Noname manuscript No.
(will be inserted by the editor)

Is a Reputation Time Series White Noise?

Peter Mitic

Received: date / Accepted: date

Abstract Special session 2: Finance and Data Mining. The plots of some reputation time series superficially resemble plots of white noise. This raises the question of whether or not the analysis of sentiment to produce a reputation index actually generates nothing more than noise. The question is answered by using the Box-Ljung statistical test to establish that the reputation time series considered in this analysis cannot be viewed as white noise. This result is supported by applying a new test based on cross-correlations of reputation time series with white noise time series.

Keywords Reputation · Reputation Index · White Noise · Box-Hjung · Cross correlation · Auto correlation

1 Introduction and Motivation

The mathematical basis of reputation measurement were described in these proceedings [1], and the process of procuring a reputation score was described in [2]. Procurement is a physical process which is well-defined and is objective, subject to possible error in the elucidation of sentiment. The result is simply a sequence of numbers. Unfortunately, such series can superficially resemble random number sequences. Figure 1 shows a genuine reputation time series (Mercedes-Benz) and a randomly generated time series from a normal distribution with mean and standard deviation approximately equal to the mean

The opinions, ideas and approaches expressed or presented are those of the author and do not necessarily reflect Santanders position. The values presented are just illustrations and do not represent Santander data.

P. Mitic
Santander UK, 2 Triton Square, Regents Place, London NW1 3AN
and Dept. Computer Science, University College London, Gower Street, London WC1E 6BT
and Laboratoire d'Excellence sur la Régulation Financière (LabEx ReFi), Paris
Tel.: +44 (0)207 756 5256
E-mail: peter.mitic@santandergcb.com

and standard deviation of the Mercedes-Benz time series. The question then arises ”Is the method of procurement of a reputation time series merely a complicated physical method of generating random numbers?”

In this paper we test the conjecture that a reputation time series is randomly generated. We then consider a more subtle problem, which is to decide whether or not a reputation time series constitutes white noise. Implicit in this discussion is the concept that reputation time series has what could be described as a “memory” - the reputation score on one day depends on previous the reputation scores. Standard statistical tests are used to test the question posed above, and an additional new statistical test is proposed that confirms the conclusion by using cross-correlations in an intuitive way.

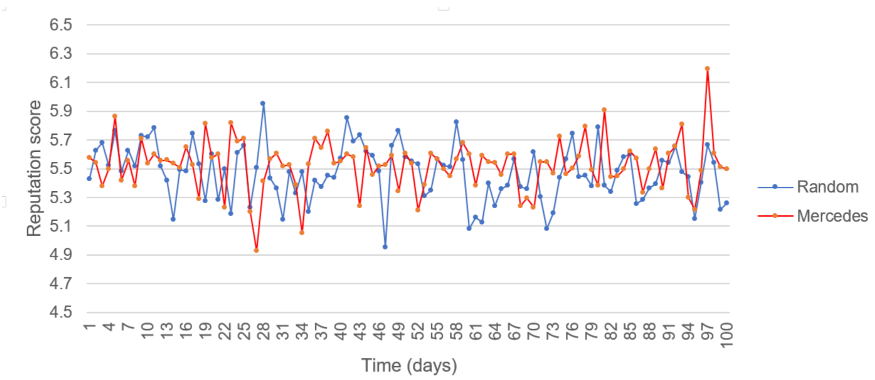

Fig. 1 Contrasting time series: reputation and random

1.1 Randomness and white noise

A precise definition of randomness is surprisingly difficult, and opinions on a suitable definition differ. The essential elements are that the outcome of any given trial should be unpredictable, that it cannot subsequently be reliably reproduced, and that a sequence of random numbers does not exhibit autocorrelation. Knuth [3] argues that a sequence of random numbers is a sequence of independent numbers with a specified distribution and that there is a specified probability of selecting a number in any given range of values.

The concept of “white noise” extends the idea of randomness and casts a more precise definition in terms of a statistical process. Specifically, if x_t is a random variable with independent variable t, each term in the sequence $\{x_t\}$ should be a stochastic term ϵ_t with mean 0 and constant variance σ^2, and any two such terms ϵ_t and ϵ_s $(t \neq s)$ are independent. An instructive illustration of white noise may be found on the Wolfram research website at https://reference.wolfram.com/language/ tutorial/AudioProcessing.html. The two sound clips defined by the following Mathematica statements show the effect of reducing ("filtering") the white noise in the broadcast of Neil Armstrong's "One small step for man..." speech on the moon. The former has a marked background hiss.

