Entropy as an A Priori Indicator of Forecastability

Peter M. Catt

Department of Computing and Information Technology, UNITEC, Auckland 1025, New Zealand;
E-Mail: peter.catt@soltius.co.nz

Abstract: The ability to accurately determine the *a priori* forecastability of a time series is an important endeavour for forecasting practitioners as it provides guidance on the potential for accurate forecasts and the associated degree of effort that is warranted. Measures of *entropy*, such as sample entropy, provide an assessment of the regularity or similarity within a time series. We posit that series with low in-sample entropy, i.e. high regularity, will positively correlate with low out-of-sample forecast error, as measured by the mean absolute scaled error (MASE). To assess this we adopt the 3003 time series used in the M3 forecasting competition spanning micro, industry, macro, finance, and demographic series. We calculate the in-sample sample entropy and out-of-sample MASE for all 3003 series using a common univariate forecasting method. We demonstrate that the *a priori* sample entropy is indeed a useful predictor of out-of-sample forecast performance, subject to the well-researched problem of structural breaks. We recommend that forecasting practitioners adopt such entropy measures, alongside well-established tests for seasonality and trend, to better understand the likelihood of successful forecasting outcomes.

Keywords: sample entropy; forecasting; time series analysis

1. Introduction

The ability to assess the forecastability of a time series is one of major interest to forecasting practitioners [1]. *A priori* assessment of forecastability equips forecasters with the ability to determine the likely outcome of forecasting a given time series and the corresponding degree of forecasting effort that is warranted. Large forecast errors naturally motivate efforts to improve the forecasting approach. However, if a given time series simply cannot be accurately forecast, then spending time and resources on attempts to reduce forecast error will be ineffective. In such situations the practitioner should instead investigate mitigating the effects of poor forecasts [2]. The forecastability of a time series is related to the regularity of the data, with a less regular series typically more difficult to forecast and achieve an acceptable level of forecast accuracy [3].

Commonly adopted methods to assess forecastable components within a time series include well established seasonal and trend approaches such as autocorrelation and spectral analysis for seasonality and periodicity and the estimation of trends with unit root tests, e.g. Dickey-Fuller [4]. However, the use of entropy metrics can augment such approaches by providing an estimate of time series regularity.

The ability to quantify the degree of regularity in a time series is an essential task in understanding the behaviour of a system [5], with such systems or underlying data generating processes being broadly characterised as deterministic, chaotic, complex, or random [6]. Measures of time series entropy have now seen widespread adoption in biomedicine for time series analysis, e.g. epilepsy studies, anesthesia, cognitive neuroscience and heart rate variability [7-10], but have yet to be shown to be useful in the domain of time series forecasting.

The remainder of this paper is organised as follows. Section 2 introduces the concept of entropy in the context of information theory and details the sample entropy metric, along with providing an illustration. Next, Section 3 describes the datasets and performance metrics used in this paper and outlines the methodology used to assess sample entropy. Section 4 reports and discusses the results obtained and Section 5 concludes the paper.

2. Entropy

The notion of entropy as a physical measure was originally introduced in the 19th century by Clausius as the Second Law of Thermodynamics, which suggests that the entropy of an isolated system tends to increase continuously until it reaches thermodynamic equilibrium, a state with maximum entropy. This was later advanced by Boltzmann as a quantification of the entropy of a system based on the number of possible microstates of a macroscopic system and therefore describes the disorder within a given system. Shannon [11] proposed his information content formula to describe the level of knowledge, or alternatively, level of ignorance associated with a sequence where several outcomes are possible, and their probabilities duly assigned. Therefore, a sequence with high regularity, e.g. a sine wave, represents low entropy and a sequence with a uniform distribution, e.g. white noise, represents high entropy. Consequently, the history and future of a sequence with high entropy are independent.

