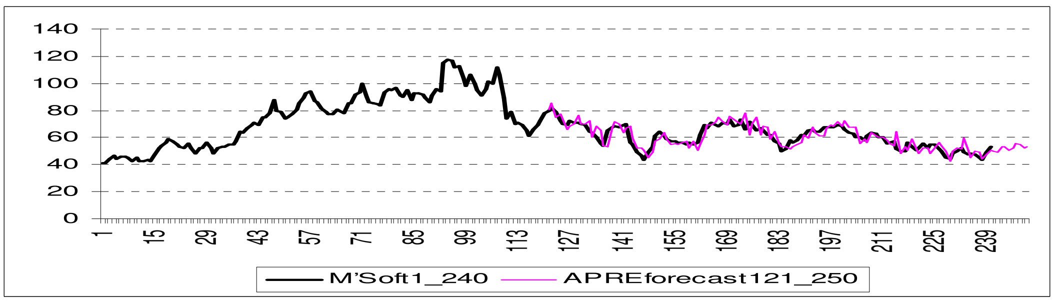


Fig. 4. Forecasts using the APRE method

Fig. 5 shows the autocorrelation function for forecasting errors for the APRE method, indicating no serial correlation and implying that the algorithm shows no bias.

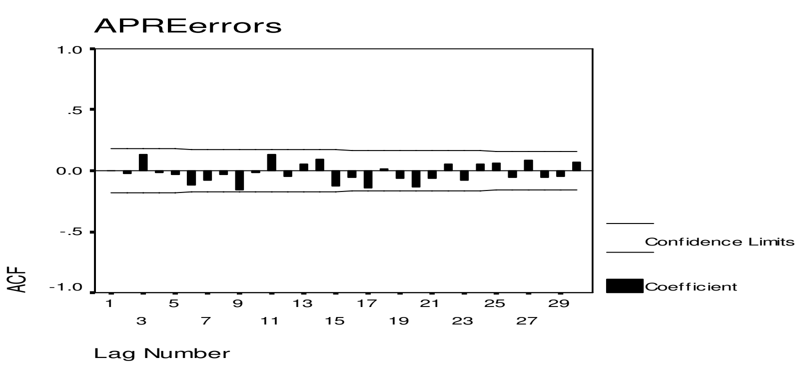


Fig. 5. Autocorrelation values of forecasting errors obtained using the APRE method

As indicated above, extrapolated observations of most of the forecasting methods do not have the same 'appearance' as the original series that they try to approximate. The fundamental question is how to provide a rigorous measure of the appearance of a time series?

7. Measuring the Appearance of Forecasts

Unfortunately the methodology in this area is still somewhat obscure. In the context of fractal measurements, Mandelbrot uses lacunarity [14] as a measure of texture, although there are no rigorous proofs indicating how different levels of lacunarity are related to various patterns. Another approach is to measure the spectral densities but this could lead to further confusion, as it is often difficult, from just a power spectrum, to differentiate a random from a deterministic, yet chaotic series.

One of the metrics that could be used is the fractal dimension, and there are numerous ways to measure fractal dimension. The Hausdorff-Besicovitch dimension is often equated to fractal dimension, which indicates that some kind of a capacity measurement, such as box-counting method, could be used. However, we decided to use the correlation dimension, as it is one of the most rigorous quantifiers of fractal dimension. We used the Grassberger-Procaccia algorithm [8] for calculating the correlation dimension.

The correlation dimension is calculated as the ratio between the correlation integral and some small distance that defines the number of pairs of vectors created by the method of delay coordinates (embedding dimension)[6]. It is, in fact, the slope of the logarithmic values of the correlation integral and the logarithmic values for different radii (embedding dimension of the recurrence plot) [17]. In a way, the correlation dimension quantifies the amount of self-similarity in a chaotic series, where low dimensional chaos is present [14]. For random, periodic or quasi-periodic series the correlation dimension will show either integer values or infinity. Chaotic series show a non-integer value greater than 1. It is important to note that the correlation dimension does not measure temporal, but spatial properties of the series.