- a = ExampleData["Audio", "Apollo11SmallStep", "Audio"]
- WienerFilter[a, 25]

Noise sounds like hiss, and converys no information. We would like to think that a reputation time series is not just hiss!

2 Tests for randomness and white noise: Review

In this section we review some standard tests for white noise and randomness, prior to introducing a new test that makes auto-correlations central and intuitive to the decision process. The way in which the tests are used is discussed in section 3.

2.1 The Runs test

The *Runs test* [4] examines randomness by counting runs of similar numbers. Having centred data about the data mean, a *run* is a set of sequential values that are either all above the mean (a "positive run") or below the mean (a "negative run"). The number of positive and negative runs, n_+ and n_- respectively, are counted. The total number of runs, $n = n_+ + n_-$, has the following Normal distribution for large n:

$$n \sim N(1 + \frac{2n_+n_-}{n_+ + n_-}, \frac{2n_+n_-(2n_+n_- - n_+ - n_-)}{(n_+ + n_- - 1)(n_+ + n_-)^2}) \qquad (1)$$

The test asks if the number of positive and negative runs are distributed equally in time. The null and alternative hypotheses are (respectively H_0 and H_a):

- H_0: The data are random
- H_a: The data are not random

2.2 The Box-Hjung test

The Box-Ljung test [5] is an unusual statistical test because it is used to test the *lack* of fit of a time series model rather than how well a model fits the data. It is a specific test for white noise. The test is applied by first fitting an ARMA model to the data and then calculating residuals. Autocorrelations covering a range of lags, rather than any one lag, of the residuals are then examined. For this reason, it is often referred to as a "portmanteau" test. The Ljung-Box test is formulated with the following null and alternative hypotheses.

- H_0: The time series exhibits autocorrelation
- H_a: The time series does not exhibit autocorrelation

The Box-Ljung test statistic, Q_k, equation (2), is

$$Q_k = n(n+2)\sum_{i=1}^{k} \frac{r_i^2}{n-i}, \tag{2}$$

where n is the time series length, k is the maximum number of lags considered, and r_i is the autocorrelation at lag i. Q_k has a $\chi^2(k)$ distribution. At significance level α, H_0 is rejected if the calculated Q_k is less than the $(1-\alpha)\%$ value of $\chi^2(k)$. A loose interpretation is that small values of Q_k indicate "not white noise".

2.3 Cross-correlation

Cross-correlation is an alternative way to assess serial autocorrelation, and is discussed in [6]. For large sample size n, the distribution of the cross-correlation coefficient for two time series $\{x_t\}$ and $\{y_t\}$ at lag k, $\rho_{xy}(k)$, is Normal, provided that at least one of the time series is white noise. Specifically,

$$\rho_{xy}(k) \sim N(0, \frac{1}{n}) \tag{3}$$

The adaptation described in section 3 gives an amendment to (3) due to the Central Limit Theorem. The null and alternative hypotheses are:

- H_0: The mean cross-correlation coefficient $= 0$
- H_a: The mean cross-correlation coefficient $\neq 0$

2.4 Other tests

This sub-section has a brief overview of some other tests for white noise. They are not implemented in this study because they are more complex, and the simpler tests used in this study suffice.

Bartlett's formula [7] tests for white noise using a matrix of autocorrelations, and assumes a stationary time series with IID noise. Autocorrelations are shown to be normally distributed for large sample size. The method is known to be inaccurate for non-linear time series because it assumes that all autocorrelations with lag greater than 1 are near zero.

Francq and Zacoïan [8] discuss an amendment based on the Bartlett formula, applicable for the asymptotic distribution of the sample autocorrelations of nonlinear processes. The first term is the same as the Bartlett linear formula. A second term introduces the kurtosis of the linear innovation process and the autocorrelation function of its square.

All the tests discussed so far are based in the time domain. An alternative is to use the frequency domain. A set of tests, starting with that of Hong [9] are based on a comparison between the spectral density of the time series and the spectral density of a white noise time series. There have been subsequent further developments of this approach.

3 Tests for randomness and white noise: Application

The way in which the tests described in the previous section are applied is described in this section. Particular attention is paid to a comparison of reputation data with randomly generated data. There was no special adaptation of the Runs Test.