2.1. Sample Entropy

This section introduces the Sample Entropy metric as proposed by Richman & Moorman [8]. Pincus [7] proposed Approximate Entropy (ApEn) to measure the probability that two segments of data within a time series that are similar will remain similar when the length of both segments is incremented by one. Smaller ApEn values indicate a greater chance that a set of data will be followed by similar data (regularity). Conversely, a larger value of ApEn indicates a lower chance of similar data being repeated (irregularity). Hence, larger values convey more disorder, randomness and system complexity. The ApEn algorithm counts each sequence as matching itself to avoid the occurrence of ln(0) in the calculations. This step might cause bias of ApEn and this bias causes ApEn to have two poor properties in practice; First, ApEn is heavily dependent on the record length and is uniformly lower than expected for short records. Second, it lacks relative consistency. That is, if the ApEn of one data set is higher than that of another, it should, but does not, remain higher for all conditions tested [12]. Sample entropy (SampEn) was proposed by Richman and Moorman [8] as an alternative to ApEn that displays relative consistency and less dependence on data length.

First of all, we consider only the first $N - m$ vectors of a length of *m* of Equation (1), ensuring that, for $1 \leq i \leq N - m, \mathrm{X}_i^m$ and X_i^{m+1} are defined. Then, we define $B_i^m(r)$ as $(N - m - 1)^{-1}$ times the number of vectors, X_j^m, within a distance, *r*, from X_i^m, where $j = 1, 2, \dots, N - m$, but also $j \neq i$, excluding in this way the self-matches, and set:

$$B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r) \qquad (1)$$

Correspondingly, we define $A_i^m(r)$ as $(N-m-1)^{-1}$ times the number of vectors, X_j^{m+1}, within a distance, *r*, from X_i^{m+1}, where $j = 1, 2, \ldots, N-m$ with $j \neq i$, and:

$$A^m(r) = (N-m)^{-1} \sum_{i=1}^{N-m} A_i^m(r) \tag{2}$$

The SampEn (*m*; *r*) is then defined as:

$$SampEn\ (m, r) = \lim_{N \to \infty} \{-\ln[A^m(r)/B^m(r)]\} \tag{3}$$

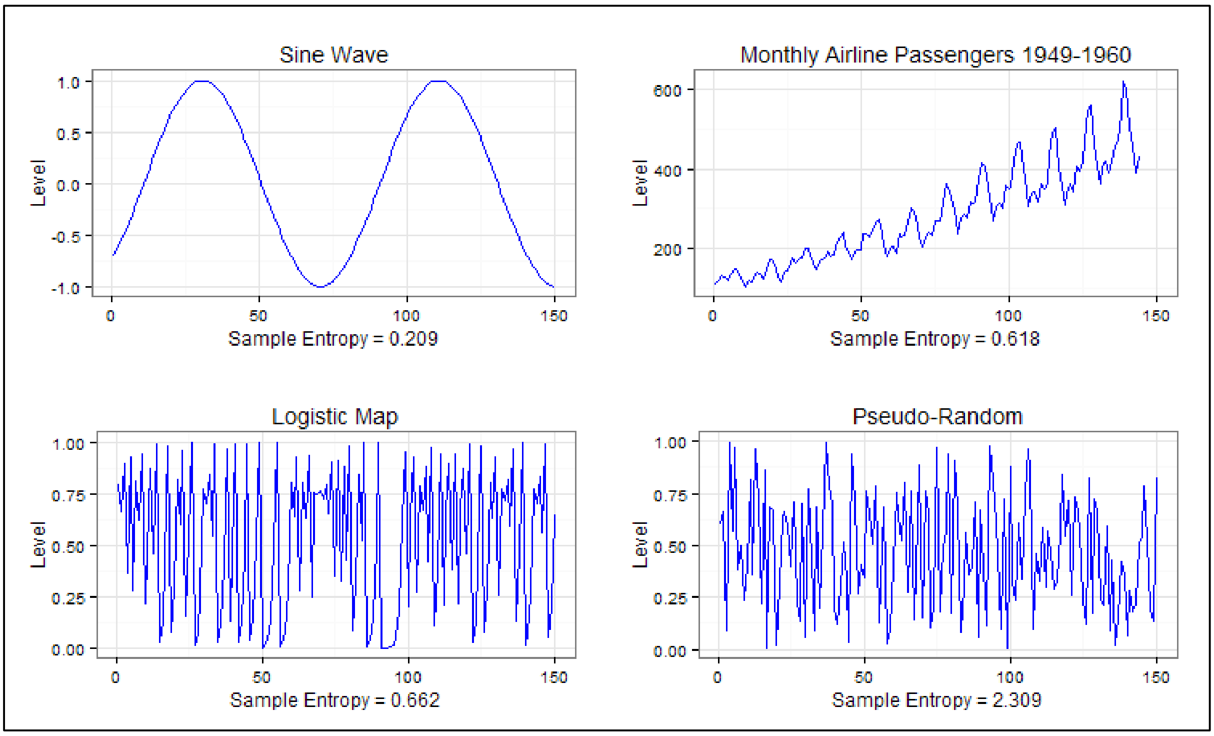
which, for finite time series, can be estimated by the statistic:

$$SampEn\ (m, r, N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{4}$$

2.2. Sample Entropy Illustration

To provide an illustration, Figure 1 presents a continuum of time series from entirely deterministic to random, demonstrated with four time series; the entirely deterministic sine wave and monthly airline passenger numbers 1949-1960 [13], the chaotic logistic map and a pseudo-random sequence generated in the R statistical programming language [14].

Figure 1. Continuum of Time Series and Corresponding Sample Entropies



The SampEn of the sine wave is the lowest of the four series with a value of 0.209 (highest regularity) and the pseudo-random series highest with a SampEn of 2.309 (lowest regularity). Notably, the SampEn correctly identified the relatively high regularity of the chaotic logistic map series, a degree of regularity not immediately apparent upon inspection of the chart.

3. Method: Datasets and Performance Metrics

This section describes the datasets and performance metrics used in this study. We subsequently outline the methodology used to assess the metrics built on these datasets.

3.1. Dataset Characteristics

We employ the M3 forecasting competition data set [15] as a well-researched and freely available public domain data set. The M3-Competition time series include yearly, quarterly, monthly and other time series. Time series also included the following domains: micro, industry, macro, finance, demographic, and other (telecommunications and utilities series). Table 1 shows the number of time series based on both time interval and domain. To ensure sufficient data were available to develop an accurate forecasting model, minimum thresholds were set for the number of observations: 14 for yearly series, 16 for quarterly series, 48 for monthly series and 60 for other series (see Table 2).

Table 1. The classification of the 3003 time series used in the M3-Competition

Time interval between successive observations	*Types of time series data*						
	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	146	102	83	58	245	11	645
Quarterly	204	83	336	76	57		756
Monthly	474	334	312	145	111	52	1428
Other	4			29		141	174
Total	828	519	731	308	413	204	3003

Table 2. Data length of the 3003 time series used in the M3-Competition

Time interval between successive observations	*Data length*					
	Total	Min	Median	Mean	Max	Forecast Horizon
Yearly	645	14	19	22	41	6
Quarterly	756	16	44	41	64	8
Monthly	1428	48	115	99	126	18
Other	174	60	63	69	96	8

3.2. Performance Evaluation

Each series was divided into an estimation (in-sample) data set for estimating the model and out-of-sample (holdout) data set. The out-of-sample data set provided for an actual versus forecast comparison of six periods for yearly, eight for quarterly, 18 for monthly and eight for the category 'other' [15].

Our primary aim was the *a priori* assessment of SampEn as an indicator of forecastability, not an attempt to identify the best forecast model for each of the 3003 series of the M3-Competition. As simple univariate methods have shown to perform well in large scale forecasting competitions [15,16], Holt's linear trend method was utilised for this study. Holt's method considers trend in the time series by calculating an estimate of level and an estimate of trend in the series. The two estimates are then combined to form the forecast value [17].

3.2.1 Holt's Linear Trend Method:

$$F_{t+m} = L_t + b_t m \quad (5)$$

where $L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$ (Level) (6)

$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$ (Trend) (7)

Y_t = actual value in period t

F_t = forecast value in period t computed when period t-1

m = periods ahead

α = smoothing of level

β = smoothing of trend

Forecasts were calculated for in-sample fitted (one-step ahead) forecasts and then one-step ahead forecasts were calculated for the entire out-of-sample period for each of the 3003 time series of the M3-Competition. The out-of-sample forecasts applied the parameters from the prior in-sample fitting and were not re-estimated out-of-sample. Estimation of the forecast model was performed using the log-likelihood approach described in Hyndman and Khandakar [18].