8. Final Validation

In order to make further comparisons possible, another extrapolation of the same data set was rendered. This forecast started from the 181st observation and 120 new values were generated, i.e. the first 60 were ex-post forecasts based on the actual data and the last 60 were based completely on forecasted values produced by the APRE method. Fig. 6 contains basic statistics taken for this new series and compared with the original data.

	Original Microsoft series	New series of 120 forecasts
Mean Value	67.79	56.23
Variance	299.67	59.49
Correlation dimension[7]	1.54	2.07

Fig. 6. Statistical properties of the original and forecasted data set using the APRE method

It is quite evident that some variability exists, i.e. the variance for the forecasted series is much smaller, indicating some conservatism in methodology. In the Appendix we separated the original Microsoft series into two segments, the first one between the 1st to 120th observation and the second one between the 121st and the 240th observation. Various measures are displayed comparing these two segments with the two different APRE forecasts (the one between 121st and 250th observation and the other one between 181st and 300th observation). It is quite evident that the variations between the original series and the APRE method forecasts do not differ more than the internal variations between the first part and the second part of the original series. The conclusions are that the Microsoft series exhibits great dynamism and variability, and that despite some discrepancies between the original series and the forecasts, APRE method is performing quite well and consistently.

The correlation dimensions provide the foundation for some interesting speculations. In general, the correlation dimension provides a measure of complexity for the underlying attractor of the system, but we used it here only as a comparative measure between the two series. In this paper we are using it primarily as an indication of the texture of the series. The above numbers show that the APRE method forecasts ex-

[6] Another approach is to think about the correlation integral as the probability of finding a point in the series within some i-th subcube of phase space which is divided into a number of subcubes of some arbitrary side length.

[7] The time delay τ corresponds to the first zero autocorrelation for each series (40 for the original series and 15 for the forecast) and the embedding dimensions are selected on the basis of calculating false nearest neighbours (in both cases the embedding dimension was selected to be 2).

hibit similar space-filling capabilities as the original series. In fact, the forecasts show even greater level of complexity than the original series. Fig. 7 shows the original series and the 60-observation ex-ante forecasting horizon.

Comparing the spectral densities, the autocorrelation and partial autocorrelation functions of the original Microsoft series and the APRE forecast, given in the Appendix, demonstrate that these two time series do not exhibit great differences in character and that they both belong to the same family of processes.

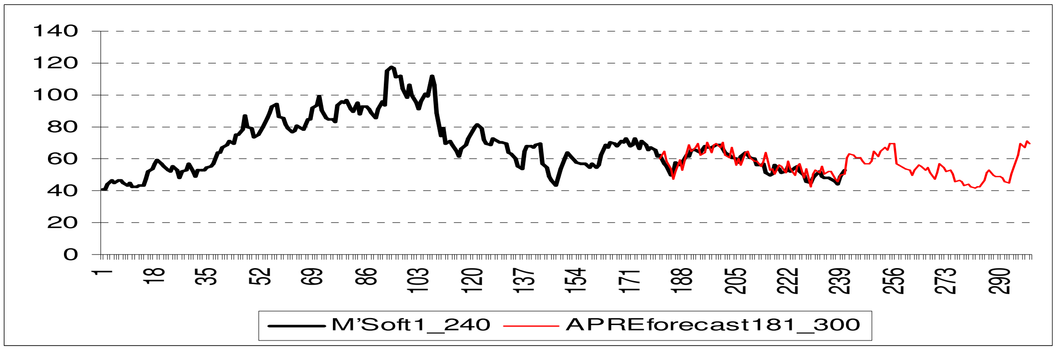


Fig. 7. Long range forecast (60 future observations) using the APRE method

9. Conclusion

The APRE method departs from the usual rule-based approach to time series analysis and introduces a particular form of case-based reasoning. Using the traditional error measurement, the results rendered by the APRE method are good but in this particular example somewhat inferior to conventional time series analysis methods. Although the method shows reasonable accuracy, the issue of the method's precision needs to be addressed. Most of the practical time series are of limited length and the number of cases (patterns) that could be identified and stored is limited. This implies that there is more than a strong probability that certain reference cases will not find a match, either pattern-wise, or in terms of the case magnitude (the value). However, if the historical case pattern is identified, but not the magnitude, a distance between the reference case and the historical case can be measured. The probabilities for various instances can be calculated and weighted by the distances, which could serve as the foundation for establishing conditional probabilities. On this basis, a Bayesian inference could be applied to improve the method's precision and accuracy.