3.1 The Box-Hjung test: Application

An implementation in the *R* statistical language, the function *Box.test*, was used to return the p-value of the embedded χ^2 test. Rather than just apply this test to the reputation data alone, it was also applied to randomly generated data (using an ARIMA(0,0,0) process) that resembles the reputation data. The Box-Hjung test was applied to 250 instances of this stochastic process and the mean p-value was calculated. The results for the reputation data and the randomly generated data were then compared.

3.2 The Cross-correlation test: Application

As it stands, equation (3) applies for each cross-correlation value in isolation. In order to derive a single-figure measure of the cross-correlation across all lags, the cross correlation test was applied 1000 times (i.e. enough to provide a consistent result), and the mean cross-correlation at each lag was calculated. Then, if the number of lags is L, the mean cross-correlation, R_{xy}, has the following distribution by the Central Limit Theorem.

$$R_{xy} \sim N(0, \frac{1}{nL}) \tag{4}$$

A measure of the significance of the empirical value of the statistic R_{xy} (the p-value, p) can then be obtained using:

$$p = 2(1 - \Phi(|m|), 0, \frac{1}{nL}), \tag{5}$$

where Φ is the right-hand tail of the standard Normal distribution and m is the empirical mean cross-correlation value. In (5) the term $|m|$ accounts for both positive and negative values of m. The implication of the hypotheses for this test is that one time series either is, or is not, a predictor of the other. In particular, we will test whether or not a white noise time series can be a predictor of a reputation time series. If it can, we would consider a reputation time series to be white noise.

4 Results

Seventeen reputation time series were considered. Ten of them comprised reputation data for a 720-day period, and the remaining seven covered a 400-day period. Both are sufficiently long to provide reliable conclusions. The results of applying the tests described in the previous section are remarkably consistent. Therefore there is no need to present each indvidually. General comments for each test suffice.

Runs test results
With the null hypothesis that the input data are random, a "small" p-value indicates that the null hypothesis should be rejected (i.e. the data are not random). This is the case for all 17 reputational time series except three, for which the p-values were 0.314, 0.159 and 0.421. The p-values for the other 14 time series were all less than 0.006. The conclusion is that 14 of the time series are *most likely* not random, and 3 of them *could be*.

Box-Hjung test results
Using the Box-Hjung test on the 17 reputation time series resulted in p-values of less than 10^{-5} in all cases. This indicates that the reputation time series all exhibit serial correlation. In contrast, applying the Box-Hjung test to 250 instances of random data produced Normally distributed p-values with mean 0.498, and standard deviation 0.293. This indicates a lack of serial correlation. The conclusion is that none of the reputation time series corresponds to white noise. Figure 2 shows the lag plot for the Ford reputation scores - illustrating the correlation between the scores with no lag and the scores lagged by 1 day. The plot resembles a random scatter, indicating very low correlation. Plots for other lags and other organisations are similar.

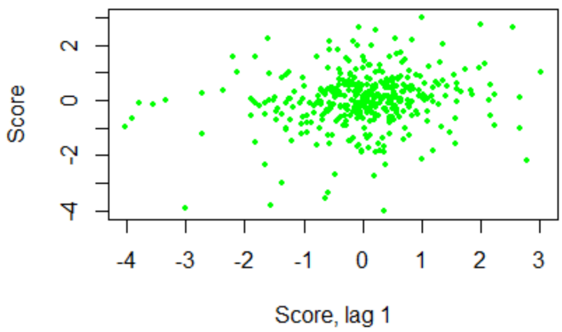

Fig. 2 Typical Lag Plot - Ford

Cross-correlation test results
The output of the cross-correlation test is a p-value which measures the degree to which one input time series is a predictor of the other. Using any of the

available reputation time series with a gaussian white noise time series, the p-values obtained were nearly always non-significant values in the range (0.942, 0.995). The only exception was one reputation time series for which the p-value was 0.792, also not significant This indicates that a white noise time series cannot be used as a predictor of a reputation time series (and vice versa). Using two gaussian white noise time series the results were much the same. The p-values obtained were always non-significant, in the range (0.940, 0.999). Clearly, one white noise time series should not be a predictor of another. Figure 3 shows a typical cross-correlogram (Ford against white noise) in which two (very marginal!) breaches of the 5% confidence limits (in blue) are apparent. It should be noted that these confidence limits apply only if the underlying time series data are normally distributed (which applies in the case of Ford). When many such results are run and averaged, the result is as shown in figure 4: no significant breaches of the 5% confidence limits.