3.2.2. Mean Absolute Scaled Error:

The mean absolute scaled error (MASE) was proposed by Hyndman and Koehler [19] as a forecast accuracy measure that is not subject to the relatively common problem of infinite or undefined values found in other widely adopted measures.

The mean absolute scaled error is given by:

$$\text{MASE} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{|e_t|}{\frac{n}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|}\right) = \frac{\sum_{t=1}^{n}|e_t|}{\frac{n}{n-1}\sum_{i=2}^{n}|Y_i - Y_{i-1}|} \quad (8)$$

where the numerator e_t is the forecast error for a given period, defined as the actual value (Y_t) minus the forecast value (F_t) for that period: $e_t = Y_t - F_t$, and the denominator is the average forecast error of the one-step "naive forecast method", which uses the actual value from the prior period as the forecast: $F_t = Y_t-1$. MASE is a normalized measure, that is, MASE < 1 means that the prediction error was, on the average, smaller than the error of a random-walk forecast on the same data. Conversely, MASE > 1 means that the prediction method did worse, on average, than the random-walk method. The MASE was chosen as it is easy to interpret and allows comparison of forecast accuracy across series on different scales [19]. For the purposes of this study we substituted the naïve (random-walk) forecast benchmark of the MASE with the mean of the in-sample series so as to achieve a greater relativity between Holt's linear trend method and the benchmark model. The naïve method is known to perform particularly well in certain situations, e.g. in cases of structural breaks [20], and would therefore not necessarily maximise relative error.

Forecast performance was determined by calculating the MASE of Holt's method for all 3003 time series, both in-sample and out-of-sample. SampEn was calculated, using the 'pracma' package in R [21], for each of the 3003 in-sample time series and the corresponding coefficients of determination produced for the MASE of both the in-sample and out-of-sample Holt's method. The SampEn r value (similarity criterion) was determined by a simple grid search to maximise the in-sample coefficient of determination with the in-sample MASE. No attempt was made to mitigate or remove series with structural breaks and/or changes in out-of-sample coefficient of variation relative to in-sample, as such departures from the in-sample data accurately represent the reality of real-world forecasting for the practitioner.

4. Results and Discussion

Figure 2 shows a scatterplot for all series of the SampEn (in-sample) and the Mean Absolute Scaled Error (in-sample) with a strong coefficient of determination of R^2=0.87. Followed by Figure 3 showing the scatterplot for all series of SampEn and Mean Absolute Scaled Error (out-of-sample) with a moderate coefficient of determination of R^2=0.56.

Figure 2. Scatterplot of Sample Entropy and Mean Absolute Scaled Error (in-sample)