Despite the fact that there is some room for improvements, the APRE method provides a valid alternative to most commonly used rule based or model free methods for time series analysis. The forecasting horizon is not limited by the properties of the method, as with most rule-based methods, but by the richness of cases and decision-making context. The ex-ante forecasts produced using the APRE method preserve the texture and the complexity of the original data set, making it unique among the methods.

References

1. Aamodt Agnar & Enric Plaza: *Case Based Reasoning: Foundational Issues, Methodological Variations and Systems Approach,* AICom – Artificial Intelligence Communications, Vol. 7, No. 1, March 1994

2. Altrock von Constantin: *Fuzzy Logic and Neuro Fuzzy Applications Explained*, Prentice Hall, Englewood Cliffs, 1995.

3. Barnsley M.F., Devaney R.L., Mandelbrot B.B., Peitgen H.-O., Saupe D. and Voss R. F.: *The Science of Fractal Images*, Springer-Verlag, New York, 1988.

4. Box E.P. George & Gwilym M. Jenkins: *Time Series Analysis: Forecasting and Control*, Holden Day, San Francisco, 1970.

5. Brown Robert: *Statistical Forecasting for Inventory Control*, McGraw Hill Book Co., New York, 1958.

6. Dooley Kevin: “Complexity in Time Series Modelling”, Society for Chaos Theory in Psychology and Life Sciences, 2(3), 1994.

7. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

8. Grassberger Peter & Procaccia Itamar: “Measuring the Strangeness of Strange Attractors”, *Physica* 9D, pp. 189-208, 1983.

9. Harvey C. Andrew: *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, 1989.

10. Holt C.C.: *Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages*, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957.

11. Jensen V. Finn: *An Introduction to Bayesian Networks*, UCL Press, London, 1996.

12. Kalman R.E.: “A new Approach to Linear Filtering and Prediction Problems”, *Journal of Basic Engineering*, D. 82, 1960.

13. Mandelbrot B. Benoit: *The Fractal Geometry of Nature*, W H Freeman, New York, 1983.

14. Kantz Holger & Thomas Schreiber: *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, 2000.

15. McLaughlin R.L.: “Time Series Forecasting”, *Marketing Research Technique*, Series No. 6, AMA, 1962.

16. Rumellhart, Hinton & Williams: "Learning representations by back-propagating errors", *Nature*, Vol. 323, 1986.

17. Schroeder M.: *Fractals, Chaos, Power Laws*, W H Freeman, New York, 1991.

18. Singh Sameer: “Multiple Forecasting using Local Approximation”, *Pattern Recognition*, Vol. 33, 2000.

19. Sket-Motnikar B., Pisanski T. & Cepar D.: “Time-series forecasting by pattern imitation”, *OR Spektrum* 18, pp. 43-49, 1996.

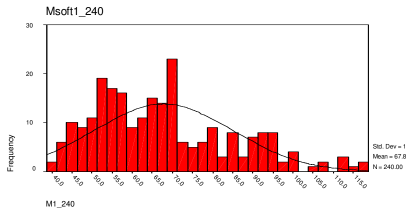
20. Tsoukalas H. Lefteri & Uhrig E. Robert: *Fuzzy and Neural Approaches in Engineering*, John Wiley, New York, 1997.

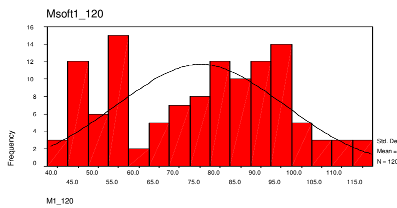
21. Winters R. Peter: ”Forecasting Sales by Exponentially Weighted Moving Averages”, *Management Science*, April 1960.