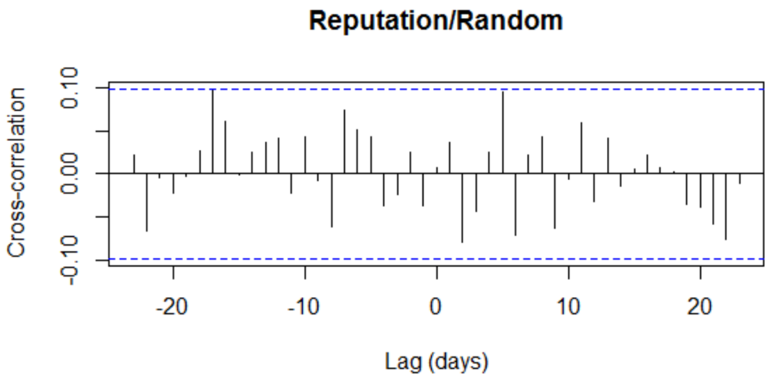

Fig. 3 Typical single cross-correlation: Ford

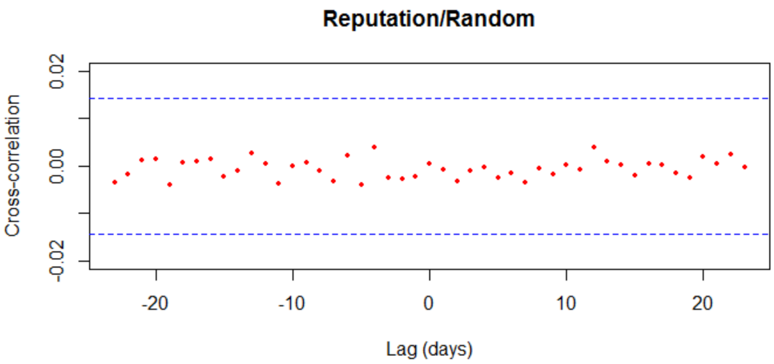

Fig. 4 Mean cross-correlations: 1000 trials: Ford

5 Discussion

This work was prompted by the observation that some reputation time series resemble white noise, and also because some of them exhibit very weak auto-correlations. Typically, the reputation time series that fall into this category tend to have empirical normal distributions with mean greater than the nominal neutral score. These are the suspect ones. Others exhibit distinct negative skews (i.e. there is a tail of low value scores). For these, correlations are more apparent, although they can be weak.

As a preliminary, the Runs Test indicates that only a few of the reputation time series can be considered as random. The resemblance shown in figure 1 is therefore superficial. The Box-Ljung test shows very clearly that there is a very distinct difference between a white noise signal and a reputation signal: the former has no auto-correlation and the latter does. This view is confirmed by an examination of cross correlations between white noise signals and reputation/noise pairings. A cross correlation plot such as figure 3 provides a good visual impression of a series of correlations at different lags, and the new test developed in sub-section 3.2 quantifies that impression.

Overall, the answer to the question "is a reputation time series the same as white noise?" is a very clear "no". Establishing that implies that a reputation time series conveys useful information. Ways to interpret that are a matter for current study.

Acknowledgements I am grateful for the support of Alva-Group for their continued interest, support, and assistance in the preparation of this paper.

References

1. Mitic,P., Standardised Reputation Measurement, Proc. IDEAL2017 Guilin China, 2017
2. Mitic, P. Reputation Risk: Measured, Proc. Complex Systems May 2017, Wessex Institute of Technology Press, 2017.
3. Knuth, D., The Art of Computer Programming, Vol. 2, 2nd ed., Addison-Wesley, 1981
4. Bradley,J.V., Distribution-Free Statistical Tests Chapter 12, Prentice-Hall, 1968
5. Ljung, G. M. and Box, G. E. P., On a Measure of a Lack of Fit in Time Series Models, Biometrika 65(2), pp297303. doi:10.1093/biomet/65.2.297.
6. Shumway,R.H. and Stoffer,D.S., Time Series Analysis and its Applications 3rd edition Chapter 1, Springer, 2010
7. Bartlett,M.S., On the theoretical specification and sampling properties of autocorrelated time series, Supplement to J. Royal Statistical Society 8, pp 27-41, 1946
8. Francq,C. and Zacoïan,J-M., Bartlett's formula for a general class of nonlinear processes, J. Time Series Analysis 30(4), pp. 449-465, 2009
9. Hong, Y., Consistent testing for serial correlationof unknown form, Econometrica 64, pp. 837864, 1996