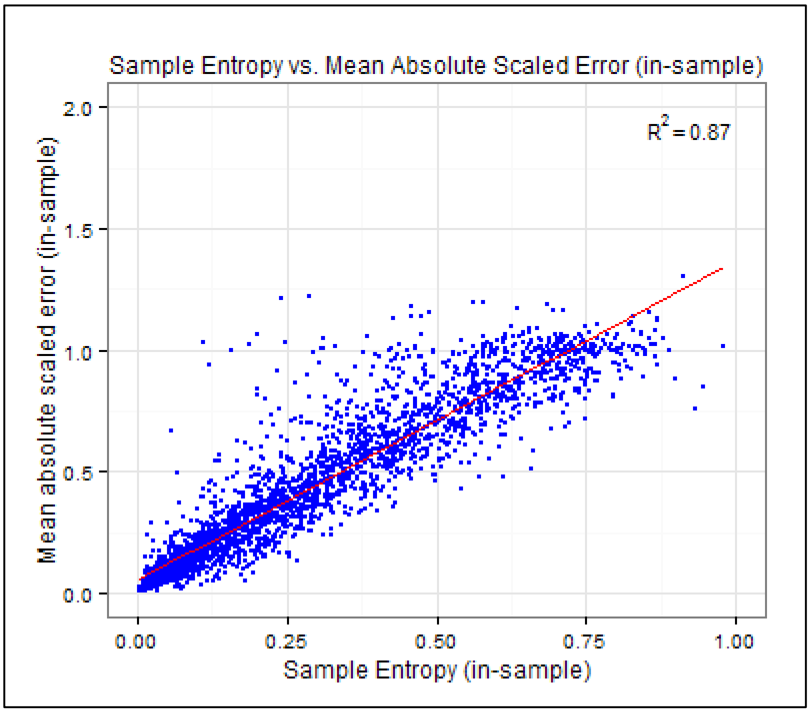


Figure 3. Scatterplot of Sample Entropy and Mean Absolute Scaled Error (out-of-sample)

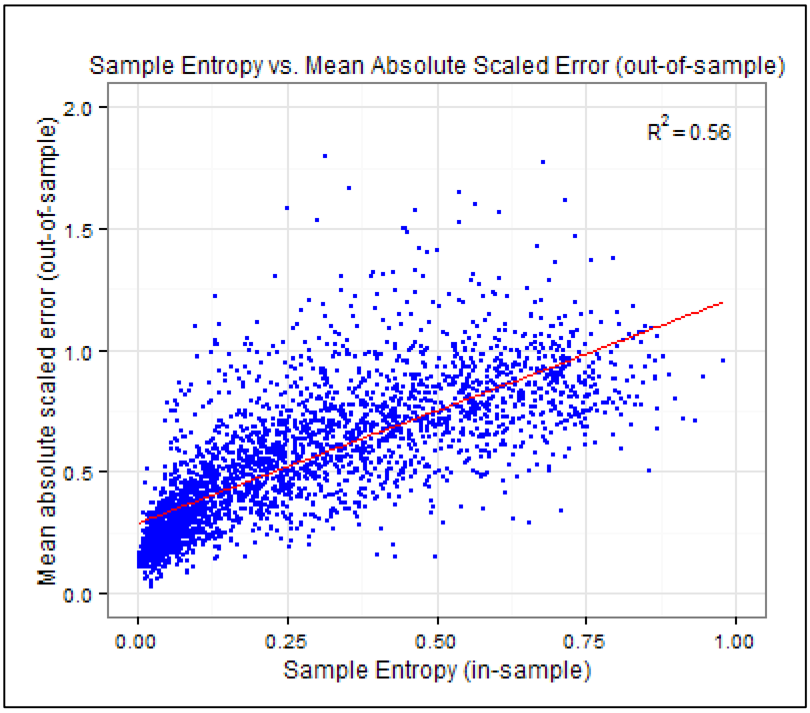


Table 2 contains the in-sample and out-of-sample sample sizes, coefficients of determination, correlation coefficients and corresponding 95% confidence intervals for the SampEn and Mean Absolute Scaled Error correlations, by time interval. The in-sample R^2 ranges from 0.77 to 0.95 and the out-of-sample R^2 ranges from 0.25 to 0.75. Table 3 contains the in-sample and out-of-sample sample size, coefficients of determination, correlation coefficients and corresponding 95% confidence intervals for the SampEn and Mean Absolute Scaled Error correlations, by series type. The in-sample R^2 ranges from 0.75 to 0.98 and the out-of-sample R^2 ranges from 0.33 to 0.85.

The lowest out-of-sample coefficients of determination apply to the "yearly" interval series (R^2=0.25) and "micro" type series (R^2=0.33), whereas the highest out-of-sample coefficients of determination related to the "other" interval series (R^2=0.75) and "other" type series (R^2=0.85), which included telecommunications and utilities series. Notably, the "other" type series of telecommunications and utilities data was characterised by continuously trending series over both the in-sample and out-of-sample periods. Such trending lends itself to successful forecasting with Holt's method.