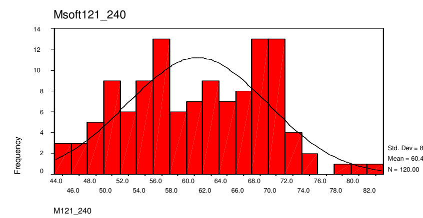
22. Zadeh A. Lotfi, Fu King-Sun, Tanaka Kokichi & Shimura Masamichi (eds): *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, Academic Press, New York, 1975.

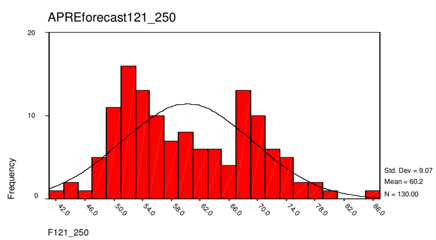
Appendix

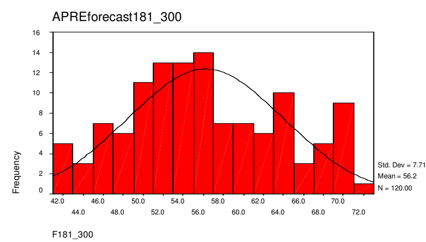
1. Distribution properties of the full 240 observation long series, the first and the second half of the series, APRE generated forecast from the 121st to the 250th observation (10 ex-ante forecasts) and APRE generated forecast from the 181st to 300th observation (60 ex-ante forecasts).





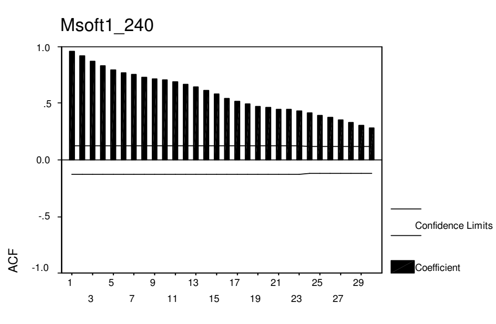




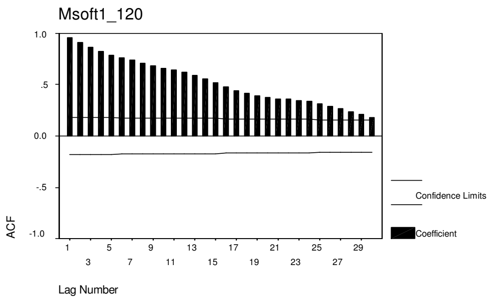


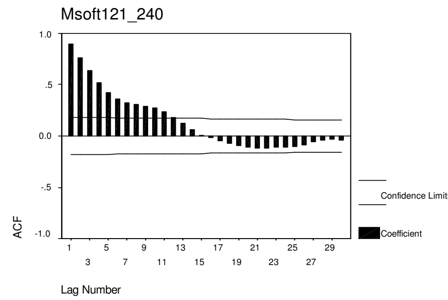
Statistic	Msoft1_240	Msoft1_120	Msoft121_240	Msoft121_250	Msoft181_300
Mean	67.79	75.18	60.40	60.23	56.23
Standard Error	1.12	1.87	0.78	0.80	0.70
Median	65.34	77.91	60.73	58.94	55.13
Mode	46.06	46.06	60.31	69.56	53.37
Standard Deviatic	17.31	20.47	8.53	9.07	7.71
Sample Variance	299.67	418.99	72.73	82.18	59.49
Kurtosis	-0.11	-1.02	-0.71	-0.74	-0.83
Skewness	0.72	-0.03	0.02	0.35	0.16
Range	76.54	76.54	38.62	43.21	29.36
Minimum	40.90	40.90	43.38	42.29	41.75
Maximum	117.44	117.44	82.00	85.50	71.11
Sum	16268.93	9021.24	7247.69	7830.13	6747.50
Count	240	120	120	130	120