Table 2. Sample Entropy and Mean Absolute Scaled Error correlations by time interval

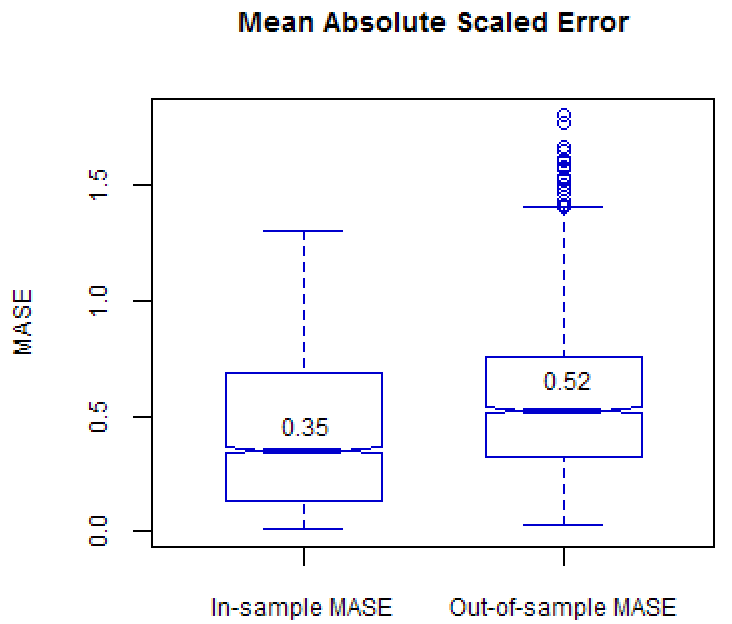
Series	*N*	*Sample*	*R^2*	*R*	*95% CI*
All Series	3003	In-sample	0.87	0.93	0.93-0.94
		Out-of-sample	0.56	0.75	0.73-0.76
Yearly	645	In-sample	0.79	0.89	0.87-0.90
		Out-of-sample	0.25	0.50	0.45-0.56
Quarterly	756	In-sample	0.77	0.88	0.86-0.90
		Out-of-sample	0.55	0.74	0.71-0.77
Monthly	1428	In-sample	0.90	0.95	0.95-0.96
		Out-of-sample	0.65	0.80	0.78-0.82
Other	174	In-sample	0.95	0.98	0.97-0.98
		Out-of-sample	0.75	0.87	0.82-0.90

Table 3. Sample Entropy and Mean Absolute Scaled Error correlations by series type

Series	*N*	*Sample*	*R^2*	*R*	*95% CI*
Micro	828	In-sample	0.79	0.89	0.87-0.90
		Out-of-sample	0.33	0.57	0.53-0.62
Industry	519	In-sample	0.84	0.92	0.90-0.93
		Out-of-sample	0.40	0.63	0.58-0.68
Macro	731	In-sample	0.82	0.91	0.89-0.92
		Out-of-sample	0.67	0.82	0.79-0.84
Finance	308	In-sample	0.87	0.93	0.91-0.95
		Out-of-sample	0.42	0.65	0.58-0.71
Demographic	413	In-sample	0.75	0.87	0.84-0.89
		Out-of-sample	0.44	0.66	0.60-0.71
Other	204	In-sample	0.98	0.99	0.98-0.99
		Out-of-sample	0.85	0.92	0.90-0.94

Figure 4 shows a box plot of the in-sample MASE and the out-of-sample MASE. The out-of-sample MASE has a higher median value (0.52 vs. 0.35) and exhibits multiple outliers. Such a deterioration in out-of-sample forecast performance is common and can be attributed to a number of factors, but is often due to structural breaks [22]. In practice, many real-world time series are non-stationary and can result from complex and/or nonlinear data-generating processes thereby resulting in series that are not constant over time [23]. Clements and Hendry [20] state that economic time series, both micro and macro, are frequently subject to major political, institutional, financial, legal, fashion, and technological changes which manifest as structural breaks in forecasting models relative to the underlying data-generating processes. This can result in models that appear to provide good in-sample fit, but do not effectively extrapolate out-of-sample. Clements and Hendry [20] also identify the following as further potential sources of forecast failure, model mis-specification, inaccurate forecast-origin data and inefficient estimation. However, despite the out-of-sample forecast deterioration, the in-sample SampEn did achieve a notable correlation with the out-of-sample forecast MASE over this comprehensive real-world data set.