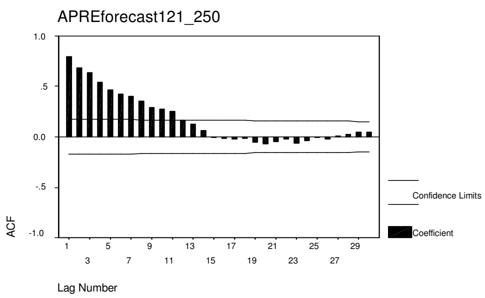
2. Autocorrelation properties of the full 240 observation long series, the first and the second half of the series, APRE generated forecast from the 121st to the 250th observation (10 ex-ante forecasts) and APRE generated forecast from the 181st to 300th observation (60 ex-ante forecasts).

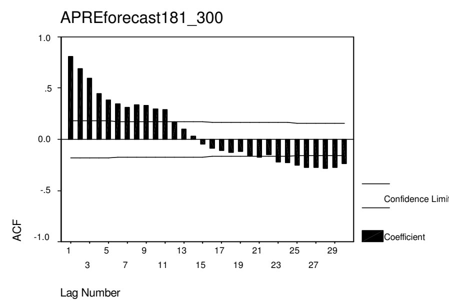


The first 40 ACF are non-zero.

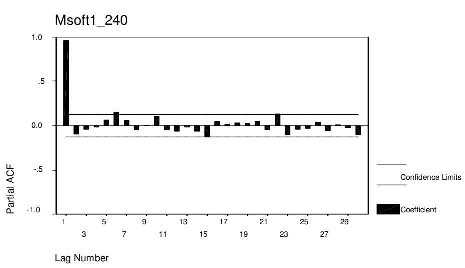


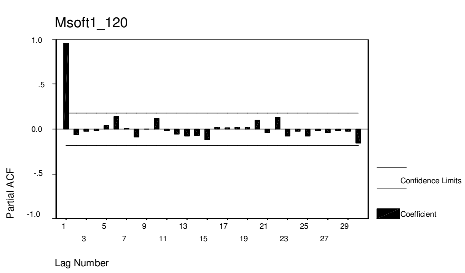


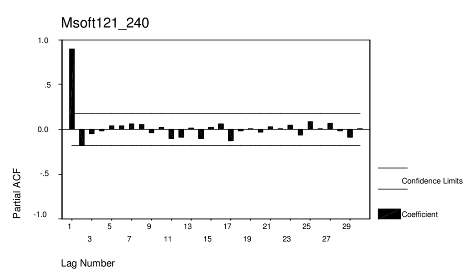


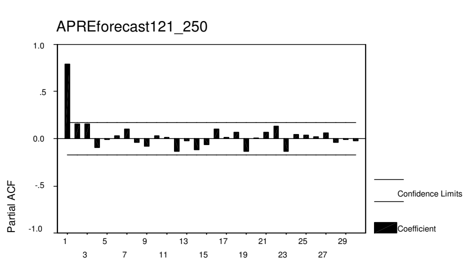


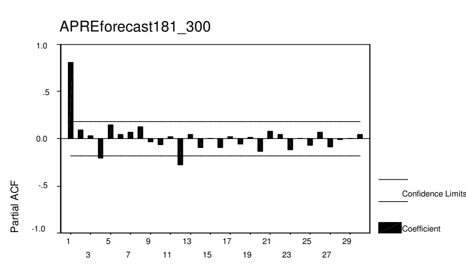
3. Partial autocorrelation properties of the full 240 observation long series, the first and the second half of the series, APRE generated forecast from the 121st to the 250th observation (10 ex-ante forecasts) and APRE generated forecast from the 181st to 300th observation (60 ex-ante forecasts).











4. Spectral density properties of the full 240 observation long series, the first and the second half of the series, APRE generated forecast from the 121st to the 250th observation (10 ex-ante forecasts) and APRE generated forecast from the 181st to 300th observation (60 ex-ante forecasts).