Figure 4. Box plot of MASE in-sample and out-of-sample



5. Conclusions

In this study we have applied the sample entropy metric to the entire M3-competition data set of 3003 time series to determine its usefulness as an indicator of forecastability. Sample entropy was shown to have a strong correlation with in-sample model fit (R^2=0.87) and a moderate correlation (R^2=0.56) with out-of-sample forecast accuracy using Holt's method, across all 3003 series. Notably, one dataset ("other", containing telecommunications and utilities series), showed a particularly strong out-of-sample coefficient of determination of R^2=0.85. We therefore conclude that the study has successfully demonstrated that sample entropy is a useful indicator of forecastability in the domain of a simple univariate forecasting method. Sample entropy is also adept at assessing regularity in chaotic data-generating processes and could prove valuable in indicating the appropriateness of nonlinear forecasting techniques. We recommend that forecasting practitioners consider its use, alongside well-established tests for seasonality and trend, in an effort to better understand the likelihood of successful forecasting outcomes. In situations where the *a priori* sample entropy is low, yet the out-of-sample forecasting error is high, then consideration should be given to the impact of out-of-sample structural breaks as well as the suitability of the forecasting method in regard to its ability to exploit the identified regularity in the series.

A limitation of this study is that a single univariate model, Holt's method, was applied to all series and more sophisticated univariate or multivariate, nonlinear and/or judgmental approaches could conceivably deliver higher out-of-sample forecast accuracy for a particular time series.

References

1. Boylan, J.E. Towards a more precise definition of forecastability. *Foresight: the International Journal of Applied Forecasting* **2009**, 34–40.
2. Kolassa, S. How to assess forecastability. *Foresight: the International Journal of Applied Forecasting* **2009**, 41–45.
3. Tashman, L.J. Special feature on forecastability: Preview. *Foresight: the International Journal of Applied Forecasting* **2009**, 23.
4. Tsay, R.S. *Analysis of financial time series*. Third ed.; Wiley: 2010.
5. Wu, S.D.; Wu, C.W.; Lin, S.G.; Wang, C.C.; Lee, K.Y. Time series analysis using composite multiscale entropy. *Entropy* **2013**, *15*, 1069–1084.
6. Catt, P.M. Forecastability: Insights from physics, graphical decomposition and information theory. *Foresight: The International Journal of Applied Forecasting* **2009**, 24–33.
7. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences* **1991**, *88*, 2297–2301.
8. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *American journal of physiology. Heart and circulatory physiology* **2000**, *278*, H2039–2049.
9. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Physical review letters* **2002**, *88*, 174102.
10. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy* **2012**, *14*, 1553–1577.
11. Shannon, C.E. *A mathematical theory of communication*. American Telephone and Telegraph Company: New York, 1948; p 80 p.
12. Yentes, J.M.; Hunt, N.; Schmid, K.K.; Kaipust, J.P.; McGrath, D.; Stergiou, N. The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of biomedical engineering* **2013**, *41*, 349–365.
13. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C. *Time series analysis, forecasting and control. Third edition.* . Holden-Day: 1976.
14. R Development Core Team. *R: A language and environment for statistical computing. R foundation for statistical computing*, Vienna, Austria, 2008.
15. Makridakis, S.; Hibon, M. The m3-competition: Results, conclusions and implications. *International Journal of Forecasting* **2000**, *16*, 451–476.
16. Clemen, R.T. Simple versus complex methods. *International Journal of Forecasting* **2001**, *17*, 549–550.
17. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* **2004**, *20*, 5–10.
18. Hyndman, R.J.; Khandakar, Y. "Automatic time series forecasting: The forecast package for r". *Journal of Statistical Software* **2008**, *26*.
19. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *International Journal of Forecasting* **2006**, *22*, 679–688.
20. Clements, M.P.; Hendry, D.F. Explaining the results of the m3 forecasting competition. *International Journal of Forecasting* **2001**, *14*, 550–554.
21. Borchers, H.W. *Pracma: Practical numerical math functions*, 1.7.0; http://CRAN.R-project.org/package=pracma, 2014.
22. Clements, M.P.; Hendry, D.F. Forecasting economic processes. *International Journal of Forecasting* **1998**, *14*, 111–131.
23. Strogatz, S.H. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering,* . Addison-Wesley: Reading, MA, 1994.